Gamma-minimax prediction in exponential families with quadratic variance function

Agata Boratyńska

Institute of Mathematics Polish Acad. Sc.
Department of Mathematical Statistics
E-MAIL: aborata@sgh.waw.pl

Abstract. *The problem of prediction of a random variable Y, distribution of which belongs to the one parameter exponential family with the quadratic variance function is considered. The knowledge about priors is introduced by a class Γ of all distributions, where the first two moments are within some given convex and compact set. The Γ-minimax predictor under the squared error loss is obtained.*

1. Introduction and notation. The problem considered in this paper belongs to a class of estimation problems for which the aim is to predict the value of a random variable Y on the basis of the observation of a random variable $X = (X_1, X_2, \ldots, X_n)$, where X and Y have a distribution dependent on the same unknown parameter θ. We consider the quadratic loss function

$$L(y, a) = (y - a)^2,$$

where y is a value of Y and a is a value of a predictor.

A gamma-minimax approach is used which allows to take into account vague prior information on the distribution of the unknown parameter θ. The uncertainty about a prior is assumed by introducing a class Γ of priors. If prior information is scarce, the class Γ under consideration is large and a decision is close to a minimax decision. In the extreme case when no information is available the Γ-minimax setup is equivalent to the usual minimax setup. If, on the other hand, the statistician has an exactly prior information and the class Γ contains a single prior, then the Γ-minimax decision is an usual Bayes decision. So it is a middle ground between the subjective Bayes setup and full minimax. For a discussion on the Γ-minimax approach see for example Berger (1985), Brown (1994), Strawderman (2000), Vidakovic (2000).

In this paper a class $\Gamma_{\mathcal{G}}$ of priors consists of all distributions, where the first two moments are within some given convex and compact set $\mathcal{G}$. We deal with a one-parameter exponential family with quadratic variance functions, which has been characterized by Morris (1982). The popular exponential families like normal, Poisson, Gamma, binomial are examples. This family and problems of estimation of a parameter were considered by Eichenauer-Herrmann (1991), Boratyńska (1997) among others.

Let v_1, v_2, v_3 be fixed real numbers such that the set $\Theta^+ = \{\theta \in \mathcal{R} : v_1\theta^2 + v_2\theta + v_3 > 0\}$ is non-empty. Let h and q be continuously differentiable real-valued functions satisfying the differential equations

$$\frac{h'(\theta)}{h(\theta)} = \frac{-\theta}{v_1\theta^2 + v_2\theta + v_3} \qquad \text{and} \qquad q'(\theta) = \frac{1}{v_1\theta^2 + v_2\theta + v_3}.$$

Let Θ be an interval (θ_0, θ_1) such that $\Theta \subset \{\theta \in \Theta^+ : \ h(\theta) > 0\}$. Let $\{P_\theta : \ \theta \in \Theta\}$ be a one parameter exponential family of probability measures on $\mathcal{R}$ with densities of the form

$$l(z|\theta) = B(z)h(\theta)e^{q(\theta)z}, \quad z \in \mathcal{R}$$

with respect to some σ-finite measure on $\mathcal{R}$. Note that if Z has a distribution P_θ then the expected value is $E_\theta Z = \theta$ and the variance is $Var_\theta Z = v_1\theta^2 + v_2\theta + v_3$.

Let $X_1, X_2, \ldots, X_n, Y$ be i.i.d. random variables with a distribution P_θ. The vector $X = (X_1, X_2, \ldots, X_n)$ is observed. A random variable Y is predicted under the quadratic loss function, thus the risk function of a predictor $d = d(X)$ is equal

$$R(\theta, d) = E_\theta(Y - d(X))^2,$$

where the operator E_θ emphasizes the expectation with respect to the joint probability distribution of random variables X and Y, if the value of a parameter is θ.

Let

$$\mathcal{M} = \left\{m = (m_1, m_2) \in \mathcal{R}^2 : \ m_1 \in \Theta \wedge m_2 \geq m_1^2 \wedge v_1 m_2 + v_2 m_1 + v_3 > 0\right\}$$

and $\mathcal{G}$ be non-empty convex and compact subset of $\mathcal{M}$. Let Θ^* be a family of all probability distributions on Θ and

$$\Gamma_{\mathcal{G}} = \left\{\Pi \in \Theta^* : E_\Pi \theta = m_1 \wedge E_\Pi \theta^2 = m_2 \wedge m = (m_1, m_2) \in \mathcal{G}\right\}$$

be a family of priors of a parameter θ. If $m_2 = m_1^2$, then Π is a one-point measure on m_1. The Bayes risk of a predictor d under a prior Π is

$$r(\Pi, d) = E(Y - d(X_1, X_2, \ldots, X_n))^2 = EVar(Y|\theta) + E(d - E(Y|\theta))^2$$

$$= v_1 m_2 + v_2 m_1 + v_3 + E(d - E(Y|\theta))^2,$$

where $E_\Pi \theta = m_1$, $E_\Pi \theta^2 = m_2$ and the operator E emphasizes the expectation with respect to the joint probability distribution of all random variables θ, X, Y.

Our objective is to find the $\Gamma_{\mathcal{G}}$-minimax predictor, thus the predictor $d_{\Gamma_{\mathcal{G}}}$ satisfying

$$\inf_{d \in D} \sup_{\Pi \in \Gamma_{\mathcal{G}}} r(\Pi, d) = \sup_{\Pi \in \Gamma_{\mathcal{G}}} r(\Pi, d_{\Gamma_{\mathcal{G}}}),$$

where D is a class of all predictors d (measurable functions of the observed random variables) with finite risk function $R(\theta, d)$.

2. The $\Gamma_{\mathcal{G}}$-minimax predictor. The probability distribution on the interval Θ with the Lebesgue density

$$\pi_{\alpha,\beta}(\theta) = C^{-1} h^{\alpha}(\theta) e^{\beta q(\theta)}, \qquad \theta \in \Theta,$$

where $C = \int_{\Theta} h^{\alpha}(\theta) e^{\beta q(\theta)} d\theta$ and α, β are real parameters satisfying

$$\alpha > 3v_1 \quad \text{and} \quad \beta > \theta_0(\alpha - 2v_1) - v_2 \quad \text{and} \quad \beta < \theta_1(\alpha - 2v_1) - v_2$$

(see Chen i in. (1991)) is a conjugate prior. For every pair $(m_1, m_2) \in \mathcal{M}$, $m_2 > m_1^2$, there exists a conjugate prior $\Pi_{\alpha,\beta}$ satisfying conditions $E_{\Pi_{\alpha,\beta}}\theta = m_1$ i $E_{\Pi_{\alpha,\beta}}\theta^2 = m_2$, and α and β are equal

$$\alpha = 2v_1 + \frac{v_1 m_2 + v_2 m_1 + v_3}{m_2 - m_1^2},$$

$$\beta = -v_2 + m_1 \frac{v_1 m_2 + v_2 m_1 + v_3}{m_2 - m_1^2}$$

(see Lemmas 3 and 4 in Chen (1991)). Denote the conjugate prior corresponding to $m = (m_1, m_2)$ as Π_m, note, that if $m_2 = m_1^2$ then Π_m satisfies $\Pi_m\{m_1\} = 1$. If $X = x$ then the Bayes predictor with respect to the prior Π_m is equal to the Bayes estimator of a parameter θ under the quadratic loss function and for $m_2 > m_1^2$

$$d_m^B = E_{\Pi_m}(\theta|x) = \frac{\beta + v_2 + \sum_{i=1}^n x_i}{\alpha + n - 2v_1}$$

$$= \frac{m_1(v_1 m_2 + v_2 m_1 + v_3) + (m_2 - m_1^2)\sum_{i=1}^n x_i}{v_1 m_2 + v_2 m_1 + v_3 + n(m_2 - m_1^2)}.$$

Its Bayes risk under a prior Π is equal

$$r(\Pi, d_m^B) = \Xi(m)\left[\left(\Omega(m) + \frac{v_1}{\Xi(m)}\right) E_{\Pi}\theta^2 + \left(\Phi(m) + \frac{v_2}{\Xi(m)}\right) E_{\Pi}\theta + \Lambda(m)\right] + v_3,$$

where

$$\Omega(m) = v_1 n(m_2 - m_1^2)^2 + (v_1 m_2 + v_2 m_1 + v_3)^2,$$

$$\Phi(m) = v_2 n(m_2 - m_1^2)^2 - 2m_1(v_1 m_2 + v_2 m_1 + v_3)^2,$$

$$\Lambda(m) = v_3 n(m_2 - m_1^2)^2 + m_1^2(v_1 m_2 + v_2 m_1 + v_3)^2,$$

$$\Xi(m) = (n(m_2 - m_1^2) + v_1 m_2 + v_2 m_1 + v_3)^{-2},$$

and

$$r(\Pi_m, d_m^B) = \frac{(v_1 m_2 + v_2 m_1 + v_3)((n+1)(m_2 - m_1^2) + v_1 m_2 + v_2 m_1 + v_3)}{n(m_2 - m_1^2) + v_1 m_2 + v_2 m_1 + v_3}.$$

Let

$$\rho(m) = \frac{1}{r(\Pi_m, d_m^B)}$$

$$= \frac{1}{n+1}\left(\frac{n}{v_1 m_2 + v_2 m_1 + v_3} + \frac{1}{(n+1)(m_2 - m_1^2) + v_1 m_2 + v_2 m_1 + v_3}\right).$$

Straightforward calculations show that ρ is strictly convex.

To find $\Gamma_{\mathcal{G}}$-minimax predictor we will apply two following theorems: the first one is the well known theorem about a saddle point in the statistical game $(\Gamma_{\mathcal{G}}, D, r)$, and the second one is the Fan theorem about minimax inequality. The method of the proof of Theorem 3 is similar to the method presented in Eichenauer-Herrmann (1991).

Theorem 1. *The following conditions are equivalent:*
a) (Π^*, d^*) *is a saddle point in* $(\Gamma_{\mathcal{G}}, D, r)$*;*
b) $\inf_{d\in\mathcal{D}} r(\Pi^*, d) \geq \sup_{\Pi\in\Gamma_{\mathcal{G}}} r(\Pi, d^*)$*;*
c) (Π^*, d^*) *is a saddle point and* d^* *is a Bayes decision rule with respect to the prior* Π^**;*
d) Π^* *is the least favourable prior,* d^* *is the* Γ*-minimax and the Bayes decision rule with respect to* Π^* *and the game* $(\Gamma_{\mathcal{G}}, D, r)$ *has a value.* ■

Theorem 2. (Fan (1972)) *Let* $\mathcal{X}$ *be a compact convex set in a Hausdorff topological vector space. Let* f *be a real valued function defined on* $\mathcal{X} \times \mathcal{X}$ *such that:*
(a) for every $x \in \mathcal{X}$, $f(x, y)$ *is a continuous function of* y,
(b) for every $y \in \mathcal{X}$, $f(x, y)$ *is a quasi-concave function of* x.

Then

$$\min_{y\in\mathcal{X}} \sup_{x\in\mathcal{X}} f(x, y) \leq \sup_{x\in\mathcal{X}} f(x, x).$$

■

We prove the following theorem.

Theorem 3.

Let $\tilde{m} \in \mathcal{G}$. Then $(\Pi_{\tilde{m}}, d^B_{\tilde{m}})$ is a saddle point in the statistical game $(\Gamma_{\mathcal{G}}, D, r)$ iff

$$\rho(\tilde{m}) = \min_{m \in \mathcal{G}} \rho(m). \tag{$*$}$$

There exists exactly one $\tilde{m} \in \mathcal{G}$ satisfying $(*)$ and the predictor $d^B_{\tilde{m}}$ is the $\Gamma_{\mathcal{G}}$-minimax predictor.

Proof. $(\Longrightarrow)$ If $(\Pi_{\tilde{m}}, d^B_{\tilde{m}})$ is a saddle point, then

$$r(\Pi_{\tilde{m}}, d^B_{\tilde{m}}) = \sup_{\Pi \in \Gamma_{\mathcal{G}}} \inf_{d \in D} r(\Pi, d) \geq \sup_{m \in \mathcal{G}} r(\Pi_m, d^B_m) \geq r(\Pi_{\tilde{m}}, d^B_{\tilde{m}}).$$

Since

$$\rho(m) = \frac{1}{r(\Pi_m, d^B_m)},$$

thus we obtain $(*)$.

$(\Longleftarrow)$ First we show that there exists $\tilde{m}$ such that $(\Pi_{\tilde{m}}, d^B_{\tilde{m}})$ is a saddle point. Let

$$E_\Pi \theta = \mu_1, \qquad E_\Pi \theta^2 = \mu_2, \qquad \mu = (\mu_1, \mu_2),$$

$$Z(m, \mu) = \left(\Omega(m) + \frac{v_1}{\Xi(m)}\right) \mu_2 + \left(\Phi(m) + \frac{v_2}{\Xi(m)}\right) \mu_1.$$

Then

$$r(\Pi, d^B_m) = \Xi(m) \left(Z(m, \mu) + \Lambda(m)\right) + v_3$$

and (Π_m, d^B_m) is a saddle point iff

$$r(\Pi_m, d^B_m) = \sup_{\Pi \in \Gamma_{\mathcal{G}}} r(\Pi, d^B_m) = \Xi(m) \left(\max_{\mu \in \mathcal{G}} Z(m, \mu) + \Lambda(m)\right) + v_3$$

$$\Longleftrightarrow Z(m, m) = \max_{\mu \in \mathcal{G}} Z(m, \mu). \tag{$**$}$$

The existence of a solution of the equation $(**)$ follows from Theorem 2. The set $\mathcal{G}$ is non-empty, convex and compact. Define $f : \mathcal{G} \times \mathcal{G} \longrightarrow R$ and $f(m, \mu) = Z(m, \mu) - Z(m, m)$. The function f satisfies conditions (a) and (b) of Theorem 2 and $f(m, m) = 0$. Thus there exists $\tilde{m}$ such that $\sup_{\mu \in \mathcal{G}} f(\tilde{m}, \mu) \leq 0$, hence $\max_{\mu \in \mathcal{G}} Z(\tilde{m}, \mu) = Z(\tilde{m}, \tilde{m})$.

Assume that m_1 satisfies $(*)$. The pair $(\Pi_{\tilde{m}}, d^B_{\tilde{m}})$ is a saddle point, hence $\tilde{m}$ satisfies $(*)$ too, and the function ρ has only one minimum on $\mathcal{G}$ (ρ is strictly convex). Thus $m_1 = \tilde{m}$. The remaining part of the proof follows from Theorem 1. ■

Example. Let $X_1, X_2, \ldots, X_n, Y$ be i.i.d. random variables from Poisson distribution $Poiss(\theta)$. Then $v_1 = v_3 = 0$ and $v_2 = 1$. Let $\mathcal{G} = [m_{1,1}, m_{1,2}] \times [m_{2,1}, m_{2,2}] \cap \mathcal{M}$ and

$$0 < m_{1,1} < m_{1,2} < +\infty, \qquad 0 < m_{2,1} < m_{2,2} < +\infty, \qquad m_{1,1} < \sqrt{m_{2,2}}.$$

The function ρ is equal

$$\rho(m_1, m_2) = \frac{1}{n+1}\left(\frac{n}{m_1} + \frac{1}{(n+1)(m_2 - m_1^2) + m_1}\right)$$

and

$$\frac{\partial}{\partial m_1}\rho(m_1, m_2) = \frac{1}{n+1}\left(-\frac{n}{m_1^2} + \frac{2(n+1)m_1 - 1}{((n+1)(m_2 - m_1^2) + m_1)^2}\right),$$

$$\frac{\partial}{\partial m_2}\rho(m_1, m_2) = -\frac{1}{((n+1)(m_2 - m_1^2) + m_1)^2} < 0$$

and $\frac{\partial \rho}{\partial m_1}$ is an increasing function of m_1. Table 1 present $\Gamma_{\mathcal{G}}$-minimax predictors for different values of $m_{i,j}$, $i, j = 1, 2$.

Table 2 presents the $\Gamma_{\mathcal{G}}$-minimax estimators of a parameter θ in the same model (for details see Chen (1991)). Note that the Bayes estimator of θ and the Bayes predictor of Y are equal, but the $\Gamma_{\mathcal{G}}$-minimax rules are not equal. They are equal if $m_{1,1} = m_{1,2}$. It is a situation, when we know the expected value of a prior and the variance belongs to the fixed interval.

Table 1.

Conditions for $m_{i,j},\ i,j=1,2$	$d_{\Gamma_{\mathcal{G}}}$
$\frac{(2(n+1)m_{1,1}-1)m_{1,1}^2}{((n+1)(m_{2,2}-m_{1,1}^2)+m_{1,1})^2} \geq n$	$\frac{m_{1,1}^2+(m_{2,2}-m_{1,1}^2)\sum_{i=1}^n x_i}{m_{1,1}+n(m_{2,2}-m_{1,1}^2)}$
$\frac{(2(n+1)m_{1,2}-1)m_{1,2}^2}{((n+1)(m_{2,2}-m_{1,2}^2)+m_{1,2})^2} \leq n$ and $m_{2,2} \geq m_{1,2}^2$	$\frac{m_{1,2}^2+(m_{2,2}-m_{1,2}^2)\sum_{i=1}^n x_i}{m_{1,2}+n(m_{2,2}-m_{1,2}^2)}$
$\exists \nu \in [m_{1,1}, \min(m_{1,2}, \sqrt{m_{2,2}})]$ such that $\frac{(2(n+1)\nu-1)\nu^2}{((n+1)(m_{2,2}-\nu^2)+\nu)^2} = n$	$\frac{\nu^2+(m_{2,2}-\nu^2)\sum_{i=1}^n x_i}{\nu+n(m_{2,2}-\nu^2)}$
$m_{2,2} < 0.25$ and $m_{2,2} < m_{1,2}^2$	$\sqrt{m_{2,2}}$

Table 2.

Conditions for $m_{i,j},\ i,j=1,2$	$\Gamma_{\mathcal{G}}$-minimax estimator of θ
$\frac{n}{2} \leq m_{1,1}(\frac{m_{1,1}}{m_{2,2}-m_{1,1}^2})^2$	$\frac{m_{1,1}^2+(m_{2,2}-m_{1,1}^2)\sum_{i=1}^n x_i}{m_{1,1}+n(m_{2,2}-m_{1,1}^2)}$
$\frac{n}{2} \geq m_1^2(\frac{m_{1,2}}{m_{2,2}-m_{1,2}^2})^2$	$\frac{m_{1,2}^2+(m_{2,2}-m_{1,2}^2)\sum_{i=1}^n x_i}{m_{1,2}+n(m_{2,2}-m_{1,2}^2)}$
$\exists \nu \in [m_{1,1}, \min(m_{1,2}, \sqrt{m_{2,2}})]$ such that $n(m_{2,2}-\nu^2)^2 = 2\nu^3$	$\frac{\nu^2+(m_{2,2}-\nu^2)\sum_{i=1}^n x_i}{\nu+n(m_{2,2}-\nu^2)}$

References

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

Boratyńska, A. (1997), Stability of Bayesian inference in exponential families, *Statist. Prob. Letters* **36**, 173-178.

Brown, L.D. (1994), Minimaxity, more or less, In *Statistical Decision Theory and Related Topics* **5**, S.S. Gupta i J.O. Berger (Eds.), Springer-Verlag, New York, 1-18.

Chen, L., Eichenauer-Hermann, J., Hofmann, H. and Kindler, J. (1991), Gamma-minimax estimators in the exponential family, *Diss. Math.* **308**.

Eichenauer-Herrmann, J. (1991), Gamma-minimax estimation in exponential families with quadratic variance functions, *Statist. and Decisions* **9**, 319-326.

Fan Ky (1972), A minimax inequality and applications, In Shisha O. (Ed.) *Inequalities III,* Academic Press, New York and London, 103-113.

Morris, C.N. (1982), Natural exponential families with quadratic variance functions, *Ann. Statist.* **10**, 65-80.

Strawderman, W.E. (2000), Minimaxity, *J. Amer. Statist. Assoc.* **95**, 1364-1368.

Vidakovic, B. (2000), Γ-minimax: a paradigm for conservative robust Bayesians, w *Robust Bayesian Analysis,* D.R. Insua, F. Ruggeri (red.), *Lectures Notes in Statistics,* Springer-Verlag, New York, 241-259.