AUSTRIAN JOURNAL OF STATISTICS
Volume 31 (2002), Number 2&3, 177-188

Minimax Robustness of Bayesian Forecasting under Functional Distortions of Probability Densities

Alexey Kharin
Belarussian State University, Minsk

Abstract: The problems of robustness in Bayesian forecasting are considered under distortions of the hypothetical probability densities. The expressions for the guaranteed upper risk functional are obtained and the robust prediction statistics under certain types of distortions are constructed.

Keywords: Bayesian Forecasting, Robustness, Minimax, Distortions, Risk.

1 Introduction

One of the main branches of applied Data Analysis is statistical forecasting. In statistical forecasting a prior information is oftenly available in addition to the information contained in observations (Viertl, 1996). The Bayesian approach is a method for involving a prior information into forecasting procedures (West and Harrison, 1989). Procedures derived from the Bayesian approach are proved to be optimal w.r.t. the minimal risk criterion (Berger, 1985). Also, nowadays PCs are powerful enough to realize the necessary computations using new numerical methods and the needed software is available (Pilz, 2000). That is why the Bayesian approach becomes to be used widely.

The optimality property of the Bayesian approach is valid only if a hypothetical model is adequate to a considered process. In practice such an assumption is more than optimistic (Dutter, 1996; Rieder, 1994). Hence there are at least two classes of problems to appear actual (Huber, 1981; Hampel, E.M. Ronchetti, P.J. Rousseeuw and W. Stahel, 1986): (i) robustness analysis of the classical Bayesian prediction statistic (p.s.) under distortions of the hypothetical model; (ii) construction of new robust p.s. using the Bayesian approach.

Many papers are devoted to the problems of Bayesian robustness for model parameter estimation (Berger et al., 1996). It is known that Bayesian forecasting is not reducible to Bayesian parameter estimation. Some analytic results on robustness of Bayesian forecasting under the Tukey–Huber distortions can be found in Galinskij and Kharin (1998). That paper also contains the results of Monte-Carlo simulations for the case of distortions in the weighted C-metric. Here we give the solutions for the problems (i), (ii) under certain types of functional distortions indicated below.

2 Hypothetical Model and Optimal Forecast

Let on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$ be defined three random elements (Figure 1): 1) an unknown parameter vector $\theta = (\theta_1, \ldots, \theta_m) \in \Theta \subseteq \mathbb{R}^m$, with a hypothetical probability density function (p.d.f.) $\pi^0(\theta)$; 2) an observation vector $x = (x_t) \subseteq X \in \mathbb{R}^T$ stochastically dependent on θ with a hypothetical conditional p.d.f. $p^0(x \mid \theta)$; 3) a value to be forecasted $y \in Y \subseteq \mathbb{R}$, stochastically dependent on θ, x, and distributed according to a hypothetical conditional p.d.f. $g^0(y \mid x, \theta)$. Suppose that p.d.f.s $\pi^0(\theta)$, $p^0(x \mid \theta)$, $g^0(y \mid x, \theta)$ are continuous functions. The problem is to construct a prediction statistic (p.s.) $\hat{y} = f(x)$ for y using x, $f(\cdot)$: $X \longrightarrow Y$ is a Borel function of T variables. To simplify the notation let us suppose that the hypothetical p.d.f. of an observation vector has no zeroes:

$$p^0(x) = \int_X p^0(x \mid \theta)\pi^0(\theta)d\theta > 0,\ x \in X. \tag{1}$$

Let X, Θ be compact sets.

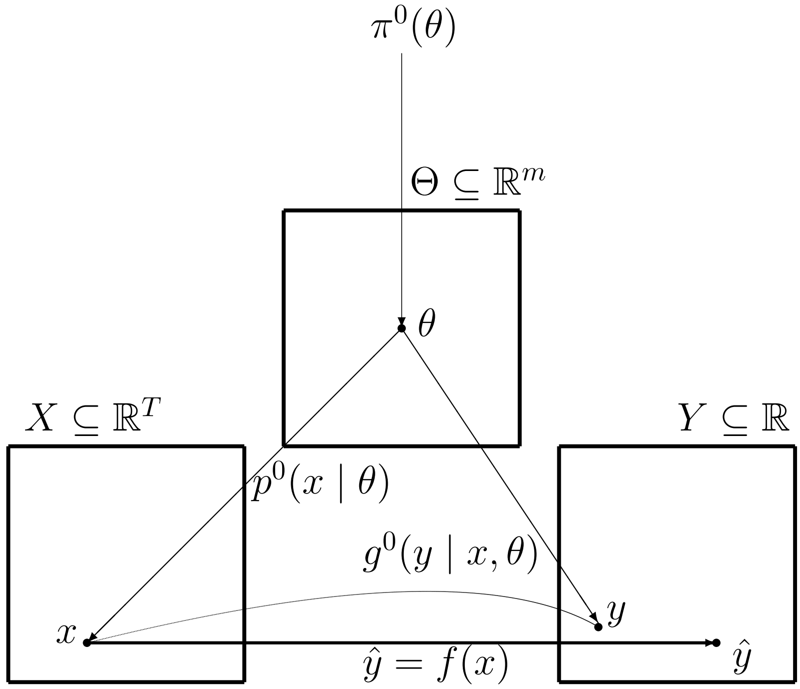

Figure 1: Hypothetical model of Bayesian forecasting

The optimality criterion for construction of a p.s. $f(\cdot)$ is the minimum of the Bayesian risk functional:

$$r_0(f) = \int_X \int_Y p^0(x)q^0(y \mid x)(f(x) - y)^2 dy dx \longrightarrow \min, \tag{2}$$

where

$$q^0(y \mid x) = \int_\Theta g^0(y \mid x, \theta)\pi^0(\theta \mid x)d\theta \tag{3}$$

is the hypothetical Bayesian prediction density, and

$$\pi^0(\theta \mid x) = p^0(x \mid \theta)\pi^0(\theta)/p^0(x).$$

Introduce the family of admissible p.s.s $F = \{f(\cdot) :\ r_0(f) < \infty\}$.

It is well known that the optimal p.s. (w.r.t. the criterion (2)) is the Bayesian one (posterior mean value of y)

$$\hat{y} = f_0(x) = \int_Y y q^0(y \mid x) dy, x \in X. \tag{4}$$

3 Robustness Characteristics under Distortions of the Hypothetical Model

Let the hypothetical model which is presented in Figure 1 be distorted: $\pi^\varepsilon(\theta) \in \Pi$, $p^\varepsilon(x \mid \theta) \in P$, $v^\varepsilon(x,\theta) \in V$ are distorted versions of the p.d.f.s $\pi^0(\theta)$, $p^0(x \mid \theta)$, $v^0(x,\theta) = p^0(x \mid \theta)\pi^0(\theta)$ respectively, where Π, P, V are any sets of admissible distorted densities. Under distortions the risk functional is:

$$r(f; s^\varepsilon) = \int_X \int_Y s^\varepsilon(x,y)(f(x) - y)^2 dy dx, \tag{5}$$

where $s^\varepsilon(x,y) = p^\varepsilon(x) q^\varepsilon(y \mid x)$, and $p^\varepsilon(x)$, $q^\varepsilon(y \mid x)$ are defined analogously to (1), (3) through the distorted p.d.f.s.

Let us define two functionals on the set of admissible p.s.s: 1) an upper risk $r_+(f)$ as an upper bound of the value set of the functional (5):

$$r(f; s^\varepsilon) \le r_+(f),\ s^\varepsilon(\cdot) \in S,$$

where S is defined by the sets Π, P and V; 2) the guaranteed upper risk functional

$$r_*(f) = \sup_{s^\varepsilon(\cdot) \in S} r(f; s^\varepsilon) \le r_+(f). \tag{6}$$

An admissible p.s. $f^*(\cdot)$ is called the r_+-robust p.s. if it minimizes the upper risk functional $r_+(\cdot)$:

$$r_+(f^*) = \inf_{f(\cdot) \in F} r_+(f).$$

Let us define the robust p.s. by minimax criterion as

$$f_*(\cdot):\ r_*(f_*) = \inf_{f(\cdot) \in F} r_*(f).$$

4 Robustness under Distortions of Prior Probability Density

Let us introduce the distance between two p.d.f.s $h_1(\cdot)$, $h_2(\cdot)$ in the space $\mathrm{C}(U)$ of continuous functions with a weight function $1/w(u)$, $w(u) > 0$, $u \in U$:

$$\rho^w_{C(U)}(h_1(\cdot), h_2(\cdot)) = \sup_{u \in U} \left(|h_1(u) - h_2(u)|/w(u)\right). \tag{7}$$

Condition 1. *Suppose that the set* Π *be the* ε_+*-neighborhood of* $\pi^0(\cdot)$ *in the metric (7) for some* $\varepsilon_+ \ge 0$*, with the weight function* $1/\pi^0(\theta)$*,* $\pi^0(\theta) \ne 0$*,* $\theta \in \Theta$*.*

Theorem 1. *Under Condition 1 for an admissible p.s.* $\hat{y} = f(x)$ *the guaranteed upper risk functional (6) satisfies the inequality:*

$$r_*(f) \leq r_+(f) = (1 + \varepsilon_+) r_0(f). \tag{8}$$

Proof. From Condition 1 it follows the inequality:

$$|\Delta\pi(\theta)| \leq \varepsilon_+ \pi^0(\theta), \tag{9}$$

where

$$\Delta\pi(\theta) = \pi^\varepsilon(\theta) - \pi^0(\theta). \tag{10}$$

The risk functional (5) takes the form:

$$r(f; \pi^\varepsilon) = \int_X \int_Y \int_\Theta (\pi^0(\theta) + \Delta\pi(\theta)) p^0(x \mid \theta) g^0(y \mid x, \theta)(f(x) - y)^2 d\theta dy dx.$$

Now, using (9) we come to (8). □

Corollary 1. *Under Condition 1 the Bayesian p.s. (4) is* r_+*-robust:*

$$\hat{y} = f_0(x) = f^*(x), \ x \in X.$$

Corollary 2. *The upper risk functional (8) is non-decreasing on* ε_+.

Now let us extend the set of admissible distorted prior p.d.f.s: $\pi^\varepsilon(\cdot)$ may have ordinary discontinuity points. For this case the distance between two p.d.f.s can be formally calculated according to (7). Introduce the notation:

$$s^0(x, y \mid \theta) = g^0(y \mid x, \theta) p^0(x \mid \theta), \ r_1(f(\cdot); \theta) = \int\limits_X \int\limits_Y s^0(x, y \mid \theta)(f(x) - y)^2 dy dx,$$

$$\Theta_z = \{\theta \in \Theta : \ r_1(f(\cdot); \theta) \geq z\}, \ z \in \mathbb{R}_+, \tag{11}$$

$$z^* = \min\left\{ z \in \mathbb{R}_+ : \int\limits_{\Theta_z} \pi^0(\theta) d\theta = \min\left\{ \frac{1}{1 + \varepsilon_+}, \frac{1}{2} \right\} \right\},$$

and let $\mathbf{1}_V(u)$, $u \in U$, be the indicator function of a set $V \subseteq U$.

Theorem 2. *If under Condition 1 only ordinary discontinuities are admitted for densities from* Π, *then for an admissible p.s.* $\hat{y} = f(x)$ *the guaranteed upper risk functional (6) is*

$$r_*(f) = r_0(f) + \varepsilon_+ \int\limits_{\Theta_{z^*}} r_1(f(\cdot); \theta) \pi^0(\theta) d\theta - \min\{\varepsilon_+, 1\} \int\limits_{\Theta \setminus \Theta_{z^*}} r_1(f(\cdot); \theta) \pi^0(\cdot) d\theta; \tag{12}$$

this supremum of risk is reached at the distorted p.d.f.

$$\pi^*(\theta) = \pi^0(\theta) \left(1 + \varepsilon_+ \mathbf{1}_{\Theta_{z^*}}(\theta) - \min\{\varepsilon_+, 1\} \mathbf{1}_{\Theta \setminus \Theta_{z^*}}(\theta)\right), \ \theta \in \Theta. \tag{13}$$

Proof. Let us consider the problem of variation calculus under notation (10), (11):

$$\begin{cases} \int\limits_{\Theta} r_1(f(\cdot);\theta)\Delta\pi(\theta)d\theta \longrightarrow \max\limits_{\Delta\pi(\theta),\ \varepsilon\in[0,\varepsilon_+]}, \\ -\min\{\varepsilon,1\}\pi^0(\theta) \le \Delta\pi(\theta) \le \varepsilon\pi^0(\theta),\ \theta\in\Theta, \\ \int\limits_{\Theta} \Delta\pi(\theta)d\theta = 0. \end{cases} \tag{14}$$

To obtain the guaranteed upper risk functional (6) is equivalent to solve (14). First let us solve (14) for a fixed $\varepsilon \in [0, \varepsilon_+]$. According to Kuhn–Tucker theorem for convex programming in an arbitrary space (Alexeev, 1979), the solution of the extremum problem (14) is

$$\Delta\pi(\theta) = \pi^0(\theta)\left(\varepsilon\mathbf{1}_{\Theta_{z^*}}(\theta) - \min\{\varepsilon,1\}\mathbf{1}_{\Theta\setminus\Theta_{z^*}}(\theta)\right).$$

Putting this expression into the objective function in (14) and then maximizing it w.r.t. $\varepsilon \in [0, \varepsilon_+]$, using the hypothetical risk definition (2), we come to (13), (12). □

Using Theorem 2 one can find the maximal increment of the risk functional under the considered distortions. To minimize the functional (12) w.r.t. p.s. the classical methods of optimization can not be applied. But comparing (12) and (8) it is easy to conclude that the both of increments, $|r_+(f) - r_0(f)|$ and $|r_*(f) - r_0(f)|$ have the first order w.r.t. the distortion level ε_+, so it is possible to talk on the "closeness" of properties of robustness and r_+-robustness for this case of distortions.

Some simulation results for autoregressive time series forecasting under distortions in the weighted C-metric are presented in Galinskij and Kharin (1998).

5 Robustness under Distortions of Joint Probability Density

5.1 Distortions in the Weighted C-metric

Theorem 3. *Let for some* $\varepsilon_+ \ge 0$ *the set* V *of admissible distorted joint p.d.f.s be the* ε_+*-neighborhood of* $v^0(x,\theta)$ *in the weighted* C*-metric (7):*

$$\Pi = \{v^\varepsilon(\cdot) :\ \rho^{v^0}_{C(X\times\Theta)}(v^0(\cdot), v^\varepsilon(\cdot)) \le \varepsilon_+\}. \tag{15}$$

Then the functional

$$r_+(f) = (1+\varepsilon_+)r_0(f) \tag{16}$$

is an upper risk functional.

Proof. Let us use the definition of risk functional (5) in the form

$$r(f; v^\varepsilon) = \int\limits_X \int\limits_Y \int\limits_\Theta g^0(y \mid x,\theta)v^\varepsilon(x,\theta)d\theta(f(x)-y)^2 dydx. \tag{17}$$

Applying to the p.d.f. $v^\varepsilon(x,\theta)$ and to the expression (17) the transformations analogous to those used in the proof of Theorem 1 we come to the statement of the theorem. □

Corollary 3. *The Bayesian p.s. (4) is r_+-robust under the conditions of Theorem 3 for the upper risk functional (16):*

$$f_0(x) = f^*(x), \ x \in X.$$

Consider the case, important for applications, where distortions (15) are caused by distortions of $p^0(x \mid \theta)$, $\pi^0(\theta)$ at different distortion levels.

Theorem 4. *Let for some $\varepsilon_{1+} \geq 0$, $\varepsilon_{2+} \geq 0$, the sets P, Π be the neighborhoods of the hypothetical p.d.f.s $p^0(x \mid \theta)$, $\pi^0(\theta)$ in the weighted* C*-metric:*

$$\Pi = \{\Pi_{\varepsilon_1} : \ 0 \leq \varepsilon_1 \leq \varepsilon_{1+}\}, \ \Pi_{\varepsilon_1} = \left\{\pi^\varepsilon(\cdot) : \ \rho^{\pi^0}_{\mathrm{C}(\Theta)}(\pi^0(\cdot), \pi^\varepsilon(\cdot)) = \varepsilon_1\right\}, \tag{18}$$

$$P = \{P_{\varepsilon_2} : \ 0 \leq \varepsilon_2 \leq \varepsilon_{2+}\}, \ P_{\varepsilon_2} = \left\{p^\varepsilon(\cdot \mid \cdot) : \ \rho^{p^0}_{\mathrm{C}(X \times \Theta)}(p^0(\cdot \mid \cdot), p^\varepsilon(\cdot \mid \cdot)) = \varepsilon_2\right\}.$$

Then the guaranteed upper risk satisfies the inequality:

$$r_*(f) \leq r_+(f) = (1 + \varepsilon_{1+})(1 + \varepsilon_{2+}) r_0(f).$$

Proof. It is analogous to the proof of Theorems 1, 3. It is enough to use (18) in (17) and to present $v^\varepsilon(x, \theta)$ in the form

$$v^\varepsilon(x, \theta) = p^\varepsilon(x \mid \theta)\pi^\varepsilon(\theta), \ x \in X, \ \theta \in \Theta.$$

□

Let us denote that under Theorem 4 conditions the statements analogous to Corollaries 1, 2 are also valid.

5.2 Distortions in χ^2-metric

Condition 2. *Let $v^0(x, \theta) \neq 0$, $x \in X$, $\theta \in \Theta$, and for some $\varepsilon_+ \geq 0$ the set of admissible distorted joint p.d.f.s of x, θ be the ε_+-neighborhood of $v^0(\cdot)$ in the χ^2-metric:*

$$V = \Big\{ v^\varepsilon(\cdot) : \ \ v^\varepsilon(x, \theta) \geq 0, \ x \in X, \ \theta \in \Theta; \\ \int_X \int_\Theta v^\varepsilon(x, \theta) d\theta dx = 1; \ \int_X \int_\Theta \frac{\left(v^\varepsilon(x,\theta) - v^0(x,\theta)\right)^2}{v^0(x,\theta)} d\theta dx \leq \varepsilon_+^2 \Big\}. \tag{19}$$

Let the subindex 0 or ε at the operators $\mathbf{E}$, $\mathbf{D}$ indicates which distribution is used: hypothetical or distorted.

For an admissible p.s. $\hat{y} = f(x)$ introduce the functional:

$$r_{x,\theta}(f; x, \theta) = \mathbf{E}_0\{(f(x) - y)^2 \mid x, \theta\}.$$

Then the risk functional (5) takes the form:

$$r_\varepsilon(f; v^\varepsilon) = \mathbf{E}_\varepsilon\{r_{x,\theta}(f; x, \theta)\}, \ \ v^\varepsilon(\cdot) \in V. \tag{20}$$

Let us denote:

$$\overset{\circ}{r}_{x,\theta}(f;x,\theta) = r_{x,\theta}(f;x,\theta) - \mathbf{E}_0\{r_{x,\theta}(f;x,\theta)\},$$

$$\varepsilon^* = \frac{\sqrt{\mathbf{D}_0\{r_{x,\theta}(f;x,\theta)\}}}{\sup_{x\in X,\theta\in\Theta} |\overset{\circ}{r}_{x,\theta}(f;x,\theta)|},$$

$$v^{*\varepsilon_+}(f;x,\theta) = v^0(x,\theta)\left(1+\varepsilon_+\frac{\overset{\circ}{r}_{x,\theta}(f;x,\theta)}{\sqrt{\mathbf{D}_0\{r_{x,\theta}(f;x,\theta)\}}}\right), \qquad (21)$$

$$R(x) = \frac{1}{\int_\Theta v^0(x,\theta)d\theta}\cdot\int_\Theta \mathbf{E}_0\{y \mid x,\theta\}v^0(x,\theta)\times$$
$$\left(\overset{\circ}{r}_{x,\theta}(f_0;x,\theta) - \int_\Theta \overset{\circ}{r}_{x,\theta}(f_0;x,\tilde\theta)\pi^0(\tilde\theta \mid x)d\tilde\theta\right)d\theta.$$

Theorem 5. *If the hypothetical model is distorted according to Condition 2 for some $\varepsilon_+ \in [0,\varepsilon^*]$, then:*

(i) for an admissible p.s. $\hat y = f(x)$ the guaranteed upper risk functional (6) is

$$r_*(f) = r_0(f) + \varepsilon_+\sqrt{\mathbf{D}_0\{r_{x,\theta}(f;x,\theta)\}}; \qquad (22)$$

(ii) the robust p.s. satisfies the following integral equation:

$$f_*(x) = \frac{1}{\int_\Theta v^{*\varepsilon_+}(f_*;x,\theta)d\theta}\int_\Theta \mathbf{E}_0\{y \mid x,\theta\}v^{*\varepsilon_+}(f_*;x,\theta)d\theta, \quad x\in X. \qquad (23)$$

Proof. **(i)** 1) Consider the maximization problem of variation calculus

$$\begin{cases} r_\varepsilon(f(\cdot);v^\varepsilon(\cdot)) \longrightarrow \max\limits_{v^\varepsilon(\cdot),\ 0\le\varepsilon\le\varepsilon_+}, \\ \int\limits_X\int\limits_\Theta v^\varepsilon(x,\theta)d\theta dx = 1, \\ \int\limits_X\int\limits_\Theta \frac{\left(v^\varepsilon(x,\theta)-v^0(x,\theta)\right)^2}{v^0(x,\theta)}d\theta dx = \varepsilon^2, \end{cases} \qquad (24)$$

$$v^\varepsilon(x,\theta)\ge 0,\ x\in X,\ \theta\in\Theta, \qquad (25)$$

which is equivalent to obtain the guaranteed upper risk.

To solve (24), for a fixed distortion level $\varepsilon\in[0,\varepsilon_+]$, let us use the Lagrange method with undefined multipliers λ_1, λ_2:

$$F(v^\varepsilon(x,\theta);\lambda_1,\lambda_2) = \int\limits_X\int\limits_\Theta v^\varepsilon(x,\theta)r_{x,\theta}(f;x,\theta)d\theta dx +$$

$$+\lambda_1\left(\int\limits_X\int\limits_\Theta v^\varepsilon(x,\theta)d\theta dx - 1\right) + \lambda_2\left(\int\limits_X\int\limits_\Theta \frac{\left(v^\varepsilon(x,\theta)-v^0(x,\theta)\right)^2}{v^0(x,\theta)}d\theta dx - \varepsilon^2\right). \qquad (26)$$

To satisfy the necessity condition, for an arbitrary variation $\delta v^{\varepsilon}(x,\theta)$, we get

$$\int\limits_X \int\limits_\Theta \Big(r_{x,\theta}(f;x,\theta)d\theta + \lambda_1 + 2\lambda_2 \cdot \frac{v^{\varepsilon}(x,\theta) - v^0(x,\theta)}{v^0(x,\theta)} \Big) \delta v^{\varepsilon}(x,\theta) d\theta dx \equiv 0.$$

The extremum function is

$$v^{*\varepsilon}_{\lambda_1,\lambda_2}(x,\theta) = v^0(x,\theta) \left(1 - \frac{1}{2\lambda_2}(\lambda_1 + r_{x,\theta}(f;x,\theta)) \right), \; x \in X, \; \theta \in \Theta. \tag{27}$$

Using the restrictions in (24), we have

$$\begin{aligned} \lambda_1^* &= -\int\limits_X \int\limits_\Theta v^0(x,\theta) r_{x,\theta}(f;x,\theta) d\theta dx, \\ (\lambda_2^*)^{-1} &= -2\varepsilon \left(\int\limits_X \int\limits_\Theta v^0(x,\theta) \left(\lambda_1^* + r_{x,\theta}(f;x,\theta) \right)^2 d\theta dx \right)^{-\frac{1}{2}}. \end{aligned} \tag{28}$$

Putting (28) into (27), we come to (21).

2) For a fixed p.s. $f(\cdot)$, if $\varepsilon \in [0, \varepsilon^*]$, then (25) holds, and (21) is a p.d.f.

3) Maximizing the objective function in (24) w.r.t. $\varepsilon \in [0, \varepsilon_+]$, we come to (22).

(ii) Minimizing the guaranteed upper risk

$$r_*(f) = \int\limits_X \int\limits_\Theta v^0(x,\theta) r_{x,\theta}(f;x,\theta) d\theta dx +$$
$$\varepsilon_+ \sqrt{\int\limits_X \int\limits_\Theta v^0(x,\theta) \left(r_{x,\theta}(f;x,\theta) - \int\limits_X \int\limits_\Theta v^0(\tilde{x},\tilde{\theta}) r_{x,\theta}(f;\tilde{x},\tilde{\theta}) d\tilde{\theta} d\tilde{x} \right)^2 d\theta dx} \longrightarrow \inf_{f(\cdot)},$$

with the Lagrange method, we get for any $\delta f(\cdot)$

$$\int\limits_X \int\limits_Y \int\limits_\Theta (f(x) - y) g^0(y \mid x,\theta) v^0(x,\theta) \left(1 + \frac{\varepsilon_+ \overset{\circ}{r}_{x,\theta}(f;x,\theta)}{\sqrt{\mathbf{D}_0\{r_{x,\theta}(f;x,\theta)\}}} \right) \delta f(x) d\theta dy dx \equiv 0,$$

which leads to (23). □

Note. *The equation (23) does not give an analytic expression for the robust p.s. Nevertheless, it gives the possibility to construct iteration procedures for its computing.*

Corollary 4. *Under the Theorem 5 conditions the one-step approximation for the robust p.s. is*

$$f_{(1)}(x) = \frac{1}{\int_\Theta v^{*\varepsilon_+}(f_0;x,\theta)d\theta} \int_\Theta \mathbf{E}_0\{y \mid x,\theta\} v^{*\varepsilon_+}(f_0;x,\theta) d\theta, \;\; x \in X. \tag{29}$$

Corollary 5. *The guaranteed upper risk functional (22) is non-decreasing function w.r.t. the parameter* $\varepsilon_+ \in [0, \varepsilon^*]$.

Theorem 6. *If the distortions of the hypothetical model are given by (19) for some* $\varepsilon_+ \in [0, \inf_{f(\cdot)\in F} \varepsilon^*]$ *and* $\mathrm{mes}_T(\mathrm{supp}\, R(\cdot)) > 0$, *then* *(i) the absolute deviation of one-step approximation (29) from the Bayesian p.s. (4) is of the first order w.r.t. ε_+:*

$$|f_{(1)}(x) - f_0(x)| = \mathcal{O}(\varepsilon_+), \ \ x \in X; \tag{30}$$

(ii) the profit in the guaranteed upper risk value is of the second order:

$$r_*(f_{(1)}) - r_*(f_0) = \mathcal{O}(\varepsilon_+^2) < 0. \tag{31}$$

Proof. **(i)** Let $D_0 = \mathbf{D}_0\{r_{x,\theta}(f_0; x, \theta)\}$. First, we get the asymptotic expansion:

$$\begin{gathered}\left(\int\limits_{\Theta} v^0(x,\theta)\left(1+\varepsilon_+\frac{\overset{\circ}{r}_{x,\theta}(f_0;x,\theta)}{\sqrt{D_0}}\right)d\theta\right)^{-1} = \\ \frac{1}{p^0(x)}\Bigg(1-\frac{\varepsilon_+}{\sqrt{D_0}}\int\limits_{\Theta}\overset{\circ}{r}_{x,\theta}\,(f_0;x,\theta)\pi^0(\theta\mid x)d\theta + \\ \frac{\varepsilon_+^2}{D_0}\left(\int\limits_{\Theta}\overset{\circ}{r}_{x,\theta}\,(f_0;x,\theta)\pi^0(\theta\mid x)d\theta\right)^2 + \mathcal{O}(\varepsilon_+^3)\Bigg), \ x \in X.\end{gathered} \tag{32}$$

Putting (32) into (29), we find

$$\begin{gathered} f_{(1)}(x) = \int\limits_{\Theta}\int\limits_{Y} yh^0(y,\theta\mid x)\left(1+\varepsilon_+\frac{\overset{\circ}{r}_{x,\theta}(f_0;x,\theta)}{\sqrt{D_0}}\right)\times \\ \Bigg(1-\frac{\varepsilon_+}{\sqrt{D_0}}\int\limits_{\Theta}\overset{\circ}{r}_{x,\theta}\,(f_0;x,\tilde{\theta})\pi^0(\tilde{\theta}\mid x)d\tilde{\theta} + \\ \frac{\varepsilon_+^2}{D_0}\left(\int\limits_{\Theta}\overset{\circ}{r}_{x,\theta}\,(f_0;x,\tilde{\theta})\pi^0(\tilde{\theta}\mid x)d\tilde{\theta}\right)^2 + \mathcal{O}(\varepsilon_+^3)\Bigg)dyd\theta. \end{gathered} \tag{33}$$

After simplifying, we come to

$$\begin{gathered} f_{(1)}(x) - f_0(x) = \varepsilon_+ \cdot \frac{1}{\sqrt{D_0}}R(x) - \\ \varepsilon_+^2 \cdot \frac{1}{D_0}\int\limits_{\Theta}\overset{\circ}{r}_{x,\theta}\,(f_0;x,\theta)\pi^0(\theta\mid x)d\theta R(x) + \mathcal{O}(\varepsilon_+^3), \ x \in X, \end{gathered} \tag{34}$$

which provides (30).

(ii) The second statement can be proven in two steps. 1) We construct the asymptotic expansion w.r.t. ε_+ for the functional $r_{x,\theta}(f_{(1)}; x, \theta)$ using the expansion of $f_{(1)}(x)$ obtained in (i):

$$\begin{gathered} r_{x,\theta}(f_{(1)};x,\theta) = r_{x,\theta}(f_0;x,\theta) + 2\varepsilon_+\cdot\frac{R(x)(f_0(x)-\mathbf{E}_0\{y|x,\theta\})}{\sqrt{D_0}} + \\ \varepsilon_+^2\frac{1}{D_0}\left(R^2(x) + 2R(x)\left(F_0(x,\theta)-f_0(x)\right)\int\limits_{\Theta}\overset{\circ}{r}_{x,\tilde{\theta}}\,(f_0;x,\tilde{\theta})\pi^0(\tilde{\theta}\mid x)d\tilde{\theta}\right) + \mathcal{O}(\varepsilon_+^3). \end{gathered} \tag{35}$$

2) Using the result from the first stage we obtain the asymptotic expansion for $r_*(f_{(1)})$. Using (35) and the risk functional definition, we get

$$r_0(f_{(1)}(\cdot)) = r_0(\cdot) + \frac{\varepsilon_+^2}{D_0}\int\limits_{X} p^0(x)R^2(x)dx + \mathcal{O}(\varepsilon_+^3). \tag{36}$$

Introduce the notation:

$$\Delta_1(f_{(1)}(\cdot), f_0(\cdot)) = r_0(f_{(1)}(\cdot)) - r_0(f_0(\cdot)), \tag{37}$$

$$\Delta_2(f_{(1)}(\cdot), f_0(\cdot)) = \varepsilon_+ \left(\sqrt{\mathbf{D}_0\{r_{x,\theta}(f_{(1)}; x,\theta)\}} - \sqrt{D_0}\right), \tag{38}$$

$$Q(x,\theta) = R(x)(f_0(x) - \mathbf{E}_0\{y \mid x,\theta\}),\ x \in X,\ \theta \in \Theta,$$

$$\overset{\circ}{Q}(x,\theta) = Q(x,\theta) - \int_X \int_\Theta v^0(\tilde{x},\tilde{\theta}) Q(\tilde{x},\tilde{\theta}) d\tilde{\theta} d\tilde{x}.$$

According to Theorem 5,

$$r_*(f_{(1)}(\cdot)) - r_*(f_0(\cdot)) = \Delta_1(f_{(1)}(\cdot), f_0(\cdot)) + \Delta_2(f_{(1)}(\cdot), f_0(\cdot)). \tag{39}$$

Putting (36) into (37), we have

$$\Delta_1(f_{(1)}(\cdot), f_0(\cdot)) = \frac{\varepsilon_+^2}{D_0} \int_X p^0(x) R^2(x) dx + \mathcal{O}(\varepsilon_+^3). \tag{40}$$

By the definition

$$\mathbf{D}_0\{r_{x,\theta}(f_{(1)}; x,\theta)\} = \int_X \int_\Theta v^0(x,\theta)(\overset{\circ}{r}_{x,\theta}(f_{(1)}; x,\theta))^2 d\theta dx. \tag{41}$$

After using (35) and performing transformations,

$$\overset{\circ}{r}_{x,\theta}(f_{(1)}; x,\theta) = \overset{\circ}{r}_{x,\theta}(f_0; x,\theta) + 2\varepsilon_+ \cdot \frac{1}{\sqrt{D_0}} \overset{\circ}{Q}(x,\theta) + \mathcal{O}(\varepsilon_+^2),\ x \in X,\ \theta \in \Theta. \tag{42}$$

Putting (42) into (41), we get

$$\mathbf{D}_0\{r_{x,\theta}(f_{(1)}; x,\theta)\} = D_0 +$$
$$4\varepsilon_+ \cdot \tfrac{1}{\sqrt{D_0}} \overset{\circ}{r}_{x,\theta}(f_0; x,\theta) \int_X \int_\Theta v^0(x,\theta) \overset{\circ}{r}_{x,\theta}(f_0; x,\theta) \overset{\circ}{Q}(x,\theta) d\theta dx + \mathcal{O}(\varepsilon_+^2),$$

and, after simplifying,

$$\mathbf{D}_0\{r_{x,\theta}(f_{(1)}; x,\theta)\} = D_0 - 4\varepsilon_+ \cdot \frac{1}{\sqrt{D_0}} \int_X p^0(x) R^2(x) dx + \mathcal{O}(\varepsilon_+^2). \tag{43}$$

Using (43) in (38), we have

$$\Delta_2(f_{(1)}(\cdot), f_0(\cdot)) = -2\varepsilon_+^2 \cdot \frac{1}{\sqrt{D_0}} \mathbf{E}_0\{R^2(x)\} + \mathcal{O}(\varepsilon_+^3). \tag{44}$$

Finally, putting (40), (44) into (39), we come to

$$r_*(f_{(1)}) - r_*(f_0) = -\varepsilon_+^2 \cdot \frac{\mathbf{E}_0\{R^2(x)\}}{D_0} + \mathcal{O}(\varepsilon_+^3).$$

□

Metrices, which are used here for the description of deviations of probability distributions, including the weighted C-metric, and the χ^2-metric, are discussed in Borovkov (1997). Also a discussion on different concepts of variation of hypothetical distributions in Bayesian statistical analysis is presented in Meczarski (1998).

6 Conclusion

The following main results have been obtained in the paper:

(i) The analytic expressions for the guaranteed upper risk and the upper risk functionals are obtained under distortions of prior and joint p.d.f.s in the weighted C-metric. For this type of distortions the r_+-robustness property of the Bayesian prediction statistic have been proven.

(ii) Under distortions in χ^2-metric the expression for the guaranteed upper risk is obtained, and the integral equation for the robust p.s. is derived; the one-step approximation of the robust p.s. is constructed and the asymptotic properties of this approximation are obtained.

Acknowledgments

The author is very thankful to Professor Rudolf Dutter for the invitation to submit the paper based on the talk at IDA 2000 Conference and to Professor Reinhard Viertl: without his help it would have been not possible to take part in IDA 2000.

The author is also thankful to the Referees for useful comments which helped to improve the paper.

The research was partially supported by the Belarussian National Science Foundation, grant No. F99M-108, and by INTAS, grant No. YSF-01/1-75.

References

V.M. Alexeev. *Optimal control*. Science, Moscow, 1979. In Russian.

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

J. Berger, B. Betro, and E. Moreno et al., editors. *Bayesian Robustness*, volume 29 of *IMS Lecture Notes — Monograph Series*. IMS, Hayward, 1996.

A.A. Borovkov. *Mathematical Statistics*. Science, Novosibirsk, 1997. In Russian.

R. Dutter. On robust estimation of variograms in geostatistics. In *Robust Statistics, Data Analysis, and Computer Intensive Methods*, volume 109 of *Lecture Notes in Statistics*, pages 153–172. Springer, New York, 1996.

V. Galinskij and A. Kharin. On minimax robustness of Bayesian statistical prediction. In Grigelionis B., editor, *Probability Theory and Mathematical Statistics*, pages 259–266, Vilnius/Utrecht/Tokio, 1998. VSP/TEV.

F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W. Stahel. *Robust Statistics. The Approach Based on Influence Functions*. Wiley & Sons, New York, 1986.

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

M. Meczarski. *Problemy Odpornosci w Bayesowskiej Analizie Statystycznej*. Szkola Glowna Handlowa, Warszawa, 1998. In Polish.

J. Pilz. Bayesian spatial prediction using the matern class of covariance functions. In *International Data Analysis Conference 2000: Abstracts*, Innsbruck, 2000.

H. Rieder. *Robust Asymptotic Statistics*. Springer–Verlag, New York, 1994.

R. Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton, 1996.

M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer–Verlag, New York, 1989.

Author's address:

Assistant Prof. Dr. Alexey Kharin
Department of Probability Theory & Mathematical Statistics
Belarussian State University
Fr. Skorina av., 4
220050 Minsk
Belarus

Tel. +375 17 2095070
Fax +375 17 2265548
E-mail: KharinAY@bsu.by
http://www.bsu.by/fpmi/