APPLICATIONES MATHEMATICAE
28,1 (2001), pp. 83–92

Maciej Wilczyński (Wrocław)

MINIMAX NONPARAMETRIC PREDICTION

Abstract. Let U_0 be a random vector taking its values in a measurable space and having an unknown distribution P and let $U_1, \ldots, U_n$ and $V_1, \ldots, V_m$ be independent, simple random samples from P of size n and m, respectively. Further, let $z_1, \ldots, z_k$ be real-valued functions defined on the same space. Assuming that only the first sample is observed, we find a minimax predictor $\boldsymbol{d}^0(n, U_1, \ldots, U_n)$ of the vector $\boldsymbol{Y}^m = \sum_{j=1}^m (z_1(V_j), \ldots$ $\ldots, z_k(V_j))^T$ with respect to a quadratic errors loss function.

1. Introduction. Let U_0 be a random vector taking its values in a measurable space $(\mathcal{Y}, \mathcal{B})$ and having an unknown distribution P, which is assumed to be an element of the set

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

Let $U_1, \ldots, U_n$ and $V_1, \ldots, V_m$ be independent, simple random samples from P of size n and m, respectively. Further, let $\boldsymbol{z} = (z_1, \ldots, z_k)^T$ be a measurable function on the space $(\mathcal{Y}, \mathcal{B})$ with values in $(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k})$. In the paper we consider the problem of predicting the value of a k-dimensional random vector $\boldsymbol{Y}^m = \sum_{j=1}^m \boldsymbol{z}(V_j)$ from the data $\boldsymbol{U}^n = (U_1, \ldots, U_n)$. Assuming that the loss function has the form

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

where $\boldsymbol{C} = [c_{ij}]$ is a nonnegative definite, symmetric $k \times k$ matrix, we find a minimax solution of the above prediction problem. We show that the minimax predictor $\boldsymbol{d}^0(n, \boldsymbol{U}^n)$ of $\boldsymbol{Y}^m$ is an affine (inhomogeneous linear) function of the random vector $\boldsymbol{X}^n = \sum_{j=1}^n \boldsymbol{z}(U_j)$.

The decision rule $\boldsymbol{d}^0(n, \boldsymbol{U}^n)$ has a risk function which is not constant and therefore proving its minimaxity cannot be accomplished by showing

2000 *Mathematics Subject Classification*: 62C20, 62G05.

Key words and phrases: minimax prediction, Dirichlet process, nonparametric prediction.

that this predictor is Bayes with respect to some prior on $\mathcal{P}$. Instead, we use the method proposed in Wilczyński (1992). First we show that $\boldsymbol{d}^0(n,\boldsymbol{U}^n)$ is minimax among all predictors which are affine functions of $\boldsymbol{X}^n$. Next, using some implications of this fact, we calculate the upper bound for a minimax risk of $\boldsymbol{d}^0(n,\boldsymbol{U}^n)$. Then, via nonparametric Bayes approach proposed by Ferguson (1973), we construct a sequence of priors on $\mathcal{P}$ for which the corresponding sequence of Bayes risks converges to this upper bound. From this we deduce minimaxity of $\boldsymbol{d}^0(n,\boldsymbol{U}^n)$.

2. Minimax estimate. The statement of our main result requires introducing the following notation. Let the function $\boldsymbol{z}_* : (\mathcal{Y},\mathcal{B}) \to (\mathbb{R}^k,\mathcal{B}_{\mathbb{R}^k})$ be defined by

$$\boldsymbol{z}_*(y) = \boldsymbol{C}^{1/2}\boldsymbol{z}(y), \qquad y \in Y,$$

where $\boldsymbol{C}^{1/2}$ is the square root of the matrix $\boldsymbol{C}$, i.e. $\boldsymbol{C}^{1/2}\boldsymbol{C}^{1/2} = \boldsymbol{C}$. The random vector $\boldsymbol{z}_*(U_0)$, its expected value and the sum of the variances of its components are denoted by $\boldsymbol{Z}_*$, $\boldsymbol{p}_*$ and $R_1(P)$ respectively, i.e. we put

$$\begin{aligned} \boldsymbol{Z}_* &= \boldsymbol{z}_*(U_0), \\ \boldsymbol{p}_* &= E_P\boldsymbol{Z}_*, \\ R_1(P) &= E_P\|\boldsymbol{Z}_* - E_P\boldsymbol{Z}_*\|^2 = E_P\|\boldsymbol{Z}_* - \boldsymbol{p}_*\|^2. \end{aligned} \tag{2}$$

Now, let (P_j) be any sequence of probability measures on $(\mathcal{Y},\mathcal{B})$ such that

$$\lim_{j\to\infty} R_1(P_j) = \sup_{P\in\mathcal{P}} R_1(P) \tag{3}$$

and let $(\boldsymbol{b}_j)$ be the corresponding sequence of points from $\mathbb{R}^k$ defined by

$$\boldsymbol{b}_j = E_{P_j}\boldsymbol{Z}_*. \tag{4}$$

In Theorem 1, we show that the above prediction problem has a nontrivial solution only when the vector-valued function $\boldsymbol{z}_*(y)$ is bounded on Y, i.e. when

$$M^2 := \sup_{y\in\mathcal{Y}} \|\boldsymbol{z}_*(y)\|^2 = \sup_{y\in\mathcal{Y}} \boldsymbol{z}^T(y)\boldsymbol{C}\boldsymbol{z}(y) < \infty.$$

Obviously, if M is finite then the random vector $\boldsymbol{Z}_*$ and its expected value $\boldsymbol{p}_*$ are bounded. This implies that $\sup_{P\in\mathcal{P}} R_1(P) < \infty$ and, because

$$\|\boldsymbol{b}_j\|^2 = \|E_{P_j}\boldsymbol{Z}_*\|^2 \le E_{P_j}\|\boldsymbol{Z}_*\|^2 \le M^2,$$

the sequence $(\boldsymbol{b}_j)$ takes its values in a convex compact subset $\mathcal{M}$ of $\mathbb{R}^k$, defined by

$$\mathcal{M} = \{\boldsymbol{b} \in \mathbb{R}^k : \|\boldsymbol{b}\|^2 \le M^2\}.$$

Therefore, this sequence has a cluster point $\boldsymbol{b}_0 \in \mathcal{M}$. Now, we define $\boldsymbol{a}_0$ as a vector from $\mathbb{R}^k$ which solves the equation

$$\boldsymbol{C}^{1/2}\boldsymbol{a}_0 = \boldsymbol{b}_0. \tag{5}$$

To see that (5) can be solved, we denote by $(\boldsymbol{C}^{1/2})^-$ any g-inverse of the matrix $\boldsymbol{C}^{1/2}$. Since $\boldsymbol{C}^{1/2}(\boldsymbol{C}^{1/2})^-\boldsymbol{C}^{1/2} = \boldsymbol{C}^{1/2}$, we have

$$\begin{aligned}\boldsymbol{b}_j &= E_{P_j}\boldsymbol{C}^{1/2}\boldsymbol{z}(U_0) = E_{P_j}\boldsymbol{C}^{1/2}(\boldsymbol{C}^{1/2})^-\boldsymbol{C}^{1/2}\boldsymbol{z}(U_0)\\ &= \boldsymbol{C}^{1/2}(\boldsymbol{C}^{1/2})^-E_{P_j}\boldsymbol{C}^{1/2}\boldsymbol{z}(U_0) = \boldsymbol{C}^{1/2}(\boldsymbol{C}^{1/2})^-\boldsymbol{b}_j.\end{aligned}$$

This implies that $(\boldsymbol{C}^{1/2})^-\boldsymbol{b}_0$ solves (5), because

$$\boldsymbol{b}_0 = \boldsymbol{C}^{1/2}(\boldsymbol{C}^{1/2})^-\boldsymbol{b}_0 = \boldsymbol{C}^{1/2}\boldsymbol{a}_0.$$

Now, let the number α_0 satisfy the condition

$$\alpha_0^2 n + m = (\alpha_0 n - m)^2 \quad \text{and} \quad \alpha_0 n - m < 0,$$

i.e. let

$$\alpha_0 = \begin{cases} \dfrac{nm - \sqrt{nm(n+m-1)}}{n(n-1)} & \text{if } n > 1,\\ \dfrac{m-1}{2} & \text{if } n = 1.\end{cases} \tag{6}$$

The following theorem is the main result of the paper.

THEOREM 1. *If* $\sup_{y\in\mathcal{Y}} \boldsymbol{z}(y)^T\boldsymbol{C}\boldsymbol{z}(y) < \infty$ *then*

$$\boldsymbol{d}^0(n, \boldsymbol{U}^n) = \alpha_0\boldsymbol{X}^n + (m - \alpha_0 n)\boldsymbol{a}_0 \tag{7}$$

is a minimax predictor of the unobservable vector $\boldsymbol{Y}^m$ *and its minimax risk equals*

$$\sup_{P\in\mathcal{P}} R(\boldsymbol{d}^0, P) = (\alpha_0 n - m)^2 \sup_{P\in\mathcal{P}} R_1(P).$$

If $\sup_{y\in\mathcal{Y}} \boldsymbol{z}(y)^T\boldsymbol{C}\boldsymbol{z}(y) = \infty$ *then*

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

and therefore a minimax predictor for $\boldsymbol{Y}^m$ *does not exist.*

3. Proof of the main result. Define the following two random vectors:

$$\boldsymbol{X}_*^n = \boldsymbol{C}^{1/2}\boldsymbol{X}^n, \qquad \boldsymbol{Y}_*^n = \boldsymbol{C}^{1/2}\boldsymbol{Y}^n.$$

To prove the first part of Theorem 1 it suffices to show that the decision rule $\boldsymbol{d}_*^0(n, \boldsymbol{U}^n) = \boldsymbol{C}^{1/2}\boldsymbol{d}^0(n, \boldsymbol{U}^n)$, which, by (7) and (5), has the form

$$\boldsymbol{d}_*^0(n, \boldsymbol{U}^n) = \boldsymbol{C}^{1/2}\boldsymbol{d}^0(n, \boldsymbol{U}^n) = \alpha_0\boldsymbol{X}_*^n + (m - \alpha_0 n)\boldsymbol{b}_0, \tag{8}$$

is a minimax predictor of the vector $\boldsymbol{Y}_*^m$ under the loss function

$$L_*(\boldsymbol{d}, \boldsymbol{Y}_*^m) = (\boldsymbol{d} - \boldsymbol{Y}_*^m)^T(\boldsymbol{d} - \boldsymbol{Y}_*^m) = \|\boldsymbol{d} - \boldsymbol{Y}_*^m\|^2.$$

Moreover, to complete the proof of Theorem 1 it suffices to show that the risk function for any predictor of $\boldsymbol{Y}_*^m$ is unbounded when $\sup_{y\in\mathcal{Y}} \boldsymbol{z}(y)^T\boldsymbol{C}\boldsymbol{z}(y) = \infty$.

Let $\mathcal{D}$ be the class of all predictors $\boldsymbol{d} = \boldsymbol{d}(n, \boldsymbol{U}^n)$ of the vector $\boldsymbol{Y}_*^m$. We start the proof by calculating the risk function $R(\boldsymbol{d}, P)$ of any decision rule $\boldsymbol{d}$ from $\mathcal{D}$. Since the vectors $\boldsymbol{d}(n, \boldsymbol{U}^n)$ and $\boldsymbol{Y}_*^m = \sum_{j=1}^m \boldsymbol{z}_*(V_j)$ are independent, and since

$$(9) \qquad E_P \boldsymbol{Y}_*^m = \sum_{j=1}^m E_P \boldsymbol{z}_*(V_j) = m E_P \boldsymbol{z}_*(U_0) = m E_P \boldsymbol{Z}_* = m \boldsymbol{p}_*,$$

this risk is equal to

$$R(\boldsymbol{d}, P) = E_P \|\boldsymbol{d}(n, \boldsymbol{U}^n) - \boldsymbol{Y}_*^m\|^2 = E_P \|\boldsymbol{d} - m\boldsymbol{p}_*\|^2 + E_P \|\boldsymbol{Y}_*^m - m\boldsymbol{p}_*\|^2.$$

Moreover, since $\boldsymbol{z}_*(V_1), \ldots, \boldsymbol{z}_*(V_m)$ are i.i.d. random vectors with expected value $\boldsymbol{p}_*$,

$$(10) \qquad E_P \|\boldsymbol{Y}_*^m - m\boldsymbol{p}_*\|^2 = E_P \Big\| \sum_{j=1}^m (\boldsymbol{z}_*(V_j) - \boldsymbol{p}_*) \Big\|^2 = m E_P \|\boldsymbol{Z}_* - \boldsymbol{p}_*\|^2 = m R_1(P).$$

Therefore, the risk for any predictor $\boldsymbol{d}(n, \boldsymbol{U}^n) \in \mathcal{D}$ may be rewritten as

$$(11) \qquad R(\boldsymbol{d}, P) = E_P \|\boldsymbol{d} - m\boldsymbol{p}_*\|^2 + m R_1(P).$$

Assume now that $\sup_{y \in \mathcal{Y}} \boldsymbol{z}(y)^T \boldsymbol{C} \boldsymbol{z}(y) < \infty$. According to the definition of minimaxity, to prove that the predictor $\boldsymbol{d}_*^0(n, \boldsymbol{U}^n)$ defined by (8) is minimax we have to show that

$$(12) \qquad \sup_{P \in \mathcal{P}} R(\boldsymbol{d}_*^0, P) = \inf_{d \in \mathcal{D}} \sup_{P \in \mathcal{P}} R(\boldsymbol{d}, P).$$

To do this, we denote by $\mathcal{D}_0$ the following class of affine predictors:

$$\mathcal{D}_0 = \{\boldsymbol{d}^{\boldsymbol{b}} \in \mathcal{D} : \boldsymbol{d}^{\boldsymbol{b}}(n, \boldsymbol{U}^n) = \alpha_0 \boldsymbol{X}_*^n + (m - \alpha_0 n)\boldsymbol{b},\ \boldsymbol{b} \in \mathcal{M}\},$$

where the number α_0 is defined by (6), and we prove that $\boldsymbol{d}_*^0 = \boldsymbol{d}^{\boldsymbol{b}_0}$ is minimax in $\mathcal{D}_0$, i.e.

$$(13) \qquad \sup_{P \in \mathcal{P}} R(\boldsymbol{d}_*^0, P) = \inf_{d \in \mathcal{D}_0} \sup_{P \in \mathcal{P}} R(\boldsymbol{d}, P).$$

Next, using some implication of the minimaxity of $\boldsymbol{d}_*^0$ in $\mathcal{D}_0$, we calculate the upper bound for $R(\boldsymbol{d}_*^0, P)$. Then, if $m > 1$, we construct a sequence of priors on $\mathcal{P}$ for which the corresponding sequence of Bayes risks converges to this upper bound. From this we deduce minimaxity of $\boldsymbol{d}_*^0 \in \mathcal{D}$. If $m = 1$, we use a different approach to prove that $\boldsymbol{d}_*^0$ is minimax in $\mathcal{D}$.

We start proving minimaxity of $\boldsymbol{d}_*^0$ in $\mathcal{D}_0$ by calculating its risk function. We first note that (cf. (9) and (10))

$$\begin{aligned} E_P \|\alpha_0 \boldsymbol{X}_*^n + (m - \alpha_0 n)\boldsymbol{b} - m\boldsymbol{p}_*\|^2 &= \alpha_0^2 E_P \|\boldsymbol{X}_*^n - n\boldsymbol{p}_*\|^2 + (\alpha_0 n - m)^2 \|\boldsymbol{b} - \boldsymbol{p}_*\|^2 \\ &= \alpha_0^2 n R_1(P) + (\alpha_0 n - m)^2 \|\boldsymbol{b} - \boldsymbol{p}_*\|^2. \end{aligned}$$

Since $\alpha_0^2 n + m = (\alpha_0 n - m)^2$, we conclude, by (11), that the risk function for a predictor $\boldsymbol{d^b} \in \mathcal{D}_0$, denoted for simplicity by $R(\boldsymbol{b}, P)$, is given by

$$
\begin{aligned}
R(\boldsymbol{b}, P) &= (\alpha_0^2 n + m) R_1(P) + (\alpha_0 n - m)^2 \|\boldsymbol{b} - \boldsymbol{p}_*\|^2 \\
&= (\alpha_0 n - m)^2 [R_1(P) + \|\boldsymbol{b} - \boldsymbol{p}_*\|^2].
\end{aligned} \tag{15}
$$

Furthermore, if $\sup_{y \in \mathcal{Y}} \|\boldsymbol{z}_*(y)\|^2 < \infty$ then the random vector $\boldsymbol{Z}_*$ and its expected value $\boldsymbol{p}_* = E_P \boldsymbol{Z}_*$ are bounded, and $R_1(P)$ can be rewritten as

$$
R_1(P) = E_P \|\boldsymbol{Z}_* - E_P \boldsymbol{Z}_*\|^2 = E_P \|\boldsymbol{Z}_*\|^2 - \|E_P \boldsymbol{Z}_*\|^2. \tag{16}
$$

Therefore,

$$
\begin{aligned}
R_1(P) + \|\boldsymbol{b} - \boldsymbol{p}_*\|^2 &= E_P \|\boldsymbol{Z}_* - \boldsymbol{p}_*\|^2 + \|\boldsymbol{b} - \boldsymbol{p}_*\|^2 \\
&= E_P \|\boldsymbol{Z}_*\|^2 - \|\boldsymbol{p}_*\|^2 + \|\boldsymbol{b} - \boldsymbol{p}_*\|^2 \\
&= E_P \|\boldsymbol{Z}_*\|^2 - 2\boldsymbol{b}^T p_* + \|\boldsymbol{b}\|^2 \\
&= E_P \|\boldsymbol{Z}_*\|^2 - 2\boldsymbol{b}^T E_P \boldsymbol{Z}_* + \|\boldsymbol{b}\|^2.
\end{aligned}
$$

This implies that

$$
R(\boldsymbol{b}, P) = (\alpha_0 n - m)^2 [E_P \|\boldsymbol{Z}_*\|^2 - 2\boldsymbol{b}^T E_P \boldsymbol{Z}_* + \|\boldsymbol{b}\|^2]. \tag{17}
$$

Obviously, to prove that the decision rule $\boldsymbol{d}_*^0(n, \boldsymbol{U}^n)$ defined by (8) is minimax in $\mathcal{D}_0$ it suffices to show that

$$
\sup_{P \in \mathcal{P}} R(\boldsymbol{b}_0, P) = \inf_{b \in \mathcal{M}} \sup_{P \in \mathcal{P}} R(\boldsymbol{b}, P). \tag{18}
$$

For this we note that $\mathcal{M}$ and $\mathcal{P}$ are convex sets and $\mathcal{M}$ is compact. Moreover, for each fixed $P \in \mathcal{P}$, the mapping $R(\boldsymbol{b}, P) : \mathcal{M} \times \mathcal{P} \to [0, \infty)$ is convex, continuous with respect to $\boldsymbol{b} \in \mathcal{M}$ and, for each fixed $\boldsymbol{b} \in \mathcal{M}$, it is, by (17), concave (linear) with respect $P \in \mathcal{P}$. 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

$$
\sup_{P \in \mathcal{P}} R(\underline{\boldsymbol{b}}, P) = \inf_{b \in \mathcal{M}} \sup_{P \in \mathcal{P}} R(\boldsymbol{b}, P) = \sup_{P \in \mathcal{P}} \inf_{b \in \mathcal{M}} R(\boldsymbol{b}, P). \tag{19}
$$

Now it remains to prove that $\underline{\boldsymbol{b}} = \boldsymbol{b}_0$. We first observe that, by (19) and (15), the minimax risk in $\mathcal{D}_0$ equals

$$
\inf_{b \in \mathcal{M}} \sup_{P \in \mathcal{P}} R(\boldsymbol{b}, P) = \sup_{P \in \mathcal{P}} \inf_{b \in \mathcal{M}} R(\boldsymbol{b}, P) = (\alpha_0 n - m)^2 \sup_{P \in \mathcal{P}} R_1(P), \tag{20}
$$

because, for a fixed distribution $P \in \mathcal{P}$, the convex function $R(\boldsymbol{b}, P)$ of the variable $\boldsymbol{b}$ attains its global minimum over $\mathcal{M}$ at the point $\boldsymbol{b}(P) = \boldsymbol{p}_*$. Moreover, an easy computation shows that, for each $P \in \mathcal{P}$, $0 < \beta < 1$ and $j \geq 1$,

$$
\begin{aligned}
\sup_{Q \in \mathcal{P}} R_1(Q) &\geq R_1(\beta P + (1 - \beta) P_j) \\
&= \beta R_1(P) + (1 - \beta) R_1(P_j) + \beta(1 - \beta) \|(E_{P_j} \boldsymbol{Z}_* - E_P \boldsymbol{Z}_*\|^2,
\end{aligned}
$$

because $\beta P + (1-\beta)P_j \in \mathcal{P}$ and, by (16),

$$\begin{aligned}R_1(\beta P + (1-\beta)P_j) &= E_{\beta P+(1-\beta)P_j}\|\boldsymbol{Z}_*\|^2 - \|E_{\beta P+(1-\beta)P_j}\boldsymbol{Z}_*\|^2\\ &= \beta E_P\|\boldsymbol{Z}_*\|^2 + (1-\beta)E_{P_j}\|\boldsymbol{Z}_*\|^2 - \|\beta E_P\boldsymbol{Z}_* + (1-\beta)E_{P_j}\boldsymbol{Z}_*\|^2\\ &= \beta(E_P\|\boldsymbol{Z}_*\|^2 - \|E_P\boldsymbol{Z}_*\|^2) + (1-\beta)(E_{P_j}\|\boldsymbol{Z}_*\|^2 - \|E_{P_j}\boldsymbol{Z}_*\|^2)\\ &\quad + \beta(1-\beta)\|E_{P_j}\boldsymbol{Z}_* - E_P\boldsymbol{Z}_*\|^2\\ &= \beta R_1(P) + (1-\beta)R_1(P_j) + \beta(1-\beta)\|E_{P_j}\boldsymbol{Z}_* - E_P\boldsymbol{Z}_*\|^2.\end{aligned}$$

Since $\boldsymbol{b}_0$ is a cluster point of the sequence $(\boldsymbol{b}_j)$, where $\boldsymbol{b}_j = E_{P_j}\boldsymbol{Z}_*$, and since $\lim_{j\to\infty} R_1(P_j) = \sup_{Q\in\mathcal{P}} R_1(Q)$, we conclude that

$$\sup_{Q\in\mathcal{P}} R_1(Q) \ge \beta R_1(P) + (1-\beta)\sup_{Q\in\mathcal{P}} R_1(Q) + \beta(1-\beta)\|\boldsymbol{b}_0 - E_P\boldsymbol{Z}_*\|^2.$$

Therefore,

$$\beta\sup_{Q\in\mathcal{P}} R_1(Q) \ge \beta R_1(P) + \beta(1-\beta)\|\boldsymbol{b}_0 - E_P\boldsymbol{Z}_*\|^2$$

and, since β is positive,

$$\sup_{Q\in\mathcal{P}} R_1(Q) \ge R_1(P) + (1-\beta)\|\boldsymbol{b}_0 - E_P\boldsymbol{Z}_*\|^2.$$

Letting $\beta \to 0^+$, we can see that

$$\sup_{Q\in\mathcal{P}} R_1(Q) \ge R_1(P) + \|\boldsymbol{b}_0 - E_P\boldsymbol{Z}_*\|^2 = R_1(P) + \|\boldsymbol{b}_0 - \boldsymbol{p}_*\|^2,$$

which implies, by (15), that

$$(\alpha_0 n - m)^2\sup_{Q\in\mathcal{P}} R_1(Q) \ge (\alpha_0 n - m)^2[R_1(P) + \|\boldsymbol{b}_0 - \boldsymbol{p}_*\|^2] = R(\boldsymbol{b}_0, P).$$

Because this is true for all $P\in\mathcal{P}$, it follows, by (20), that

$$\sup_{P\in\mathcal{P}} R(\boldsymbol{b}_0, P) \le (\alpha_0 n - m)^2\sup_{P\in\mathcal{P}} R_1(P) = \inf_{b\in\mathcal{M}}\sup_{P\in\mathcal{P}} R(\boldsymbol{b}, P) \le \sup_{P\in\mathcal{P}} R(\boldsymbol{b}_0, P).$$

Thus the predictor $\boldsymbol{d}_*^0(n, \boldsymbol{U}^n) = \boldsymbol{d}^{\boldsymbol{b}_0}(n, \boldsymbol{U}^n)$ is minimax in $\mathcal{D}_0$ and its minimax risk is given by

$$\sup_{P\in\mathcal{P}} R(\boldsymbol{d}_*^0, P) = (\alpha_0 n - m)^2\sup_{P\in\mathcal{P}} R_1(P). \tag{21}$$

To prove that $\boldsymbol{d}_*^0(n, \boldsymbol{U}^n)$ is minimax in $\mathcal{D}$ we assume first that $m = 1$. Then $\alpha_0 = 0$ and, for any predictor $\boldsymbol{d}\in\mathcal{D}$, we obtain, by (11) and (21),

$$\begin{aligned}\sup_{P\in\mathcal{P}} R(\boldsymbol{d}, P) &\ge m\sup_{P\in\mathcal{P}} R_1(P) = \sup_{P\in\mathcal{P}} R_1(P) = (\alpha_0 n - m)^2\sup_{P\in\mathcal{P}} R_1(P)\\ &= \sup_{P\in\mathcal{P}} R(\boldsymbol{d}_*^0, P),\end{aligned}$$

which implies minimaxity of $\boldsymbol{d}_*^0(n, \boldsymbol{U}^n)$ in the case where $m = 1$.

Now we assume that $m > 1$. Then $\alpha_0 > 0$ and to show that $\boldsymbol{d}^0_*(n, \boldsymbol{U}^n)$ is minimax in $\mathcal{D}$ we make use of the nonparametric Bayes approach proposed in Ferguson (1973). The structure of the argument will be the same as in Wilczyński (1992).

For each $j \geq 1$ we denote by Π_j a Dirichlet prior process on $(\mathcal{Y}, \mathcal{B})$ with parameter $\beta_j = [(m - \alpha_0 n)/\alpha_0]P_j$, where (P_j) is a sequence defined by (3). From Ferguson (1973), Example b, the Π_j Bayes nonparametric estimator of $m\boldsymbol{p}_* = mE_P\boldsymbol{Z}_*$, and therefore, by (11), the Π_j Bayes nonparametric predictor of $\boldsymbol{Y}^m_*$ is given by

$$m\left[\frac{(m-\alpha_0 n)/\alpha_0}{n + (m - \alpha_0 n)/\alpha_0} E_{P_j}\boldsymbol{Z}_* + \frac{n}{n + (m-\alpha_0 n)/\alpha_0} \cdot \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{z}_*(U_j)\right]$$
$$= m\left[\frac{m - \alpha_0 n}{m}\boldsymbol{b}_j + \frac{\alpha_0}{m}\boldsymbol{X}^n_*\right] = \boldsymbol{d}^{\boldsymbol{b}_j}(n, \boldsymbol{U}^n).$$

Moreover, the Bayes risk $\varrho(j)$ for this decision rule has the form

$$\begin{aligned}\varrho(j) &:= E_{\Pi_j} R(\boldsymbol{b}_j, P) = (\alpha_0 n - m)^2[E_{P_j}\|\boldsymbol{Z}_*\|^2 - \|\boldsymbol{b}_j\|^2] \\ &= (\alpha_0 n - m)^2[E_{P_j}\|\boldsymbol{Z}_*\|^2 - \|E_{P_j}\boldsymbol{Z}_*\|^2] = (\alpha_0 n - m)^2 R_1(P_j),\end{aligned}$$

because, by (17),

$$R(\boldsymbol{b}_j, P) = (\alpha_0 n - m)^2[E_P\|\boldsymbol{Z}_*\|^2 - 2\boldsymbol{b}_j^T E_P\boldsymbol{Z}_* + \|\boldsymbol{b}_j\|^2]$$

and, by Ferguson (1973), Theorem 3,

$$E_{\Pi_j}[E_P\|\boldsymbol{Z}_*\|^2] = E_{P_j}\|\boldsymbol{Z}_*\|^2 \quad \text{and} \quad E_{\Pi_j}[E_P\boldsymbol{Z}_*] = E_{P_j}\boldsymbol{Z}_* = \boldsymbol{b}_j. \tag{22}$$

As $j \to \infty$, the Bayes risk $\varrho(j)$ converges to $(\alpha_0 n - m)^2 \sup_{P\in\mathcal{P}} R_1(P)$, which, by (21), is the upper bound for the risk of $\boldsymbol{d}^0_*(n, U^n)$. This implies that $\boldsymbol{d}^0_*(n, \boldsymbol{U}^n)$ is a minimax predictor of $\boldsymbol{Y}^m_*$ (see Ferguson 1967, Theorem 2, p. 91) and thus the proof of the first part of Theorem 1 is complete.

We now turn to the proof of the second part. Since we assume that $\sup_{y\in\mathcal{Y}} \boldsymbol{z}(y)^T\boldsymbol{C}\boldsymbol{z}(y) = \infty$, there exists a sequence $(y_j) \subset \mathcal{Y}$ such that

$$\lim_{j\to\infty} \|\boldsymbol{z}_*(y_j)\|^2 = \infty. \tag{23}$$

Let the distribution P_j of U_0 be defined by

$$P_j(U_0 = y_1) = P_j(U_0 = y_j) = 0.5.$$

Then $\lim_{j\to\infty} R_1(P_j) = \sup_{P\in\mathcal{P}} R_1(P) = \infty$, because an easy calculation shows that

$$R_1(P_j) = \frac{\|\boldsymbol{z}_*(y_j) - \boldsymbol{z}_*(y_1)\|^2}{4},$$

and, by the triangle inequality and (23),

$$\|\boldsymbol{z}_*(y_j) - \boldsymbol{z}_*(y_1)\| \geq \|\boldsymbol{z}_*(y_j)\| - \|\boldsymbol{z}_*(y_1)\| \to \infty.$$

Therefore, the sequence of Bayes risks $\varrho(j)$ defined above converges to ∞. This implies, in turn, that the risk of any predictor $d(n, \boldsymbol{U}^n) \in \mathcal{D}$ is unbounded, because

$$\sup_{P \in \mathcal{P}} R(\boldsymbol{d}, P) \geq E_{\Pi_j} R(\boldsymbol{d}, P) \geq \varrho(j) \to \infty.$$

The proof of Theorem 1 is complete.

The first part of Theorem 1 can be slightly generalized. For this we denote by $\boldsymbol{I}$ the k-dimensional identity matrix and we put $H = (\boldsymbol{C}^{1/2})^{-}\boldsymbol{C}^{1/2}$. Since $\boldsymbol{C}^{1/2}(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0}$, we have the following result:

Corollary 1. *If* $\sup_{y \in \mathcal{Y}} \boldsymbol{z}(y)^T \boldsymbol{C} \boldsymbol{z}(y) < \infty$ *then*, *for each* $\boldsymbol{c}_0 \in \mathbb{R}^k$, *the decision rule*

$$\boldsymbol{d}^0(n, \boldsymbol{U}^n) = \alpha_0 \boldsymbol{X}^n + (m - \alpha_0 n)\boldsymbol{a}_0 + (\boldsymbol{I} - \boldsymbol{H})c_0$$

is a minimax predictor of $\boldsymbol{Y}^m$.

4. Examples. As an application of the results obtained we consider the following three examples.

Example 1. Suppose that the set $\mathcal{Y}$ is centrosymmetric about $\boldsymbol{0}$ and that, for each $y \in \mathcal{Y}$, $\boldsymbol{z}_*(y) = -\boldsymbol{z}_*(-y)$. Let (P_j) be a sequence for which (3) holds and let P_j^- denote the distribution of the random vector $-U_0$ whenever U_0 is distributed according to P_j. Then the sequence (P_j'), with $P_j' = (1/2)(P_j + P_j^-)$, satisfies (3), because

$$\begin{aligned} R_1(P_j') &= E_{P_j'} \|\boldsymbol{Z}_*\|^2 - \|E_{P_j'} \boldsymbol{Z}_*\|^2 = E_{P_j} \|\boldsymbol{Z}_*\|^2 - \|\boldsymbol{0}\|^2 \\ &\geq E_{P_j} \|\boldsymbol{Z}_*\|^2 - \|E_{P_j} \boldsymbol{Z}_*\|^2 = R_1(P_j). \end{aligned}$$

Therefore, we may assume that $\boldsymbol{b}_j = E_{P_j'} \boldsymbol{Z}_* = \boldsymbol{0}$, which implies that $\boldsymbol{b}_0 = \boldsymbol{a}_0 = \boldsymbol{0}$ and thus the decision rule

$$\boldsymbol{d}^0(n, \boldsymbol{U}^n) = \alpha_0 \boldsymbol{X}^n$$

is a minimax predictor of the unobservable vector $\boldsymbol{Y}^m$.

Example 2. Suppose that $\boldsymbol{C} = [c_{ij}]$ is a diagonal matrix and that there exist two sequences $\{\overline{y}_j\}$ and $\{\overline{\overline{y}}_j\}$ in $\mathcal{Y}$ such that, for each $1 \leq i \leq k$,

$$\lim_{j \to \infty} z_i(\overline{y}_j) = \inf_{y \in Y} z_i(y) > -\infty, \qquad \lim_{j \to \infty} z_i(\overline{\overline{y}}_j) = \sup_{y \in Y} z_i(y) < \infty.$$

Let the distribution P_j of U_0, $j \geq 1$, be defined by

$$P_j(U_0 = \overline{y}_j) = P_j(U_0 = \overline{\overline{y}}_j) = 0.5.$$

Then it is easy to verify that for each $1 \le i \le k$,

$$\sup_{P\in\mathcal{P}}[E_P(z_i(U_0))^2 - (E_P z_i(U_0))^2] = \lim_{j\to\infty}[E_{P_j}(z_i(U_0))^2 - (E_{P_j} z_i(U_0))^2] = \lim_{j\to\infty}\frac{|z_i(\overline{y}_j) - z_i(\overline{\overline{y}}_j)|^2}{4}.$$

This implies that (P_j) is a sequence of distributions as in (3), because $\boldsymbol{C}$ is assumed to be a diagonal matrix and thus

$$R_1(P) = \sum_{i=1}^{k} c_{ii}[E_P(z_i(U_0))^2 - (E_P z_i(U_0))^2].$$

Since the function $\boldsymbol{z}(y)$ is bounded on $\mathcal{Y}$,

$$\boldsymbol{C}^{1/2}a_0 = \boldsymbol{b}_0 = \lim_{j\to\infty} E_{P_j}\boldsymbol{C}^{1/2}\boldsymbol{z}(U_0) = \boldsymbol{C}^{1/2}\lim_{j\to\infty} E_{P_j}\boldsymbol{z}(U_0) = \boldsymbol{C}^{1/2}\lim_{j\to\infty}\frac{\boldsymbol{z}(\overline{y}_j) + \boldsymbol{z}(\overline{\overline{y}}_j)}{2}.$$

Therefore, the coordinates of the point $\boldsymbol{a}_0 = (a_{01}, a_{02}, \ldots, a_{0k})^T$ are given by

$$a_{0i} = \lim_{j\to\infty}\frac{z_i(\overline{y}_j) + z_i(\overline{\overline{y}}_j)}{2} = \frac{\inf_{y\in\mathcal{Y}} z_i(y) + \sup_{y\in\mathcal{Y}} z_i(y)}{2}, \qquad 1 \le i \le k,$$

and

$$\boldsymbol{d}^0(n, \boldsymbol{U}^n) = \alpha_0\boldsymbol{X}^n + (m - \alpha_0 n)\boldsymbol{a}_0$$

is a minimax predictor of the unobservable vector $\boldsymbol{Y}^m$.

EXAMPLE 3. Let $\{A_i\}_{i=1}^k$ be a measurable partition of $\mathcal{Y}$, i.e. let $A_1, \ldots$ $\ldots, A_k$ be measurable, pairwise disjoint subsets of $\mathcal{Y}$ whose union equals $\mathcal{Y}$. Furthermore, let $z_i(y) = \mathbf{1}_{A_i}(y)$, $1 \le i \le k$, be the indicator functions. Then the random vectors $\boldsymbol{Z} = \boldsymbol{z}(U_0)$, $\boldsymbol{X}^n$ and $\boldsymbol{Y}^m$ have $(1, \boldsymbol{p})$, $(n, \boldsymbol{p})$ and $(m, \boldsymbol{p})$ multinomial distributions, respectively, in which the parameter $\boldsymbol{p} = E_P\boldsymbol{Z}$ takes its values in the simplex S defined by

$$S = \{(s_1, \ldots, s_k) : \text{for all } 1 \le i \le k,\ s_i \ge 0, \text{ and } s_1 + \ldots + s_k = 1\}.$$

Furthermore, it is easy to calculate that

$$R_1(P) = \boldsymbol{c}^T\boldsymbol{p} - \boldsymbol{p}^T\boldsymbol{C}\boldsymbol{p},$$

where $\boldsymbol{c} = (c_{11}, c_{22}, \ldots, c_{kk})^T$ stands for the diagonal of the matrix $\boldsymbol{C} = [c_{ij}]$. This function attains its maximum over $\mathcal{P}$ at the distribution P_0 for which $E_{P_0}\boldsymbol{Z} = \boldsymbol{p}_0$, where

$$\boldsymbol{c}^T\boldsymbol{p}_0 - \boldsymbol{p}_0^T\boldsymbol{C}\boldsymbol{p}_0 = \max_{p\in S}[\boldsymbol{c}^T\boldsymbol{p} - \boldsymbol{p}^T\boldsymbol{C}\boldsymbol{p}].$$

Therefore, $\boldsymbol{b}_0 = E_{P_0}\boldsymbol{Z}_* = E_{P_0}\boldsymbol{C}^{1/2}\boldsymbol{Z} = \boldsymbol{C}^{1/2}E_{P_0}\boldsymbol{Z} = \boldsymbol{C}^{1/2}\boldsymbol{p}_0$ and

$$\boldsymbol{d}^0(n, \boldsymbol{U}^n) = \alpha_0\boldsymbol{X}^n + (m - \alpha_0 n)\boldsymbol{p}_0$$

is a minimax predictor of the unobservable vector $\boldsymbol{Y}^m$.

References

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

T. S. Ferguson (1967), *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.

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

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@neyman.im.pwr.wroc.pl

Received on 10.4.2000;
revised version on 27.9.2000 (1537)