Title	Sample Size of Readings in Time Study (III)
Author(s)	Kase, Shigeo; Nishiyama, Noriyuki
Editor(s)	
Citation	Bulletin of University of Osaka Prefecture. Series A, Engineering and natural sciences. 1964, 13(1), p.179-184
Issue Date	1964-10-31
URL	http://hdl.handle.net/10466/8077
Rights	

Sample Size of Readings in Time Study (III)

Shigeo KASE* and Noriyuki NISHIYAMA*

(Received June 30, 1964)

Based upon the opinion that N' in time study is distributed doubly exponentially, a step in the right direction for determining an accurate N' is proposed in this paper. After the discussion on optimality of required N', the minimum-variance unbiased linear order-statistics estimator is applied to determine the mode of N'. The validity of this proposed method is successfully verified by the simulation tests with aid of a digital computer.

Introduction

The fact that N', the total number of observations to be made in order to provide the desired confidence level and confidence interval, follows the doubly exponential law has been proved graphically and numerically by these authors[1,2]. This makes it possible for us to proceed to the next study on the reasonable determination of the practical number of observations in time study.

Notwithstanding each N' value calculated by a statistical technique, the rigorous statistical treatment of N's population and therefore a method of condensing the information contained in a set of N's has never been researched. This results in a rather indefinite determination of N', for example, as follwos[3]:
Calculate N' from the formula

$$N' = [40\sqrt{N\Sigma X^2-(\Sigma X)^2}/\Sigma X]^2 . \tag{1}$$

If N' is equal to or less than the number of readings recorded, then the average for that element is probably (95 chances out of 100) within $\pm$ 5 per cent of the correct representative average. If N' is greater than the number of readings taken, then the study does not meet criterion of expected reliability, and a new study with N' or more readings of this element should be taken and rechecked.

A random sample of a population contains only a limited amount of information about the population and the operation of chance may have produced an extreme sample rather unrepresentative of the population. In order to improve these circumstances one should take the behavior of N' into consideration and construct a method of estimating the optimum number of observations or N'. This paper proposes in answer to such a request a new method based on the minimum-variance unbiased linear order-statistics estimator.

Optimality of determining N'

Generally in the problem of condensing and summarizing the information contained in the frequency distribution of a set of observations, certain functions of the distribution

* Department of Industrial Engineering, College of Engineering.

are useful. For most purposes, two salient characteristics of the distribution are the arithmetic mean and the standard deviation. This is sepecially true for a normal distribution, because it has two corresponding parameters m and σ.

However, the N' values calculated by Eq. (1) or the formula[4]

$$N' = [40RN/(d_2 \Sigma X)]^2 \tag{2}$$

are not distributed according to the normal law but to doubly exponential one[1,2]

$$y = (N' - \hat{N}')/\sigma', \quad F(y) = P = \exp[-\exp(-y)]. \tag{3}$$

This is the cumulative (or ogive) form of the distribution, which expresses the chance (100 P-per cent) that an observed number of readings will not exceed N' in value. The parameters of the distribution shown by Eq. (3) are $\hat{N}'$ and σ'. The quantity $\hat{N}'$ is the mode or highest point of the (frequency) distribution. The quantity σ' is a scale parameter, analogous to the standard deviation σ in the case of the normal distribution. In fact, σ' equals $\sqrt{6}/\pi$ times the standard deviation of the doubly exponential distribution.

Although the two parameters $\hat{N}'$ and σ' completely specify the distribution, it is desirable to introduce from Eq. (3) another quantity

$$\xi_P = \hat{N}' + \sigma' y_P$$

which is a linear combination of the parameters $\hat{N}'$ and σ'. The reason why this new parameter ξ_P is recommendable consists in the advantage that it can estimate $\hat{N}'$ and σ' simultaneously, rather than in terms of two separate problems. Thus if ξ_P can be estimated as $a + by_P$ with a and b known, then the values $\hat{N}' = a$ and $\sigma' = b$ are able to be read off at once.

In consideration of the practical purpose of time study, it may be not necessarily to obtain the accurate value of the spread or dispersion beyond the mode of the observed N'. At least it should be understood that only one value of N' is insufficient from the implicit recognition of the dispersion or σ'.

The above comes to the conclusion that an optimum representation of the actual number of readings becomes the **mode** of the several results each of which is calculated by Eq. (1) or (2) from 10 readings. And that the estimate of mode can be obtained by $\hat{N}' = a$ in the minimum-variance unbiased linear order-statistics estimator which shall be described later.

Estimators for mode in the doubly exponential distribution

Lieblein[5] considered an estimator of

$$\xi_P = \hat{N}' + \sigma' y_P$$

of the form

$$L = \sum_{i=1}^{n} (a_i + b_i y_P)(N')_i \tag{4}$$

where $(N')_1 \leqq (N')_2 \leqq \cdots \leqq (N')_n$ are the order statistics of a sample of n from Eq. (3), and sought to find the values a_i and b_i which minimize $Var(L)$ subject to

$$E(L) = \xi_P . \tag{5}$$

From Eq. (3),

$$N' = \hat{N}' + \sigma' y$$

where y is the reduced variate and N' the observed variable. From this the following relations for the order statistics $(N')_i$ and y_i are apparent:

$$(N')_i = \hat{N}' + \sigma' y_i\,, \qquad i = 1, 2, \cdots, n \tag{6}$$

$$(N')_1 \leqq N'_2 \leqq \cdots \leqq N'_n$$

$$y_1 \leqq y_2 \leqq \cdots \leqq y_n$$

$$E((N')_i) = \hat{N}' + \sigma' E(y_i)\,. \tag{7}$$

The values $E(y_i)$ may be obtained with the aid of the table in Lieblein's brochure.

Eqs. (4), (5) and (7) give

$$E(L) = \sum_{i=1}^{n} (a_i + b_i y_P)\,[\hat{N}' + \sigma' E(y_i)] = \xi_P = \hat{N}' + \sigma' y_P\,.$$

This is required to be an identity for all values of the parameters $\hat{N}'$ and σ'. Equating their coefficients gives the two conditions on the weights

$$\left.\begin{aligned} &\sum_{i=1}^{n} (a_i + b_i y_P) = 1\,, \\ &\sum_{i=1}^{n} (a_i + b_i y_P) E(y_i) = y_P \end{aligned}\right\} \tag{8}$$

where the numerical values $E(y_i)$ may be easily obtained as already indicated.

Turning to the variance, there is obtained in view of Eq. (4)

$$\sigma^2(L) = \sum_{i=1}^{n} (a_i + b_i y_P)^2 \sigma^2((N')_i) + \sum_{j=1}^{n} \sum_{\substack{i=1 \\ i \neq j}}^{n} (a_i + b_i y_P)(a_j + b_j y_P)\,\sigma((N')_i (N')_j)\,.$$

From Eq. (6) and the properties of the variances and covariances of linear estimators,

$$\sigma^2((N')_i) = (\sigma')^2 \sigma^2(y_i) = (\sigma')^2 \sigma_i^2$$

$$\sigma((N')_i (N')_j) = (\sigma')^2 \sigma(y_i y_j) = (\sigma')^2 \sigma_{ij}$$

making an obvious simplification in notation, whence

$$\begin{aligned} Var(L) &= [\Sigma\, w_i^2 \sigma_i^2 + \Sigma\,\Sigma' w_i w_j \sigma_{ij}](\sigma')^2 \\ &= \text{minimum subject to Eq. (8).} \end{aligned} \tag{9}$$

Use of Lagrange multipliers gives, after differentiations, the conditions on the weights

$$w_k \sigma_k^2 + \sum_{\substack{i=1 \\ i \neq k}}^{n}{}' w_i \sigma_{ik} + \lambda + \mu E(y_k) = 0\,, \qquad k = 1, 2, \cdots, n\,.$$

These latter are n linear equations which, with the two in conditions Eq. (8), form a simultaneous system of $(n+2)$ equations in the $(n+2)$ unknowns, w_1, $w_2, \cdots, w_n$, λ and μ. The right-hand sides of these $(n+2)$ equations are $1, y_P, 0, \cdots, 0$ and the solutions w_i, λ and μ are linear combinations of these with numerical coefficients which involve only σ_i^2, σ_{ij} and $E(y_i)$, but not y_P. The solutions are shown in Table 1 with the variances Q_n for sample size $n=2$ to 5.

Table 1. Weights for minimum-variance unbiased linear order-statistics estimator of percentage point ξ_P and variance $Var(\xi_P)=Q_n$ for $n=2$ to 5.

n		$(N')_1$	$(N')_2$	$(N')_3$	$(N')_4$	$(N')_5$
2	a_i	0.91637	0.08363			
	b_i	−0.72135	0.72135			
	Q_2	$(0.71186\ y_P^2-0.12864\ y_P+0.65955)\ (\sigma')^2$				
3	a_i	0.65632	0.25571	0.08797		
	b_i	−0.63054	0.25582	0.37473		
	Q_3	$(0.34472\ y_P^2+0.04954\ y_P+0.40286)\ (\sigma')^2$				
4	a_i	0.51100	0.26394	0.15368	0.07138	
	b_i	−0.55862	0.08590	0.22392	0.24880	
	Q_4	$(0.22528\ y_P^2+0.06938\ y_P+0.29346)\ (\sigma')^2$				
5	a_i	0.41893	0.24628	0.16761	0.10882	0.05835
	b_i	−0.50313	0.00653	0.13045	0.18166	0.18448
	Q_5	$(0.16665\ y_P^2+0.06798\ y_P+0.23140)\ (\sigma')^2$				

Since the present problem is released from the estimation of σ' and that $y_P=0$ for the mode $\hat{N}'$ as easily known, b_is in Table 1 can be disregarded. Two places below the decimal point of the coefficients a_is would be sufficient for actual estimation of the mode of N'. Finally, the recommendable estimators for mode are as follows:

$$n=2:\quad \hat{N}'=0.92(N')_1+0.08(N')_2$$
$$n=3:\quad \hat{N}'=0.66(N')_1+0.25(N')_2+0.09(N')_3$$
$$n=4:\quad \hat{N}'=0.51(N')_1+0.27(N')_2+0.15(N')_3+0.07(N')_4$$
$$n=5:\quad \hat{N}'=0.42(N')_1+0.24(N')_2+0.17(N')_3+0.11(N')_4+0.06(N')_5$$

where, in all equations

$$(N')_1\leqq(N')_2\leqq(N')_3\leqq(N')_4\leqq(N')_5\,.$$

Proposed method of determining N'

Substituting the values of w_i given in Table 1 into Eq. (9) yields an expression of the form

$$Var_{n,\,min}=Q_n=(A_n y_P^2+B_n y_P+C_n)(\sigma')^2$$

whose coefficients are seen in Table 1. In this equation the substitution of $y_P=0$ gives the coefficients in variances of the estimators for mode as 0.65955, 0.40286, 0.29346, 0.23140, when $n=2$, 3, 4, 5, respectively. From the above it is legitimate that the larger sample size is used for estimation of mode the more accurate estimate can be obtained.

In the practical time study, however, the performance of the operator, an irregular element or delay during the performance, etc. are apt to make N' calculations inaccurate. There is little benefit by increasing values of n, with correspondingly increasing difficulty, so that the sample size in the proposed method is truncated at $n=5$. It should be noted that the upper limit of $n=5$ is concerned with N' calculations which are gained from Eq. (1) or (2) by at least 10 readings.

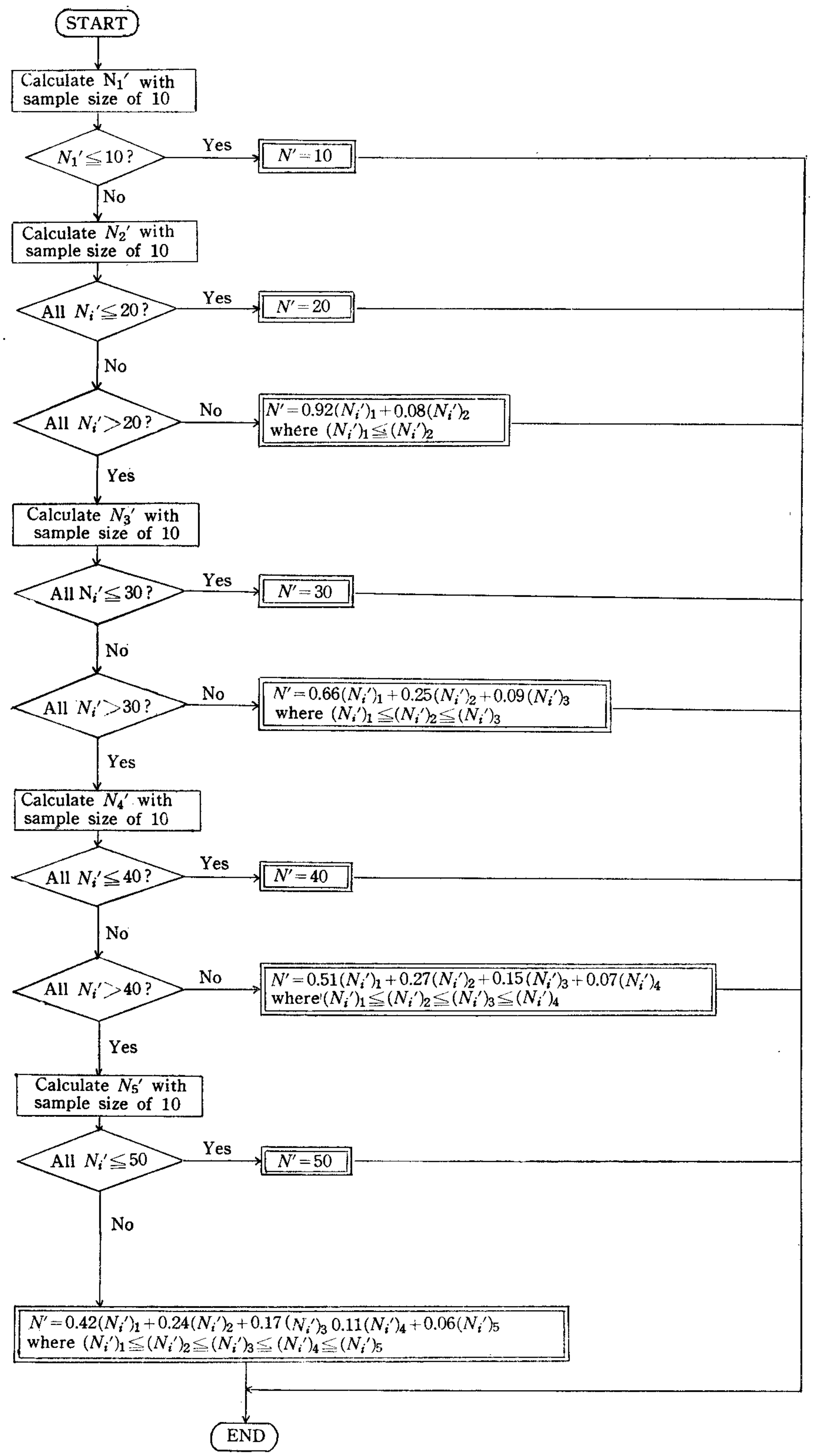

Figure 1. Flow Chart of The Proposed Method.

This size of the initial sample exceeds often the estimated mode of N', then the required number of readings is set down as $\hat{N'}=10$. Thus the results by the proposed method whose flow chart is shown in Fig. 1 might not be the correct representative mode for the comparatively small N', but at least they are on the safe, conservative side. Although such a situation is imperative throughout the method, it is rather an advantage of the method from the viewpoint of diminishing the relative error and making the calculation easier.

The validity of this new method has been verified by two simulation tests. In one of the tests the simulated 500 data of N' are constructed from the same way as in the previous paper[2]. That is, they come from 500 sets of ten numbers each of which is independent value from the normal distribution with mean 4×10^3 and variance 1×10^6, or $N(4\times10^3, 1\times10^6)$. In another test 500 data of N' are calculated by use of each ten random numbers from $N(8\times10^3, 1\times10^6)$. The modes $\hat{N'}$ estimated from these data are as shown in Table 2.

The means of the results which are yielded by substitution of the successive values of N' into the flow chart shown by Figure 1 are 67.03, 79.07, 18.55, 20.63, respectively, in that order of Table 2. The comparison of all four pairs justifies the above discussion

Table 2. Estimated modes of the simulated 500 values which are obtained from Eqs. (1) and (2).

Distribution	$N(4\times10^3, 1\times10^6)$		$N(8\times10^3, 1\times10^6)$	
Method	Eq. (1)	Eq. (2)	Eq. (1)	Eq. (2)
Mode $\hat{N'}$	68.46	80.57	17.07	20.01

in consideration of the conservative side for smaller N'. Thus the verification is satisfied, leading to the conclusion that the proposed method is recommendable in the precise analysis as well as the routine study.

References

1) S. Kase and N. Nishiyama, Bulletin of University of Osaka Prefecture, Series A, **12**, No. 2, 115 (1964).
2) S. Kase and N. Nishiyama, ibid., **12**, No. 2, 121 (1964).
3) M. E. Mundel, "Motion and Time Study, Principles and Practice," 3rd Ed., Prentice-Hall, Inc., 1960.
4) G. Nadler, "Motion and Time Study," McGraw-Hill Book Co., Inc., New York, 1955.
5) J. Lieblein, "A New Method of Analyzing Extreme-Value Data," Technical Note 3053, National Advisory Committee for Aeronautics, Washington, D. C., 1954.