APPLICATIONES MATHEMATICAE
30,1 (2003), pp. 1–10

Maciej Wilczyński (Wrocław)

GAMMA MINIMAX NONPARAMETRIC ESTIMATION

Abstract. Let Y be a random vector taking its values in a measurable space and let $\boldsymbol{z}$ be a vector-valued function defined on that space. We consider gamma minimax estimation of the unknown expected value $\boldsymbol{p}$ of the random vector $\boldsymbol{z}(Y)$. We assume a weighted squared error loss function.

1. Introduction. Let Y be a random variable (or vector) taking its values in a measurable space $(\mathcal{Y},\mathcal{B})$, whose unknown distribution P is assumed to be an element of the set

$$\mathcal{P}=\{\text{all probability measures on } (\mathcal{Y},\mathcal{B})\}.$$

Further, let $\boldsymbol{Y}^n=(Y_1,\ldots,Y_n)$ be a random sample of size n from P and let $\boldsymbol{z}=(z_1,\ldots,z_k)^T$ be a bounded, measurable function on $(\mathcal{Y},\mathcal{B})$ with values in $(\mathbb{R}^k,\mathcal{B}_{\mathbb{R}^k})$. Consider estimation of the unknown vector $\boldsymbol{p}$ defined as the expected value of the random vector $\boldsymbol{Z}=\boldsymbol{z}(Y)$, i.e.

$$\boldsymbol{p}=E_P\boldsymbol{Z}.$$

We assume that the loss function, which describes the loss to the statistician if he estimates $\boldsymbol{p}$ by $\boldsymbol{d}$, has the form

$$L(\boldsymbol{d},P)=(\boldsymbol{d}-\boldsymbol{p})^T\boldsymbol{C}(\boldsymbol{d}-\boldsymbol{p}), \tag{1}$$

where the $k\times k$ matrix $\boldsymbol{C}=[c_{ij}]$ is symmetric and nonnegative definite. To choose a reasonable decision rule $\boldsymbol{d}\in\mathcal{D}$, where

$$\mathcal{D}=\{\text{all estimators } \boldsymbol{d}=\boldsymbol{d}(\boldsymbol{Y}^n) \text{ of the unknown vector } \boldsymbol{p}\},$$

we can use different principles. If we have no prior information on the unknown probability P then we can use the minimax principle. Let $R(\boldsymbol{d},P)$ be the risk function of an estimator $\boldsymbol{d}\in\mathcal{D}$, i.e.

$$R(\boldsymbol{d},P)=E_P[L(\boldsymbol{d}(\boldsymbol{Y}^n),P)].$$

2000 *Mathematics Subject Classification*: 62C20, 62G05.
Key words and phrases: gamma minimax estimation, nonparametric estimation, Dirichlet process.

Then a decision rule $\boldsymbol{d}_0$ is said to be *minimax* if it minimizes the maximum expected loss, i.e.

$$\sup_{P\in\mathcal{P}} R(\boldsymbol{d}_0, P) = \inf_{d\in\mathcal{D}} \sup_{P\in\mathcal{P}} R(\boldsymbol{d}, P).$$

In Wilczyński (1992) it was proved that the minimax estimator of $\boldsymbol{p}$ under the loss function (1) has the form

$$\boldsymbol{d}_0(\boldsymbol{Y}^n) = \frac{\boldsymbol{X}^n + \sqrt{n}\,\boldsymbol{p}_0}{n+\sqrt{n}}, \quad \text{where} \quad \boldsymbol{X}^n = \sum_{j=1}^{n} \boldsymbol{z}(Y_j).$$

For the definition of the vector $\boldsymbol{p}_0$ see Wilczyński (1992). Note that $\boldsymbol{d}_0$ is affine (inhomogeneous linear) with respect to $\boldsymbol{X}^n$ and thus easy to evaluate and handle analytically.

If we know, on the other hand, that P is a random probability measure chosen according to a known prior distribution $\pi \in \Pi$, where

$$\Pi = \{\text{all priors on the space of all probability mesures on } (\mathcal{Y}, \mathcal{B})\},$$

we can use the (nonparametric) Bayes principle. Let $r(\boldsymbol{d}, \pi)$ be the π-*Bayes risk* of an estimator $\boldsymbol{d}$, i.e. the expected value of the risk function $R(\boldsymbol{d}, P)$ with respect to the prior π:

$$r(\boldsymbol{d}, \pi) = E_\pi R(\boldsymbol{d}, P).$$

Then a decision rule $\boldsymbol{d}_\pi$ is said to be π-Bayes if it minimizes the π-Bayes risk, i.e.

$$r(\boldsymbol{d}_\pi, \pi) = \inf_{d\in\mathcal{D}} r(\boldsymbol{d}, \pi).$$

Unfortunately, finding a workable prior distribution π defined on the space of all probability measures on a given sample space is not an easy task. Ferguson (1973) stated that there are two desirable, but antagonistic, properties of a prior distribution for nonparametric problems: its support should be large and the posterior distribution given a sample of observations from the true probability distribution should be manageable analytically. The simplest priors which have the latter property are the Dirichlet processes introduced by Ferguson (1973). There are a large number of such processes, one for each finite nonnull measure on $(\mathcal{Y}, \mathcal{B})$. Suppose that π is the Dirichlet process corresponding to a measure AQ, where A is a positive number and $Q \in \mathcal{P}$. Then, from Ferguson (1973) (example b), the π-Bayes nonparametric estimator of $\boldsymbol{p}$ has the form

$$\boldsymbol{d}_\pi(\boldsymbol{Y}^n) = \frac{\boldsymbol{X}^n + A\boldsymbol{q}}{n+A},$$

where $\boldsymbol{q} = E_Q \boldsymbol{Z}$.

There is an intermediate approach between the Bayes and the minimax principles, the Γ-*minimax principle*, which is appropriate in the following situation. Suppose that P is a random probability measure chosen according to an unknown prior distribution π which belongs to a given subset Γ of Π. Then a decision rule $\boldsymbol{d}_\Gamma$ is said to be *Γ-minimax* if it minimizes the maximum Bayes risk with respect to the elements of Γ, i.e.

$$\sup_{\pi\in\Gamma} r(\boldsymbol{d}_\Gamma,\pi) = \inf_{d\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\boldsymbol{d},\pi). \tag{2}$$

In this paper we consider Γ-minimax estimation of an unknown vector $\boldsymbol{p}$ under the loss function (1). We assume that the set Γ has the form

$$\Gamma = \{\pi \in \Pi : (\boldsymbol{\nu}_1(\pi), \nu_2(\pi)) \in \mathcal{G}\},$$

where $\mathcal{G}$ is a given convex subset of $\mathbb{R}^{k+1}$ and $\boldsymbol{\nu}_1(\pi)$ and $\nu_2(\pi)$ denote the first moments of $E_P\boldsymbol{Z}$ and $E_P\boldsymbol{Z}^T\boldsymbol{C}\boldsymbol{Z}$ with respect to the prior $\pi \in \Pi$, i.e.

$$\boldsymbol{\nu}_1(\pi) = E_\pi(E_P\boldsymbol{Z}), \quad \nu_2(\pi) = E_\pi(E_P\boldsymbol{Z}^T\boldsymbol{C}\boldsymbol{Z}).$$

We prove that the Γ-minimax estimator $\boldsymbol{d}_\Gamma$ is an affine transformation of the random vector $\boldsymbol{X}^n$.

As is well known, a decision rule which is Γ-minimax for $\Gamma = \Pi$ has another optimal property—it is also minimax. Thus, we generalize our previous result concerning minimax nonparametric estimation (cf. Wilczyński (1992)), because $\Gamma = \Pi$ when $\mathcal{G} = \mathbb{R}^{k+1}$.

The problem of Γ-minimax estimation has been considered by many authors. In particular, the case where Γ consists of all distributions whose first two moments are within some given bounds has been described by Jackson *et al.* (1970), Robbins (1964), Eichenauer *et al.* (1988) and Chen and Eichenauer-Hermann (1990). A similar set of priors has been chosen by Chen *et al.* (1991) who have explicitly determined the Γ-minimax estimator for the unknown parameter θ of a one-parameter exponential family. This general result has been obtained under the assumption that there exists an unbiased statistic for θ with variance which is quadratic in the parameter. Next, this result has been strengthened by Eichenauer (1991) who has assumed that the set Γ of priors consists of all distributions whose first two moments are within some given convex compact set $\mathcal{G}$. A set Γ determined by certain moment-type conditions has also been considered in the paper of Magiera (2001), where the aim is to estimate unknown parameters of Markov-additive processes from the data observed up to a random stopping time.

In all the references given above the problem of estimation is parametric and the observed random variable (or vector) Y has a distribution which depends on the unknown parameter θ, which takes its values in a finite-dimensional Euclidean space. In contrast, we consider the nonparametric version of the problem of Γ-minimax estimation. We assume that the unknown distribution of Y can be described by any probability measure P defined on the measurable space $(\mathcal{Y},\mathcal{B})$ in which Y takes its values.

2. Gamma minimax estimate. Since the vector-valued function $\boldsymbol{z}$ is assumed to be bounded there exists a positive number M such that $\sup_{y\in\mathcal{Y}}\|\boldsymbol{z}(y)\|\le M$, where $\|\ \|$ denotes the standard norm in $\mathbb{R}^k$. This implies that the random vector $\boldsymbol{Z}:=\boldsymbol{z}(Y)$ is bounded and takes its values in the convex compact subset $\mathcal{M}$ of $\mathbb{R}^k$ defined by

$$\mathcal{M}=\{\boldsymbol{x}\in\mathbb{R}^k:\|x\|\le M\}.$$

We denote by (π_j) a sequence of priors from Γ for which

$$\lim_{j\to\infty}(\nu_2(\pi_j)-\boldsymbol{\nu}_1^T(\pi_j)\boldsymbol{C}\boldsymbol{\nu}_1(\pi_j))=\sup_{\pi\in\Gamma}(\nu_2(\pi)-\boldsymbol{\nu}_1^T(\pi)\boldsymbol{C}\boldsymbol{\nu}_1(\pi)).\tag{3}$$

Since $\boldsymbol{Z}\in\mathcal{M}$, the corresponding sequence

$$(\boldsymbol{\nu}_1(\pi_j))=(E_{\pi_j}(E_P\boldsymbol{Z}))\tag{4}$$

takes its values in $\mathcal{M}$ and therefore has a cluster point $\boldsymbol{p}_\Gamma\in\mathcal{M}$. The following theorem is the main result of the paper:

THEOREM 1. *The Γ-minimax estimator of the unknown vector $\boldsymbol{p}$ under the loss function* (1) *has the form*

$$\boldsymbol{d}_\Gamma(\boldsymbol{Y}^n)=\frac{\boldsymbol{X}^n+\sqrt{n}\,\boldsymbol{p}_\Gamma}{n+\sqrt{n}},\tag{5}$$

and its Γ-minimax risk equals

$$\sup_{\pi\in\Gamma}r(\boldsymbol{d}_\Gamma,\pi)=\sup_{\pi\in\Gamma}\frac{\nu_2(\pi)-\boldsymbol{\nu}_1^T(\pi)\boldsymbol{C}\boldsymbol{\nu}_1(\pi)}{(\sqrt{n}+1)^2}.\tag{6}$$

Proof. We will use a method analogous to that in Wilczyński (1992). First we will show that $\boldsymbol{d}_\Gamma(\boldsymbol{Y}^n)$ is Γ-minimax if the class of estimators is restricted to a subset $\mathcal{D}_0\subset\mathcal{D}$ defined by

$$\mathcal{D}_0=\left\{\boldsymbol{d}^{\boldsymbol{b}}\in\mathcal{D}:\boldsymbol{d}^{\boldsymbol{b}}(\boldsymbol{Y}^n)=\frac{\boldsymbol{X}^n+\sqrt{n}\,\boldsymbol{b}}{n+\sqrt{n}},\ \boldsymbol{b}\in\mathcal{M}\right\}.$$

Then, using some implications of this fact, we will find the least upper bound for the Bayes risk of $\boldsymbol{d}_\Gamma(\boldsymbol{Y}^n)$. Finally, we will construct a sequence (π_j^*) of priors from Γ and a sequence of π_j^*-Bayes estimators $(\boldsymbol{d}_{\pi_j^*})$ for which the corresponding sequence of Bayes risks $(r(\boldsymbol{d}_{\pi_j^*},\pi_j^*))$ approaches that upper bound. This will complete the proof of the theorem.

We first calculate the risk function for an estimator $\boldsymbol{d}^{\boldsymbol{b}}\in\mathcal{D}_0$. We note that $\boldsymbol{z}(Y_1),\ldots,\boldsymbol{z}(Y_n)$ are i.i.d. random vectors distributed as $\boldsymbol{Z}=\boldsymbol{z}(Y)$. Therefore,

$$\begin{aligned}E_P(\boldsymbol{X}^n-n\boldsymbol{p})^T\boldsymbol{C}(\boldsymbol{X}^n-n\boldsymbol{p})&=\sum_{j=1}^n E_P(\boldsymbol{z}(Y_j)-\boldsymbol{p})^T\boldsymbol{C}(\boldsymbol{z}(Y_j)-\boldsymbol{p})\\&=nE_P(\boldsymbol{Z}-\boldsymbol{p})^T\boldsymbol{C}(\boldsymbol{Z}-\boldsymbol{p})=n(E_P\boldsymbol{Z}^T\boldsymbol{C}Z-\boldsymbol{p}^T\boldsymbol{C}\boldsymbol{p}).\end{aligned}$$

This implies that the risk function of an estimator $\boldsymbol{d}^{\boldsymbol{b}}$ from $\mathcal{D}_0$ has the form

$$R(\boldsymbol{d}^{\boldsymbol{b}}, P) = \frac{n(E_P \boldsymbol{Z}^T \boldsymbol{C}\boldsymbol{Z} - \boldsymbol{p}^T \boldsymbol{C}\boldsymbol{p}) + (\sqrt{n})^2 (\boldsymbol{b}-\boldsymbol{p})^T \boldsymbol{C}(\boldsymbol{b}-\boldsymbol{p})}{(n+\sqrt{n})^2}$$
$$= \frac{(E_P \boldsymbol{Z}^T \boldsymbol{C}\boldsymbol{Z} - \boldsymbol{p}^T \boldsymbol{C}\boldsymbol{p}) + (\boldsymbol{b}-\boldsymbol{p})^T \boldsymbol{C}(\boldsymbol{b}-\boldsymbol{p})}{(\sqrt{n}+1)^2}$$
$$= \frac{E_P \boldsymbol{Z}^T \boldsymbol{C}\boldsymbol{Z} - 2\boldsymbol{b}^T \boldsymbol{C}\boldsymbol{p} + \boldsymbol{b}^T \boldsymbol{C}\boldsymbol{b}}{(\sqrt{n}+1)^2}.$$

Moreover, for any prior $\pi \in \Pi$ the π-Bayes risk of $\boldsymbol{d}^{\boldsymbol{b}}$ is

$$r(\boldsymbol{d}^{\boldsymbol{b}}, \pi) = \frac{E_\pi E_P \boldsymbol{Z}^T \boldsymbol{C}\boldsymbol{Z} - 2\boldsymbol{b}^T \boldsymbol{C} E_\pi \boldsymbol{p} + \boldsymbol{b}^T \boldsymbol{C}\boldsymbol{b}}{(\sqrt{n}+1)^2} \tag{7}$$
$$= \frac{\nu_2(\pi) - 2\boldsymbol{b}^T \boldsymbol{C}\boldsymbol{\nu}_1(\pi) + \boldsymbol{b}^T \boldsymbol{C}\boldsymbol{b}}{(\sqrt{n}+1)^2}.$$

Let the function $r_1 : \mathcal{M} \times \mathit{\Gamma} \to [0, \infty)$ be defined by

$$r_1(\boldsymbol{b}, \pi) := r(\boldsymbol{d}^{\boldsymbol{b}}, \pi).$$

Note that $\mathcal{M}$ and $\mathit{\Gamma}$ are convex sets and $\mathcal{M}$ is compact. Moreover, for each fixed $\pi \in \mathit{\Gamma}$, $r_1(\boldsymbol{b}, \pi)$ is convex, continuous with respect to $\boldsymbol{b} \in \mathcal{M}$, and for each fixed $\boldsymbol{b} \in \mathcal{M}$, $r_1(\boldsymbol{b}, \pi)$ is concave (linear) with respect $\pi \in \mathit{\Gamma}$. This means that all the assumptions of the Nikaido theorem (see Aubin (1980), p. 217) are fulfilled and thus there exists a point $\underline{\boldsymbol{b}}$ for which

$$\sup_{\pi \in \mathit{\Gamma}} r_1(\underline{\boldsymbol{b}}, \pi) = \inf_{b \in \mathcal{M}} \sup_{\pi \in \mathit{\Gamma}} r_1(\boldsymbol{b}, \pi) = \sup_{\pi \in \mathit{\Gamma}} \inf_{b \in \mathcal{M}} r_1(\boldsymbol{b}, \pi).$$

The last equality implies that the $\mathit{\Gamma}$-minimax risk in $\mathcal{D}_0$ equals

$$\inf_{b \in \mathcal{M}} \sup_{\pi \in \mathit{\Gamma}} r_1(\boldsymbol{b}, \pi) = \sup_{\pi \in \mathit{\Gamma}} \inf_{b \in \mathcal{M}} r_1(\boldsymbol{b}, \pi) = \sup_{\pi \in \mathit{\Gamma}} \frac{\nu_2(\pi) - \boldsymbol{\nu}_1^T(\pi) \boldsymbol{C}\boldsymbol{\nu}_1(\pi)}{(\sqrt{n}+1)^2}, \tag{8}$$

because, for a fixed distribution $\pi \in \mathit{\Gamma}$, the convex function $r_1(\boldsymbol{b}, \pi)$ of the variable $\boldsymbol{b}$ attains its global minimum over $\mathcal{M}$ at the point $\boldsymbol{b}(\pi) = \boldsymbol{\nu}_1(\pi)$. Now it remains to prove that $\underline{\boldsymbol{b}} = \boldsymbol{p}_{\mathit{\Gamma}}$. Set, for simplicity,

$$r_2(\pi) = \nu_2(\pi) - \boldsymbol{\nu}_1^T(\pi) \boldsymbol{C}\boldsymbol{\nu}_1(\pi). \tag{9}$$

Let (π_j) be a sequence of priors satisfying (3). Since the functions $\boldsymbol{\nu}_1(\pi)$ and $\nu_2(\pi)$ are linear in π, an easy calculation shows that for any $\pi \in \mathit{\Gamma}$ and $0 < \beta < 1$,

$$\sup_{\overline{\pi} \in \mathit{\Gamma}} r_2(\overline{\pi}) \geq r_2(\beta\pi + (1-\beta)\pi_j) = \beta r_2(\pi) + (1-\beta) r_2(\pi_j)$$
$$+ \beta(1-\beta)(\boldsymbol{\nu}_1(\pi_j) - \boldsymbol{\nu}_1(\pi))^T \boldsymbol{C}(\boldsymbol{\nu}_1(\pi_j) - \boldsymbol{\nu}_1(\pi)).$$

This implies that

$$\sup_{\overline{\pi} \in \mathit{\Gamma}} r_2(\overline{\pi}) \geq \beta r_2(\pi) + (1-\beta) \sup_{\overline{\pi} \in \mathit{\Gamma}} r_2(\overline{\pi}) + \beta(1-\beta)(\boldsymbol{p}_{\mathit{\Gamma}} - \boldsymbol{\nu}_1(\pi))^T \boldsymbol{C}(\boldsymbol{p}_{\mathit{\Gamma}} - \boldsymbol{\nu}_1(\pi)),$$

because $\boldsymbol{p}_\Gamma$ is a cluster point of the sequence $(\boldsymbol{\nu}_1(\pi_j))$, and $\lim_{j\to\infty} r_2(\pi_j) = \sup_{\pi\in\Gamma} r_2(\pi)$ by (3). Therefore,

$$\beta \sup_{\overline{\pi}\in\Gamma} r_2(\overline{\pi}) \geq \beta r_2(\pi) + \beta(1-\beta)(\boldsymbol{p}_\Gamma - \boldsymbol{\nu}_1(\pi))^T \boldsymbol{C}(\boldsymbol{p}_\Gamma - \boldsymbol{\nu}_1(\pi)),$$

and since β is positive,

$$\sup_{\overline{\pi}\in\Gamma} r_2(\overline{\pi}) \geq r_2(\pi) + (1-\beta)(\boldsymbol{p}_\Gamma - \boldsymbol{\nu}_1(\pi))^T \boldsymbol{C}(\boldsymbol{p}_\Gamma - \boldsymbol{\nu}_1(\pi)).$$

Letting $\beta \to 0^+$, we can see by (9) that

$$\begin{aligned}\sup_{\overline{\pi}\in\Gamma} r_2(\overline{\pi}) &\geq r_2(\pi) + (\boldsymbol{p}_\Gamma - \boldsymbol{\nu}_1(\pi))^T \boldsymbol{C}(\boldsymbol{p}_\Gamma - \boldsymbol{\nu}_1(\pi)) \\ &= \nu_2(\pi) - 2\boldsymbol{p}_\Gamma^T \boldsymbol{C}\boldsymbol{\nu}_1(\pi) + \boldsymbol{p}_\Gamma^T \boldsymbol{C}\boldsymbol{p}_\Gamma,\end{aligned}$$

which implies by (7) that

$$\sup_{\overline{\pi}\in\Gamma} \frac{\nu_2(\overline{\pi}) - \boldsymbol{\nu}_1^T(\overline{\pi})\boldsymbol{C}\boldsymbol{\nu}_1(\overline{\pi})}{(\sqrt{n}+1)^2} \geq \frac{\nu_2(\pi) - 2\boldsymbol{p}_\Gamma^T \boldsymbol{C}\boldsymbol{\nu}_1(\pi) + \boldsymbol{p}_\Gamma^T \boldsymbol{C}\boldsymbol{p}_\Gamma}{(\sqrt{n}+1)^2} = r(\boldsymbol{d}_\Gamma, \pi).$$

Because this is true for all $\pi \in \Gamma$, it follows from (8) that

$$\begin{aligned}\sup_{\pi\in\Gamma} r(\boldsymbol{d}_\Gamma, \pi) &\leq \sup_{\pi\in\Gamma} \frac{\nu_2(\pi) - \boldsymbol{\nu}_1^T(\pi)\boldsymbol{C}\boldsymbol{\nu}_1(\pi)}{(\sqrt{n}+1)^2} = \inf_{b\in\mathcal{M}} \sup_{\pi\in\Gamma} r_1(\boldsymbol{b}, \pi) \\ &= \inf_{b\in\mathcal{M}} \sup_{\pi\in\Gamma} r(\boldsymbol{d}^{\boldsymbol{b}}, \pi) = \inf_{d\in\mathcal{D}_0} \sup_{\pi\in\Gamma} r(\boldsymbol{d}, \pi).\end{aligned} \tag{10}$$

This implies that $\boldsymbol{d}_\Gamma(\boldsymbol{Y}^n)$ is Γ-minimax if the class of estimators is restricted to $\mathcal{D}_0$ of $\mathcal{D}$.

To complete the proof we will construct a sequence (π_j^*) of priors from Γ and a sequence $(\boldsymbol{d}_{\pi_j^*})$ of π_j^*-Bayes estimators for which

$$\lim_{j\to\infty} r(\boldsymbol{d}_{\pi_j^*}, \pi_j^*) = \sup_{\overline{\pi}\in\Gamma} \frac{\nu_2(\overline{\pi}) - \boldsymbol{\nu}_1^T(\overline{\pi})\boldsymbol{C}\boldsymbol{\nu}_1(\overline{\pi})}{(\sqrt{n}+1)^2} = \sup_{\pi\in\Gamma} r(\boldsymbol{d}_\Gamma, \pi).$$

Let (π_j) be a sequence of priors from Γ satisfying (3) and let (P_j) be a sequence of probability measures on $(\mathcal{Y}, \mathcal{B})$ such that

$$\bigwedge_{A\in\mathcal{B}} P_j(A) = E_{\pi_j}(P(A)).$$

For each $j \geq 1$ we denote by π_j^* a Dirichlet prior process on $(\mathcal{Y}, \mathcal{B})$ with parameter $\beta_j = \sqrt{n}\, P_j$. To prove that $\pi_j^* \in \Gamma$ we note first that by Ferguson (1973) (Theorems 3 and 4),

$$E_{\pi_j^*}[E_P \boldsymbol{Z}^T \boldsymbol{C}\boldsymbol{Z}] = E_{P_j} \boldsymbol{Z}^T \boldsymbol{C}\boldsymbol{Z}, \qquad E_{\pi_j^*} \boldsymbol{p} = E_{\pi_j^*}[E_P \boldsymbol{Z}] = E_{P_j} \boldsymbol{Z}.$$

Since by the definition of the probability measure P_j,

$$E_{P_j} \boldsymbol{Z} = E_{\pi_j}[E_P \boldsymbol{Z}] = \boldsymbol{\nu}_1(\pi_j), \qquad E_{P_j} \boldsymbol{Z}^T \boldsymbol{C}\boldsymbol{Z} = E_{\pi_j}[E_P \boldsymbol{Z}^T \boldsymbol{C}\boldsymbol{Z}] = \nu_2(\pi_j),$$

we deduce that

$$\boldsymbol{\nu}_1(\pi_j^*) = E_{\pi_j^*}[E_P \boldsymbol{Z}] = \boldsymbol{\nu}_1(\pi_j), \quad \nu_2(\pi_j^*) = E_{\pi_j^*}[E_P \boldsymbol{Z}^T \boldsymbol{C} \boldsymbol{Z}] = \nu_2(\pi_j).$$

This obviously implies that $\pi_j^* \in \Gamma$, because $\pi_j \in \Gamma$. Moreover, from Ferguson (1973) (example b), the π_j^*-Bayes nonparametric estimator of $\boldsymbol{p} = E_P \boldsymbol{Z}$ has the form

$$\boldsymbol{d}_{\pi_j^*}(\boldsymbol{Y}^n) = \frac{\sqrt{n}}{n+\sqrt{n}} E_{P_j} \boldsymbol{Z} + \frac{n}{n+\sqrt{n}} \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{z}(Y_j) = \frac{\boldsymbol{X}^n + \sqrt{n}\,\boldsymbol{\nu}_1(\pi_j)}{n+\sqrt{n}},$$

because $E_{P_j} \boldsymbol{Z} = \boldsymbol{\nu}_1(\pi_j)$. To calculate the π_j^*-Bayes risk $r(\boldsymbol{d}_{\pi_j^*}, \pi_j^*)$ we note that $\boldsymbol{d}_{\pi_j^*}(\boldsymbol{Y}^n) = \boldsymbol{d}^{\boldsymbol{b}}(\boldsymbol{Y}^n) \in \mathcal{D}_0$ with $\boldsymbol{b} = \boldsymbol{\nu}_1(\pi_j)$. Thus, by (7),

$$\begin{aligned} r(\boldsymbol{d}_{\pi_j^*}, \pi_j^*) &= \frac{\nu_2(\pi_j^*) - 2\boldsymbol{\nu}_1^T(\pi_j)\boldsymbol{C}\boldsymbol{\nu}_1(\pi_j^*) + \boldsymbol{\nu}_1{}^T(\pi_j)\boldsymbol{C}\boldsymbol{\nu}_1(\pi_j)}{(\sqrt{n}+1)^2} \\ &= \frac{\nu_2(\pi_j) - \boldsymbol{\nu}_1^T(\pi_j)\boldsymbol{C}\boldsymbol{\nu}_1(\pi_j)}{(\sqrt{n}+1)^2}, \end{aligned}$$

because $\boldsymbol{\nu}_1(\pi_j^*) = \boldsymbol{\nu}_1(\pi_j)$ and $\nu_2(\pi_j^*) = \nu_2(\pi_j)$. Therefore, by (10),

$$\begin{aligned} \inf_{d\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\boldsymbol{d}, \pi) &\ge \lim_{j\to\infty} r(\boldsymbol{d}_{\pi_j^*}, \pi_j^*) = \lim_{j\to\infty} \frac{\nu_2(\pi_j) - \boldsymbol{\nu}_1^T(\pi_j)\boldsymbol{C}\boldsymbol{\nu}_1(\pi_j)}{(\sqrt{n}+1)^2} \\ &= \sup_{\pi\in\Gamma} \frac{\nu_2(\pi) - \boldsymbol{\nu}_1^T(\pi)\boldsymbol{C}\boldsymbol{\nu}_1(\pi)}{(\sqrt{n}+1)^2} \ge \sup_{\pi\in\Gamma} r(\boldsymbol{d}_\Gamma, \pi) \ge \inf_{d\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\boldsymbol{d}, \pi), \end{aligned}$$

which implies that the estimator $\boldsymbol{d}_\Gamma(\boldsymbol{Y}^n)$ is Γ-minimax and its Γ-minimax risk is given by (6). This completes the proof of Theorem 1. ■

3. Generalization. In this section we present a slight generalization of Theorem 1. Instead of assuming that the function $\boldsymbol{z}$ is bounded on $\mathcal{Y}$, we suppose that a weaker condition is fulfilled: $\sup_{y\in\mathcal{Y}} \|\boldsymbol{C}^{1/2}\boldsymbol{z}(y)\| < \infty$, where $\boldsymbol{C}^{1/2}$ is the square root of the matrix $\boldsymbol{C}$, i.e. $\boldsymbol{C}^{1/2}\boldsymbol{C}^{1/2} = \boldsymbol{C}$. Then the random vector

$$\boldsymbol{Z}^* := \boldsymbol{C}^{1/2}\boldsymbol{z}(Y) = \boldsymbol{C}^{1/2}\boldsymbol{Z}$$

is bounded, which implies that for each affine estimator $\boldsymbol{d}^{\boldsymbol{b}} \in \mathcal{D}_0$ its risk function $R(\boldsymbol{d}^{\boldsymbol{b}}, P)$ is bounded for $P \in \mathcal{P}$. Let $\boldsymbol{\nu}_1^*(\pi)$ and $\nu_2^*(\pi)$ denote the first moments of $E_P \boldsymbol{Z}^*$ and $E_P(\boldsymbol{Z}^*)^T \boldsymbol{Z}^* = E_P\|\boldsymbol{Z}^*\|^2$ with respect to a prior $\pi \in \Pi$, i.e.

$$\boldsymbol{\nu}_1^*(\pi) = E_\pi(E_P \boldsymbol{Z}^*), \quad \nu_2^*(\pi) = E_\pi(E_P\|\boldsymbol{Z}^*\|^2),$$

and let (π_j) be a sequence of priors from Γ for which

$$\lim_{j\to\infty}(\nu_2^*(\pi_j) - \|\boldsymbol{\nu}_1^*(\pi_j)\|^2) = \sup_{\pi\in\Gamma}(\nu_2^*(\pi) - \|\boldsymbol{\nu}_1^*(\pi)\|^2).$$

Then, by the same arguments as in the previous section, the sequence $(\boldsymbol{\nu}_1^*(\pi_j))$, where

$$\boldsymbol{\nu}_1^*(\pi_j) = E_{\pi_j}(E_P\boldsymbol{Z}^*), \quad j \ge 1, \tag{11}$$

has a cluster point $\boldsymbol{p}_\Gamma^*$. Since $\boldsymbol{p}_\Gamma^*$ belongs to the linear space generated by the columns of the matrix $\boldsymbol{C}^{1/2}$, there exists a vector $\boldsymbol{p}_\Gamma$ for which

$$\boldsymbol{C}^{1/2}\boldsymbol{p}_\Gamma = \boldsymbol{p}_\Gamma^*. \tag{12}$$

The following theorem generalizes the results of the previous section.

THEOREM 2. *Suppose that* $\sup_{y\in\mathcal{Y}}\|\boldsymbol{C}^{1/2}\boldsymbol{z}(y)\| < \infty$. *Then the* Γ*-minimax estimator of the unknown vector* $\boldsymbol{p}$ *under the loss function* (1) *has the form*

$$\boldsymbol{d}_\Gamma(\boldsymbol{Y}^n) = \frac{\boldsymbol{X}^n + \sqrt{n}\,\boldsymbol{p}_\Gamma}{n+\sqrt{n}}, \tag{13}$$

where $\boldsymbol{p}_\Gamma$ *is any solution of* (12). *Moreover*, *the* Γ*-minimax risk for* $\boldsymbol{d}_\Gamma$ *is*

$$\sup_{\pi\in\Gamma} r(\boldsymbol{d}_\Gamma,\pi) = \sup_{\pi\in\Gamma}\frac{\nu_2^*(\pi) - \|\boldsymbol{\nu}_1^*(\pi)\|^2}{(\sqrt{n}+1)^2}. \tag{14}$$

Proof. Let the random vector $\boldsymbol{X}^{*n}$ be defined by

$$\boldsymbol{X}^{*n} := \boldsymbol{C}^{1/2}\boldsymbol{X}^n = \sum_{j=1}^n \boldsymbol{z}^*(Y_j).$$

As can easily be seen, it suffices to show that the decision rule $d_\Gamma^*(\boldsymbol{Y}^n) = \boldsymbol{C}^{1/2}\boldsymbol{d}_\Gamma(\boldsymbol{Y}^n)$, which by (13) and (12) has the form

$$\boldsymbol{d}_\Gamma^*(\boldsymbol{Y}^n) = \frac{\boldsymbol{X}^{*n} + \sqrt{n}\,\boldsymbol{p}_\Gamma^*}{n+\sqrt{n}},$$

is the Γ-minimax estimator of the vector $\boldsymbol{p}^* = \boldsymbol{C}^{1/2}\boldsymbol{p} = E_P\boldsymbol{Z}^*$ under the loss function

$$L^*(\boldsymbol{d}^*, P) = (\boldsymbol{d}^* - \boldsymbol{p}^*)^T(\boldsymbol{d}^* - \boldsymbol{p}^*) = \|\boldsymbol{d}^* - \boldsymbol{p}^*\|^2.$$

This, however, can be easily deduced from Theorem 1. Moreover, since

$$\bigwedge_{P\in\mathcal{P}} L^*(\boldsymbol{d}_\Gamma^*, P) = L(\boldsymbol{d}_\Gamma, P),$$

the estimators $\boldsymbol{d}_\Gamma^*$ and $\boldsymbol{d}_\Gamma$ have the same risk functions, and (6) yields (14). ∎

4. Example. Finding analytically the cluster point $\boldsymbol{p}_\Gamma$ is not an easy task. However, in the following example this can easily be done.

EXAMPLE. Suppose that the set $\mathcal{Y}$ is centrosymmetric about $\mathbf{0}$ and that

$$\boldsymbol{z}(y) = -\boldsymbol{z}(-y),\ y \in \mathcal{Y}, \qquad (\boldsymbol{\nu}_1, \nu_2) \in \mathcal{G} \ \Leftrightarrow\ (-\boldsymbol{\nu}_1, \nu_2) \in \mathcal{G}. \tag{15}$$

Let P^- stand for the distribution of the random vector $-Y$, whenever Y is distributed according to P. For any prior $\pi \in \Pi$ we denote by π^- its modified version in which each probability distribution P chosen by π is replaced by P^-. The assumption (15) implies that $\pi \in \Gamma \Leftrightarrow \pi^- \in \Gamma$, because

$$\nu_2(\pi^-) = \nu_2(\pi), \qquad \boldsymbol{\nu}_1(\pi^-) = -\boldsymbol{\nu}_1(\pi).$$

Now, let (π_j) be a sequence of priors from Γ satisfying (3). Then for each $j \geq 1$, the prior $\overline{\pi}_j = \frac{1}{2}(\pi_j + \pi_j^-)$ belongs to Γ, because $\overline{\pi}_j \in \Gamma$, $\pi_j^- \in \Gamma$ and the set Γ is convex. Moreover, since $\nu_2(\overline{\pi}_j) = \nu_2(\pi_j)$ and $\boldsymbol{\nu}_1(\overline{\pi}_j) = \mathbf{0}$, we conclude that

$$\nu_2(\overline{\pi}_j) - \boldsymbol{\nu}_1^T(\overline{\pi}_j)\boldsymbol{C}\boldsymbol{\nu}_1(\overline{\pi}_j) = \nu_2(\pi_j) \geq \nu_2(\pi_j) - \boldsymbol{\nu}_1^T(\pi_j)\boldsymbol{C}\boldsymbol{\nu}_1(\pi_j).$$

This implies that the sequence $(\overline{\pi}_j)$ also satisfies (3). Therefore, the estimator

$$\boldsymbol{d}_\Gamma(\boldsymbol{Y}^n) = \frac{\boldsymbol{X}^n}{n + \sqrt{n}}$$

is Γ-minimax, because $\boldsymbol{p}_\Gamma = \lim_{j\to\infty} \boldsymbol{\nu}_1(\overline{\pi}_j) = \mathbf{0}$.

Acknowledgements. The author wishes to thank the referee for his helpful comments which improved the original version of the paper.

References

J. P. Aubin (1980), *Mathematical Methods of Game and Economic Theory*, North-Holland.

L. Chen (1989), *Γ-minimax estimation of the parameters of the multinomial distribution*, J. Math. Res. Exposition 9, 489–494.

L. Chen and J. Eichenauer-Herrmann (1990), *Gamma-minimax estimation of a multivariate normal mean*, Metrika 37, 1–6.

L. Chen, J. Eichenauer-Herrmann, H. Hoffmann and J. Kindler (1991), *Gamma-minimax estimators in the exponential family*, Dissertationes Math. 308.

J. Eichenauer-Herrmann, J. Lehn and S. Rettig (1988), *A gamma-minimax result in credibility theory*, Insurance Math. Econom. 7, 49–57.

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

T. S. Ferguson (1973), *A Bayesian analysis of some nonparametric problems*, Ann. Statist. 1, 209–230.

D. A. Jackson, T. M. O'Donovan, W. J. Zimmer and J. J. Deely (1970), *G_2-minimax estimators in the exponential family*, Biometrika 57, 439–443.

R. Magiera (2001), *Γ-minimax sequential estimation for Markov-additive processes*, Appl. Math. (Warsaw) 28, 467–485.

H. Robbins (1964), *The empirical Bayes approach to statistical decision problem*, Ann. Math. Statist. 36, 1–20.
M. Wilczyński (1992), *Minimax Nonparametric Estimation*, Sankhyā Ser. A 54, 425–431.

Institute of Mathematics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: wilczyn@im.pwr.wroc.pl

Received on 21.9.2001;
revised version on 8.7.2002 (1592)