ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN—POLONIA
VOL. V, 6 SECTIO A 1951

Z Zakładu Statystyki Matematycznej Wydziału Matematyczno-Przyrodniczego UMCS Kierownik: z. prof. dr M. Olekiewicz

M. OLEKIEWICZ

On certain improved estimates of the mean

O pewnych ulepszonych ocenach średniej arytmetycznej

Об улучшенных оценках средней арифметической

Let X be a random variable with finite mean μ and variance σ^2.

Let a be its parameter to be estimated from the sample.

Let $a^* = a^* (X_1, X_2, \ldots, X_n)$ be a function of n observations on X in a sample, and let a^* have finite mean, $E(a^*)$, and finite variance $D^2(a^*)$, given by

$$\left.\begin{aligned} E(a^*) &= \int_{-\infty}^{\infty} a^* \, d F(a^*) \\ D^2(a^*) &= \int_{-\infty}^{\infty} (a^* - E(a^*))^2 dF(a^*) \end{aligned}\right| \tag{1}$$

where the right sides are Lebesgue-Stieltjes integrals, $F(a^*)$ being (cumulative) distribution function of random variable a^*.

The general problem in point estimation is this: which of all possible functions a^* is the best estimate of a?

Criteria of the best estimate. There is a widely accepted requirement believed to be a necessary condition for a good estimate: that a^* should be unbiased

$$\text{bias } b(a^*) = E(a^* - a) = 0\,, \quad \text{i. e.} \quad E(a^*) = a \tag{2}$$

This requirement is subject to serious objections. On its own merit condition (2) can be accepted as necessary only in cases, where estimation is a kind of "fair game" played between two parties in which positive errors are appraised as gains from the point of view of one party and as losses from the point of view of the other, or vice versa, and it is desired to obtain a neutralized effect in the long run. Such situations are possible in economic relation, but in most applications of estimation, as for instance in technology, science, etc., there seems to be no purpose in neutralizing positive and negative errors of estimates. In typical estimation cases deviations from the "true values" are considered losses irrespectively of their signs. From this point of view condition (2) is not only unnecessary, but faulty [1]), for it is inconsistent with the natural requirement that mean absolute error of estimate should be as small as possible

$$\epsilon(a^*) = E\,|a^* - a| = \text{minimum}, \tag{3}$$

or with a more tractable though less natural requirement that mean square error of estimate shold be as small as possible

$$\mathcal{E}^2(a^*) = E\,(a^* - a)^2 = \text{minimum}. \tag{4}$$

Condition (2) is never considered sufficient. It is followed by the requirement that variance of an unbiased estimate should be as small as possible. Out of all unbiased estimates the one with least variance is usually considered the best. Clearly, such a criterion is weaker than criterion (4).

In a former paper [2]) some other criteria were discussed, among them the principle of least squares, and the principle of Maximum Likelihood, both of which were found weaker than (4). It may be added here that Bayes' principle of estimation, aiming at the average value of a parameter in all populations, instead of at its value in a given population, is also weaker than (4).

Because of poor tractability of (3), (4) will be used in this paper as the working criterion of the best estimate.

Given an estimate a^*, we shall restrict curselves to finding an optimum value of the coefficient A in the expression Aa^*, such that $\mathcal{E}^2(Aa^*)$ should be minimum. This optimum value of A we shall

---

1) In some cases unbiased estimates are useful as a device of correcting some easily obtained estimates for gross error, i. e., of freeing them from a large admixture of some other systematic influences than the parameter in question, as for instance when subranges are corrected for bias in estimating standard deviation.

2) M. Olekiewicz, „On the Efficiency of Biased Estimates", Annales UMCS, Sectio A, v. III, 3, 1949.

denote by A_0. The corresponding estimate will be denoted by $a^*_0 = A_0 a^*$. Mean square error of estimate Aa^* can be expresed as follows

$$\mathcal{E}^2(Aa^*) = D^2(Aa^*) + b^2(Aa^*)$$
$$= A^2 D^2(a^*) + (AE(a^*) - a)^2 \qquad (5)$$

Keeping a^* fixed, and letting A vary, we find by minimizing (5) with respect to A

$$A_0 = \frac{a\,E(a^*)}{E^2(a^*) + D^2(a^*)} \qquad (6)$$

and the optimum estimate

$$a_0^* = A_0\, a^* = \frac{a\,E(a^*)}{E^2(a^*) + D^2(a^*)} \cdot a^* \qquad (7)$$

Mean square error of this estimate is

$$\mathcal{E}^2(a_0^*) = \frac{a^2 D^2(a^*)}{E^2(a^*) + D^2(a^*)} = A_0 D^2(a^*) \cdot \frac{a}{E(a^*)} \qquad (8)$$

When A_0 is independent of a, a_0^* can be taken for the best estimate of a of the form Aa^*. When A_0 depends on a, a^*_0 will be no estimator, since number A_0 will remain unknown. In any case, however, there exists a certain number A_0, known or unknown, such that when a^* is multiplied by it, a better estimate is obtained than any that could be got through multiplying a^* by a different number. If a^* happens to be an unbiased efficient estimate in the sense defined by H. Cramér[1]) (for a fixed n), it still can be improved through multiplying by $A_0 = \frac{a^2}{a^2 + D^2(a^*)}$. In the former paper this fact was utilized for a regular estimation case, vis., for estimating variance in normal population, where it was found that

$$\frac{\Sigma(X - \overline{X})^2}{n+1} \text{ is better than } \frac{\Sigma(X - \overline{X})^2}{n-1}.$$

In the present paper certain (regular and irregular) cases of estimating mean (μ) in various populations will be considered, and improved estimates of the mean over those commonly used will be given.

A general case when $\mu^* = \hat{\mu} = \overline{X} = \frac{1}{n}\Sigma X$

When a^* is unbiased ($a^* = \hat{a}$), we shall have

---

[1]) Cramér, H. Mathematical Methods of Statistics. Princeton University Press, 1946.

$$\left.\begin{aligned} A_0 &= \frac{a^2}{a^2 + D^2(\hat{a})} \\ \overset{*}{a}_0 &= A_0 \hat{a} = \frac{a^2}{a^2 + D^2(\hat{a})} \cdot \hat{a} \\ \mathcal{E}^2(a_0^*) &= \frac{a^2 D^2(\hat{a})}{a^2 + D^2(\hat{a})} = A_0 D^2(\hat{a}) \end{aligned}\right\} \tag{9}$$

According to (9) the optimum estimate of μ of the form $A\bar{X}$ is

$$\mu_0^* = A_0 \bar{X} = \frac{n}{n + \left(\frac{\sigma}{\mu}\right)^2} \cdot \bar{X} \tag{10}$$

From (10) it can be seen that the larger is σ as compared to $|\mu|$, the smaller will be A_0. This fact expresses a peculiar property of large deviations.

If $\frac{\sigma}{|\mu|}$ is known, μ and σ being unknown, μ_0^* will serve as the best estimator of μ of the form $A\bar{X}$, and if, besides, $\bar{X}$ is efficient, μ_0^* will be "linearly efficient", i. e. will represent the best of all possible "linearly biased" or unbiased estimates of μ [1]).

If only lower bound for $\frac{\sigma}{|\mu|}$ can be ascertained, we shall be able to determine such limits for, A, that any $A\bar{X}$ complying with them will be better estimates of μ than $\bar{X}$ is.

For this purpose we shall make use of the fact that when A is allowed to vary continuously, $\mathcal{E}^2(A\bar{X})$ will have its only critical point (minimum) at $A = A_0$, and there will be two values of A at which $\mathcal{E}^2(A\bar{X})$ will equal $\mathcal{E}^2(\bar{X}) = D^2(\bar{X})$, vis.

$$A = 1, \text{ and } A = 2A_0 - 1 \tag{11}$$

Thus all estimates $A\bar{X}$ for which A satisfies the inequality

$$2A_0 - 1 < A < 1 \tag{12}$$

will have smaller mean square errors of estimate than $\bar{X}$ has. Now, when it is known that $\frac{\sigma}{|\mu|} \geqslant l$, l being a known number, then

$$A_0 \leqslant \frac{n}{n + l^2}, \text{ and } 2A_0 - 1 \leqslant \frac{n - l^2}{n + l^2}; \tag{13}$$

[1]) Cf. M. Olekiewicz, op. cit.

and therefore it can be stated that all estimates, $A\overline{X}$, for which

$$\frac{n-l^2}{n+l^2}<A<1 \tag{14}$$

are better than $\overline{X}$ is.

If we are ignorant of the lower bound for $\frac{\sigma}{|\mu|}$ but σ is known, we may be able to determine a confidence interval for $|\mu|$, and taking its upper limit, find the lower limit for $\frac{\sigma}{|\mu|}$ with an acceptable risk of error. Then our statement about improved estimates of μ made in the form of (14) should be qualified with an appriopriate probability clause.

For instance, if X is normal, we shall have for estimated lower bound of $\frac{\sigma}{|\mu|}$ to be inserted in (14)

$$\frac{\sigma}{|\overline{X}|+\frac{\sigma U_{2P}}{\sqrt{n}}} \tag{15}$$

where U_{2P} is the normal deviate corresponding to one-tailed risk of error, P.

If neither $\frac{\sigma}{|\mu|}$ nor σ is known, it may still be possible to estimate a lower bound for $\frac{\sigma}{|\mu|}$ with an ecceptable risk of error to be inserted in (14).

In the case of distribution of Poisson, since $\sigma^2=\mu$, (10) becomes

$$\overset{*}{\mu}_0=\frac{n}{n+\frac{1}{\mu}}\overline{X} \tag{16}$$

The knowledge of an upper bound for μ will suffice in this case to determine a range of better estimates of μ than $\overline{X}$ is. If it is known that $\mu \leqslant K$, where K is a known number, then all estimates $A\overline{X}$ for which A satisfies the inequality

$$\frac{nK-1}{nK+1}<A<1 \tag{17}$$

are better than $\overline{X}$ is.

If we do not know anything obout μ, we still be able to take the upper limit of confidence interwal for μ,[1]) (if $n\bar{X}$ is sufficiently large)

$$K = \bar{X} + \frac{U_{2P}^2}{2n} + U_{2P} \sqrt{\frac{\bar{X}}{n} + \frac{U_{2P}^2}{4n^2}} \tag{18}$$

determined with an acceptable risk of error, P, and substituting it into (17), make our probality statement about the range of improved estimates compared to $\bar{X}$.

In the case of binomial distribution with two alternatives we have $\mu = p$, $\sigma^2 = pq$, where p is probability of success and $q = 1 - p$. Hence (10) becomes

$$\mu_0^* = p_0^* = \frac{n}{n + \frac{q}{p}} \cdot \hat{p} \tag{19}$$

where $\hat{p} = \bar{X}$ is an observed fraction of successes. If $\frac{q}{p}$ is unknown, but and upper bound K for p can be ascertained that is smaller than unity, then all estimates $A\bar{X}$ for which A satisfies the inequality

$$\frac{(n+1)K - 1}{(n-1)K + 1} < A < 1 \tag{20}$$

are better estimates of p than $\hat{p}$ is. If upper bound for p is not known. then for determinig a probable range of improved estimates K may be estimated from the formula [1]) (if $n\hat{p}$ is sufficiently large)

$$K = \frac{n}{n + U_{2P}^2} \left(\hat{p} + \frac{U_{2P}^2}{2n} + U_{2P} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{U_{2P}^2}{4n^2}} \right) \tag{21}$$

In connection with (19) it may be noted that the best effect will be obtained when $p < q$. The reason for this is the following. The mean square error of $q^* = 1 - p_0^*$ as an estimate of q is equal to mean square error of p_0^* as an estimate of p, vis., it equals $\frac{p^2 q}{np+q}$ Now, mean square error od c_0^* as an estimate of q equals $\frac{pq^2}{nq+p}$. This, of course, is the smallest mean square error that can be obtained for any estimate of the form $A\hat{q}$, but it is to be noted that q^* cannot be reduced to this form, and if $p < q$, then q^* turns out to have a smaller

[1]) Cf. Cramér, H., op. cit. pp. 514—515.

mean square error than q_0^* has, which can be seen from the following relation

$$\frac{\mathcal{E}^2(q^*)}{\mathcal{E}^2(q_0^*)} = \frac{\mathcal{E}^2(p_0^*)}{\mathcal{E}^2(q_0^*)} = \frac{npq + p^2}{npq + q^2} \tag{22}$$

It follows that if $p < q$, then p_0^* is better than $p^* = 1 - q_0^*$; but if $p > q$, then p^* will be better than p_0^*.

Some special cases of rectangular and triangular distributions

When probability density for a rectangular distribution is

$$f(x) = \begin{cases} \dfrac{1}{2\mu}, & 0 \leqslant x \leqslant 2\mu \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

then $\sigma^2 = \dfrac{\mu^2}{3}$, $D^2(\overline{X}) = \dfrac{\mu^2}{3n}$. Putting $\hat{\mu} = \overline{X}$, (10) becomes

$$\mu_0^* = \frac{3n}{3n+1} \cdot \overline{X},$$

so that μ_0^* is the best estimate of μ of the form $A\overline{X}$. Its mean square error of estimate is

$$\mathcal{E}^2(\mu_0^*) = \frac{\mu^2}{3n+1} \tag{25}$$

which is smaller than $\mathcal{E}^2(\hat{\mu}) = D^2(\overline{X})$.

The considered case is not a regular case of estimation, and $\overline{X}$ is not the best unbiased estimate of μ. The best unbiased estimate is known to be [1])

$$\tilde{\mu} = \frac{n+1}{2n} \cdot X_{(n)} \tag{26}$$

where $X_{(n)}$ is the largest observed value in the sample.
The mean square error of this estimate is

$$\mathcal{E}^2(\tilde{\mu}) = D^2(\tilde{\mu}) = \frac{\mu^2}{n(n+2)} \tag{27}$$

Applying (9) we obtain the best estimate of the form $A\tilde{\mu}$

$$\tilde{\mu}_0 = \frac{n(n+2)}{n(n+2)+1} \cdot \frac{n+1}{2n} X_{(n)} \tag{28}$$

Its mean square error of estimate is

$$\mathcal{E}^2(\tilde{\mu}_0) = \frac{\mu^2}{n(n+2)+1} \tag{29}$$

---

[1]) Cf. Davis, R. C. „On minimum variance in nonregular estimation". *The Annals of Mathematical Statistics*, vol. 22, 1951.

When frequency function for a triangular distribution is

$$f(x) = \begin{cases} \frac{8x}{9\mu^2} \text{ in } \left(0, \frac{3\mu}{2}\right) \\ 0 \text{ otherwise} \end{cases} \tag{30}$$

then best unbiased estimate of μ is

$$\tilde{\mu} = \frac{2n+1}{3n} X_{(n)} \tag{31}$$

Its mean square error of estimate is

$$\mathcal{E}^2(\tilde{\mu}) = D^2(\tilde{\mu}) = \frac{\mu^2}{4n(n+1)} \tag{32}$$

Applying (9) we obtain the best estimate of μ of the form $A\tilde{\mu}$

$$\tilde{\mu}_0 = \frac{4n(n+1)}{4n(n+1)+1} \cdot \frac{2n+1}{3n} X_{(n)} \tag{33}$$

Its mean square error of estimate is

$$\mathcal{E}^2(\tilde{\mu}_0) = \frac{\mu^2}{4n(n+1)+1} \tag{34}$$

Similarly the unbiased estimates of the mean can be improved in other special cases of distributions.

References

1. Cramér, H. Mathematical Methods of Statistics. Princeton University Press, 1946.

2. Davis, R. C. On minimum variance in non-regular estimation. AMS, Vol. 22, 1951.

3. Olekiewicz, M. On the Efficiency of Biased Estimates. Annales UMCS, v. III, 3, 1949.

Streszczenie

Autor podaje ulepszone oceny średniej arytmetycznej, posługując się kryterium najmniejszego średniego błędu kwadratowego. Rozpatrzony jest przypadek ogólny oraz przypadki szczególne dla rozkładów: dwumiennego, Poissona, jednostajnego i trójkątnego.

Резюме

Автор даёт улучшенные оценки средней арифметической пользуясь принципом наименьшей средней погрешности. Рассмотрен общий случай и некоторые частные случаи для распределений: биномияльного, Пуассона, равномерного и треугольного.