SOME PROBLEMS OF SIMULTANEOUS MINIMAX ESTIMATION

By Stanisław Trybuła

Institute of Mathematics, Polish Academy of Sciences, Wroclaw

1. Summary. In this paper, we give minimax estimates of the parameters of the multivariate hypergeometric distribution and of the multinomial distribution, and of some parameters of an unspecified distribution with known range. We use as loss a weighted linear combination of squared differences between the true and the estimated values of the parameters. Some properties of the minimax estimates obtained are discussed.

2. Introduction. For our purpose, it is sufficient to define the estimation problem in a fixed sample size experiment as follows ([3], [4]). The random variable X is distributed in the space $\mathfrak{X}$ according to the distribution F belonging to the family $\mathfrak{F}$. We want to estimate $\omega(F)$ where ω is a function, the values of which belong to some space Ω, defined on $\mathfrak{F}$. (In the following we assume that X and $\omega(F)$ are vector valued.) An estimate is a statistic $f(X)$ having values in Ω. The nonnegative function $L[\omega(F), f(x)]$ is the loss resulting if, when F obtains, the estimate $f(x)$ is made. Define the risk by

$$R(f, F) = E\{L[\omega(F), f(X)]|F\} \tag{1}$$

and call $v(f) = \sup_{F\varepsilon\mathfrak{F}} R(f, F)$ the guaranteed value for the estimate f. We seek the minimax estimate f^0, that is, the estimate whose guaranteed value is minimal. Obviously, such an estimate does not always exist. It is our aim to derive minimax estimates in some specific problems.

3. Problem 1. In practice, we often meet the following situation. A lot consisting of N units of a product has been produced. The units are classified into l categories, the ith category containing U_i units ($i = 1, \cdots, l$). A sample of size n is taken from the lot in which $k_1, \cdots, k_l$ units of categories $1, \cdots, l$ are cbserved. The problem is to estimate $U_1, \cdots, U_l$.

This leads to the estimation of the parameter $U = (U_1, \cdots, U_l)$ of a multivariate hypergeometric distribution. Thus, let

$$P(X_1 = k_1, \cdots, X_l = k_l) = \frac{\binom{U_1}{k_1} \cdots \binom{U_l}{k_l}}{\binom{N}{n}}. \tag{2}$$

It is known that,

$$m_i = E(X_i \mid U) = n\frac{U_i}{N}, \tag{3}$$

Received April 12, 1957; revised July 11, 1957.

$$\sigma_i^2 = E\{[X_i - E(X_i \mid U)]^2 \mid U\} = \frac{n(N-n)}{N^2(N-1)} U_i(N - U_i). \tag{4}$$

Suppose that the loss is

$$L(U, f) = \sum_{i=1}^{l} c_i[f_i(X) - U_i]^2 \qquad (c_i \geqq 0), \tag{5}$$

where $f = (f_1, \cdots, f_l)$ is the estimate of U and $X = (X_1, \cdots, X_n)$ is the sample. The risk is then

$$R(f, U) = E[L(U, f) \mid U] = E\left\{\sum_{i=1}^{l} c_i[f_i(X) - U_i]^2 \mid U\right\}. \tag{6}$$

If we study estimates of the form

$$f_i(X) = aX_i + b_i \qquad (i = 1, \cdots, l),$$

then

$$\begin{aligned} R(f, U) &= \sum_{i=1}^{l} c_i E\{[aX_i + b_i - U_i]^2 \mid U\} \\ &= \sum_{i=1}^{l} c_i[(am_i + b_i - U_i)^2 + a^2\sigma_i^2] \\ &= \sum_{i=1}^{l} c_i\left[\left(a\frac{nU_i}{N} + b_i - U_i\right)^2 + a^2\frac{n(N-n)}{N^2(N-1)} U_i(N - U_i)\right]. \end{aligned} \tag{7}$$

Let the constant a assume a value such that the terms quadratic in U vanish. For this, it suffices to put

$$a = \frac{N}{n + \sqrt{n\dfrac{N-n}{N-1}}}.$$

If, moreover, we put

$$b_i = \frac{s_i N\sqrt{n\dfrac{N-n}{n-1}}}{n + \sqrt{n\dfrac{N-n}{N-1}}},$$

then (7) may be written

$$R(f, U) = \frac{nN\dfrac{N-n}{N-1}}{\left(n + \sqrt{n\dfrac{N-n}{N-1}}\right)^2} \sum_{i=1}^{l} c_i[Ns_i^2 + (1 - 2s_i)U_i]. \tag{8}$$

Without loss of generality, we may assume $c_1 \geqq c_2 \geqq \cdots \geqq c_l \geqq 0$. For the present, assume also that $c_2 \neq 0$. Let l_0 be the greatest index i for which $c_i \neq 0$

and let

$$L = \max_s \left[s \leqq l_0, \sum_{i=1}^{s} 1/c_i > \frac{s-2}{c_s} \right]. \tag{9}$$

The above assumptions being satisfied, we prove the following lemma:

If $L \leqq l$ then

$$\delta = \frac{L-2}{\sum_{j=1}^{L} 1/c_j} \geqq c_i \text{ for } i = L+1, L+2, \cdots, l. \tag{10}$$

PROOF. First, observe that a proof of the inequality is necessary only for $i = L + 1$. If $c_{L+1} = 0$, then the lemma obviously holds. If $c_{L+1} \neq 0$, it follows from the definition of L that

$$L - 1 \geqq c_{L+1} \sum_{j=1}^{L+1} \frac{1}{c_j} = 1 + c_{L+1} \sum_{j=1}^{L} 1/c_j .$$

The lemma is a direct consequence of this inequality.

Now put

$$s_i = \begin{cases} \frac{1}{2}\left(1 - \frac{\delta}{c_i}\right), & \text{when } i \leqq L, \\ 0, & \text{when } i > L. \end{cases} \tag{11}$$

Observe that $i \leqq L$, $0 < s_i \leqq \frac{1}{2}$. We shall show that the estimate

$$f^0 = (f_1^0, f_2^0, \cdots, f_l^0),$$

where

$$f_i^0(X) = N \frac{X_i + s_i \sqrt{n \frac{N-n}{N-1}}}{n + \sqrt{n \frac{N-n}{N-1}}}, \tag{12}$$

is the minimax estimate sought.

From (8) and (11) we have

$$R(f^0, U) = \frac{Nn \frac{N-n}{N-1}}{\left(n + \sqrt{n \frac{N-n}{N-1}}\right)^2} \left\{ \sum_{i=1}^{L} \left[c_i \frac{N}{4}\left(1 - \frac{\delta}{c_i}\right)^2 + \delta U_i \right] + \sum_{i=L+1}^{l} c_i U_i \right\}. \tag{13}$$

Observe that for

$$U_{L+1} = U_{L+2} = \cdots = U_l = 0, \tag{14}$$

$R(f^0, U) = c$, where c is a constant. By the lemma, $R(f^0, U) \leqq c$. Thus, by theorem 2.1 of [4], it is sufficient to prove that a distribution of the random variable U exists which satisfies (14) and for which f^0 is the Bayes estimate.

We seek for such a distribution among those of the form

$$P(U_{L+1} = \cdots = U_l = 0) = 1 \tag{15}$$

$$P(U_1 = u_1, \cdots, U_L = u_L) = C\,\frac{\Gamma(a_1 + u_1) \cdots \Gamma(a_L + u_L)}{u_1! \cdots u_L!}. \tag{16}$$

Let

$$r(f, P) = E[R(f, U)] = \sum_{i=1}^{l} c_i E\{E[f_i(X) - U_i]^2 \mid U\}. \tag{17}$$

It follows from (15) that the expected risk does not depend on f_i if $k_j \neq 0$ for at least one $j > L$. Thus, any estimate which minimizes (17) throughout the region $k_{L+1} = k_{L+2} = \cdots = k_l = 0$ is a Bayes estimate. Now, if

$$k_{L+1} = k_{L+2} = \cdots = k_l = 0$$

then, as is well-known, the expression (17) attains its minimum value for

$$\begin{aligned} &f_i(k_1, \cdots, k_L, 0, \cdots, 0) \\ &\quad = E(U_i \mid X_1 = k_1, \cdots, X_L = k_L;\, X_{L+1} = \cdots = X_l = 0) \\ &\quad = \begin{cases} 0 & \text{for } i > L; \\ \dfrac{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1\geq k_1,\cdots,u_L\geq k_L}} u_i \prod_{j=1}^{L} \binom{u_j}{k_j} \frac{\Gamma(a_j + u_j)}{u_j!}}{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1\geq k_1,\cdots,u_L\geq k_L}} \prod_{j=1}^{L} \binom{u_j}{k_j} \frac{\Gamma(a_j + u_j)}{u_j!}} & \text{otherwise.} \end{cases} \end{aligned} \tag{18}$$

The second part of Eq. (18) reduces to

$$\begin{aligned} &\frac{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1\geq k_1,\cdots,u_L\geq k_L}} u_i \prod_{j=1}^{L} \frac{\Gamma(a_j + u_j)}{(u_j - k_j)!}}{\displaystyle\sum_{\substack{u_1+\cdots+u_L=N \\ u_1\geq k_1,\cdots,u_L\geq k_L}} \prod_{j=1}^{L} \frac{\Gamma(a_j + u_j)}{(u_j - k_j)!}} \\ &\qquad = \frac{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geq 0,\cdots,v_L\geq 0}} [(a_i + k_i + v_i) - a_i] \prod_{j=1}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}}{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geq 0,\cdots,v_L\geq 0}} \prod_{j=1}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}} \\ &\qquad = \frac{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geq 0,\cdots,v_L\geq 0}} \Gamma(a_i + k_i + v_i + 1) \prod_{\substack{j=1 \\ j\neq i}}^{L} \frac{\Gamma(a_j + k_j + v_j)}{v_j!}}{\displaystyle\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geq 0,\cdots,v_L\geq 0}} \prod_{j=1}^{L} \frac{\Gamma(a_i + k_j + v_j)}{v_j!}} - a_i \\ &\qquad = \frac{L_i}{M_i} - a_i. \end{aligned}$$

Observe that

$$\sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geqq 0,\cdots,v_L\geqq 0}} \frac{(N-n)!}{v_1!\cdots v_L!} \frac{\Gamma(b_1+v_1)\cdots\Gamma(b_L+v_L)}{\Gamma\left(N-n+\sum_{j=1}^{L} b_j\right)}$$

(19)
$$= \int\cdots\int_{\substack{p_1+\cdots+p_L=1 \\ p_1\geqq 0,\cdots,p_L\geqq 0}} p_1^{b_1-1}\cdots p_L^{b_L-1} \sum_{\substack{v_1+\cdots+v_L=N-n \\ v_1\geqq 0,\cdots,v_L\geqq 0}} \frac{(N-n)!}{v_1!\cdots v_L!} \cdot p_1^{v_1}\cdots p_L^{v_L}\,dp_1\cdots dp_L$$

$$= \int\cdots\int_{\substack{p_1+\cdots+p_L=1 \\ p_1\geqq 0,\cdots,p_L\geqq 0}} p_1^{b_1-1}\cdots p_L^{b_L-1}\,dp_1\cdots dp_L = \frac{\Gamma(b_1)\cdots\Gamma(b_L)}{\Gamma\left(\sum_{i=1}^{L} b_i\right)}.$$

Applying (19) to L_i with $b_j = a_j + k_j$ for $j \neq i$ and

$$b_i = a_i + k_i + 1 (i = 1, \cdots, L)$$

and to M_i with $b_j = a_j + k_j$ we obtain

(20)
$$\frac{L_i}{M_i} - a_i = \frac{(a_i+k_i)\left(N+\sum_{j=1}^{L} a_j\right)}{n+\sum_{j=1}^{L} a_j} - a_i$$

$$= \frac{\left(N+\sum_{j=1}^{L} a_j\right)k_i + (N-n)a_i}{n+\sum_{j=1}^{L} a_j} = f_i^0(k_1, \cdots, k_L, 0, \cdots, 0)$$

for

$$a_i = s_i \frac{N\sqrt{n\frac{N-n}{N-1}}}{N-n-\sqrt{n\frac{N-n}{N-1}}}.$$

Thus f^0 is minimax whenever $a_i > 0$; that is, when $N > n + 1$. For $N = n$ this result is immediate. For $N = n + 1$ it is a consequence of the fact that f^0 is Bayes for the a priori distribution of U defined by

$$P(U_{L+1} = \cdots = U_l = 0) = 1,$$

$$P(U_1 = u_1, \cdots, U_L = u_L) = \frac{N!}{u_1!\cdots u_L!} s_1^{u_1}\cdots s_L^{u_L}.$$

Up to this point, we have assumed $c_2 \neq 0$. Consider now the remaining cases. If all $c_i = 0$ then, obviously, every estimate is minimax. If alone $c_1 \neq 0$ then the problem may be considered as that of finding a minimax estimate of the parameter U in a one-dimensional hypergeometric distribution for the loss $L = (f - U)^2$. In this case, the formula for the minimax estimate is (see [4])

$$f(X) = N \frac{X + \frac{1}{2}\sqrt{n \frac{N-n}{N-1}}}{n + \sqrt{n \frac{N-n}{N-1}}}. \tag{21}$$

It is easy to verify that the estimate (12) satisfies the condition

$$\sum_{i=1}^{L} f_i^0 = N.$$

Observe that we are actually dealing with only $l - 1$ independent parameters since one parameter, say U_l, may be computed from

$$U_1 + \cdots + U_l = N. \tag{22}$$

If we consider the problem of finding a minimax estimate for $U_1, \cdots, U_{l-1}$ under the loss

$$\bar{L}(U, f) = \sum_{i=1}^{l-1} \bar{c}_i(f_i - U_i), \tag{23}$$

the same estimate as above for $U_1, \cdots, U_{l-1}$ results as is seen by identifying c_i in the above with $\bar{c}_i (i = 1, \cdots, l - 1)$ and putting $c_l = 0$.

In solving our problem, we have restricted ourselves to the case $c_i \geqq 0$. If, however, some $c_j < 0$ then for $f_j \to \pm\infty$ the loss tends to $-\infty$ and, consequently, the problem becomes trivial.

In the special case $c_1 = c_2 = \cdots = c_l > 0$, formula (12) takes the form

$$f_i^0(x) = N \frac{X_i + \frac{1}{l}\sqrt{n \frac{N-n}{N-1}}}{n + \sqrt{n \frac{N-n}{N-1}}}. \tag{24}$$

4. Corollaries for the multinomial case. For $N \to \infty$, the distribution of X converges to the multinomial distribution defined by

$$P(X_1 = k_1, \cdots, X_l = k_l) = \frac{n!}{k_1! \cdots k_l!} p_1^{k_1} \cdots p_l^{k_l},$$

$$0 \leqq k_i, 0 \leqq p_i \leqq 1, i = 1, \cdots, l'$$

$$\sum_1^l k_i = n;$$

and

$$\lim_{N\to\infty} \frac{f_i^0(X)}{N} = \frac{X_i + s_i\sqrt{n}}{n + \sqrt{n}} = g_i^0(x).$$

We shall prove[1] that $g^0 = (g_1^0, \cdots, g_l^0)$ is really a minimax estimate of the parameter $p = (p_1, \cdots, p_l)$ for the loss

$$L(g, p) = \sum_{i=1}^{l} c_i(g_i - p_i)^2, \qquad c_i \geqq 0. \tag{25}$$

When L, δ and s_i are defined by (9), (10), and (11), respectively, the loss is

$$R(g^0, p) = E\{L(g^0, p) \mid p\} = \frac{1}{(\sqrt{n} + 1)^2}\left[\sum_{i=1}^{L} (c_i s_i^2 + \delta p_i) + \sum_{i=L+1}^{l} c_i p_i\right], \tag{26}$$

which for $p_{L+1} = \cdots = p_l = 0$ is constant and, by the lemma of Sec. 2, maximum in p.

As is easy to verify, g^0 is Bayes for the a priori distribution $G(p)$ defined by

$$dG(p) = \begin{cases} Cp_1^{\sqrt{n}s_1 - 1} \cdots p_L^{\sqrt{n}s_L - 1}, & \text{when } p_{L+1} = \cdots = p_l = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

By theorem 2.1 of [4] it follows that g^0 is minimax.

For $c_1 = \cdots = c_l > 0$, $s_i = 1/l$ and the minimax estimate g^0 takes the form

$$g_i^0(X) = \frac{X_i + \frac{1}{l}\sqrt{n}}{n + \sqrt{n}}.$$

This case was previously solved by H. Steinhaus in [5].

5. Problem 2. We shall prove the following theorem:

THEOREM. *Let X be a random variable distributed according to the unknown distribution F on the measurable space A. Let $g_1, \cdots, g_m$ be such bounded measurable functions on A that there exist two points x' x'' ε A such that each of these functions attains its minimum in x' and its maximum in x''. Let $X_1, \cdots, X_n$ be a random sample from F, and let $\lambda_i = E[g_i(X)]$. If the loss is given by*

$$L(f, \lambda) = \sum_{i=1}^{m} c_i(f_i - \lambda_i)^2, \tag{29}$$

where $f = (f_1, \cdots, f_m)$ is an estimate of $\lambda = (\lambda_1, \cdots, \lambda_m)$, then the minimax estimate of λ is given by

$$f_i^0(X_1, \cdots, X_n) = \frac{\sum_{j=1}^{n} g_i(X_j)}{n + \sqrt{n}} + \frac{s_i}{\sqrt{n} + 1} \tag{30}$$

(s_i is the arithmetic mean of the maximum and minimum values of $g_i(x)$).

PROOF. If $\bar{f}_i(X_1, \cdots, X_n) = a \sum_{j=1}^{n} g_i(X_j) + b_i$, then the risk may be

[1] While this paper was being written, Joseph Dubay communicated to me a result similar to this but not in its full generality.

written

$$
(31)\quad \begin{aligned} R(\bar{f}, F) &= E\left[\sum_{i=1}^{m} c_i(f_i - \lambda_i)^2 \mid F\right] = \sum_{i=1}^{m} c_i E\left\{\left[a\sum_{j=1}^{n} g_i(X_j) + b_i - \lambda_i\right]^2 \middle| F\right\} \\ &= \sum_{i=1}^{m} c_i\{[(1 - an)\lambda_i - b_i]^2 + na^2E\{[g_i(X) - \lambda_i]^2 \mid F\}\}. \end{aligned}
$$

Let

$$
\alpha_i = \min_{x\varepsilon A} g_i(x) = g_i(x'), \qquad \beta_i = \max_{x\varepsilon A} g_i(x) = g_i(x'').
$$

It is easy to prove that

$$
(32)\qquad E\{[g_i(X) - \lambda_i]^2 \mid F\} \leqq (\beta_i - \lambda_i)(\lambda_i - \alpha_i).
$$

Thus

$$
(33)\qquad R(\bar{f}, F) \leqq \sum_{i=1}^{m} c_i\{[(1 - an)\lambda_i - b_i]^2 + na^2(\beta_i - \lambda_i)(\lambda_i - \alpha_i)\}.
$$

Putting

$$
a = \frac{1}{n + \sqrt{n}}, \qquad b_i = \frac{s_i}{\sqrt{n} + 1},
$$

we obtain

$$
(34)\qquad R(f^0, F) \leqq \frac{1}{4(\sqrt{n} + 1)^2} \sum_{i=1}^{m} c_i(\beta_i - \alpha_i)^2 = c.
$$

Observe that if a distribution $\bar{F}$ of the random variable X is defined by

$$
(35)\qquad \begin{aligned} P(X = x') &= 1 - p, \\ P(X = x'') &= p. \end{aligned}
$$

Then $\lambda_i = \alpha_i + (\beta_i - \alpha_i)p$, and equality obtains in (32); i.e.

$$
(36)\qquad R(f^0, \bar{F}) = c.
$$

The distribution F depends on the parameter p. Since (34) and (36) hold, it is sufficient to show (as in Sec. 3) that there exists a distribution G of p for which (30) is Bayes—that is, a distribution G such that (30) minimizes the expected risk

$$
\begin{aligned} r(f, G) &= E[R(f, \bar{F})] = \sum_{i=1}^{m} c_i\, E\{E[(\lambda_i - f_i)^2 \mid \bar{F}]\} \\ &= \sum_{i=1}^{m} c_i E\{E\{[\alpha_i + (\beta_i - \alpha_i)p - f_i]^2 \mid p\}\}. \end{aligned}
$$

It is easy to verify that this happens for the distribution $G^0(p)$ defined by equation

$$dG^0(p) = C(pq)^{(\sqrt{n}/2)-1}\, dp \qquad (q = 1 - p). \tag{37}$$

This completes the proof.

6. In this paper, we have used the loss $L = \sum_{i=1}^m c_i(f_i - \omega_i)^2$. This loss has been extensively investigated ([2], [4], [5], [6]). For many special problems, other loss functions might be used, for example,

$$L = \sum_{i=1}^{m} c_i \mid f_i - \omega_i \mid,$$

about which little is known at present.

Problems considered in this paper were suggested to me by L. J. Savage and H. Steinhaus. I am indebted to J. A. Dubay and L. J. Savage for help and suggestions made during the preparation of this paper.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, Inc., New York, 1954.

[2] M. A. Girshick and L. J. Savage, "Bayes and minimax estimates for quadratic loss functions," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 53–73.

[3] P. R. Halmos, "The theory of unbiased estimation," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 34–43.

[4] J. L. Hodges, Jr., and E. L. Lehmann, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 182–197.

[5] H. Steinhaus, "The problem of estimation," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 633–648.

[6] S. Trybuła, "On minimaksowej estymacji parametrów w rozkładzie wielomianowym," to be published.

[7] A. Wald, "Contributions to the theory of statistical estimation and testing hypothesis," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 229–325.

[8] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, Inc., New York, 1950.