

ALICJA JOKIEL-ROKITA (Wrocław)

MINIMAX PREDICTION UNDER RANDOM SAMPLE SIZE

Abstract. A class of minimax predictors of random variables with multinomial or multivariate hypergeometric distribution is determined in the case when the sample size is assumed to be a random variable with an unknown distribution. It is also proved that the usual predictors, which are minimax when the sample size is fixed, are not minimax, but they remain admissible when the sample size is an ancillary statistic with unknown distribution.

1. Introduction. In this paper we derive the minimax predictor of a random variable Y on the basis of the observation of a random variable X , in the case when X and Y have the multinomial or multivariate hypergeometric distribution with the same unknown parameter. We assume that the loss function is given by (2) below and the sample size N is an *ancillary statistic* (i.e. a statistic whose distribution does not depend on the unknown parameter). Moreover, we assume that this distribution is unknown. A widely held notion about ancillary statistics is that their distribution should be irrelevant to statistical inference. Therefore the minimax predictor for the case of fixed sample size may seem to be the best candidate for the minimax predictor when the sample size is random. However, we prove that it is not minimax in our case. The first example of this *ancillarity paradox* was given by Brown (1990). Further examples were presented by He (1990) and Amrhein (1995).

There are a lot of practical situations which emphasize the importance of considering a random sample size. For instance, there are so-called non-response models that take into account that for certain units in the sample, it may not be clear to which stratum they belong. Non-response typically

2000 *Mathematics Subject Classification*: Primary 62F10.

Key words and phrases: admissibility, ancillarity paradox, loss function, minimax predictor, multinomial distribution, multivariate hypergeometric distribution, risk function.

occurs in surveys concerning sensitive data, such as drug abuse or tax evasion. To model this so-called phenomenon of non-at-homes, we assume that selected units, independent of each other and independent of their strata, fail to answer with the same probability. The effective sample size is then an ancillary statistic with a binomial distribution with an unknown success probability. (For other examples, see Chaudhuri and Stenger (1992), Chap. 11.)

2. Minimax prediction for the multinomial distribution. Let $X = (X_1, \dots, X_r)^T$ be an observation such that $X | N = n$, $n = 0, 1, 2, \dots$, has the multinomial distribution with parameters (n, p) , i.e.,

$$P_p(X = (x_1, \dots, x_r)^T) = \begin{cases} \frac{n!}{x_1! \dots x_r!} p_1^{x_1} \dots p_r^{x_r} & \text{if } x_i \in \{0, 1, \dots, n\} \text{ and } x_1 + \dots + x_r = n, \\ 0 & \text{otherwise,} \end{cases}$$

where $p \in \mathcal{P} = \{p = (p_1, \dots, p_r)^T : p_i \geq 0, i = 1, \dots, r; p_1 + \dots + p_r = 1\}$ is unknown. We observe the value of these random variables X, N and using an estimate $d(X, N) = (d_1(X, N), \dots, d_r(X, N))^T$, we want to predict the value of the random variable $Y = (Y_1, \dots, Y_r)^T$ which has the multinomial distribution with parameters (m, p) . We assume that m is known, that the unknown distribution of N given by the probability function

$$(1) \quad f(n) = P(N = n)$$

does not depend on p , that X and Y are independent and that the loss connected with the estimator $d(X, N)$ is of the form

$$(2) \quad L(Y, d) = (d - Y)^T C (d - Y) = \sum_{i,j=1}^r c_{ij} (d_i - Y_i)(d_j - Y_j),$$

where $C = (c_{ij})$ is a nonnegative definite, symmetric matrix.

Denote by

$$R(p, d, f) = E^N \{E^{(X,Y)} [L(Y, d(X, N)) | N]\}$$

the risk function connected with the predictor $d(X, N)$. We are interested in finding the minimax predictor of Y , say

$$d^*(X, N) = (d_1^*(X, N), \dots, d_r^*(X, N))^T,$$

for which

$$\sup_{p \in \mathcal{P}, (f(n)) \in \mathcal{F}} R(p, d^*, f) = \inf_{d \in \mathcal{D}} \sup_{p \in \mathcal{P}, (f(n)) \in \mathcal{F}} R(p, d, f),$$

where \mathcal{D} is the set of all predictors for which the risk function is finite, and $\mathcal{F} = \{(f(n)) : f(i) \geq 0, i = 0, 1, \dots, \sum_{i=0}^{\infty} f(i) = 1\}$.

Set $c = (c_{11}, \dots, c_{rr})^T$. The following theorem determines the minimax predictor of Y in the case when the distribution of the sample size N is known (Jokiel-Rokita (1997)).

THEOREM 1. *Suppose that the distribution of the sample size N given by (1) is known. Define a predictor of the random variable Y by*

$$d^0(X, N) = \begin{cases} m\beta_0 & \text{if } N = 0, \\ m\left(\frac{1}{N + \alpha_0} X + \frac{\alpha_0}{N + \alpha_0} \beta_0\right) & \text{if } N > 0, \end{cases}$$

whenever

$$(3) \quad m\left[f(0) - \sum_{i=1}^{\infty} \frac{f(i)}{i}\right] - 1 < 0,$$

and

$$d^0(X, N) = \begin{cases} m\beta_0 & \text{if } N = 0, \\ \frac{m}{N} X & \text{if } N > 0, \end{cases}$$

whenever

$$(4) \quad m\left[f(0) - \sum_{i=1}^{\infty} \frac{f(i)}{i}\right] - 1 \geq 0,$$

where α_0 is a solution to the equation

$$mE^N\left[\frac{\alpha^2 - N}{(N + \alpha)^2}\right] - 1 = 0,$$

and β_0 is the point $(\beta_1^0, \dots, \beta_r^0)^T$ for which

$$(5) \quad c^T \beta_0 - \beta_0^T C \beta_0 = \max_{\beta \in \mathcal{P}} (c^T \beta - \beta^T C \beta).$$

Then $d^0(X, N)$ is minimax under the loss function given by (2) with C nonnegative definite. ■

The following theorem determines a minimax predictor of Y in the case when the distribution of the sample size N is unknown.

THEOREM 2. *Under the loss function given by (2) with C nonnegative definite, the predictor of Y defined by*

$$(6) \quad d^*(X, N) = \begin{cases} m\beta_0 & \text{if } N = 0, \\ \frac{m}{N} X & \text{if } N > 0, \end{cases}$$

where β_0 is a solution to equation (5), is minimax.

Proof. The predictor $d^*(X, N)$ coincides with the minimax predictor $d^0(X, N)$ derived under the assumption that the inequality (3) is false. The risk function $R(p, d^*, f)$ of $d^*(X, N)$ is

$$R(p, d^*, f) = f(0)[(m^2 - m)p^T Cp + m^2 \beta_0^T C \beta_0 - 2m^2 \beta_0^T Cp + mc^T p] \\ + (c^T p - p^T Cp) \sum_{i=1}^{\infty} (m^2/i + m)f(i),$$

and

$$R(p, d^*, f) \leq f(0)[(m^2 - m)p^T Cp + m^2 \beta_0^T C \beta_0 - 2m^2 \beta_0^T Cp + mc^T p] \\ + (c^T p - p^T Cp)(m^2 + m) \sum_{i=1}^{\infty} f(i) \\ \leq (m^2 - m)p^T Cp + m^2 \beta_0^T C \beta_0 - 2m^2 \beta_0^T Cp + mc^T p \\ = R(p, m\beta_0, f_0),$$

where f_0 is the probability function of the distribution concentrated at zero. Thus the risk function of $d^*(X, N)$ is bounded from above by the risk function of $m\beta_0$, the minimax predictor of Y in the case when the ancillary statistic N has distribution given by f_0 (Jokiel-Rokita (1998)). Therefore, this bound must be attained by any decision used to predict Y under the unknown distribution of N . ■

REMARK 1. The minimax predictor $d_0(X)$ of Y , in the case when the sample size is fixed (Wilczyński (1985)), given by

$$(7) \quad d_0(X) = \begin{cases} m \left(\frac{1}{n + \alpha_0} X + \frac{\alpha_0}{n + \alpha_0} \beta_0 \right) & \text{if } n > 0, \\ m\beta_0 & \text{if } n = 0, \end{cases}$$

where

$$\alpha_0 = \frac{n + \sqrt{m^2 n + mn(n-1)}}{m-1} \quad \text{if } n > 0,$$

and β_0 is a solution to equation (5), is not minimax in the case when the sample size n is a random variable with an unknown distribution.

REMARK 2. In contrast to the case when the distribution of the random sample size N is known (but is not concentrated at a point) (Jokiel-Rokita (1997)), the minimax predictor $d_0(X)$ based on a fixed sample size, given by (7), remains admissible in the case when the distribution of N is unknown. This follows from the fact that $d_0(X)$ is the unique Bayes predictor (Jokiel-Rokita (1998)) when the ancillary statistic N is concentrated at a point.

3. Minimax prediction for the multivariate hypergeometric distribution. Let $X = (X_1, \dots, X_r)^T$ be an observation such that $X | N = n$, $n = 0, 1, \dots, W$, has the multivariate hypergeometric distribution with parameters (W, M, n) , i.e.,

$$P_p(X = (x_1, \dots, x_r)^T) = \begin{cases} \frac{\binom{M_1}{x_1} \cdots \binom{M_r}{x_r}}{\binom{W}{n}} & \text{if } x_i \in \{0, 1, \dots, M_i\}, i = 1, \dots, r, \\ & x_1 + \dots + x_r = n, \\ 0 & \text{otherwise,} \end{cases}$$

where $M_1 + \dots + M_r = W$, $W > 2$ is known, $r \geq 2$, $0 < n \leq W$, and $p = (M_1/W, \dots, M_r/W)^T$ is an unknown parameter. Moreover, suppose that N has an unknown distribution given by the probability function $f(n) = P(N = n)$ which does not depend on the unknown parameter p . We want to find the minimax predictor $d^*(X, N) = (d_1^*(X, N), \dots, d_r^*(X, N))^T$ of the random variable $Y = (Y_1, \dots, Y_r)^T$ which has the multivariate hypergeometric distribution with parameters (W, M, m) . We assume that X and Y are independent, and that m is known.

The following theorem determines the minimax predictor of Y in the case when the distribution of the sample size N is known (Jokieli-Rokita (1997)).

THEOREM 3. *Suppose that the distribution of the sample size N given by (1) is known. Define a predictor of the random variable Y by*

$$d^0(X, N) = \begin{cases} m\beta_0 & \text{for } N = 0, \\ m \left[\frac{W + \alpha_0}{W(N + \alpha_0)} X + \frac{\alpha_0(W - N)}{W(N + \alpha_0)} \beta_0 \right] & \text{for } N > 0, \end{cases}$$

whenever

$$\begin{cases} m \left[f(0) - \sum_{i=1}^W \frac{f(i)(W - i)}{i(W - 1)} \right] - \frac{W - m}{W - 1} < 0, \\ m \sum_{i=1}^W f(i) \frac{(W - i)(W - i - 1)}{W(W - 1)} - \frac{W - m}{W - 1} > 0; \end{cases}$$

by

$$d^0(X, N) = \begin{cases} m\beta_0 & \text{for } N = 0, \\ \frac{m}{W} [X + (W - N)\beta_0] & \text{for } N > 0, \end{cases}$$

whenever

$$m \sum_{i=1}^W f(i) \frac{(W-i)(W-i-1)}{W(W-1)} - \frac{W-m}{W-1} = 0;$$

and by

$$d^0(X, N) = \begin{cases} m\beta_0 & \text{for } N = 0, \\ \frac{m}{N} X & \text{for } N > 0, \end{cases}$$

whenever

$$(8) \quad m \left[f(0) - \sum_{i=1}^W \frac{f(i)(W-i)}{i(W-1)} \right] - \frac{W-m}{W-1} \geq 0,$$

where α_0 is a solution to the equation

$$mE^N \left[\frac{(W-N)^2(\alpha^2 - N) - (W-N)(N + \alpha)^2}{W(W-1)(N + \alpha)^2} \right] - \frac{W-m}{W-1} = 0,$$

and β_0 is a solution to equation (5). Then $d^0(X, N)$ is minimax under the loss function given by (2) with C nonnegative definite. ■

The following theorem determines a minimax predictor of Y in the case when the distribution of the sample size N is unknown.

THEOREM 4. *Under the loss function given by (2) with C nonnegative definite, the predictor of Y defined by*

$$(9) \quad d^*(X, N) = \begin{cases} m\beta_0 & \text{for } N = 0, \\ \frac{m}{N} X & \text{for } N > 0, \end{cases}$$

where β_0 is a solution to equation (5), is minimax.

Proof. The predictor $d^*(X, N)$ coincides with the minimax predictor $d^0(X, N)$ derived under the assumption that the distribution of the random sample size satisfies (8). The risk function $R(p, d^*, f)$ of $d^*(X, N)$ is

$$\begin{aligned} R(p, d^*, f) &= f(0) \left[\left(m^2 - m \frac{W-m}{W-1} \right) p^T C p + m^2 \beta_0^T C \beta_0 - 2m^2 \beta_0^T C p + m \frac{W-m}{W-1} c^T p \right] \\ &\quad + (c^T p - p^T C p) \sum_{i=1}^W \left(\frac{m^2(W-i)}{i(W-1)} + m \frac{W-m}{W-1} \right) f(i), \end{aligned}$$

and

$$\begin{aligned}
 & R(p, d^*, f) \\
 & \leq f(0) \left[\left(m^2 - m \frac{W-m}{W-1} \right) p^T C p + m^2 \beta_0^T C \beta_0 - 2m^2 \beta_0^T C p + m \frac{W-m}{W-1} c^T p \right] \\
 & \quad + (c^T p - p^T C p) \left(m^2 + m \frac{W-m}{W-1} \right) \sum_{i=1}^W f(i) \\
 & \leq \left(m^2 - m \frac{W-m}{W-1} \right) p^T C p + m^2 \beta_0^T C \beta_0 - 2m^2 \beta_0^T C p + m \frac{W-m}{W-1} c^T p \\
 & = R(p, m\beta_0, f_0),
 \end{aligned}$$

where f_0 is the probability function of the distribution concentrated at zero. Therefore the theorem follows from the same arguments as in the proof of Theorem 2. ■

REMARK 3. The minimax predictor $d_0(X)$ in the case when the sample size is fixed (Jokiel-Rokita (1998)), given by

$$(10) \quad d_0(X) = \begin{cases} m \left[\frac{W + \alpha_0}{W(n + \alpha_0)} X + \frac{\alpha_0(W - n)}{W(n + \alpha_0)} \beta_0 \right] & \text{if } n > 0, \\ m\beta_0 & \text{if } n = 0, \end{cases}$$

where

$$\begin{aligned}
 \alpha_0 &= \frac{n[m(W - n) + W(W - m)] + (W - n)\sqrt{\Delta}}{m(W - n)(W - n - 1) - W(W - m)}, \\
 \Delta &= m^2 n(W - n)(W - 1) + mnW(n - 1)(W - m)
 \end{aligned}$$

and β_0 is a solution to equation (5), is not minimax in the case when the sample size n is a random variable with an unknown distribution.

REMARK 4. Just as in Remark 2, the minimax predictor $d_0(X)$ based on a fixed sample size, given by (10), remains admissible in the case when the distribution of the random sample size N is unknown.

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

EXAMPLE 1 (multinomial distribution). Suppose that during a certain period we observe s independently working devices, and each of them can undergo failure with an unknown probability θ . Each of the failures can appear due to one of r possible reasons. Then the observed values $X_i, i = 1, \dots, r$, represent the number of devices which have been damaged as a consequence of the i th reason, and N is the number of devices which have suffered damage. Hence the sample size N is an ancillary statistic which has the binomial distribution $b(s, \theta)$. The purpose is to predict the value

of $Y = (Y_1, \dots, Y_r)^T$, where Y_i is the unknown number among m devices which we should expect to be destroyed as a result of the i th reason. Using Theorem 2, under the loss function given by (2) with $C = I$, we see that the minimax predictor of Y is

$$d^*(X, N) = \begin{cases} m\beta_0 & \text{if } N = 0, \\ \frac{m}{N} X & \text{if } N > 0, \end{cases}$$

where $\beta_0 = (1/r, \dots, 1/r)^T$.

EXAMPLE 2 (multivariate hypergeometric distribution). Suppose that we randomly choose (without replacement) s elements from a population which contains a subset of W elements. The values M_i , $i = 1, \dots, r$, representing for example the number of elements of the i th quality category are assumed to be unknown ($\sum_{i=1}^r M_i = W$). The observed values X_i , $i = 1, \dots, r$, represent the number of elements from the subset of W elements which are from the i th quality category. Hence the sample size N is an ancillary statistic which has the hypergeometric distribution. We are interested in predicting the value of $Y = (Y_1, \dots, Y_r)^T$, where Y_i is the unknown number of elements of the i th category which we should expect to be in our new sample of size m . One can give an analogous example relating to a voting model. In these cases we can use Theorem 4.

References

- P. Amrhein (1995), *Minimax estimation of proportions under random sample size*, J. Amer. Statist. Assoc. 90, 1107–1111.
- L. D. Brown (1990), *An ancillarity paradox which appears in multiple linear regression*, Ann. Statist. 18, 471–493.
- A. Chaudhuri and H. Stenger (1992), *Survey Sampling—Theory and Methods*, Dekker, New York.
- K. He (1990), *An ancillarity paradox in the estimation of multinomial probabilities*, J. Amer. Statist. Assoc. 85, 824–828.
- A. Jokieli-Rokita (1997), *Minimax prediction for the multinomial and multivariate hypergeometric distributions under random sample size*, technical report, Inst. Math., Wrocław Univ. of Technology.
- A. Jokieli-Rokita (1998), *Minimax prediction for the multinomial and multivariate hypergeometric distributions*, Appl. Math. (Warsaw) 25, 271–283.
- M. Wilczyński (1985), *Minimax estimation for the multinomial and multivariate distributions*, Sankhyā 47, 128–132.

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

Received on 11.7.2001;
revised version on 4.2.2002

(1585)