ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XVI, 2 (1978)

S. TRYBUŁA (Wroclaw)

MINIMAX PREDICTION OF A SAMPLE DISTRIBUTION FUNCTION

1. Let us suppose that the random variable X_0 is distributed according to an unknown distribution function $F(x)$. We choose two independent random samples from F,

$$X = (X_1, X_2, \ldots, X_m) \quad \text{and} \quad Y = (Y_1, Y_2, \ldots, Y_n),$$

but only the values of the first sample are known. In the paper the problem of minimax prediction of the sample distribution function in the second sample from the values of the first sample is considered.

2. Let $\hat{F}(t)$ be the sample distribution function in the second sample,

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n}\delta_{Y_i}(t),$$

where

$$\delta_{Y_i}(t) = \begin{cases} 1 & \text{if } Y_i \leqslant t, \\ 0 & \text{if } Y_i > t. \end{cases}$$

Let $\Phi(t) = \Phi(t, X)$ be a predictor of $F(t)$. Let us suppose that the loss connected with the estimate $\Phi(t)$ is of the form

$$L(\hat{F}, \Phi) = \int_{-\infty}^{\infty} [\hat{F}(t) - \Phi(t)]^2 dW(t),$$

where W defines a non-zero finite measure on the σ-field B of Borel sets on the straight line R. The risk connected with the predictor Φ takes the form

$$R(F, \Phi) = \mathrm{E}_F[L(\hat{F}, \Phi)] = \int_{-\infty}^{\infty} \mathrm{E}_F[\hat{F}(t) - \Phi(t)]^2 dW(t), \tag{1}$$

where $\mathrm{E}_F(\cdot)$ is the operator of the expected value.

3. We now prove

THEOREM 1. *Let $\Phi_1(t)$ be a predictor of the form*

$$\Phi_1(t) = a + \sum_{i=1}^{m} b_i \delta_{X_i}(t). \tag{2}$$

The risk $R(F, \Phi_1)$ is independent of $F(t)$ if and only if

$$\sum_{i=1}^{m} b_i^2 - \Big(1 - \sum_{i=1}^{m} b_i\Big)^2 + \frac{1}{n} = 0, \tag{3}$$

$$2a = 1 - \sum_{i=1}^{m} b_i. \tag{4}$$

Proof. We have

$$\begin{aligned} R(F, \Phi_1) &= \int_{-\infty}^{\infty} \mathrm{E}_F\big(\hat{F}(t) - \Phi_1(t)\big)^2 dW(t) \\ &= \int_{-\infty}^{\infty} \mathrm{E}_F\Big(\frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i}(t) - a - \sum_{i=1}^{m} b_i \delta_{X_i}(t)\Big)^2 dW(t) \\ &= \int_{-\infty}^{\infty} \mathrm{E}_F\Big(\frac{1}{n}\sum_{i=1}^{n} \big(\delta_{Y_i}(t) - F(t)\big) - \sum_{i=1}^{m} b_i\big(\delta_{X_i}(t) - F(t)\big) - a + \\ &\quad + \Big(1 - \sum_{i=1}^{m} b_i\Big) F(t)\Big)^2 dW(t) \\ &= \int_{-\infty}^{\infty} \Big(\frac{1}{n} F(t)\big(1 - F(t)\big) + \sum_{i=1}^{m} b_i^2 F(t)\big(1 - F(t)\big) + a^2 + \\ &\quad + \Big(1 - \sum_{i=1}^{m} b_i\Big)^2 F^2(t) - 2a\Big(1 - \sum_{i=1}^{m} b_i\Big) F(t)\Big) dW(t), \end{aligned} \tag{5}$$

since

$$\mathrm{E}_F\big(\delta_{X_i}(t)\big) = \mathrm{E}\big(\delta_{Y_i}(t)\big) = F(t),$$

$$\mathrm{E}_F\big(\delta_{X_i}(t) - F(t)\big)^2 = \mathrm{E}_F\big(\delta_{Y_i}(t) - F(t)\big)^2 = F(t)\big(1 - F(t)\big).$$

The risk $R(F, \Phi_1)$ is independent of $F(t)$ if and only if the coefficients of $F^2(t)$ and $F(t)$ on the right-hand side of formula (5) are equal to zero, which leads to equations (3) and (4). Thus Theorem 1 is proved.

In this case the risk $R(F, \Phi_1)$ is of the form

$$R(F, \Phi_1) = a^2 \int_{-\infty}^{\infty} dW(t). \tag{6}$$

Let us suppose now that $b_i = b$ $(i = 1, 2, \ldots, m)$. Equations (2) and (3) take now the form

$$mb^2 - (1-mb)^2 + \frac{1}{n} = 0, \tag{7}$$

$$2a = 1 - mb. \tag{8}$$

Assume that $0 \leqslant \Phi_1(t) \leqslant 1$. It leads to the condition

$$0 \leqslant a + mb \leqslant 1. \tag{9}$$

The only solution of equations (7) and (8) satisfying condition (9) is the following:

$$b = \begin{cases} \dfrac{1-\sqrt{1/m+1/n-1/mn}}{m-1} & \text{if } m > 1, \\ \dfrac{n-1}{2n} & \text{if } m = 1, \end{cases}$$

$$a = \begin{cases} \dfrac{m\sqrt{1/m+1/n-1/mn}-1}{2(m-1)} & \text{if } m > 1, \\ \dfrac{n+1}{4n} & \text{if } m = 1. \end{cases} \tag{10}$$

We now prove that the predictor

$$\Phi_0(t) = a + b\sum_{i=1}^{m} \delta_{X_i}(t),$$

where the constants a and b are defined by (10), is a minimax predictor.

4. The considered problem of finding a minimax predictor of the sample distribution function we can view as the problem of determining the optimal strategy in a game against nature. The nature chooses a distribution function $F(t)$, the statistician chooses a predictor $\Phi(t)$, the pay-off function is given by (1). Let us define now the sequence $\{\tau_k\}$ of mixed strategies for nature (*a priori* distributions of F) which will be used in the proof of optimality of the strategy $\Phi_0(t)$.

We define the *a priori* distribution τ_k in the following manner. We choose the parameter p according to the β-distribution with density

$$g(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1}(1-p)^{\beta-1} \qquad (0 \leqslant p \leqslant 1), \tag{11}$$

where

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

and then, for the given p, we choose the distribution function $F(t)$ of the form

$$F(t) = \begin{cases} 0 & \text{if } t < -k, \\ p & \text{if } -k \leqslant t < k, \\ 1 & \text{if } t \geqslant k. \end{cases} \tag{12}$$

For the *a priori* distribution τ_k defined in such a manner the expected risk takes the form

$$r(\tau_k, \Phi) \overset{\text{df}}{=} \mathrm{E}_{\tau_k}[R(f, \Phi)] = \mathrm{E}_{\tau_k}\mathrm{E}_F[L(\hat{F}, \Phi)] = \int_{-\infty}^{\infty} \mathrm{E}_{\tau_k}\mathrm{E}_F\big(\hat{F}(t) - \Phi(t)\big)^2\,dW(t). \tag{13}$$

Let us consider the problem of determining the minimum of the functional $r(\tau_k, \Phi)$, The expected risk $r(\tau_k, \Phi)$ attains its minimum if for each t we choose $\Phi(t)$ so that

$$\mathrm{E}_{\tau_k}\mathrm{E}_F\big(\hat{F}(t) - \Phi(t)\big)^2 \tag{14}$$

attains its minimum. But

$$\mathrm{E}_F\big(\hat{F}(t) - \Phi(t)\big)^2 = \mathrm{E}_F\big(\hat{F}(t) - F(t)\big)^2 + \mathrm{E}_F\big(F(t) - \Phi(t)\big)^2. \tag{15}$$

The first component on the right-hand side of (15) is independent of $\Phi(t)$. Thus expression (14) attains its minimum if the function

$$\varrho_t(\tau_k, \Phi) = \mathrm{E}_{\tau_k}\mathrm{E}_F\big(\hat{F}(t) - \Phi(t)\big)^2$$

attains its minimum.

The following theorem was proved by Phadia [2] who considered the problem of minimax estimation of the distribution function $F(t)$.

THEOREM 2. *Let $X_1, X_2, \ldots, X_m$ be a random sample from the unknown distribution function F and let τ_k be the a priori distribution function defined on the set of distribution functions and given by* (11) *and* (12). *Then*

$$\min_{\Phi} \varrho_t(\tau_k, \Phi) = \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+n)}\, 1_{[-k,k)}(t),$$

where $1_A(t)$ denotes the characteristic function of the set A.

It follows from (13), (14), and (15) that in order to determine the minimum of $r(\tau_k, \Phi)$ it is necessary to find the value of the functional

$$\varrho_t(\tau_k) = \mathrm{E}_{\tau_k}\mathrm{E}_F\big(\hat{F}(t) - F(t)\big)^2.$$

But

$$(16)\quad \mathrm{E}_F\big(\hat{F}(t)-F(t)\big)^2 = \mathrm{E}_F\left(\frac{1}{n}\sum_{i=1}^{n}\big(\delta_{Y_i}(t)-F(t)\big)\right)^2 = \frac{F(t)\big(1-F(t)\big)}{n}.$$

Then, by the definition of the *a priori* distribution τ_k and by (16), for $-k \leqslant t < k$ we have

$$(17)\quad \varrho_t(\tau_k) = \frac{1}{nB(\alpha,\beta)}\int_{-\infty}^{\infty} p^{\alpha}(1-p)^{\beta}\,dp$$

$$= \frac{B(\alpha+1,\beta+1)}{nB(\alpha,\beta)} = \frac{\alpha\beta}{n(\alpha+\beta)(\alpha+\beta+1)}$$

and

$$(18)\quad \varrho_t(\tau_k) = 0 \quad \text{if } \tau < -k \text{ and } \tau \geqslant k.$$

Equations (17) and (18) can be rewritten in the form

$$(19)\quad \varrho_t(\tau_k) = \frac{\alpha\beta}{n(\alpha+\beta)(\alpha+\beta+1)}\,1_{[-k,k)}(t).$$

It follows from (19) and Theorem 2 that for the *a priori* distribution given by (11) and (12) the Bayes risk is given according to the formula

$$(20)\quad \min_{\Phi} r(\tau_k, \Phi) = \int_{\infty}^{\infty}\big(\min \varrho_t(\tau_k, \Phi) + \varrho_t(\tau_k)\big)\,dW(t)$$

$$= \frac{\alpha\beta(\alpha+\beta+m+n)}{n(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+m)}\int_{-\infty}^{\infty} 1_{[-k,k)}(t)\,dW(t).$$

5. In this section we prove that $\Phi_0(t)$ is a minimax predictor. For this purpose we introduce the notion of a Bayes predictor.

Let Ω be the set of all distribution functions on R and let τ be the *a priori* distribution on Ω. The predictor $\Phi_\tau(t)$ for which the expected risk $r(\tau, \Phi)$ attains its minimum is a Bayes predictor.

The following lemma which we formulate in a form useful for us is well known in the theory of decision functions.

LEMMA. *Let $\{\Phi_k(t)\}$ be a sequence of Bayes predictors of $F(t)$ for the a priori distributions τ_k, let $\{r(\tau_k, \Phi_k)\}$ be the corresponding sequence of Bayes risks defined for a given loss function, and let c be a constant. If $r(\tau_k, \Phi_k) \to c$ as $k \to \infty$ and Φ is a predictor such that $R(F, \Phi) \leqslant c$ for each $F \in \Omega$, then Φ is a minimax predictor.*

THEOREM 3. *Let $X = (X_1, X_2, \ldots, X_m)$ and $Y = (Y_1, Y_2, \ldots, Y_n)$ be independent random samples from an unknown distribution F on R, let $\hat{F}(t)$ be the sample distribution function in the second sample, and let Θ*

be the set of all predictors $\Phi(t) = \Phi(t, X)$ *of* $F(t)$. *If the loss function is of the form*

$$L(\hat{F}, \Phi) = \int_{-\infty}^{\infty} (\hat{F}(t) - \Phi(t))^2 dW(t),$$

where W *is a non-zero finite measure on* (R, B), *then the minimax predictor* $\Phi_0(t)$ *takes the form*

$$\Phi_0(t) = a + b \sum_{i=1}^{m} \delta_{X_i}(t),$$

where the constants a *and* b *are defined by* (10).

Proof. Since equation (6) holds for the predictor $\Phi_0(t)$, by the Lemma it is sufficient to prove that there exist constants α and β in (11) such that

$$\lim_{k\to\infty} r(\tau_k, \Phi_k) = a^2 \int_{-\infty}^{\infty} dW(t), \tag{21}$$

where a is given by (10). But by (20) we have

$$r(\tau_k, \Phi_k) = \frac{\alpha\beta(\alpha+\beta+m+n)}{n(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+m)} \int_{-\infty}^{\infty} 1_{[-k,k)}(t) dW(t). \tag{22}$$

Suppose that $n > 1$ and let

$$\alpha = \beta = \frac{a}{b} = \frac{m+\sqrt{mnx}}{2(n-1)},$$

where $x = m+n-1$. The coefficient of the integral in equation (22) reduces to the form

$$A = \frac{\alpha(2\alpha+m+n)}{2n(2\alpha+1)(2\alpha+m)} = \frac{(m+\sqrt{mnx})(nx+\sqrt{mnx})}{4n(x+\sqrt{mnx})(mn+\sqrt{mnx})}$$

$$= \frac{1}{4}\left(\frac{m+\sqrt{mnx}}{mn+\sqrt{mnx}}\right)^2 = \begin{cases} \left(\dfrac{\sqrt{mnx}-n}{2n(m-1)}\right)^2 = a^2 & \text{if } m > 1, \\ \left(\dfrac{n+1}{4n}\right)^2 = a^2 & \text{if } m = 1. \end{cases}$$

Then for $n > 1$ condition (21) holds and $\Phi_0(t)$ is a minimax predictor.

Let $n = 1$. In this case, according to (10), the predictor $\Phi_0(t)$ is independent of X. Let us define the *a priori* distribution τ_k as follows. We choose, with probability 1, the distribution function

$$F(t) = \begin{cases} 0 & \text{if } t < -k, \\ \frac{1}{2} & \text{if } -k \leqslant t < k, \\ 1 & \text{if } t \geqslant k. \end{cases}$$

For the *a priori* distribution defined in such a manner the Bayes predictor $\Phi_k(t) = F(t)$ and the Bayes risk takes the form

$$r(\tau_k, \Phi_k) = \int_{-\infty}^{\infty} \mathrm{E}_F\big(\hat{F}(t) - F(t)\big)^2 dW(t) = \frac{1}{4} \int_{-\infty}^{\infty} 1_{[-k,k)}(t)\, dW(t)$$

$$= a^2 \int_{-\infty}^{\infty} 1_{[-k,k)}(t)\, dW(t).$$

Condition (21) is fulfilled also in this case, which completes the proof of the theorem.

6. If we suppose that the measure $W(t)$ is concentrated at one point, say t_0, then the considered problem reduces to the problem of determining the minimax predictor of the random variable Y/n from the values of X/m, where X and Y have binomial distributions with parameters $m, p; n, p$, respectively, and $p = F(t_0)$. This problem was solved by Hodges and Lehmann in [1].

If $n \to \infty$, then $a \to 1/(m+\sqrt{m})$, $b \to 1/2(\sqrt{m}+1)$ and in the limit the predictor $\Phi_0(t)$ is identical with the minimax estimate of the distribution function $F(t)$ for the loss function

$$L(F', \Phi) = \int_{-\infty}^{\infty} \big(F(t) - \Phi(t)\big)^2 dW(t).$$

This estimate was found by Phadia in [2]. In this paper we applied a similar method and some of his results.

References

[1] J. L. Hodges and E. L. Lehmann, *Some problems in minimax point estimation*, Ann. Math. Statist. 21 (1950), p. 182-196.

[2] E. G. Phadia, *Minimax estimation of a cumulative distribution function*, Ann. Statist. 1 (1973), p. 1149-1157.

INSTITUTE OF MATHEMATICS
TECHNICAL UNIVERSITY
50-370 WROCŁAW

Received on 2. 1. 1976

S. TRYBUŁA (Wrocław)

MINIMAKSOWA PROGNOZA DYSTRYBUANTY EMPIRYCZNEJ

STRESZCZENIE

Załóżmy, że cecha X_0 populacji generalnej jest zmienną losową o nie znanej nam dystrybuancie $F(x)$. Z populacji tej wybieramy dwie niezależne próby proste $X = (X_1, X_2, \ldots, X_m)$ oraz $Y = (Y_1, Y_2, \ldots, Y_n)$, lecz tylko wartości z pierwszej próby są nam znane. Niech $\hat{F}(t)$ będzie dystrybuantą empiryczną w drugiej próbie. W pracy rozpatrzono problem minimaksowej prognozy statystyki $\hat{F}(t)$ na podstawie wartości z pierwszej próby, gdy funkcja straty $L(\hat{F}, \Phi)$, związana z oceną $\Phi(t) = \Phi(t, X)$ funkcji $\hat{F}$, jest postaci

$$L(\hat{F}, \Phi) = \int_{-\infty}^{\infty} (\hat{F}(t) - \Phi(t))^2 dW(t),$$

gdzie $W(t)$ jest niezerową skończoną miarą na σ-ciele zbiorów borelowskich na prostej. Niech

$$\delta_{X_i}(t) = \begin{cases} 1, & \text{gdy } X_i \leqslant t, \\ 0, & \text{gdy } X_i > t. \end{cases}$$

W pracy udowodniono, że minimaksowy predyktor statystyki $F(t)$ jest postaci

$$\Phi_0(t) = a + b \sum_{i=1}^{m} \delta_{X_i}(t),$$

gdzie stałe a i b określone są wzorem (10).