MINIMAX MUTUAL PREDICTION

Abstract. The problems of minimax mutual prediction are considered for binomial and multinomial random variables and for sums of limited random variables with unknown distribution. For the loss function being a linear combination of quadratic losses minimax mutual predictors are determined where the parameters of predictors are obtained by numerical solution of some equations.

1. Introduction. Suppose that a number of statisticians are observing some random variables. Assume that the \(i\)th statistician is observing a random variable \(X_i\). He wants to predict the random variables of his partners. Let \(d_{ij}(X_i)\) be the predictor he applies to predict \(X_j\). Let the loss connected with this prediction be \(L_{ij}(X_j, d_{ij}(X_i))\). Then the total loss of all statisticians is

\[
L(X, d) = \sum_{i,j=1 \atop i \neq j}^l L_{ij}(X_j, d_{ij}(X_i))
\]

where \(X = (X_1, \ldots, X_l)\),

\[
d = \begin{bmatrix}
- & d_{12} & \cdots & d_{1l} \\
d_{21} & - & \cdots & d_{2l} \\
\vdots & \vdots & \ddots & \vdots \\
d_{l1} & d_{l2} & \cdots & -
\end{bmatrix} =: [d_{ij}]^l_1.
\]

Suppose that the random variable \(X\) has distribution depending on an un-
known parameter $\mu$. The risk function is defined as

\begin{equation}
R(\mu, d) = E_{\mu}(L(X, d)) = \sum_{i,j=1 \atop i \neq j}^l E_{\mu}(L_{ij}(X_j, d_{ij}(X_i)))
\end{equation}

where $E_{\mu}(\cdot)$ denotes the expected value.

A mutual predictor $d_0$ is minimax if

$$\sup_{\mu} R(\mu, d_0) = \inf_{d} \sup_{\mu} R(\mu, d).$$

In this paper we shall find minimax mutual predictors in some situations.

2. Mutual predictors for binomial random variables. Let the variables $X_1, \ldots, X_l$ be independent and have binomial distributions

$$f_i(x_i \mid p) = \binom{n_i}{x_i} p^{x_i} (1 - p)^{n_i - x_i},$$

where the $f_i$ are densities with respect to the counting measure. Let the losses $L_{ij}$ be quadratic. Then

$$L(X, d) = \sum_{i,j=1 \atop i \neq j}^l k_{ij}(d_{ij}(X_i) - X_j)^2$$

where $k_{ij} \geq 0$, $\sum_{i \neq j} k_{ij} > 0$. Hence the risk function $R(p, d)$ can be represented in the form

\begin{equation}
R(p, d) = \sum_{i,j=1 \atop i \neq j}^l k_{ij} E_p(d_{ij}(X_i) - X_j)^2
= \sum_{i,j=1 \atop i \neq j}^l k_{ij} [E_p(d_{ij}(X_i) - n_j p)^2 + n_j p(1 - p)].
\end{equation}

We shall look for minimax mutual predictors. Consider predictors of the form

\begin{equation}
d_{ij}(X_i) = n_j \frac{X_i + \alpha}{n_i + \gamma}, \quad \alpha > 0, \ \gamma > 0.
\end{equation}

In this case

\begin{equation}
R(p, d) = \sum_{i,j=1 \atop i \neq j}^l k_{ij} \left[ n_j^2 E_p \left( \frac{X_i + \alpha}{n_i + \gamma} - p \right)^2 + n_j p(1 - p) \right]
\end{equation}
Minimax mutual prediction

\[
= \sum_{i,j=1 \atop i \neq j}^l k_{ij} \left[ \frac{n_j^2}{(n_i + \gamma)^2} (n_i p(1 - p) + (\alpha - \gamma p)^2) + n_j p(1 - p) \right].
\]

The risk \(R(p, d)\) will be constant if

\[
\sum_{i,j=1 \atop i \neq j}^l k_{ij} \left[ \frac{n_j^2}{(n_i + \gamma)^2} (-n_i + \gamma^2) - n_j \right] = 0
\]

and

\[
\sum_{i,j=1 \atop i \neq j}^l k_{ij} \left[ \frac{n_j^2}{(n_i + \gamma)^2} (n_i - 2\alpha\gamma) + n_j \right] = 0.
\]

For \(\gamma = 0\) the left side of the first equation is negative, and as \(\gamma \to \infty\) it tends to

\[
A = \sum_{i,j=1 \atop i \neq j}^l k_{ij} n_j (n_j - 1) \geq 0.
\]

Moreover it is an increasing function of the parameter \(\gamma\). Therefore if \(A > 0\) there always exists a unique solution of equation (6).

Suppose that there exists a solution \(\gamma\) of (6). In this case there exists a solution \(\alpha\) of (7) and

\[
\alpha = \gamma/2.
\]

Equation (6) can be solved numerically.

When \(n_1 = \ldots = n_l =: n > 1\), the solution of (6) is

\[
\gamma = \frac{n}{n-1} (\sqrt{2n-1} + 1)
\]

and it is independent of \(k_{ij}\).

When all \(k_{ij} = 0\) except, say \(k_{12}\), equation (6) has a solution

\[
\gamma = \frac{n_1}{n_2 - 1} \left( n_2 \left( \frac{1}{n_1} + \frac{1}{n_2} - \frac{1}{n_1 n_2} \right) + 1 \right).
\]

Suppose that there exists a solution \(\gamma > 0\) of (6). It is easy to prove that the predictors given by (4) are Bayes with respect to the a priori distribution of the parameter \(p\) given by the density

\[
g(p) = \frac{1}{B(\alpha, \gamma - \alpha)} p^{\alpha-1} (1 - p)^{\gamma - \alpha - 1} \quad \text{for } 0 < p < 1.
\]

Thus the mutual predictor \(d = [d_{ij}]_1^l\) defined by (4), being a constant risk Bayes predictor, is minimax if \(\alpha = \gamma/2\), where \(\gamma\) is a solution of (6).
When \( A = 0 \) a minimax mutual predictor is given by
\[
d_{ij}(X_i) = n_j/2.
\]
It is obtained by letting \( \gamma \to \infty, \alpha = \gamma/2 \) in (4).

The problem of minimax prediction when only one \( k_{ij} > 0 \) was solved by Hodges and Lehmann in [1].

When \( A > 0 \) the minimax risk is
\[
R(p, d_0) = \frac{1}{4} \sum_{i,j=1}^{l} \frac{k_{ij}n_j^2}{(n_i + \gamma)^2} \gamma^2
\]
where \( d_0 \) satisfies (4), (6) and (9).

When \( A = 0 \) the minimax risk can be obtained by letting \( \gamma \to \infty \) in formula (14):
\[
R(p, d_0) = \frac{1}{4} \sum_{i,j=1}^{l} \frac{k_{ij}n_j^2}{(n_i + \gamma)^2} \gamma^2
\]
where \( d_0 \) is given by (13).

3. Minimax mutual predictors for multinomial random variables. Let now \( X_i = (X_{i1}, \ldots, X_{ir}) \), \( i = 1, \ldots, l \), be independent random variables distributed according to multinomial laws
\[
f_i(x_i | p) = \frac{n_i!}{x_{i1}! \cdots x_{ir}!} p_{i1}^{x_{i1}} \cdots p_{ir}^{x_{ir}},
\]
where \( x_i = (x_{i1}, \ldots, x_{ir}) \) is the value of \( X_i \). Let the loss function be of the form
\[
L(X, d) = \sum_{i,j=1}^{l} \sum_{k=1}^{r} k_{ij}(d_{ij}^{(k)}(X_i) - X_{jk})^2.
\]
Let us consider the predictors
\[
d_{ij}^{(k)}(X_i) = n_j \frac{X_{ik} + \alpha_k}{n_i + \gamma}, \quad i, j = 1, \ldots, l, \quad i \neq j, \quad k = 1, \ldots, r.
\]
It is easy to show that for the loss function given by (16) with some \( k_{ij} > 0 \) these Bayes predictors satisfy the equations
\[
\sum_{k=1}^{r} d_{ij}^{(k)}(X_i) = n_j, \quad j = 1, \ldots, l.
\]
All these equations will surely be satisfied by \( d_{ij}^{(k)} \) given in (17) when
\[
\alpha_k = \gamma/r, \quad k = 1, \ldots, r.
\]
For predictors satisfying (17) and (19) the risk function will take the form

\[ R(p, d) = \sum_{i,j=1 \atop i \neq j}^{l} \sum_{k=1}^{r} k_{ij} \left[ \frac{n_j^2}{(n_i + \gamma)^2} (n_i p_k (1 - p_k) + (\alpha_k - \gamma p_k)^2) + n_j p_k (1 - p_k) \right]. \]

Notice that in (20) the coefficients of \( p_k^2 \) for \( k = 1, \ldots, r \) are the same. They are zero when \( \gamma \) satisfies (6).

Let \( \alpha_k \) satisfy (19). In this case for \( \gamma \) given by (6) the expression (20) will take the form

\[ R(p, d) = \sum_{i,j=1 \atop i \neq j}^{l} \sum_{k=1}^{r} k_{ij} \left[ \frac{n_j^2}{(n_i + \gamma)^2} (n_i - \frac{\gamma^2}{r}) + n_j \right] \]

\[ \overset{(6)}{=} \sum_{i,j=1 \atop i \neq j}^{l} k_{ij} \left[ \frac{n_j^2}{(n_i + \gamma)^2} (n_i - \frac{\gamma^2}{r}) \right. \]

\[ + n_j + \frac{n_j^2}{(n_i + \gamma)^2} (-n_i + \gamma^2) - n_j \]

\[ = \frac{r - 1}{r} \sum_{i,j=1 \atop i \neq j}^{l} k_{ij} \frac{n_j^2}{(n_i + \gamma)^2} \gamma^2. \]

From the above it follows that \( d = [d_{ij}^{(1)}, \ldots, d_{ij}^{(r)}]_1 \), where \( d_{ij}^{(k)} \) are given by (17), (19) and (6), is a constant risk mutual predictor. Let \( p = (p_1, \ldots, p_r) \) be a random variable. The expression

\[ E(E_p(d_{ij}^{(k)}(X_i) - X_{jk})^2) = E(E_p(d_{ij}^{(k)}(X_i) - n_j p_k)^2 + n_j p_k (1 - p_k)) \]

attains its minimum when

\[ d_{ij}^{(k)}(X_i) = n_j E(p_k | X_i) = n_j E(p_k | (X_{i1}, \ldots, X_{ir})). \]

Here \( E(p_k | X_i) \) denotes the conditional expectation of the random variable \( p_k \) under the condition that \( X_i \) is given. For the a priori distribution given by the density

\[ g(p_1, \ldots, p_r) = \frac{\Gamma(\gamma)}{\Gamma(\gamma/r)^r} (p_1 \ldots p_r)^{\gamma/r - 1} \]

we find that

\[ d_{ij}^{(k)}(X_i) = n_j E(p_k | X_i) = n_j \frac{X_{ik} + \gamma/r}{n_i + \gamma} \]

is a Bayes predictor.
We have shown that the mutual predictor \(d\) given by (17) and (19) is a Bayes predictor. Thus for \(\gamma\) satisfying (6) it is a minimax predictor for the loss function (16).

Equation (6) has a solution \(\gamma\) when \(A > 0\) (see (8)). When \(A = 0\) it is easy to show that a minimax mutual predictor is independent of \(X = (X_1, \ldots, X_l)\) and is given by

\[
d^{(k)}(X_i) = n_j/r, \quad i, j = 1, \ldots, l, \ i \neq j, \ k = 1, \ldots, r,
\]

and

\[
R(p, d) = \frac{r - 1}{r} \sum_{i, j = 1 \atop i \neq j}^l k_{ij} n_j^2.
\]

If only one \(k_{ij} \neq 0\), the results of this section follow from the paper of Wilczyński [4].

4. Minimax predictors of limited random variables. Suppose that the \(i\)th statistician is observing \(n_i\) random variables \(X_{i1}, \ldots, X_{in_i}\), with values in the interval \([0, 1]\) and let \(X_i = \sum_{k=1}^{n_i} X_{ik}\), \(X = (X_1, \ldots, X_l)\), where the \(X_{ik}\) are independent. The statistician wants to predict the random variables \(X_j, j = 1, \ldots, l, j \neq i\). Let the total loss function of all statisticians be of the form

\[
L(X, d) = \sum_{i, j = 1 \atop i \neq j}^l k_{ij} (d_{ij}(X_i) - X_j)^2.
\]

Let the random variables \(X_{ik}, i = 1, \ldots, l, k = 1, \ldots, n_i\), have the same distribution function \(F\) and let \(\mu = E_F(X_{ik})\). The risk function is

\[
R(\mu, d) = \sum_{i, j = 1 \atop i \neq j}^l k_{ij} E_F(d_{ij}(X_i) - X_j)^2
= \sum_{i, j = 1 \atop i \neq j}^l k_{ij} [E_F(d_{ij}(X_i) - n_j \mu)^2 + E_F(X_j - n_j \mu)^2].
\]

We look for minimax predictors.

Let us try predictors of the form

\[
d_{ij} = n_j \frac{X_i + \alpha}{n_i + \gamma}.
\]
Then

\[ R(\mu, d) = \sum_{i,j=1}^{l} k_{ij} \frac{n_{ij}^2}{(n_i + \gamma)^2} \left[ E_F(X_i - n_i\mu)^2 + (\alpha - \gamma\mu)^2 + E_F(X_j - n_j\mu)^2 \right] \]

\[ \leq \sum_{i,j=1}^{l} k_{ij} \left[ \frac{n_{ij}^2}{(n_i + \gamma)^2} (n_i\mu(1 - \mu) + (\alpha - \gamma\mu)^2) + n_j\mu(1 - \mu) \right]. \]

Let equations (6) and (7) be satisfied. Then \( \alpha = \gamma/2 \) and

\[ R(\mu, d) = \frac{1}{4} \sum_{i,j=1}^{l} k_{ij} \frac{n_{ij}^2}{(n_i + \gamma)^2} \gamma^2 = c \]

and for any random variable \( X = (X_1, \ldots, X_l) \) satisfying the conditions given at the beginning of this section,

\[ R(\mu, d) \leq c. \]

Let \( X_{ik} \) have two-point distribution

\[ P(X_{ik} = 0) = 1 - p, \quad P(X_{ik} = 1) = p. \]

Then \( \mu = p \) and equality holds in (31). From formulae (31)-(34) and the results of Section 2 it follows that for \( \gamma \) satisfying (6) and \( \alpha = \gamma/2 \), the predictor \( d = [d_{ij}]_1 \) given by (30) is a minimax mutual predictor. This holds when \( A > 0 \), where \( A \) is defined in (8).

For \( A = 0 \) a minimax mutual predictor is given by the formula

\[ d_{ij}(X_i) = n_j/2. \]

When only one \( k_{ij} \neq 0 \) the problem considered in this section was solved by Hodges and Lehmann in [1].

Analogous results can also be obtained for mutual prediction of sample cumulative distribution functions for a properly chosen loss function. For minimax estimation of cumulative distribution functions, see Phadia [2].

For minimax estimators and predictors of many parameters and random variables, see [3], [4].

References


Stanisław Trybula
Institute of Mathematics
Technical University of Wrocław
50-370 Wrocław, Poland

Received on 30.11.1999;
revised version on 17.3.2000