[12] J. L. Doob, *Heuristic approach to the Kolmogorov-Smirnov theorems*, Annals of Mathematical Statistics 20 (1949), p. 393-403.

[13] Б. В. Гнеденко, В. С. Королюк, *О максимальном расхождении двух эмпирических распределений*, Доклады Академии Наук СССР 80 (1951), p. 525-528.

[14] N. Centsov, *Weak convergence of stochastic processes whose realizations have no discontinuities of second kind*, Теория Вероятностей и ее Применения 1 (1956), p. 155-161.

[15] Yu. V. Prohorov, *Convergence of random processes and limit theorems in probability theory*, Теория Вероятностей и ее Применения 1 (1956), p. 177-238.

[16] M. L. Donsker, *Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems*, Annals of Mathematical Statistics 23 (1952), p. 277-281.

MATHEMATICAL INSTITUTE OF THE POLSIH ACADEMY OF SCIENCES

Reçu par la Rédaction le 17. 3. 1959

COLLOQUIUM MATHEMATICUM

VOL. VII 1960 FASC. 2

ON SOME LOSS FUNCTIONS

BY

S. TRYBUŁA (WROCŁAW)

In this paper we shall deal with some questions concerning the Wald theory of decision functions. For some known distributions depending on a parameter we shall find a loss function such that the minimax estimate of that parameter is unbiased. We shall see that the least favourable prior distribution of the estimated parameter is the uniform one.

1. Definitions. Let $F(x|\omega)$ be a distribution function defined on a Euclidean space $\mathcal{X}$ which depends on a parameter $\omega \in \Omega$. In the sequel we shall assume that ω is a vector. Each estimate of ω is a measurable function $f(x)$ with values belonging to Ω. Let $L[f(x), \omega_0]$ be the loss to the statistician if he applies the estimate $f(x)$ when x is the observed value of X, and ω_0 is the value of the parameter ω. If we establish the function $f(x)$ and the value of ω, then we can find the expected value of the loss L, i. e.

$$R(f, \omega) = \int_{\mathcal{X}} L[f(x), \omega]\, dF(x|\omega) \overset{\mathrm{df}}{=} E\{L[f(X), \omega]|\omega\}; \tag{1.1}$$

here X is a random variable with distribution function $F(x|\omega)$.

The function $R(f, \omega)$ will be called the *risk*.

The estimate f^0 is called *minimax* if

$$\sup_{\omega\in\Omega} R(f^0, \omega) = \inf_f \sup_\omega R(f, \omega). \tag{1.2}$$

Let the prior distribution of the parameter ω be given by a distribution function $G(\omega)$. The expected risk $r(f, G)$ is

$$r(f, G) = \int_{\Omega} R(f, \omega)\, dG(\omega) \overset{\mathrm{df}}{=} E_G[R(f, \omega)]. \tag{1.3}$$

According to Wald the estimate $f_G(x)$ minimizing for a given G the function $r(f, G)$ will be called the *Bayes estimate* for G. The distribution G^0 for which

$$\inf_f r(f, G^0) = \sup_G \inf_f r(f, G) \tag{1.4}$$

holds, is defined to be the least favourable distribution.

In the sequel the loss function will in general be of the form

$$L[f(x), \omega] = \sum_{i=1}^{k} \frac{[f_i(x)-\omega_i]^2}{\omega_i} \tag{1.5}$$

where $f = (f_1, \ldots, f_k)$ and $\omega = (\omega_1, \ldots, \omega_n)$. The distributions of X which will be dealt with are the multinomial and multivariate hypergeometric.

2. The multinomial distribution. It is known (see [1]) that if the random variable X is binomial and the loss function L is given by

$$L[f(x), p] = \frac{[f(x)-p]^2}{p(1-p)},$$

then the minimax estimate f^0 is given by the formula $f^0(x) = x/n$. The following question arises: which loss function should be chosen in the multinomial case in order to obtain a similar result? Let us set

$$P(X_1 = m_1, \ldots, X_n = m_n) = \frac{n!}{m_1! \ldots m_n!} p_1^{m_1} \ldots p_n^{m_n}. \tag{2.1}$$

We shall prove

THEOREM 1. *If the random variable X is distributed according to* (2.1), *the loss function L is given by*

$$L[f(x), p] = \sum_{i=1}^{k} \frac{[f_i(x)-p_i]^2}{p_i} \tag{2.2}$$

and we restrict ourselves to estimates $f = (f_1, \ldots, f_k)$ with $\sum_{i=1}^{k} f_i = 1$, then the estimate $f^0 = (f_1^0, \ldots, f_k^0)$ defined by $f_i^0(x) = x_i/n$ is the unique minimax estimate of parameter $p = (p_0, \ldots, p_k)$. Furthermore the uniform distribution is the least favourable one.

Proof. It is known that

$$\mu_i = E(X_i|p) = np_i,$$

$$\sigma_i^2 = E[(X_i-\mu_i)^2|p] = np_i(1-p_i) \qquad (i = 1, 2, \ldots, k).$$

We shall show first that for the loss function (2.2) and $f \equiv f^0$ the risk R is constant, namely

$$R(f^0, p) = E\left[\sum_{i=1}^{k} \frac{\left(\frac{x_i}{n}-p_i\right)^2}{p_i} \Big| p\right] = \sum_{i=1}^{k} \frac{\sigma_i^2}{n^2 p_i} = \frac{1}{n}\sum_{i=1}^{k}(1-p_i) = \frac{k-1}{n}.$$

Furthermore we shall show that the estimate f^0 is a Bayes estimate for the uniform distribution G^0 of the parameter $p = (p_1, \ldots, p_k)$. We have

$$dG^0(p) = c\,dp_1 \ldots dp_k,$$

$$r(f, G^0) = c \overbrace{\int \ldots \int}^{k}_{\substack{p_1+\ldots+p_k=1 \\ p_1\geqslant 0, \ldots, p_k \geqslant 0}} \sum_{m_1+\ldots+m_k=n} \frac{n!}{m_1! \ldots m_k!} p_1^{m_1} \ldots$$

$$\ldots, p_k^{m_k} \sum_{i=1}^{k} \frac{[f_i(m_1, \ldots, m_k)-p_i]^2}{p_i} dp_1 \ldots dp_k$$

$$= c \sum_{m_1+\ldots+m_k=n} \frac{n!}{m_1! \ldots m_k!} \int_{\substack{p_1+\ldots+p_k=1 \\ p_1\geqslant 0, \ldots, p_k \geqslant 0}} \ldots \int p_1^{m_1} \ldots \tag{2.3}$$

$$\ldots p_k^{m_k} \left[\sum_{i=1}^{k} \frac{f_i^2(m_1, \ldots, m_k)}{p_i} - 1\right] dp_1 \ldots dp_k$$

$$= c \sum_{m_1+\ldots+m_k=n} \frac{n!}{m_1! \ldots m_k!} \sum_{i=1}^{k} f_i^2(m_1, \ldots, m_k) \times$$

$$\times \int_{\substack{p_1+\ldots+p_k=1 \\ p_1\geqslant 0, \ldots, p_k \geqslant 0}} \ldots \int p_1^{m_1} \ldots p_{i-1}^{m_{i-1}} p_i^{m_i-1} p_{i+1}^{m_{i+1}} \ldots p_k^{m_k} dp_1 \ldots dp_k + C,$$

where C is finite and does not depend on the value of $f_i(m_1, \ldots, m_k)$.

It is known that

$$\int_{\substack{p_1+\ldots+p_k=1 \\ p_1\geqslant 0, \ldots, p_k \geqslant 0}} \ldots \int p_1^{a_1} \ldots p_k^{a_k} dp_1 \ldots dp_k$$

is finite if and only if $a_1 > -1, \ldots, a_k > -1$. Thus the expected risk will be finite only if

$$f_i(m_1, \ldots, m_k) = 0 \quad \text{for} \quad m_i = 0 \quad (i = 1, 2, \ldots, k). \tag{2.4}$$

There exists an estimate (e. g. f^0) for which the expression (2.3) is finite. This implies that the values $f_i(m_1, \ldots, m_k)$ which minimize (2.3) must satisfy condition (2.4). Assuming this condition to be satisfied we obtain

$$R(f, G^0) = c \sum_{m_1+\ldots+m_k=n} \frac{n!}{m_1!\ldots m_k!} \sum_{\substack{i \\ m_i\neq 0}} f_i^2(m_1, \ldots, m_k) \times \tag{2.5}$$

$$\times \int_{\substack{p_1+\ldots+p_k=1 \\ p_1\geqslant 0,\ldots,p_k\geqslant 0}} \ldots \int p_1^{m_1}\ldots p_{i-1}^{m_{i-1}} p_i^{m_i-1} p_{i+1}^{m_{i+1}}\ldots p_k^{m_k}\, dp_1\ldots dp_k + C.$$

In order to determine the values $f_i(m_1, \ldots, m_k)$ for which (2.5) attains its minimum it is necessary to find for each system $(m_1, \ldots, m_k)$ the values x_i which minimize the quadratic form

$$\varphi = \sum_{\substack{i \\ m_i\neq 0}} x_i^2 \overbrace{\int\ldots\int}^{k} p_1^{m_1}\ldots p_{i-1}^{m_{i-1}} p_i^{m_i-1} p_{i+1}^{m_{i+1}}\ldots p_k^{m_k}\, dp_1\ldots dp_k, \tag{2.6}$$

where x_i satisfy the condition $\sum_{i;m_i\neq 0} x_i = 1$. Let the numbers $m_1, \ldots, m_k$ be arbitrary. Without loss of generality we can assume that $m_k \neq 0$. Putting in formula (2.6)

$$x_n = 1 - \sum_{\substack{i:m_i\neq 0 \\ i\neq k}} x_i$$

and observing that the form (2.6) is positively determined, we shall find its minimum since the partial derivatives vanish there:

$$\frac{\partial\varphi}{\partial x_i} = 0 \quad (i: m_i \neq 0,\ i \neq k). \tag{2.7}$$

Thus we have

$$\frac{(n+k-2)!}{(m_i-1)!\prod\limits_{\substack{j=1 \\ j\neq i}}^{k} m_j!}\, x_i = \frac{(n+k-2)!}{(m_k-1)\prod\limits_{j=1}^{k-1} m_j!}\, x_k$$

or

$$m_k x_i - m_i x_k = 0. \tag{2.8}$$

This formula has been proved for those i for which $m_i \neq 0$. But it follows from condition (2.4) that (2.8) holds also for the remaining values of i.

Summing equalities (2.8), where $i = 1, 2, \ldots, k$, we obtain

$$m_k - n x_k = 0,$$

i. e.

$$x_k = \frac{m_k}{n}.$$

Substituting this result in (2.8) we obtain

$$x_i = \frac{m_i}{n} \quad (i = 1, 2, \ldots, k). \tag{2.9}$$

The estimate f^0 minimizes $r(f, G^0)$ and thus it is by definition the Bayes estimate for G^0.

We shall now use the following lemma given in paper [3]:

LEMMA. *The Bayes estimate f_G for which the risk does not depend on parameter ω is the minimax estimate. If, furthermore, the estimate f_G is the unique Bayes estimate for G, then it is the unique minimax estimate.*

From this lemma, and from the above results, it follows that the estimate f^0 is the minimax one. Its uniqueness follows from the fact that the numbers $x_i = m_i/n$ which minimize (2.3) were uniquely determined.

Since for $f = f^0$ the risk $R(f, p)$ is constant, it follows that the expected risk $r(f^0, G)$ does not depend on G. This implies

$$\min_f r(f, G) \leqslant r(f, G^0) = r(f^0, G^0) = \min_f r(f, G^0).$$

The uniform distribution G^0 is thus the least favourable.

3. The multivariate hypergeometric distribution. In an urn there are N balls, M_1 of them denoted by 1, M_2 of them denoted by 2, ..., and M_k of them denoted by k. If we take out of the urn n balls then the probability that there are among them m_1 denoted by 1, ..., m_k denoted by k $(\sum_{i=1}^{k} m_i = n)$ is given by

$$P(X_1 = m_1, \ldots, X_n = m_k) = \frac{\binom{M_1}{m_1}\binom{M_2}{m_2}\ldots\binom{M_k}{m_k}}{\binom{N}{n}}. \tag{3.1}$$

This distribution depends on the parameter $M = (M_1, \ldots, M_k)$, which is often unknown in practice and must be estimated from a sample. We shall prove for this distribution a theorem analogous to theorem 1.

THEOREM 2. *If the random variable $X = (X_1, \ldots, X_k)$ is distributed accordingly to (3.1) and we restrict ourselves to estimates $f = (f_1, \ldots, f_k)$*

of the parameter M, which satisfy the condition $\sum_{i=1}^{k} f_i = N$, then the estimate $f^0 = (f_1^0, \ldots, f_k^0)$, where $f_i^0(x) = Nx_i/n$ is the minimax estimate for the loss function L of the form

$$L(f, M) = \sum_{i=1}^{k} K[f_i(x), M_i], \tag{3.2}$$

where

$$K(u, v) = \frac{(u-v)^2}{v} \quad \textit{for} \quad v > 0,$$

$$K(u, 0) = \infty \quad \textit{for} \quad u > 0,$$

$$K(0, 0) = \frac{N(N-n)}{n(N-1)}.$$

For the loss function (3.2) the estimate f^0 is the unique minimax estimate, and the uniform distribution is the least favourable one.

Proof. It is known that

$$\mu_i = E(X_i|M) = \frac{n}{N} M_i, \tag{3.3}$$

$$\sigma_i^2 = E[(X_i-\mu_i)^2|M] = \frac{n(N-n)}{N^2(N-1)} M_i(N-M_i) \quad (i = 1, 2, \ldots, k). \tag{3.4}$$

Applying formulas (3.3) and (3.4) we obtain

$$R(f^0, M) = \sum_{i=1}^{k} E\left[K\left(\frac{N}{n}X_i, M_i\right)\Big|M\right] = \sum_{\substack{i\\ M_i\neq 0}} E\left[\frac{\left(\frac{N}{n}X_i - M_i\right)^2}{M}\Big|M\right] + \tag{3.5}$$
$$+ \sum_{\substack{i\\ M_i=0}} \frac{N(N-n)}{n(N-1)} = \frac{N^2}{n^2}\sum_{\substack{i\\ M_i\neq 0}} \frac{\sigma_i^2}{M_i} + \sum_{\substack{i\\ M_i=0}} \frac{N(N-n)}{n(N-1)} = \frac{N(N-n)}{n(N-1)}(k-1).$$

Thus for $f \equiv f^0$ the risk $R(f, M)$ is constant.

Let us denote by $P^0(M)$ the prior distribution of the parameter M, which is determined as follows:

$$P^0[(M_1, \ldots, M_k)] = \text{const} = c \left(M_1 \geqslant 0, \ldots, M_k \geqslant 0;\ \sum_{i=1}^{k} M_i = N\right). \tag{3.6}$$

The expected risk $r(f, P^0)$ then takes the form

$$r(f, P^0) \tag{3.7}$$
$$= c \sum_{M_1+\ldots+M_k=N} \sum_{\substack{m_1+\ldots+m_k=n\\ m_1\leqslant M_1,\ldots,m_k\leqslant M_k}} \frac{\binom{M_1}{m_1}\cdots\binom{M_k}{m_k}}{\binom{N}{n}} \sum_{i=1}^{k} K[f_i(m_1, \ldots, m_k), M],$$

which can also be written as follows:

$$r(f, P^0) = \tag{3.8}$$
$$\frac{c}{\binom{N}{n}} \sum_{m_1+\ldots+m_k=n} \sum_{\substack{n_1+\ldots+n_k=N-n\\ n_1\geqslant 0,\ldots,n_k\geqslant 0}} \binom{n_1+m_1}{m_1}\cdots\binom{n_k+m_k}{m_k} \times$$
$$\times \sum_{i=1}^{k} K[f_i(m_1, \ldots, m_k), n_i+m_i].$$

It is easily seen that (3.8) is finite if

$$f_i(m_1, \ldots, m_k) = 0 \quad \text{for} \quad m_i = 0 \quad (i = 1, 2, \ldots, k). \tag{3.9}$$

From formula (3.5) it follows that $r(f, P^0)$ is finite at least for one f (namely for $f \equiv f^0$). Thus if some estimate f minimizes (3.8) then it must satisfy (3.9). Let us observe furthermore that, for each system $(m_1, \ldots, m_k)$, $r(f, P^0)$ attains its minimum if

$$\varphi = \sum_{\substack{n_1+\ldots+n_k=N-n\\ n_1\geqslant 0,\ldots,n_k\geqslant 0}} \binom{n_1+m_1}{m_1}\cdots\binom{n_k+m_k}{m_k} \sum_{i=1}^{k} K[f_i(m_1, \ldots, m_k), n_i+m_i] \tag{3.10}$$

attains its minimum. Let us fix a system $(m_1, \ldots, m_k)$. By putting $f_i(m_1, \ldots, m_k) = 0$ if $m_i = 0$ we can rewrite (3.10) in the form

$$\varphi = \sum_{\substack{n_1+\ldots+n_k=N-n\\ n_1\geqslant 0,\ldots,n_k\geqslant 0}} \binom{n_1+m_1}{m_1}\cdots\binom{n_k+m_k}{m_k} \sum_{\substack{i\\ m_i\neq 0}} \frac{[f_i(m_1, \ldots, m_k)-n_i-m_i]^2}{n_i+m_i} + C. \tag{3.11}$$

Without loss of generality we can assume that $m_k \neq 0$. Proceeding as in the multinomial case, that is by putting

$$f_i(m_1, \ldots, m_k) = x_i \ (i < k), \qquad f_k(m_1, \ldots, m_k) = N - \sum_{\substack{i: m_i\neq 0\\ i\neq k}} x_i$$

in (3.11), and taking derivatives with respect to x_i, we find that it will attain its minimum if

$$\sum_{\substack{n_1+\ldots+n_k=N-n\\ n_1\geqslant 0,\ldots,n_k\geqslant 0}} \binom{n_1+m_1}{m_1}\ldots\binom{n_k+m_k}{m_k}\left(\frac{f_i(m_1,\ldots,m_k)}{n_i+m_i}-\frac{f_k(m_1,\ldots,m_k)}{n_k+m_k}\right)=0.$$

But

$$\sum_{\substack{n_1+\ldots+n_k=N-n\\ n_1\geqslant 0,\ldots,n_k\geqslant 0}} \frac{1}{n_i+m_i}\binom{n_1+m_1}{m_1}\ldots\binom{n_k+m_k}{m_k}$$

$$=\frac{1}{m_i}\sum_{\substack{n_1+\ldots+n_k=N-n\\ n_1\geqslant 0,\ldots,n_k\geqslant 0}} \binom{n_1+m_1}{m_1}\ldots\binom{n_i+m_i-1}{m_i-1}\ldots\binom{m_k+n_k}{m_k}$$

$$=\frac{(N+k-2)!}{(N-n)!\,m_1!\ldots m_k!}\sum_{\substack{n_1+\ldots+n_k=N-n\\ n_1\geqslant 0,\ldots,n_k\geqslant 0}}\frac{(N-n)!}{n_1!\ldots n_k!}\overbrace{\int\ldots\int}^{k}_{\substack{p_1+\ldots+p_k=1\\ p_1\geqslant 0,\ldots,p_k\geqslant 0}} p_1^{n_1+m_1}\ldots p_i^{n_i+m_i-1}\ldots p_k^{n_k+m_k}\,dp_1\ldots dp_k$$

$$=\frac{(N+k-2)!}{(N-n)!\,m_1!\ldots m_k!}\int\ldots\int_{\substack{p_1+\ldots+p_k=1\\ p_1\geqslant 0,\ldots,p_k\geqslant 0}} p_1^{m_1}\ldots p_i^{m_i-1}\ldots p_k^{m_k}\left(\sum_{\substack{n_1+\ldots+n_k=N-n\\ n_1\geqslant 0,\ldots,n_k\geqslant 0}}\frac{(N-n)!}{n_1!\ldots n_k!}p_1^{n_1}\ldots p_k^{n_k}\right)dp_1\ldots dp_k$$

$$=\frac{(N+k-2)!}{(N-n)!\,m_1!\ldots m_k!}\int\ldots\int_{\substack{p_1+\ldots+p_k=1\\ p_1\geqslant 0,\ldots,p_k\geqslant 0}} p_1^{m_1}\ldots p_i^{m_i-1}\ldots p_k^{m_k}\,dp_1\ldots dp_k$$

$$=\binom{N+k-2}{n+k-2}\frac{1}{m_i}.$$

Thus we have

$$m_k f_i(m_1,\ldots,m_k)-m_i f_k(m_1,\ldots,m_k)=0 \tag{3.12}$$

for those i for which $m_i\neq 0$. Since for $m_i=0$, $f_i=0$, formula (3.12) can, therefore, be extended over all $i=1,2,\ldots,k$. Summing formulae (3.12) where $i=1,2,\ldots,k$, and observing that $\sum_{i=1}^{k} f_i=N$, $\sum_{i=1}^{k} m_i=n$, we obtain

$$f_i(m_1,\ldots,m_k)=N\frac{m_i}{n}\qquad (i=1,2,\ldots,k).$$

Thus the estimate f^0 is the the Bayes estimate for the distribution P^0. Considerations analogous to those in the multinomial case lead us to the conclusion that this is the unique minimax estimate and that P^0 is the least favourable distribution.

We have obtained our result for the loss function of the form

$$L(f,\omega)=\sum_{i=1}^{k}[(f_i-\omega_i)^2|\omega_i].$$

If the loss function is

$$L(f,\omega)=\sum_{i=1}^{k}\lambda_i(\omega_i)(f_i-\omega_i)^2, \tag{3.13}$$

then the minimax estimates are mostly biased. Further information about the loss function (3.13) can be found in references [1-6].

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Introduction to the Theory of Games and Statistical Decisions*, 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, Berkeley and Los Angeles 1951, p. 53-73.

[3] J. L. Hodges Jr. and E. L. Lehman, *Some problems in minimax point estimation*, Annals of Mathematical Statistics 21 (1950), p. 182-197.

[4] H. Steinhaus, *The problem of estimation*, ibidem 28 (1957), p. 633-649.

[5] S. Trybuła, *Some problems in simultaneous minimax estimation*, ibidem 29 (1958), p. 245-253.

[6] — *On a problem of prognosis*, Bulletin de l'Académie Polonaise des Sciences, Classe III, 5 (1957), p. 859-862.

Reçu par la Rédaction le 15. 4. 1958