ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
21,2 (1991), pp. 169–173

S. TRYBUŁA (Wrocław)

MINIMAX PREDICTION OF THE DIFFERENCE OF MULTINOMIAL RANDOM VARIABLES

A minimax predictor is determined, for the loss function (1), of the difference of random variables distributed according to the multinomial distributions with parameters n_1, p and n_2, p, respectively, $p = (p_1, \ldots, p_r)$. This predictor is based on the multinomial random variable with parameters m, p.

Let $X = (X_1, \ldots, X_r)$, $Y_i = (Y_{i1}, \ldots, Y_{ir})$, $i = 1, 2$, be independent random variables with multinomial distributions

$$P(X = x) = P(X_1 = x_1, \ldots, X_r = x_r) = \frac{m!}{x_1! \ldots x_r!} p_1^{x_1} \ldots p_r^{x_r},$$

$$P(Y_i = y_i) = P(Y_{i1} = y_{i1}, \ldots, Y_{ir} = y_{ir}) = \frac{n_i}{y_{i1}! \ldots y_{ir}!} p_1^{y_{i1}} \ldots p_r^{y_{ir}}, \quad i = 1, 2, \ n_1 > n_2.$$

Suppose that $X = x$ is observed and that we want to predict the difference $Y = Y_1 - Y_2$. Let $\widehat{a} = (a_1, \ldots, a_r)$ be a prediction of Y and let the loss associated with the prediction $\widehat{a}$ (the *loss function*) be

$$L(Y, \widehat{a}) = \sum_{i,j=1}^{r} c_{ij}(a_i - Y_i)(a_j - Y_j), \tag{1}$$

where $Y = (Y_1, \ldots, Y_r)$ and the matrix $C = \|c_{ij}\|_1^r$ is nonnegative definite.

A predictor $d^0(x) = (d_1^0(x), \ldots, d_r^0(x))$ of $Y = Y_1 - Y_2$ is called *minimax* if

$$\sup_p R(p, d^0) = \inf_d \sup_p R(p, d)$$

where $R(p, d) = E[L(Y, d(X))]$ is the risk function for the loss function (1).

1991 *Mathematics Subject Classification*: Primary 62F15.
Key words and phrases: minimax prediction, general quadratic loss function, multinomial distribution.

We are looking for a minimax predictor for Y. We have

$$(2)\quad E[L(Y,d(X))\mid X] = \sum_{i,j=1}^{r} c_{ij}[(d_i(X)-(n_1-n_2)p_i)(d_j(X)-(n_1-n_2)p_j) + E[(Y_{i1}-n_1p_i)(Y_{j1}-n_1p_j)] + E[(Y_{i2}-n_2p_i)(Y_{j2}-n_2p_j)]].$$

Notice that the second and third terms in the above expression are independent of d. Let

$$(3)\qquad d_i(X) = (n_1-n_2)\frac{X_i+\beta_i\alpha}{m+\alpha} = aX_i + b_i,$$

where $\alpha > 0$, $\beta_i \geq 0$, $i = 1,\ldots,r$, $\sum_{i=1}^{r}\beta_i = 1$. In this case, from (2) we obtain

$$R(p,d) = \left(\frac{n_1-n_2}{m+\alpha}\right)^2\Big\{\sum_{i,j=1}^{r} c_{ij}[(\alpha^2-m)p_ip_j + \alpha^2(\beta_i-2p_i)\beta_j] + m\sum_{i=1}^{r} c_{ii}p_i\Big\} + (n_1+n_2)\Big[-\sum_{i,j=1}^{r} c_{ij}p_ip_j + \sum_{i=1}^{r} c_{ii}p_i\Big].$$

Let $\alpha > 0$ be chosen so as to obtain

$$(n_1-n_2)^2(\alpha^2-m) - (n_1+n_2)(m+\alpha)^2 = 0.$$

This equation is surely satisfied if

$$(4)\qquad \alpha = \frac{(n_1+n_2)m + (n_1-n_2)\sqrt{(n_1+n_2)m(m-1)+(n_1-n_2)^2m}}{(n_1-n_2)^2-n_1-n_2},$$

assuming that

$$(5)\qquad (n_1-n_2)^2-n_1-n_2 > 0,$$

and for this α

$$(6)\qquad a = \frac{(n_1-n_2)m - \sqrt{(n_1+n_2)m(m-1)+(n_1-n_2)^2m}}{m(m-1)},$$

$$(7)\qquad b_i = \frac{-n_1+n_2+\sqrt{(n_1+n_2)m(m-1)+(n_1-n_2)^2m}}{m-1}\beta_i,$$

when $m > 1$, and

$$(8)\qquad a = \frac{(n_1-n_2)^2-n_1-n_2}{2(n_1-n_2)},\qquad b_i = \frac{(n_1-n_2)^2+n_1+n_2}{2(n_1-n_2)}\beta_i,$$

when $m = 1$.

Moreover, for this α

$$R(p,d) = \left[\frac{(n_1-n_2)\alpha}{m+\alpha}\right]^2 \left[\sum_{i,j=1}^{r} c_{ij}\beta_i\beta_j + \sum_{i,j=1}^{r}(c_{ii}-2c_{ij})\beta_j p_i\right]. \tag{9}$$

THEOREM 1. *If there are constants v, $\beta_1,\ldots,\beta_r$ and a set $A \subset R = \{1,\ldots,r\}$, $|A| \geq 2$, such that*

(a) $$\sum_{j\in A}(c_{ii}-2c_{ij})\beta_j = v \quad \text{for } i \in A,$$

(b) $$\sum_{j\in A}(c_{ii}-2c_{ij})\beta_j \leq v \quad \text{for } i \in R-A,$$

$\beta_j > 0$ for $j \in A$, $\beta_j = 0$ for $j \in R-A$, $\sum_{j\in A}\beta_j = 1$, then the predictor d defined by (3) *and* (4), *with β_i satisfying* (a) *and* (b), *is a minimax predictor provided that condition* (5) *holds.*

Proof. Assume (a), (b) and (5) to hold. Then from (9)

$$R(p,d) = \left[\frac{(n_1-n_2)\alpha}{m+\alpha}\right]^2 \left[\sum_{i,j\in A} c_{ij}\beta_i\beta_j + v\right] \stackrel{\mathrm{df}}{=} c$$

for $p_i = 0$ when $i \in R-A$ and $R(p,d) \leq c$ for any p. Therefore the theorem follows from the fact that the predictor defined by (3) with $\alpha_i = \beta_i\alpha$ for $i \in A$, $\alpha_i = 0$ for $i \in R-A$ is Bayes with respect to the a priori distribution of the parameter p given by the density

$$g(p_1,\ldots,p_r) = \begin{cases} \dfrac{\Gamma(\alpha)}{\Gamma(\alpha_{i_1})\ldots\Gamma(\alpha_{i_s})} p_{i_1}^{\alpha_{i_1}-1}\ldots p_{i_s}^{\alpha_{i_s}-1} & \text{if } p_{i_k} \geq 0\ (k=1,\ldots,s), \\ & \sum_{k=1}^{s} p_{i_k} = 1\ (A=\{i_1,\ldots,i_s\}), \\ 0 \quad \text{otherwise}, & \end{cases}$$

as may be deduced from (2).

THEOREM 2 (Wilczyński [6]). *If the matrix C is nonnegative definite then the constants $v, \beta_1,\ldots,\beta_r$ and the set A mentioned in Theorem* 1 *always exist.*

This result was proved while determining a minimax estimator of a multinomial parameter $p = (p_1,\ldots,p_r)$ for a general quadratic loss function.

THEOREM 3. *If*

$$(n_1-n_2)^2 - n_1 - n_2 \leq 0 \tag{10}$$

then

$$d_i(X) = (n_1-n_2)\beta_i, \quad i = 1,\ldots,r, \tag{11}$$

is a minimax predictor, where the constants β_i are determined in the same way as in Theorem 1.

Proof. For $d_i = (n_1 - n_2)\beta_i$

$$R(p,d) \stackrel{\text{df}}{=} R_0(p,\beta) = \sum_{i,j=1}^{r} c_{ij}[(n_1-n_2)^2(\beta_i - p_i)(\beta_j - p_j) - (n_1+n_2)p_ip_j] + (n_1+n_2)\sum_{i=1}^{r} c_{ii}p_i.$$

This function is convex with respect to β for fixed p and under condition (10) it is concave with respect to p for fixed β, $p = (p_1,\ldots,p_r)$, $\beta = (\beta_1,\ldots,\beta_r)$. Moreover,

$$\min_{\beta\in P_0} R_0(p,\beta) = R_0(p,p)$$

if $p \in P_0$, where

$$P_0 = \Big\{p = (p_1,\ldots,p_r) : p_i \ge 0,\ i = 1,\ldots,r,\ \sum_{i=1}^{r} p_i = 1\Big\}.$$

Now applying the method developed in [6] one can prove that predictor (11) with β_i satisfying (a) and (b) is minimax.

When $c_{ij} = 0$ for $i \neq j$ the constants β_i in Theorems 1 and 3 can be determined explicitly.

Let $c_{11} \ge c_{22} \ge \ldots \ge c_{rr} \ge 0$. We now prove that $A = \{1,\ldots,L\}$, where

$$L = \max_s \Big\{ s \le l_0 : \sum_{i=1}^{s} \frac{1}{c_{ii}} \ge \frac{s-2}{c_{ss}} \Big\},$$

and l_0 is the greatest index i for which $c_{ii} \neq 0$.

Let $c_{22} > 0$. In this case $L \ge 2$ and we obtain from (a) and (b) in Theorem 1

$$c_{ii}(1 - 2\beta_i) = v \quad \text{for } i \le L, \tag{12}$$

$$c_{ii} \le v \quad \text{for } i > L. \tag{13}$$

From (12) it follows that we must have

$$v = \frac{L-2}{\sum_{i=1}^{L} 1/c_{ii}}, \quad \beta_i = \begin{cases} 1 - v/c_{ii} & \text{if } i \le L, \\ 0 & \text{if } i > L. \end{cases}$$

We still have to prove (13). First observe that the proof is only necessary for $i = L+1$. If $c_{L+1,L+1} = 0$ then the inequality obviously holds. If $c_{L+1,L+1} \neq 0$, it follows from the definition of L that

$$L - 1 \ge c_{L+1,L+1} \sum_{j=1}^{L+1} \frac{1}{c_{jj}} = 1 + c_{L+1,L+1} \sum_{j=1}^{L} \frac{1}{c_{jj}}.$$

From this we obtain (13).

When only $c_{11} \neq 0$ the problem reduces to that of minimax prediction of the difference of binomial random variables Y_1, Y_2 for a quadratic loss function

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

$Y = Y_1 - Y_2$. It is easy to prove that in this case:

1) If $(n_1 - n_2)^2 > n_1 + n_2$, then

$$d(X) = aX + b$$

is a minimax predictor, where a is given by (6) and

$$b = \frac{-n_1 + n_2 + \sqrt{(n_1 + n_2)m(m-1) + (n_1 - n_2)^2 m}}{2(m-1)}$$

for $m > 1$, and a is given by (8) and

$$b = \frac{(n_1 - n_2)^2 + n_1 + n_2}{4(n_1 - n_2)}$$

for $m = 1$.

2) If $(n_1 - n_2)^2 \leq n_1 + n_2$, then

$$d(X) = \frac{n_1 - n_2}{2}$$

is a minimax predictor.

For minimax and Bayes estimation problems and related theory see [1]–[6].

References

[1] M. H. De Groot, *Optimal Statistical Decisions*, McGraw-Hill, New York 1970.

[2] J. L. Hodges and E. L. Lehmann, *Some problems in minimax point estimation*, Ann. Math. Statist. 21 (1950), 182–196.

[3] E. G. Phadia, *Minimax estimation of a cumulative distribution function*, Ann. Statist. 1 (1973), 1149–1157.

[4] S. Trybuła, *Some investigations in minimax estimation theory*, Dissertationes Math. 240 (1985).

[5] —, *Minimax estimation and prediction for random variables with bounded sum*, Zastos. Mat. 19 (1987), 85–91.

[6] M. Wilczyński, *Minimax estimation for multinomial and multivariate hypergeometric distribution*, Sankhyā Ser. A 47 (1985), 128–132.

STANISŁAW TRYBUŁA
INSTITUTE OF MATHEMATICS
TECHNICAL UNIVERSITY OF WROCŁAW
WYBRZEŻE WYSPIAŃSKIEGO 27
50-370 WROCLAW, POLAND

Received on 20.11.1989