ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XIV, 2 (1974)

S. TRYBUŁA (Wrocław)

TWO PROBLEMS OF MINIMAX ESTIMATION

1. Let $F(x|\omega)$ be the distribution function of the random variable X dependent on the unknown parameter $\omega \in \Omega$. Let us suppose that X is observed and that we want to estimate ω. The function $L(u, \omega)$ is connected with the estimate $u = f(x)$; it is the loss for the statistician if he applies the estimate $f(x)$, where $x \in \mathfrak{X}$ is the observed value of X. For $f(x)$ and ω established, we can find the *expected value of the loss* L, i.e.

$$(1) \qquad R(f, \omega) = \mathrm{E}\big[L(f(X), \omega)\big] \stackrel{\mathrm{df}}{=} \int_{\mathfrak{X}} L(f(x), \omega)\, dF(x|\omega).$$

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

The estimate f_0 is called *minimax* if

$$\sup_{\omega\in\Omega} R(f_0, \omega) = \inf_f \sup_{\omega\in\Omega} R(f, \omega).$$

Let the a priori distribution of the parameter ω be given by the distribution function $G(\omega)$. The *expected risk* is

$$(2) \qquad r(f, G) = \mathrm{E}_\omega[R(f, \omega)] \stackrel{\mathrm{df}}{=} \int_\Omega R(f, \omega)\, dG(\omega).$$

According to Wald, $f_G(x)$ minimizing for a given G the function $r(f, G)$ will be called the *Bayes estimate for* G. The distribution G_0, for which

$$\inf_f r(f, G_0) = \sup_G \inf_f r(f, G),$$

is defined to be the *least favourable distribution.*

Let Ω_1 be a set of values of the parameter ω. Hodges and Lehmann [2] proved that if there exists the estimate f_0 such that $R(f_0, \omega) = c$ for $\omega \in \Omega_1$ and $R(f_0, \omega) \leqslant c$ for $\omega \in \Omega - \Omega_1$ and if there exists the distribution function $G_0(\omega)$ on Ω_1 such that f_0 is the Bayes estimate for G_0, then f_0 is the minimax estimate and G_0 is the least favourable distribution.

Particularly, the Bayes constant risk estimate is minimax.

2. Let us suppose that the random variable X is distributed according to the binomial law

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (k = 0, 1, \ldots, n),$$

where p is the unknown parameter.

Let $f(x)$ be an estimate of p. Assume that the loss function is of the form

$$L(f(x), p) = c(p)(f(x)-p)^2.$$

The risk, according to (1), takes the form

$$R(f, p) = \mathrm{E}[L(f(x), p)] = \sum_{k=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} c(p)(f(k)-p)^2.$$

Let the a priori distribution of the parameter p be defined by the distribution function $G(p)$. Then, according to (2), the expected risk is

$$(3) \qquad r(f, G) = \mathrm{E}_p[R(f, p)] = \int_0^1 \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} c(p)(f(k)-p)^2 dG(p).$$

A. Hodges and Lehmann [2] proved that, for $c(p) = c > 0$, the minimax estimate of p is

$$f(X) = \frac{X + \sqrt{n}/2}{n + \sqrt{n}}$$

and that the least favourable distribution is

$$dG(p) = C[p(1-p)]^{\sqrt{n}/2-1} dp.$$

B. Blackwell and Girshick [1] proved that in case where $c(p) = 1/p(1-p)$ the minimax estimate of p is $f(X) = X/n$, and that the least favourable distribution is $dG(p) = dp$.

3. Let

$$c(p) = \frac{1}{p(1-p)+c^2} \qquad (c > 0).$$

We prove that *the minimax estimate $f_0(X)$ of the parameter p is of the form*

$$f_0(X) = aX + b$$

with

$$(4) \qquad a = \frac{1}{n + 2c\sqrt{n}/\sqrt{1+4c^2}}$$

and

$$(5) \qquad b = \frac{c\sqrt{n}/\sqrt{1+4c^2}}{n + 2c\sqrt{n}/\sqrt{1+4c^2}}.$$

We obtain

$$R(f_0, p) = \frac{1}{p(1-p)+c^2} \mathrm{E}(aX+b-p)^2.$$

But

$$\mathrm{E}(aX+b-p)^2 = \mathrm{E}[a(X-np)+b-(1-an)p]^2$$
$$= a^2p(1-p)+\big(b-(1-an)p\big)^2.$$

Let us suppose that

$$(6) \qquad 2b = 1-an.$$

Then

$$\mathrm{E}(aX+b-p)^2 = a^2np(1-p)-(1-an)^2p(1-p)+b^2,$$

and the risk

$$R(f_0, p) = \frac{[a^2n-(1-an)^2]p(1-p)+b^2}{p(1-p)+c^2}$$

for

$$(7) \qquad b^2/c^2 = a^2n-(1-an)$$

does not depend on p.

Equations (6) and (7) will be satisfied for a and b given by (4) and (5), respectively.

For the constants a and b chosen in such a way, the risk is given by the formula

$$R(f_0, p) = \frac{n/(1+4c^2)}{\big(n+2c\sqrt{n}/\sqrt{1+4c^2}\big)^2}.$$

It follows from equations (4) and (5) that if $c \to 0$, then $a \to 1/n$ and $b \to 0$, and we obtain the minimax estimate in problem B. If $c \to \infty$, then

$$a \to \frac{1}{n+\sqrt{n}} \quad \text{and} \quad b \to \frac{\sqrt{n}/2}{n+\sqrt{n}},$$

and in the limit we obtain the minimax estimate for the quadratic loss function.

Since the estimate $f_0(X) = aX+b$ with a and b defined by (4) and (5), respectively, gives the constant risk, it is sufficient to prove that it is a Bayes estimate.

Let the a priori distribution of the parameter p be defined by the density $g(p)$. From equation (3) we obtain

$$(8) \qquad r(f, G) = \int_0^1 \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} \frac{\big(f(k)-p\big)^2}{p(1-p)+c^2} g(p)\,dp.$$

The function $r(f, G)$ given by (8) is a positively defined quadratic form of variables $x_k = f(k)$ $(k = 0, 1, \ldots, n)$. It attains its minimum when the partial derivatives with respect to $f(k)$ are equal to zero. Then we obtain

$$f(k)\int_0^1 p^k(1-p)^{n-k}h(p)\,dp = \int_0^1 p^{k+1}(1-p)^{n-k}h(p)\,dp,$$

where

$$h(p) = \frac{g(p)}{p(1-p)+c^2}.$$

Since $f(k) = ak+b$, we have the equation

$$(9)\qquad (ak+b)\int_0^1 p^k(1-p)^{n-k}h(p)\,dp = \int_0^1 p^{k+1}(1-p)^{n-k}h(p)\,dp \qquad (k = 0, 1, \ldots, n).$$

Multiplying both sides of (9) by $\binom{n}{k}x^k$, adding the obtained expressions for $k = 0, 1, \ldots, n$ and taking into account the equality

$$\sum_{k=0}^{n}\binom{n}{k}k(px)^k(1-p)^{n-k} = npx(px+1-p)^{n-1},$$

we obtain

$$\int_0^1 [anpx(px+1-p)^{n-1}+b(px+1-p)^n]h(p)\,dp = \int_0^1 p(px+1-p)^n h(p)\,dp.$$

But $px+1-p = p(x-1)+1$. Then we have

$$(10)\qquad \int_0^1 \left\{an\sum_{k=1}^{n}\binom{n-1}{k-1}[p(x-1)]^k + anp\sum_{k=1}^{n}[p(x-1)]^k + b\sum_{k=0}^{n}\binom{n}{k}[p(x-1)]^k\right\}h(p)\,dp = \int_0^1 p\sum_{k=0}^{n}\binom{n}{k}[p(x-1)]^k h(p)\,dp.$$

Equality (10) holds for each value of x; thus the coefficients at $(x-1)^k$ are equal for each k, respectively. We have

$$\int_0^1 \left\{an\binom{n-1}{k-1}p^k + an\binom{n-1}{k}p^{k+1} + b\binom{n}{k}p^k\right\}h(p)\,dp = \int_0^1 \binom{n}{k}p^{k+1}h(p)\,dp \qquad (k = 0, 1, \ldots, n).$$

Taking into account the equalities

$$\binom{n-1}{k-1} = \frac{k}{n}\binom{n}{k} \quad \text{and} \quad \binom{n-1}{k} = \frac{n-k}{n}\binom{n}{k},$$

we obtain, after reduction,

$$(11) \qquad \int_0^1 akp^{k-1}[p(1-p)h(p)]dp + \int_0^1 bp^k h(p)dp = (1-an)\int_0^1 p^k[ph(p)]dp$$
$$(k = 0, 1, \ldots, n).$$

Let us suppose that $g(p)$ is differentiable in $(0, 1)$. Since $g(p)$ is a density, it follows that

$$\lim_{p\to 0+} p(1-p)g(p) = \lim_{p\to 1-} p(1-p)g(p) = 0.$$

Then also

$$\lim_{p\to 0+} p(1-p)h(p) = \lim_{p\to 1-} p(1-p)h(p) = 0.$$

Integrating by parts and taking into account the above-given equalities, we have

$$\int_0^1 kp^{k-1}[p(1-p)h(p)]dp = -\int_0^1 p^k[p(1-p)h(p)]'dp \qquad (k = 1, 2, \ldots, n)$$

which reduces equation (11) to the form

$$(12) \qquad -a\int_0^1 p^k[p(1-p)h(p)]'dp + b\int_0^1 p^k h(p)dp = (1-an)\int_0^1 p^k[ph(p)]dp$$
$$(k = 0, 1, \ldots, n).$$

Equation (12) will be certainly fulfilled if

$$[p(1-p)h(p)]' - [b-(1-an)p]h(p) = 0.$$

Then

$$p(1-p)h(p) = C[p(1-p)]^{b/a}$$

and

$$(13) \qquad g(p) = C[p(1-p)+c^2][p(1-p)]^{b/a-1}.$$

The proof that $f_0(X) = aX+b$, with a and b given by (4) and (5), respectively, is a minimax estimate is much shorter if we show that f_0 is the Bayes estimate for g given by (13). Our method gives the way to find $g(p)$.

Let us notice that if $c\to 0$, then it follows from (4) and (5) that $a\to 1/n$, $b\to 0$ and $g(p)\to 1$. If $c\to\infty$, then

$$g(p)\to C[p(1-p)]^{\sqrt{n}/2-1}.$$

4. The second problem considered in this paper is the following:

Let us suppose that the random variable X takes values from the finite interval $[c, d]$ $(c < d)$. Let $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_n)$ be two samples, let $X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n$ be independent and have the same distribution as X. The question arises how, knowing the values of the random variables $X_1, X_2, \ldots, X_m$, to predict the value of

$$Z = \sum_{j=1}^{n} g(Y_j).$$

Particularly, for $g(y) = y^k/n$, the question arises how to predict the value of the k-th sample moment of the second sample.

Let us suppose that the random variable Z takes the value z and that our prediction is z'. Let us assume that the loss connected with the pair (z', z) is $L(z', z) = (z-z')^2$. Let $z' = f(x_1, x_2, \ldots, x_m)$ be an estimate of z and let F be the distribution function of the random variable X. Then we can define the risk connected with the selection of the function f by

$$R(f, F) = \mathrm{E}\Big[L\Big(f(X_1, X_2, \ldots, X_m), \sum_{j=1}^{n} g(Y_j)\Big)\Big].$$

In this section we find the minimax estimate of Z, i.e. the estimate f_0 such that

$$\sup_F R(f_0, F) = \inf_f \sup_F R(f, F).$$

We prove the following

THEOREM. *Let X be a random variable distributed according to the unknown distribution F on the measurable space A. Let g be a bounded, measurable function on A and let two points $x', x'' \in A$ be such that g attains its minimum in x' and its maximum in x''. Let $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_n)$ be independent simple random samples from F, and let*

$$Z = \sum_{i=1}^{n} g(Y_i).$$

If the loss is given by

$$L(f, z) = (f-z)^2,$$

where f is an estimate of z, then the minimax estimate of Z is given by

$$f_0(X_1, X_2, \ldots, X_m) = a\sum_{i=1}^{m} g(X_i)+b,$$

with

$$(14)\qquad a = \begin{cases} \dfrac{n}{m-1}\left(1-\sqrt{\dfrac{1}{m}+\dfrac{1}{n}-\dfrac{1}{mn}}\right) & \text{if } m>1,\\ \dfrac{n-1}{2} & \text{if } m=1,\end{cases}$$

$$b = \frac{n-am}{2}(\beta+a)$$

and

$$a = \min_{x\epsilon A} g(x), \qquad \beta = \max_{x\epsilon A} g(x).$$

Proof. Let us write $\lambda = \mathrm{E}[g(X)]$. If

$$(15)\qquad \bar{f}(X_1, X_2, \ldots, X_m) = a\sum_{i=1}^{n} g(X_i)+b,$$

then the risk can be evaluated as follows:

$$\begin{aligned} R(\bar{f}, F) &= \mathrm{E}\Big[\sum_{j=1}^{n} g(Y_j) - a\sum_{i=1}^{m} g(X_i) - b\Big]^2 \\ &= \mathrm{E}\Big[\sum_{j=1}^{n}\big(g(Y_j)-\lambda\big)^2 - a\sum_{i=1}^{m}\big(g(X_i)-\lambda\big)+(n-am)\lambda-b\Big]^2 \\ &= \sum_{j=1}^{n}\mathrm{E}[g(Y_j)-\lambda]^2 + a^2\sum_{i=1}^{m}\mathrm{E}[g(X_i)-\lambda]^2+[(n-am)\lambda-b]^2. \end{aligned}$$

Let $a' = g(x')$ and $\beta = g(x'')$. It is easy to show that

$$(16)\qquad \mathrm{E}[g(X_i)-\lambda]^2 = \mathrm{E}[g(Y_j)-\lambda]^2 \leqslant (\beta-\lambda)(\lambda-a) \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n).$$

Thus

$$(17)\qquad R(\bar{f}, F) \leqslant (n+a^2m)(\beta-\lambda)(\lambda-a)+[(n-am)\lambda-b]^2$$
$$= (n+a^2m)[(\beta-a)-(\lambda-a)](\lambda-a)+[(n-am)(\lambda-a)+(n-am)a-b]^2.$$

The right-hand side of (16) does not depend on $\lambda-a$ if

$$(18)\qquad (n-am)^2-(n+a^2m) = 0$$

and

$$(19)\qquad (n+a^2m)(\beta-a)+2(n-am)^2a = 2(n-am)b.$$

Equations (17) and (18) are fulfilled for (14) and for

$$(20)\qquad b = \begin{cases} \dfrac{n-am}{2}(\beta+a) = \dfrac{n}{2(m-1)}\left(m\sqrt{\dfrac{1}{m}+\dfrac{1}{n}-\dfrac{1}{mn}}-1\right)(\beta+a) & \text{if } m>1,\\ \dfrac{n+1}{4}(\beta+a) & \text{if } m=1.\end{cases}$$

Let us denote by f_0 the estimate of form (15) for which the constants a and b are given by (14) and (20), respectively. Then from (17), (18) and (19) we obtain

$$R(f_0, F) \leqslant [(n-am)a-b]^2 = \frac{(n-am)^2}{4}(\beta-\alpha)^2 \stackrel{\mathrm{df}}{=} c. \tag{21}$$

Observe that if the distribution F_0 of the random variable X fulfils the conditions $\mathrm{P}(X = x') = 1-p$ and $\mathrm{P}(X = x'') = p$, then $\lambda = \alpha + (\beta-\alpha)p$ and an equality is obtained in (16), which gives

$$R(f_0, F_0) = c. \tag{22}$$

The distribution F_0 depends on the parameter p. Since (21) and (22) hold, it is sufficient to show that there exists a distribution G of p such that, for $f = f_0$, the expected risk $r(f, G)$ attains its minimum. We have

$$\begin{aligned} r(f,G) &= \mathrm{E}_p[R(f, F_0)] = \mathrm{E}_p\Big\{\mathrm{E}\Big[\sum_{j=1}^{n} g(Y_j) - f(X_1, X_2, \ldots, X_m)\Big]^2\Big\} \\ &= \mathrm{E}_p\Big\{\sum_{j=1}^{n}\mathrm{E}[g(Y_j)-\lambda]^2 + \mathrm{E}[n\lambda - f(X_1, X_2, \ldots, X_m)]^2\Big\} \\ &= \mathrm{E}_p\Big\{\sum_{j=1}^{n}\mathrm{E}\big[g(Y_j)-\big(\alpha+(\beta-\alpha)p\big)\big]^2\Big\} + \\ &\qquad + \mathrm{E}_p\big\{\mathrm{E}\big[n\big(\alpha+(\beta-\alpha)p\big) - f(X_1, X_2, \ldots, X_m)\big]^2\big\}, \end{aligned} \tag{23}$$

where we obtain the minimum if $\mathrm{E}_p\{\mathrm{E}[n(\alpha+(\beta-\alpha)p) - f(X_1, \ldots, X_m)]^2\}$ attains its minimum. This leads to the equation

$$f(X_1, X_2, \ldots, X_m) = \mathrm{E}_p\big[n\big(\alpha+(\beta-\alpha)p\big) \mid X_1, X_2, \ldots, X_m\big]. \tag{24}$$

Let k among the random variables $X_1, X_2, \ldots, X_m$ assume the value x'' and let $m-k$ of them assume the value x'. In this case equation (24) takes the form

$$f(X_1, X_2, \ldots, X_m) = \frac{\int_0^1 n\big(\alpha+(\beta-\alpha)p\big)p^k(1-p)^{m-k}dG(p)}{\int_0^1 p^k(1-p)^{m-k}dG(p)}.$$

Let

$$dG(p) = C[p(1-p)]^r dp.$$

Then

$$f(X_1, X_2, \ldots, X_m) = na + n\frac{k+r+1}{n+2r+2}(\beta - a)$$

$$= \frac{n}{m+2r+2}\big(k\beta + (m-k)a + (r+1)(\beta+a)\big)$$

$$= \frac{n}{m+2r+2}\Big(\sum_{i=1}^{m} g(X_i) + (r+1)(\beta+a)\Big) = f_0(X_1, X_2, \ldots, X_m)$$

for

$$r = \begin{cases} \dfrac{(m+2)\sqrt{1/m+1/n-1/mn}-3}{2(1-\sqrt{1/m+1/n-1/nm})} & \text{if } m > 1,\ n > 1, \\ \dfrac{3-n}{2(n-1)} & \text{if } m = 1,\ n > 1. \end{cases}$$

It is easy to verify that in both cases $r > -1$.

If $n = 1$, then $a = 0$ which means that the estimate f_0 does not depend on the observed values of the random variables $X_1, X_2, \ldots, X_m$. It is easy to verify that in this case f_0 is the Bayes estimate for the a priori distribution defined by the formula $\mathrm{P}(p = 1/2) = 1$ which completes the proof of the theorem.

References

[1] D. Blackwell and M. A. Girshick, *Introduction to the theory of games and statistical decisions*, New York 1954.

[2] J. L. Hodges and E. L. Lehmann, *Some problems in minimax point estimation*, Annals of Math. Stat. 21 (1950), p. 182-197.

WROCŁAW TECHNICAL UNIVERSITY
INSTITUTE OF MATHEMATICS AND THEORETICAL PHYSICS

Received on 18. 10. 1973

S. TRYBUŁA (Wrocław)

DWA PROBLEMY MINIMAKSOWEJ ESTYMACJI

STRESZCZENIE

Pokazano, że minimaksowy estymator parametru p w rozkładzie dwumianowym dla funkcji straty

$$L(f, p) = \frac{(f-p)^2}{p(1-p)+c^2} \quad (c > 0)$$

jest postaci $f(X) = aX+b$, gdzie

$$a = \frac{1}{n+2c\sqrt{n}/\sqrt{1+4c^2}} \quad \text{and} \quad b = \frac{c\sqrt{n}/\sqrt{1+4c^2}}{n+2c\sqrt{n}/\sqrt{1+4c^2}}.$$

Udowodniono następujące twierdzenie:

Niech X będzie zmienną losową o rozkładzie F na A. Niech g będzie ograniczoną, mierzalną funkcją na A i niech punkty $x', x'' \in A$ będą takie, że g osiąga minimum w punkcie x' i maksimum w punkcie x''. Niech $(X_1, X_2, \ldots, X_m)$ i $(Y_1, Y_2, \ldots, Y_n)$ będą niezależnymi próbami prostymi z populacji o rozkładzie F i niech

$$Z = \sum_{i=1}^{n} g(Y_i).$$

Jeżeli strata jest określona wzorem $L(f, z) = (f-z)^2$, gdzie z jest wartością zmiennej losowej Z, a f — predyktorem z, to minimaksowy predyktor zmiennej losowej Z określony jest wzorem

$$f_0(X_1, X_2, \ldots, X_m) = a\sum_{i=1}^{m} g(X_i)+b,$$

w którym

$$a = \begin{cases} \dfrac{n}{m-1}\left(1-\sqrt{\dfrac{1}{m}+\dfrac{1}{n}-\dfrac{1}{mn}}\right) & \text{dla } m > 1, \\ \dfrac{n-1}{2} & \text{dla } m = 1, \end{cases}$$

$$b = \frac{n-am}{2}(\alpha+\beta)$$

oraz

$$\alpha = \min_{x\in A} g(x), \quad \beta = \max_{x\in A} g(x).$$

Jeżeli zmienna losowa X jest ograniczona, to dla $g(x) = x^k/n$ otrzymujemy minimaksowy predyktor k-tego momentu z próby.