APPLICATIONES MATHEMATICAE
25,3 (1998), pp. 271–283

A. JOKIEL-ROKITA (Wrocław)

MINIMAX PREDICTION FOR THE MULTINOMIAL AND MULTIVARIATE HYPERGEOMETRIC DISTRIBUTIONS

Abstract. A problem of minimax prediction for the multinomial and multivariate hypergeometric distribution is considered. A class of minimax predictors is determined for estimating linear combinations of the unknown parameter and the random variable having the multinomial or the multivariate hypergeometric distribution.

1. Introduction. The problem considered in the 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, where X and Y have a distribution dependent on the same unknown parameter. The paper deals with a special form of such problems—namely, with the problem of finding a minimax predictor for the multinomial and multivariate hypergeometric distributions. In the paper of Trybuła (1958) a minimax estimator was found for estimating the parameters of the multivariate hypergeometric distribution and of the multinomial distribution under a weighted quadratic loss function. Wilczyński (1985) obtained a minimax predictor of a random variable distributed according to the multinomial law when the loss function has a more general form than in Trybuła (1958).

In this paper it is assumed that $\mathbf{X}$ and $\mathbf{Y}$ are random variables having the multinomial or multivariate hypergeometric distribution with the unknown parameter $\mathbf{p} = (p_1, \ldots, p_r)$. Assuming the loss function to be of the form (1) below, in both cases minimax predictors of linear combinations of the form $\mathbf{Z} = a\mathbf{p} + b\mathbf{Y}$ are determined. The results obtained generalize the corresponding results of Trybuła (1958) and Wilczyński (1985).

1991 *Mathematics Subject Classification*: Primary 62F15.

Key words and phrases: Bayes estimation, Bayes risk, minimax estimation, multinomial distribution, multivariate hypergeometric distribution, minimax predictor.

2. Minimax prediction for the multinomial distribution. Let $\mathbf{X} = (X_1, \ldots, X_r)$ be a random variable having the multinomial distribution with parameters $(n, \mathbf{p})$, i.e.,

$$P_{\mathbf{p}}\{\mathbf{X} = \mathbf{x} = (x_1, \ldots, x_r)\} = \begin{cases} \dfrac{n!}{x_1! \ldots x_r!} p_1^{x_1} \ldots p_r^{x_r} & \text{if } x_i \in \{0, 1, \ldots, n\} \text{ and } x_1 + \ldots + x_r = n, \\ 0 & \text{otherwise,} \end{cases}$$

where $p_i \geq 0$, $i = 1, \ldots, r$, and $p_1 + \ldots + p_r = 1$. We observe the values of this random variable and using an estimate $\mathbf{d}(\mathbf{X}) = (d_1(\mathbf{X}), \ldots, d_r(\mathbf{X}))$ we want to estimate the linear combination $\mathbf{Z} = a\mathbf{p} + b\mathbf{Y}$ of the unknown parameter $\mathbf{p}$ and the random variable $\mathbf{Y} = (Y_1, \ldots, Y_r)$ which has the multinomial distribution with parameters $(m, \mathbf{p})$. We assume that a, b, n and m are known, $\mathbf{X}$ and $\mathbf{Y}$ are independent and the loss connected with the estimator $\mathbf{d}(\mathbf{X})$ is of the form

$$L(\mathbf{d}, \mathbf{Z}) = \sum_{i,j=1}^{r} c_{ij}(d_i - Z_i)(d_j - Z_j), \tag{1}$$

where the matrix $C = (c_{ij})$ is nonnegative definite and $Z_i = ap_i + bY_i$.

We shall be interested in finding a minimax estimator of $\mathbf{Z}$, that is, an estimator $\mathbf{d}_0(\mathbf{X}) = (d_1^0(\mathbf{X}), \ldots, d_r^0(\mathbf{X}))$ for which

$$\sup_{\mathbf{Z} \in \mathcal{Z}} R(\mathbf{d}_0, \mathbf{Z}) = \inf_{\mathbf{d}} \sup_{\mathbf{Z} \in \mathcal{Z}} R(\mathbf{d}, \mathbf{Z}),$$

where $R(\mathbf{d}, \mathbf{Z}) = E_{\mathbf{p}}\{L[\mathbf{d}(\mathbf{X}), \mathbf{Z}]\}$ is the risk function and $\mathcal{Z} = \{\mathbf{Z} = (Z_1, \ldots, Z_r) \in \mathbb{R}^r : Z_1 + \ldots + Z_r = a + bm\}$.

This problem was considered by Trybuła (1958) in the case when $a = 1$ and $b = 0$, and C is an arbitrary nonnegative definite diagonal matrix, and also by Wilczyński (1985) in the case when $a = 1$ and $b = 0$ and when $a = 0$ and $b = 1$ and C is an arbitrary nonnegative definite matrix.

The following two theorems and the lemma will be used to prove the main results of the paper established in Theorems 3 and 4.

THEOREM 1 (Sion, see Aubin (1979)). *Let $g : \mathcal{P} \times \mathcal{Q} \to \mathbb{R}$. Suppose that*

(a) *$\mathcal{P}$ and $\mathcal{Q}$ are convex, compact subsets of Euclidean spaces,*
(b) *$p \mapsto g(p, q)$ is convex and continuous for each $q \in \mathcal{Q}$,*
(c) *$q \mapsto g(p, q)$ is concave and continuous for each $p \in \mathcal{P}$.*

Then there exists a saddle point $(\overline{p}, \overline{q}) \in \mathcal{P} \times \mathcal{Q}$, i.e., a point $(\overline{p}, \overline{q})$ for which

$$\inf_{p \in \mathcal{P}} \sup_{q \in \mathcal{Q}} g(p, q) = \sup_{q \in \mathcal{Q}} g(\overline{p}, q) = g(\overline{p}, \overline{q}) = \inf_{p \in \mathcal{P}} g(p, \overline{q}) = \sup_{q \in \mathcal{Q}} \inf_{p \in \mathcal{P}} g(p, q).$$

THEOREM 2 (Karmanov (1986), Theorem 3.5.4). *Let $g : \mathcal{Q} \to \mathbb{R}$ be a convex function defined on a convex subset $\mathcal{Q} = \{q = (q_1, \ldots, q_n) \in \mathbb{R}^n :$*

$a_i^T q - b_i \geq 0,\ i = 1, \ldots, m\}$ *for some* $a_1, \ldots, a_m \in \mathbb{R}^n$ *and* $b_1, \ldots, b_m \in \mathbb{R}$. *An element* $\overline{q} \in \mathcal{Q}$ *is a solution to the equation* $\inf_{q\in\mathcal{Q}} g(q) = g(\overline{q})$ *iff there exist real numbers* $u_1, \ldots, u_m$, $u_i \geq 0$, $i = 1, \ldots, m$, *for which*

$$g'(\overline{q}) = \sum_{i=1}^{m} u_i a_i, \qquad \sum_{i=1}^{m} (a_i^T \overline{q} - b_i) u_i = 0,$$

where $g'(q)$ *stands for the gradient of* g *at* q.

LEMMA 1 (Ferguson (1967)). *Let* π *be an a priori distribution of a parameter* θ *and let* $r(\pi, d) = E_\pi R(d, \theta)$ *denote the Bayes risk of an estimator* d *of* θ. *If* d_0 *is a Bayes estimator of* θ *with respect to an a priori distribution* π_0 *and*

$$\sup_{\theta\in\Theta} R(d_0, \theta) = r(d_0, \pi_0),$$

then d_0 *is minimax.*

Define $\mathbf{c} = (c_{11}, \ldots, c_{rr})$ and $\mathcal{P} = \{\mathbf{p} = (p_1, \ldots, p_r) : p_i \geq 0,\ p_1 + \ldots + p_r = 1\}$. Let $\pi_{\alpha,\beta}$ stand for the a priori Dirichlet distribution $D(\alpha\beta_1, \ldots, \alpha\beta_r)$ of $\mathbf{p}$ with density

$$h(p_1, \ldots, p_r) = \begin{cases} \dfrac{\Gamma(\alpha)}{\prod_{j\in A} \Gamma(\alpha\beta_j)} \prod_{j\in A} p_j^{\alpha\beta_j - 1} & \text{if } \sum_{i=1}^{r} p_i = \sum_{i\in A} p_i = 1, \\ & p_i \geq 0,\ i = 1, \ldots, r, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

where $\alpha > 0$, $\beta = (\beta_1, \ldots, \beta_r) \in \mathcal{P}$ and $A = \{i \in \{1, \ldots, r\} : \beta_i > 0\}$. We denote by π_β the a priori distribution of $\mathbf{p}$ defined by $P(\mathbf{p} = \beta) = 1$. The following lemma determines the Bayes predictors of $\mathbf{Z}$ with respect to the a priori distributions $\pi_{\alpha,\beta}$ and π_β of $\mathbf{p}$.

LEMMA 2. *Under the loss function given by* (1) *with* C *nonnegative definite*, *the predictors*

$$d_j^{\alpha,\beta}(\mathbf{X}) = (a + bm)\left(\frac{1}{n+\alpha} X_j + \frac{\alpha}{n+\alpha} \beta_j\right), \tag{3}$$

$$d_j^{\beta}(\mathbf{X}) = (a + bm)\beta_j \tag{4}$$

of the linear combination $\mathbf{Z}$ *are Bayes w.r.t. the a priori distributions* $\pi_{\alpha,\beta}$ *and* π_β, *respectively*, *and their Bayes risks are*

$$r(\pi_{\alpha,\beta}, \mathbf{d}^{\alpha,\beta}) = w_1 \left[\sum_{i\neq j} c_{ij} \frac{\alpha\beta_i\beta_j}{\alpha+1} + \sum_i c_{ii} \frac{\beta_i(\alpha\beta_i + 1)}{\alpha+1}\right] - w_2 \beta^T C \beta + w_3 c^T \beta, \tag{5}$$

$$r(\pi_\beta, \mathbf{d}^\beta) = b^2 m (c^T \beta - \beta^T C \beta), \tag{6}$$

where

$$w_1 = (a+bm)^2 \frac{\alpha^2 - n}{(n+\alpha)^2} - b^2 m = w_2 - w_3, \tag{7}$$

$$w_2 = (a+bm)^2 \frac{\alpha^2}{(n+\alpha)^2}, \tag{8}$$

$$w_3 = (a+bm)^2 \frac{n}{(n+\alpha)^2} + b^2 m. \tag{9}$$

Proof. For any predictor $\mathbf{d}(\mathbf{X})$, the loss function $L(\mathbf{d}, \mathbf{Z})$ can be rewritten in the form

$$\begin{aligned} L(\mathbf{d}, \mathbf{Z}) &= \sum_{i,j=1}^{r} c_{ij}(d_i - Z_i)(d_j - Z_j) \\ &= \sum_{i,j=1}^{r} c_{ij}(a+bm)^2 \left(\frac{d_i}{a+bm} - p_i\right)\left(\frac{d_j}{a+bm} - p_j\right) \\ &\quad - \sum_{i,j=1}^{r} c_{ij} b(a+bm)\left(\frac{d_i}{a+bm} - p_i\right)(Y_j - mp_j) \\ &\quad - \sum_{i,j=1}^{r} c_{ij} b(a+bm)(Y_i - mp_i)\left(\frac{d_j}{a+bm} - p_j\right) \\ &\quad + \sum_{i,j=1}^{r} c_{ij} b^2 (Y_i - mp_i)(Y_j - mp_j), \end{aligned}$$

so

$$\begin{aligned} R(\mathbf{d}, \mathbf{Z}) &= E_{\mathbf{p}}\Big[\sum_{i,j=1}^{r} c_{ij}(a+bm)^2 \left(\frac{d_i}{a+bm} - p_i\right)\left(\frac{d_j}{a+bm} - p_j\right)\Big] \\ &\quad + E_{\mathbf{p}}\Big[\sum_{i,j=1}^{r} c_{ij} b^2 (Y_i - mp_i)(Y_j - mp_j)\Big]. \end{aligned} \tag{10}$$

If we want to find an estimator $\mathbf{d}_0$ such that $r(\pi, \mathbf{d}_0) = \inf_{\mathbf{d} \in \mathcal{D}} r(\pi, \mathbf{d})$ (for any a priori distribution π), it is sufficient to find one for which the expectation

$$E_\pi\Big\{E_{\mathbf{p}}\Big[\sum_{i,j=1}^{r} c_{ij}(a+bm)^2 \left(\frac{d_i}{a+bm} - p_i\right)\left(\frac{d_j}{a+bm} - p_j\right)\Big]\Big\}$$

attains its minimum. Thus the predictor will be the product of $a + bm$ and the Bayes estimator of the parameter $\mathbf{p}$, so that it is given by (3) if $\pi = \pi_{\alpha,\beta}$, and by (4) if $\pi = \pi_\beta$. The risk function associated with the predictor $\mathbf{d}^{\alpha,\beta}(\mathbf{X})$ given by (3) is

$$R(\mathbf{d}^{\alpha,\beta},\mathbf{Z}) = w_1\mathbf{p}^TC\mathbf{p} + w_2\beta^TC\beta + w_3\mathbf{c}^T\mathbf{p} - 2w_2\beta^TC\mathbf{p}, \tag{11}$$

where w_1, w_2 and w_3 are given by (7), (8) and (9), respectively. Using the Liouville equation (Fichtenholz (1985), Vol. 3), we can show that

$$E_{\pi_{\alpha,\beta}}(p_i) = \beta_i, \quad E_{\pi_{\alpha,\beta}}(p_i^2) = \frac{\beta_i(\alpha\beta_i+1)}{\alpha+1}, \quad E_{\pi_{\alpha,\beta}}(p_ip_j) = \frac{\alpha\beta_i\beta_j}{\alpha+1},$$

so that $r(\pi_{\alpha,\beta},\mathbf{d}^{\alpha,\beta})$ is of the form (5). The risk function associated with the predictor $\mathbf{d}^{\beta}(\mathbf{X})$ given by (4) is

$$\begin{aligned} R(\mathbf{d}^{\beta}(\mathbf{X}),\mathbf{Z}) = {} & (a+bm)^2\beta^TC\beta - 2(a+bm)^2\beta^TC\mathbf{p} \\ & + [(a+bm)^2 - b^2m]\mathbf{p}^TC\mathbf{p} + b^2m\mathbf{c}^T\mathbf{p}, \end{aligned} \tag{12}$$

so that the Bayes risk $r(\pi_\beta,\mathbf{d}^\beta)$ is given by (6). ■

The following theorem determines a minimax predictor of $\mathbf{Z}$.

THEOREM 3. *Under the loss function given by* (1) *with* C *nonnegative definite*, *the predictor of* $\mathbf{Z}$ *defined by*

$$\mathbf{d}_0(\mathbf{X}) = \begin{cases} (a+bm)\left(\dfrac{1}{n+\alpha_0}\mathbf{X} + \dfrac{\alpha_0}{n+\alpha_0}\beta_0\right) & \text{if } (a+bm)^2 - b^2m > 0, \\ (a+bm)\beta_0 & \text{otherwise}, \end{cases}$$

where

$$\alpha_0 = \begin{cases} \dfrac{nb^2m + |a+bm|\sqrt{(a+bm)^2n + b^2mn(n-1)}}{(a+bm)^2 - b^2m} & \text{if } n > 1, \\ \dfrac{(a+bm)^2 + b^2m}{(a+bm)^2 - b^2m} & \text{if } n = 1, \end{cases} \tag{13}$$

and β_0 *is a point* $(\beta_1^0,\ldots,\beta_r^0)$ *for which*

$$\mathbf{c}^T\beta_0 - \beta_0^TC\beta_0 = \max_{\beta\in\mathcal{P}}(\mathbf{c}^T\beta - \beta^TC\beta), \tag{14}$$

is minimax.

Proof. Consider the Bayes predictors $\mathbf{d}^{\alpha,\beta}(\mathbf{X})$ of the linear combination $\mathbf{Z}$ with respect to the a priori distribution $\pi_{\alpha,\beta}$ of the parameter $\mathbf{p}$, which are of the form defined by (3). The risk function associated with the predictor (3) is of the form (11). If $(a+bm)^2 - b^2m > 0$, then there exists $\alpha_0 > 0$ for which $w_1 = 0$. It is easy to check that α_0 is of the form (13). Set $R_1(\beta,\mathbf{p}) = R(\mathbf{d}^{\alpha_0,\beta},\mathbf{Z})$. Taking $\alpha = \alpha_0$ yields $w_1 = 0$ and $w_2 = w_3$, and, consequently,

$$R_1(\beta,\mathbf{p}) = w_2(\beta^TC\beta + \mathbf{c}^T\mathbf{p} - 2\beta^TC\mathbf{p}). \tag{15}$$

Notice that the function $R_1(\cdot,\cdot) : \mathcal{P}\times\mathcal{P} \to \mathbb{R}$ is convex w.r.t. β, concave w.r.t. $\mathbf{p}$, continuous w.r.t. $(\beta,\mathbf{p})$, and $\mathcal{P}$ is a convex, compact subset of $\mathbb{R}^r$.

From Theorem 1 it follows that there exists a point $(\beta_0, \mathbf{p}_0) \in \mathcal{P} \times \mathcal{P}$ such that

$$\begin{aligned} \inf_{\beta\in\mathcal{P}} \sup_{\mathbf{p}\in\mathcal{P}} R_1(\beta,\mathbf{p}) &= \sup_{\mathbf{p}\in\mathcal{P}} R_1(\beta_0,\mathbf{p}) = R_1(\beta_0,\mathbf{p}_0) \\ &= \inf_{\beta\in\mathcal{P}} R_1(\beta,\mathbf{p}_0) = \sup_{\mathbf{p}\in\mathcal{P}} \inf_{\beta\in\mathcal{P}} R_1(\beta,\mathbf{p}). \end{aligned} \tag{16}$$

It is well known that for the components of the saddle point, one can choose $\mathbf{p}_0$ and β_0—independently—at which the outer extrema are attained in the following minimaxes:

$$\begin{aligned} \max_{\mathbf{p}\in\mathcal{P}} \inf_{\beta\in\mathcal{P}} R_1(\beta,\mathbf{p}) &= \inf_{\beta\in\mathcal{P}} R_1(\beta,\mathbf{p}_0), \\ \min_{\beta\in\mathcal{P}} \sup_{\mathbf{p}\in\mathcal{P}} R_1(\beta,\mathbf{p}) &= \sup_{\mathbf{p}\in\mathcal{P}} R_1(\beta_0,\mathbf{p}). \end{aligned} \tag{17}$$

Temporarily assuming that the matrix C is positive definite we see that the point $\mathbf{p}_0 = (p_1^0, \ldots, p_r^0)$ for which

$$\mathbf{c}^T\mathbf{p}_0 - \mathbf{p}_0^T C\mathbf{p}_0 = \max_{\mathbf{p}\in\mathcal{P}} (\mathbf{c}^T\mathbf{p} - \mathbf{p}^T C\mathbf{p}) \tag{18}$$

is a unique solution to (17). On the other hand, the strictly convex function $R_1(\beta, \mathbf{p}_0)$ of the variable β (because we assume that C is positive definite) attains its unique minimum at $\beta_0 = \mathbf{p}_0$. Hence, $(\mathbf{p}_0, \mathbf{p}_0)$ is the only saddle point.

In order to find $\mathbf{p}_0$ for which (18) holds, also in the case when C is nonnegative definite, we use Theorem 2. In our case:

$$\begin{gathered} a_1 = (1,0,\ldots,0), \quad a_2 = (0,1,\ldots,0), \ \ldots, \ a_r = (0,0,\ldots,1), \\ a_{r+1} = a_1 + \ldots + a_r, \quad a_{r+2} = -a_{r+1}, \\ b_i = 0, \quad i = 1,\ldots,r, \quad b_{r+1} = -b_{r+2} = 1. \end{gathered}$$

Thus $\mathbf{p}_0 = (p_1^0, \ldots, p_r^0)$ is a solution to (18) iff there exists a constant z_0 such that for all $i \in \{1, \ldots, r\}$,

$$\text{if} \quad p_i^0 > 0, \quad \text{then} \quad 2\sum_{j=1}^r c_{ij} p_j^0 - c_{ii} = z_0,$$

and

$$\text{if} \quad p_i^0 = 0, \quad \text{then} \quad 2\sum_{j=1}^r c_{ij} p_j^0 - c_{ii} \geq z_0.$$

The predictor $\mathbf{d}^{\alpha_0,\mathbf{p}_0}(\mathbf{X})$ is Bayes w.r.t. the a priori distribution $\pi_{\alpha_0,\mathbf{p}_0}$. The Bayes risk associated with the a priori distribution $\pi_{\alpha_0,\mathbf{p}_0}$ and the predictor $\mathbf{d}^{\alpha_0,\mathbf{p}_0}(\mathbf{X})$ is

$$r(\pi_{\alpha_0,\mathbf{p}_0}, \mathbf{d}^{\alpha_0,\mathbf{p}_0}) = E_{\pi_{\alpha_0,\mathbf{p}_0}}[R(\mathbf{d}^{\alpha_0,\mathbf{p}_0}, \mathbf{Z})] = R_1(\mathbf{p}_0, \mathbf{p}_0).$$

Making use of (16) with $\beta_0 = \mathbf{p}_0$ yields

$$\sup_{\mathbf{Z}\in\mathcal{Z}} R(\mathbf{d}^{\alpha_0,\mathbf{p}_0},\mathbf{Z}) = \sup_{\mathbf{p}\in\mathcal{P}} R_1(\mathbf{p}_0,\mathbf{p}) = R_1(\mathbf{p}_0,\mathbf{p}_0) = r(\pi_{\alpha_0,\mathbf{p}_0},\mathbf{d}^{\alpha_0,\mathbf{p}_0}).$$

Then $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\alpha_0,\mathbf{p}_0}(\mathbf{X})$ is minimax by Lemma 1.

In the case when $(a+bm)^2 - b^2m \le 0$ consider the Bayes predictors $\mathbf{d}^\beta(\mathbf{X}) = (a+bm)\beta$ w.r.t. the a priori distributions π_β of the parameter $\mathbf{p}$. The risk function of $\mathbf{d}^\beta(\mathbf{X})$ is given by (12). Temporarily assume $a+bm \ne 0$; then it turns out that this function is convex w.r.t. β, concave w.r.t. $\mathbf{p}$, and continuous w.r.t. $(\beta,\mathbf{p})$. In the same way as in the case $(a+bm)^2-b^2m>0$ we can show that (β_0,β_0), where β_0 is a solution to equation (14), is a saddle point of $R(\mathbf{d}^\beta,\mathbf{Z})$. Moreover,

$$\sup_{\mathbf{p}\in\mathcal{P}} R(\mathbf{d}^{\beta_0},\mathbf{Z}) = b^2m(\mathbf{c}^T\beta_0 - \beta_0^T C\beta_0) = r(\pi_{\beta_0},\mathbf{d}^{\beta_0}).$$

Now it follows from Lemma 1 that $\mathbf{d}^{\beta_0}(\mathbf{X})$ is minimax. In the case when $a+bm=0$,

$$R(\mathbf{d}^\beta(\mathbf{X}),\mathbf{Z}) = b^2m(\mathbf{c}^T\mathbf{p} - \mathbf{p}^T C\mathbf{p})$$

and

$$\sup_{\mathbf{p}\in\mathcal{P}} R(\mathbf{d}^\beta(\mathbf{X}),\mathbf{Z}) = \sup_{\mathbf{p}\in\mathcal{P}} R(\mathbf{d}^{\beta_0}(\mathbf{X}),\mathbf{Z}) = b^2m(\mathbf{c}^T\beta_0 - \beta_0^T C\beta_0) = r(\pi_{\beta_0},\mathbf{d}^{\beta_0}).$$

Thus by Lemma 1, in the case $(a+bm)^2-b^2m\le 0$ the predictor $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\beta_0}(\mathbf{X})$ is minimax. ∎

3. Minimax prediction for the multivariate hypergeometric distribution. Let $\mathbf{X} = (X_1,\ldots,X_r)$ be a random variable having the multivariate hypergeometric distribution with parameters $(W,\mathbf{W},n)$, i.e.,

$$P_{\mathbf{p}}\{\mathbf{X}=\mathbf{x}=(x_1,\ldots,x_r)\} = \begin{cases} \dfrac{\binom{W_1}{x_1}\cdots\binom{W_r}{x_r}}{\binom{W}{n}} & \text{if } x_i\in\{0,1,\ldots,W_i\},\ i=1,\ldots,r,\\ & x_1+\ldots+x_r=n,\\ 0 & \text{otherwise,}\end{cases}$$

where $W_1+\ldots+W_r = W$, $r\ge 2$, $0<n\le W$, $W>2$ and $\mathbf{p} = (W_1/W,\ldots,W_r/W)$. Suppose that n and W are known and we want to find a minimax estimator $d(\mathbf{X}) = (d_1(\mathbf{X}),\ldots,d_r(\mathbf{X}))$ of the linear combination $\mathbf{Z} = a\mathbf{p}+b\mathbf{Y}$ of the unknown parameter $\mathbf{p}$ and the random variable $\mathbf{Y} = (Y_1,\ldots,Y_r)$ which has the multivariate hypergeometric distribution with parameters $(W,\mathbf{W},m)$, where $0<m\le W$. We assume that $\mathbf{X}$ and $\mathbf{Y}$ are independent, and that a,b and m are known and satisfy one of the

following conditions:

$$(a+bm)^2(W-n)(W-n-1)-b^2m(W-m)W\ge 0,$$
$$(a+bm)^2-b^2m\frac{(W-m)}{W-1}\le 0.$$

Suppose that the loss connected with the estimator $\mathbf{d}(\mathbf{X})$ is given by (1), where the matrix $C=(c_{ij})$ is nonnegative definite. This problem was considered by Trybuła (1958) in the case when $a=1$ and $b=0$, and C is an arbitrary nonnegative definite diagonal matrix, and also by Wilczyński (1985) in the case when $a=1$ and $b=0$ and C is an arbitrary nonnegative definite matrix. Set $\mathbf{c}=(c_{11},\ldots,c_{rr})$ and $\mathcal{P}=\{\mathbf{p}=(p_1,\ldots,p_r): p_i\ge 0,\ p_1+\ldots+p_r=1\}$. The following lemma determines the Bayes predictors of $\mathbf{Z}$ w.r.t. the following a priori distributions of $\mathbf{p}$:

1) the Pólya–Eggenberger distribution $\pi_{\alpha,\beta}$ with parameters α,β ($\alpha>0$, $\beta=(\beta_1,\ldots,\beta_r)\in\mathcal{P}$) given by

$$P\{\mathbf{p}=(W_1/W,\ldots,W_r/W)\}=\begin{cases}\dfrac{W!\,\Gamma(\alpha)}{\Gamma(W+\alpha)}\displaystyle\prod_{j\in A}\frac{\Gamma(W_j+\alpha\beta_j)}{W_j\Gamma(\alpha\beta_j)} & \text{if } W_i\in\{0,\ldots,W\},\ i=1,\ldots,r,\\ & \displaystyle\sum_{i=1}^r W_i=\sum_{j\in A}W_j=W,\\ 0 & \text{otherwise;}\end{cases}$$

2) the multinomial distribution $\pi_{\infty,\beta}$ (with $\beta=(\beta_1\ldots,\beta_r)\in\mathcal{P}$) given by

$$P\{\mathbf{p}=(W_1/W,\ldots,W_r/W)\}=\begin{cases}W!\displaystyle\prod_{j\in A}\frac{\beta_j^{W_j}}{W_j!} & \text{if } W_i\in\{0,\ldots,W\}, i=1,\ldots,r,\\ & \displaystyle\sum_{i=1}^r W_i=\sum_{j\in A}W_j=W,\\ 0 & \text{otherwise;}\end{cases}$$

3) π_β defined by $P(\mathbf{p}=\beta)=1$, where $\beta=(\beta_1,\ldots,\beta_r)\in\mathcal{P}$.

Above, $A=\{i\in\{1,\ldots,r\}:\beta_i>0\}$.

Lemma 3. *Under the loss function given by* (1) *with C nonnegative definite*, *the predictors*

$$d_j^{\alpha,\beta}(\mathbf{X})=(a+bm)\left[\frac{W+\alpha}{W(n+\alpha)}X_j+\frac{\alpha(W-n)}{W(n+\alpha)}\beta_j\right], \tag{19}$$

$$d_j^{\infty,\beta}(\mathbf{X})=\frac{a+bm}{W}[X_j+(W-n)\beta_j], \tag{20}$$

$$d_j^{\beta}(\mathbf{X})=(a+bm)\beta_j \tag{21}$$

of the linear combination **Z** *are Bayes w.r.t. the a priori distributions* $\pi_{\alpha,\beta}$, $\pi_{\infty,\beta}$ *and* π_β, *respectively, and their Bayes risks are*

$$r(\pi_{\alpha,\beta}, \mathbf{d}^{\alpha,\beta}) = z_1 E_\pi(\mathbf{p}^T C\mathbf{p}) - z_2\beta^T C\beta + z_3\mathbf{c}^T\beta, \tag{22}$$

$$r(\pi_{\infty,\beta}, \mathbf{d}^{\infty,\beta}) = \left[\frac{(a+bm)^2(W-n)}{W^2} + \frac{b^2m(W-m)}{W}\right] \times [\mathbf{c}^T\beta - \beta^T C\beta], \tag{23}$$

$$r(\pi_\beta, \mathbf{d}^\beta) = b^2\frac{W-m}{W-1}(\mathbf{c}^T\beta - \beta^T C\beta), \tag{24}$$

where

$$z_1 = (a+bm)^2\left[\frac{\alpha^2(W-n)^2}{W^2(n+\alpha)^2} - \frac{n(W-n)(W+\alpha)^2}{W^2(W-1)(n+\alpha)^2}\right] - b^2m\frac{W-m}{W-1}, \tag{25}$$

$$z_2 = (a+bm)^2\left[\frac{\alpha(W-n)}{W(n+\alpha)}\right]^2, \tag{26}$$

$$z_3 = (a+bm)^2\frac{n(W-n)(W+\alpha)^2}{W^2(W-1)(n+\alpha)^2} + b^2m\frac{W-m}{W-1}. \tag{27}$$

P r o o f. Under the loss function given by (1), the risk function associated with any predictor $\mathbf{d}(\mathbf{X})$ of $\mathbf{Z}$ is given by (10). Hence (analogously to the case of the multinomial distribution) the Bayes predictor will be the product of $a + bm$ and the Bayes estimator of $\mathbf{p}$. Now it is easy to show that the Bayes predictors w.r.t. $\pi_{\alpha,\beta}$, $\pi_{\infty,\beta}$ and π_β are given by (19), (20) and (21), respectively. The associated risk functions are

$$R(\mathbf{d}^{\alpha,\beta}, \mathbf{Z}) = z_1\mathbf{p}^T C\mathbf{p} + z_2\beta^T C\beta + z_3\mathbf{c}^T\mathbf{p} - 2z_2\beta^T C\mathbf{p}, \tag{28}$$

$$\begin{aligned} R(\mathbf{d}^{\infty,\beta}, \mathbf{Z}) = {} & \left[(a+bm)^2\frac{(W-n)(W-n-1)}{W(W-1)} - b^2m\frac{W-m}{W-1}\right]\mathbf{p}^T C\mathbf{p} \\ & + \frac{(a+bm)^2(W-n)^2}{W^2}\beta^T C\beta \\ & - 2\frac{(a+bm)^2(W-n)^2}{W^2}\beta^T C\mathbf{p} \\ & + \left[\frac{(a+bm)^2(W-n)n}{W^2(W-1)} + \frac{b^2m(W-m)}{W-1}\right]\mathbf{c}^T\mathbf{p}, \end{aligned} \tag{29}$$

$$\begin{aligned} R(\mathbf{d}^\beta, \mathbf{Z}) = {} & \left[(a+bm)^2 - b^2m\frac{W-m}{W-1}\right]\mathbf{p}^T C\mathbf{p} - 2(a+bm)^2\beta^T C\mathbf{p} \\ & + (a+bm)^2\beta^T C\beta + b^2m\frac{W-m}{W-1}\mathbf{c}^T\mathbf{p}, \end{aligned} \tag{30}$$

respectively, where z_1, z_2 and z_3 are given by (25), (26) and (27). It is easy to show that the Bayes risks associated with the a priori distributions $\pi_{\alpha,\beta}$,

$\pi_{\infty,\beta}$ and π_β and the predictors $\mathbf{d}^{\alpha,\beta}(\mathbf{X})$, $\mathbf{d}^{\infty,\beta}(\mathbf{X})$ and $\mathbf{d}^{\beta}(\mathbf{X})$ are of the form (22), (23) and (24), respectively. ■

The following theorem generalizes results of Trybuła (1958) and Wilczyński (1985).

THEOREM 4. *Under the loss function given by* (1) *with* C *nonnegative definite*, *the following predictor of* $\mathbf{Z}$ *is minimax*:

$$\mathbf{d}_0(\mathbf{X}) = (a+bm)\left[\frac{W+\alpha_0}{W(n+\alpha_0)}\mathbf{X} + \frac{\alpha_0(W-n)}{W(n+\alpha_0)}\beta_0\right], \tag{31}$$

if

$$(a+bm)^2(W-n)(W-n-1) - b^2m(W-m)W > 0; \tag{32}$$

$$\mathbf{d}_0(\mathbf{X}) = \frac{a+bm}{W}[\mathbf{X} + (W-n)\beta_0], \tag{33}$$

if

$$(a+bm)^2(W-n)(W-n-1) - b^2m(W-m)W = 0; \tag{34}$$

$$\mathbf{d}_0(\mathbf{X}) = (a+bm)\beta_0, \tag{35}$$

if

$$(a+bm)^2 - b^2m\frac{W-m}{W-1} \le 0, \tag{36}$$

where

$$\alpha_0 = \frac{n[(W-n)(a+bm)^2 + b^2mW(W-m)] + |a+bm|(W-n)\sqrt{\Delta}}{(a+bm)^2(W-n)(W-n-1) - b^2mW(W-m)}, \tag{37}$$

$$\Delta = (a+bm)^2(W-n)(W-1)n + b^2m(W-m)(n-1)Wn$$

and β_0 *is a point* $(\beta_1^0,\ldots,\beta_r^0)$ *for which*

$$\mathbf{c}^T\beta_0 - \beta_0^T C\beta_0 = \max_{\beta\in\mathcal{P}}(\mathbf{c}^T\beta - \beta^T C\beta).$$

P r o o f. Consider the Bayes predictors $\mathbf{d}^{\alpha,\beta}(\mathbf{X})$ of $\mathbf{Z}$ w.r.t. the Pólya–Eggenberger a priori distribution $\pi_{\alpha,\beta}$ of the parameter $\mathbf{p}$, which are of the form (19). The associated risk function is of the form (28). If the condition (32) is satisfied, then there exists $\alpha_0 > 0$ for which $z_1 = 0$. It is easy to check that α_0 is of the form (37). Set $R_1(\beta,\mathbf{p}) = R(\mathbf{d}^{\alpha_0,\beta},\mathbf{Z})$. Putting $\alpha = \alpha_0$, we obtain $z_1 = 0$ and $z_2 = z_3$, and, consequently,

$$R_1(\beta,\mathbf{p}) = z_2(\beta^T C\beta + \mathbf{c}^T\mathbf{p} - 2\beta^T C\mathbf{p}).$$

Notice that the function $R_1(\cdot,\cdot) : \mathcal{P}\times\mathcal{P} \to \mathbb{R}$ is convex w.r.t. β, concave w.r.t. $\mathbf{p}$, continuous w.r.t. $(\beta,\mathbf{p})$, and $\mathcal{P}$ is a convex, compact subset of $\mathbb{R}^r$. In the same way as in the case of the multinomial distribution we can show that (β_0,β_0) is a saddle point of $R_1(\beta,\mathbf{p})$. Moreover,

$$R_1(\beta_0,\beta_0) = r(\pi_{\alpha_0,\beta_0},\mathbf{d}^{\alpha_0,\beta_0}),$$

so that $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\alpha_0,\beta_0}(\mathbf{X})$ given by (31) is minimax.

When condition (34) holds, consider the Bayes predictors $\mathbf{d}^{\infty,\beta}(\mathbf{X})$ w.r.t. the a priori distribution $\pi_{\infty,\beta}$ of $\mathbf{p}$. In this case,

$$R(\mathbf{d}^{\infty,\beta},\mathbf{Z}) = \frac{(a+bm)^2(W-n)^2}{W^2}\beta^T C\beta - 2\frac{(a+bm)^2(W-n)^2}{W^2}\beta^T C\mathbf{p} + \left[\frac{(a+bm)^2(W-n)n}{W^2(W-1)} + \frac{b^2m(W-m)}{W-1}\right]\mathbf{c}^T\mathbf{p}.$$

Using the same arguments as above we can show that the predictor (33) is minimax.

Analogously, under condition (36), the risk function (30) associated with the Bayes predictor $\mathbf{d}^{\beta}(\mathbf{X})$ w.r.t. the a priori distribution π_β is convex w.r.t. β, concave w.r.t. $\mathbf{p}$, continuous w.r.t. $(\beta,\mathbf{p})$, and (β_0,β_0) is its saddle point. Moreover,

$$\sup_{\mathbf{p}\in\mathcal{P}} R(\mathbf{d}^{\beta},\mathbf{Z}) = r(\pi_{\beta_0},\mathbf{d}^{\beta_0}),$$

so that $\mathbf{d}_0(\mathbf{X}) := \mathbf{d}^{\beta_0}(\mathbf{X})$ given by (35) is minimax. ∎

Let us consider some examples.

Example 1 (multinomial distribution). Suppose that each of n observed independently working devices undergoes failure due to one of r possible reasons. Then the observed values X_i, $i=1,\ldots,r$, represent the number of devices which have been damaged for the ith reason. The purpose is to estimate the value of $\mathbf{Z}=(Z_1,\ldots,Z_r)$, $Z_i = Y_i - ap_i$, where Y_i is the unknown number of devices in the group of m devices which we should expect to be destroyed for the ith reason, and ap_i is the mean value of the number of failures due to the ith reason in a group consisting of a devices. By Theorem 3, under the loss function given by (1) with $C=I$, a minimax predictor of $\mathbf{Z}$ will be of the form

$$(m-a)\left(\frac{1}{n+\alpha_0}\mathbf{X} + \frac{\alpha_0}{n+\alpha_0}\beta_0\right),$$

where

$$\alpha_0 = \frac{nm + |m-a|\sqrt{(m-a)^2n + mn(n-1)}}{(m-a)^2 - m}$$

and $\beta_0 = (1/r,\ldots,1/r)$.

Example 2 (multivariate hypergeometric distribution). A group consisting of W elements undergoes statistical quality inspection. The values W_i, $i=1,\ldots,r$, representing the number of elements of the ith quality category are assumed to be unknown. On the basis of the observations from randomly chosen n elements we want to predict the number of elements of each quality category which should appear among anew randomly chosen m elements, or we want to estimate the increments (decrements) of these numbers in comparison to the mean values of the inspection for a elements. In the first case we want to find the value of the predictor $\mathbf{Z} = \mathbf{Y}$ ($a = 0$, $b = 1$; we recall that even this special case of Theorem 4 has not been treated before), and in the second the value of the predictor $\mathbf{Z} = \mathbf{Y} - a\mathbf{p}$.

One can give an analogous example related to a voting model. In this case the values W_i, $i = 1, \ldots, r$, represent the number of persons who are inclined to vote for the ith candidate. We want to predict the numbers Y_i, $i = 1, \ldots, r$, of persons, among m voters, who would vote for the ith candidate.

By Theorem 4, under the loss function given by (1) with $C = I$, the minimax predictor of $\mathbf{Z} = \mathbf{Y} - a\mathbf{p}$ will be of the form:

(i) $$\mathbf{d}_0(\mathbf{X}) = (m-a)\left[\frac{W+\alpha_0}{W(n+\alpha_0)}\mathbf{X} + \frac{\alpha_0(W-n)}{W(n+\alpha_0)}\beta_0\right] \quad \text{if}$$
$$(m-a)^2(W-n)(W-n-1) - m(W-m)W > 0,$$

(ii) $$\mathbf{d}_0(\mathbf{X}) = \frac{m-a}{W}[\mathbf{X} + (W-n)\beta_0] \quad \text{if}$$
$$(m-a)^2(W-n)(W-n-1) - m(W-m)W = 0,$$

(iii) $$\mathbf{d}_0(\mathbf{X}) = (m-a)\beta_0 \quad \text{if}$$
$$(m-a)^2 - m\frac{W-m}{W-1} \le 0,$$

where

$$\alpha_0 = \frac{n[(W-n)(m-a)^2 + mW(W-m)] + |m-a|(W-n)\sqrt{\Delta}}{(m-a)^2(W-n)(W-n-1) - mW(W-m)},$$
$$\Delta = (m-a)^2(W-n)(W-1)n + m(W-m)(n-1)Wn$$

and $\beta_0 = (1/r, \ldots, 1/r)$.

REMARK 1. It follows from Theorem 2 that $A = \{i_1\}$ iff $c_{ij} = c_0$ for all i, j. This case is not interesting, because every predictor $\mathbf{d} = (d_1, \ldots, d_r)$ for which $d_1 + \ldots + d_r = a + bm$ is then a minimax predictor, and we may assume that $k \ge 2$.

REMARK 2. The minimax predictor of $\mathbf{Z}$ established in Theorems 3 and 4 is a linear combination of the minimax estimator of $\mathbf{p}$ and a minimax predictor of the random variable $\mathbf{Y}$ with coefficients depending on a, b, m and n.

References

J. P. Aubin (1979), *Mathematical Methods of Game and Economic Theory*, North-Holland.

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

G. M. Fichtenholz (1985), *Differential and Integral Calculus*, PWN, Warszawa (in Polish).

V. G. Karmanov (1986), *Mathematical Programming*, Nauka, Moscow.

S. Trybuła (1958), *Some problems of simultaneous minimax estimation*, Ann. Math. Statist. 29, 245–253.

M. Wilczyński (1985), *Minimax estimation for the multinomial and multivariate distributions*, Sankhyā 47, 128–132.

Alicja Jokiel-Rokita
Institute of Mathematics
Technical University of Wrocław
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: arokita@im.pwr.wroc.pl

Received on 5.3.1996;
revised version on 20.1.1998