ON DETERMINATION OF THE OPTIMAL SAMPLE SIZE IN RADIATION CONTROL

L. P. Pashchenko
Research Institute of Industrial and Marine Medicine, Yu.Gagarin
Prospect 67, St. Petersburg 196143, Russia

INTRODUCTION

Among various statistical models used for fitting occupational and environmental radiation data the lognormal distribution is the most common one. Its application became especially wide after this model has been adopted by International Commission on Radiological Protection (1, 2). Natural and occupational gamma- radon- and ultraviolet exposure, rate of global radioactive fallout, radiation contamination of the atmosphere are very often distributed lognormally (3,4). Typical problems arising when using the lognormal distribution model for purposes of radiation control are: how to estimate the average level of radiation with a given accuracy (the problem of estimation of unknown true mean of the distribution) and how to determine the minimum number of measurements by which the calculated average contamination differs from the true mean value no more than a given value of error (the problem of the number of sampling which is enough for estimating the unknown true mean with a given accuracy). The method universally practised of estimating the mean value consist of finding the arithmetic mean of the set of N measured values. Although the arithmetic mean as a statistical estimator is unbiased, it not always provides minimum variance of the estimation. Naturally arithmetic mean often differs from the true mean if the statistical distribution differs from Gaussian normal distribution. Therefore, it is necessary to know the distribution of sample mean from a lognormal population, but unfortunately this distribution is not known. All that can be said is that for a large number of samples it has an asymptotically normal distribution behaviour (5). Early, an attempt was made to obtain this distribution using some simplifications (6, 7), but uncertainty of the results caused by these simplifications, is too difficult to assess. In the present paper the sampling distribution of the lognormal population was found without any simplification.

GENERAL CONSIDERATIONS

As shown by Finney (8) the best estimation x_{EST} of the true mean value $\bar{x}$ (unbiased and given minimum variance of the differences from the true mean) is given for lognormal distribution by the formula:

$$x_{EST} = \exp(\mu)\, \psi_N(S^2) \qquad (1)$$

where μ and S^2 are the sample mean and variance, respectively, and $\psi_N(S^2)$ is a correction function depending on the sample variance S^2 and on the sampling number N. In (9) was shown that the variable L, defined as the ratio of sample mean calculated using expression (1), to its true mean value

$$L = x_{EST} / \bar{x} \qquad (2)$$

can be considered as a function of two random independent variables ξ and U, where the first is distributed normally and the second has a χ^2 distribution. Taking this into account we can express the distribution of L by multiplication of the distributions of ξ and U. Therefore the probability of fulfilling the condition $L \leq L'$ can be expressed by the integral of this product in that region of ξ and U, where this condition is fulfilled:

$$P(L \le L') \sim \iint_{L \le L'} \varphi(\xi)\, \chi^2(U)\, d\xi\, dU \qquad (3)$$

CALCULATION OF THE DISTRIBUTION

In the paper (9) the distribution of L was calculated by method of statistical testing (Monte Carlo method). But this technique is not an appropriate method if it necessary to get the results with a high precision. That is why in this investigation analytical and numerical methods were used. It gave the possibility to improves considerably the accuracy of calculations.

We have determined the distribution of L by numerical calculation of the integral (3) using Gauss' method with 64 nodes and their corresponding weight coefficients. For each combination of the sample size N (varied from 2 to 200) and sample standard deviation S (varied from 0.1 to 2.0) values for the cumulative probability, corresponding to different relative deviation D, between the true mean value and its best estimation, given by

$$D = (x_{EST} - \bar{x}) / \bar{x} \qquad (4)$$

were calculated. The D values for which calculations were made varied from 0.1 to 1.0 with a step equal to 0.1. The probability F_- and F_+ for negative and positive deviations of x_{EST} from the true mean $\bar{x}$ was calculated separately. Thus, the obtained data give a possibility to determine the probability of underestimation as well as overestimation by using x_{EST} as estimator of the true mean $\bar{x}$. Also their sum $F = F_- + F_+$ were calculated.

RESULTS AND THEIR APPLICATION

The distribution of the sampling mean were calculated and then used for the determination of the number of the sample size N that is enough for the estimation of the unknown true mean with a given accuracy. A part of the results obtained are presented in Table 1. The required number of sampling N are given in the Table as a function of the empirical variance S^2 and the confidence level P. For estimation of N it is necessary first to calculate μ and S^2. Then for a given value of error D and for a desirable confidence level P in Table 1 the number of N can be found.

REFERENCES

1. ICRP. Recommendations of the ICRP. Publication 26 (Oxford, Pergamon Press) (1977).
2. UNSCEAR. Sources and Effects of Ionizing Radiation. Appendix E (New York, United Nations) (1977).
3. Hounam RF. An application of the log-normal distribution to some air sampling results and recommendations on the interpretation of air sampling data. AERE-M 1469, Harwell, 1965.
4. Sources and effects of ionizing radiation. Report of the Scientific committee on effects of atomic radiation, United Nations, Vienna, 1978, Vol. 2, pp. 10-15.
5. Aitchison J, Brown JAC. The lognormal distribution. Cambridge university Press, 1957.
6. Sichel NS. New methods in statistical evaluation of mine sampling data, Trans. Inst. Min. Metall. (London) 61, 261-288 (1952).
7. Sichel NS. The estimation of the means and associated confidence limits for small samples from lognormal populations. Symposium on Mathematical Statistics and Computer Application in Ore Valuation. S. Afr. Inst. Min. Mettall. Johannesburg, March 1966, pp. 106-123.
8. Finney DJ. J. R. Statist. Soc. Suppl. N7, 151 (1941).
9. Pashchenko LP. On estimate of the mean by sampling from lognormal population lognormal population, Hygiene and Sanitation, N 5, p. 62-64, 1993, Moscow.

Table 1. The number of sample size N that is enough for the estimation of the mean with a given accuracy

Empirical standard deviation, S	Confidence level, P	Relative deviation, D				
		0.1	0.2	0.3	0.4	0.5
0.1	0.68	3	2	2	2	2
	0.90	5	3	3	3	3
	0.95	8	4	4	4	3
0.2	0.68	6	3	2	2	2
	0.90	14	5	4	3	3
	0.95	20	8	5	4	4
0.3	0.68	12	4	3	3	2
	0.90	28	10	5	4	4
	0.95	40	12	8	5	5
0.4	0.68	20	6	4	3	3
	0.90	50	14	8	5	4
	0.95	70	20	10	8	6
0.5	0.68	30	10	5	4	3
	0.90	80	22	12	8	5
	0.95	120	30	16	10	8
0.6	0.68	44	12	6	4	3
	0.90	120	32	16	10	8
	0.95	170	44	22	14	10
0.7	0.68	65	18	8	5	4
	0.90	170	44	20	12	10
	0.95	> 200	65	28	18	12
0.8	0.68	90	22	12	8	5
	0.90	> 200	60	28	16	12
	0.95	> 200	90	38	22	16
0.9	0.68	120	30	14	8	6
	0.90	> 200	80	36	22	14
	0.95	> 200	120	50	30	20
1.0	0.68	120	38	18	10	8
	0.90	> 200	110	46	26	18
	0.95	> 200	150	65	38	24