In : A week with elephants. Eds. J.C. Daniel and H.Datye, Bombay Natural History Society, Bombay and Oxford University Press, New Delhi. Pp.394-404. (1995)

SOME SOLUTIONS TO PROBLEMS IN ESTIMATION OF ELEPHANT DENSITY

J.A. SANTOSH AND R. SUKUMAR[1]

(With seven text-figures)

Statistical methods currently being used to estimate elephant densities and their confidence intervals are not statistically robust. A major problem in various methods has been that of incorporating variances of different parameters into the final elephant density estimate. Analytical solutions are complex and their derivations require parameters to follow particular statistical distributions. Another problem has been the estimation of the dung decay rate and its variance. Simple solutions that use Monté Carlo simulations and other procedures are suggested to solve these problems. These procedures satisfactorily incorporate parameter variances while estimating the confidence intervals of the density estimate. Since these solutions are computer intensive, a software that carries out these analyses has been prepared and is available on request.

INTRODUCTION

Density estimations of elephants pose statistical problems that are somewhat unique to the methods of sampling used and the nature of data (Barnes and Jensen 1987). Some of the problems encountered are, of course, common to studies of other vertebrates.

At an international workshop on censusing elephants held in southern India during January 1991 (Ramakrishnan *et al.* 1991), experts felt the need to improve upon techniques used to estimate elephant density and evolve standard procedures. There was also much debate about ways to increase the statistical robustness of estimates. In consultations with experts, the Asian Elephant Conservation Center has since been exploring various techniques and approaches to resolve aberrations in the analytical procedures.

Without getting into a debate over `precision vs. accuracy', it can be said that a major challenge in estimating the mean of any parameter involves developing a technique to obtain an unbiased estimate of its error. The standard error of the estimate is usually used to determine confidence intervals (CI) at a given significance level. Estimation of confidence intervals (CI) by conventional methods involves accurate estimation of the variance of the parameter. Implicit in this statement is that individual data points will also have to be accurately measured.

---

[1] Asian Elephant Conservation Centre, Centre for Ecological Sciences, Indian Institute of Science, Bangalore 560 012, India.

A major problem in estimation of elephant numbers is that it involves formulae that have several variables, each of which has a variance of its own. A major concern has been incorporating all these variances while estimating the confidence intervals for the final parameter (mean elephant density or number).

Two methods recommended by the Asian Elephant Specialist Group to estimate elephant numbers are the direct count (Karanth and Sunquist 1992) and the indirect count (Barnes and Jensen 1987), both of which use the line transect (Burnham *et al.* 1980). We shall deal with some drawbacks in the analyses of data that can potentially bias estimation of confidence intervals.

1. The Direct Count: In the direct count method, we estimate from line transects, the density of groups(D), and its CI (usually $D \pm 1.96 \times se[D]$, where $se[D]$ is the standard error of the density estimate). We extrapolate this to an estimate of the density of elephants using,

$$E = D \times H$$

where E is the elephant density, D, the group density (estimated from the transects), and H, the mean group size.

The 95% CI for the elephant density estimates are sometimes obtained by multiplying the lower and upper limits of the 95% CI of the group density into the mean group size, i.e.,

95% CI(of E), lower limit $= \{D-(1.96 \times se[D])\}\ H$ and
95% CI(of E), upper limit $= \{D+(1.96 \times se[D])\}\ H$ . . . . . (Eqn. 1)

This 95% CI of the elephant density would be appropriate if H is a constant. But this is far from the case - H is an estimate by itself, which means there is a variance attached to it. This error in measuring H is not reflected in the elephant estimate at all. This leads to a narrowing of the confidence interval, giving us a falsely over-precise estimate of the mean elephant density. Further, this could result in a reduction in the sampling effort put in on the transects (as researchers often stop sampling at an acceptable error limit) thereby leading to a less accurate estimation of the mean itself.

One suggestion at the 1992 Workshop was to determine the CI for the elephant density by multiplying the lower limit of the herd density into the lower limit of the herd size, and similarly the higher limits of both the parameters. Thus we have,

95%CI(of E), lower limit $= \{D-(1.96 \times se[D])\} \times \{H-(1.96 \times se[H])\}$
95%CI(of E), upper limit$= \{D+(1.96 \times se[D])\} \times H+(1.96 \times se[H])\}$
. . . . (Eqn. 2)

The equation 2 gives us very 'safe' estimates of the CI. The problem with this method is that it is too conservative giving us a very large CI. In addition, this situation might lead the user to put in much more effort than is actually required, thereby wasting resources.

Karanth and Sunquist (1992) have used an acceptable solution by Drummer (1987) who has suggested that the standard error of D be calculated the following way.

$$se[E^2] = se[D^2] \times \frac{se[H^2]}{n} + se[D^2] \times H^2 + \frac{se[H^2]}{n} \times [D^2] \quad \ldots \text{(Eqn. 3)}$$

where n is the sample size.

The equation 3 has been the most satisfactory approach so far, but is not without problems.

A more serious problem with estimating CIs by deterministic models is an underlying assumption of symmetry in the distributions of parameters. This is violated by many field data sets that we have examined. Because the distribution of parameters is not symmetrical, CIs that are equidistant from the mean on either side will be biased.

2. The Indirect Count: A similar problem arises in the indirect (dung) count method (Barnes and Jensen 1987). The errors of not integrating parameter variances are compounded by the fact that there are three variables used in this method. All three variables, the dung density, the dung decay rate and the defecation rate, are estimates with variances attached to them. The equation used in this method is,

$$E = \frac{Y\,r}{D}$$

where E is the elephant density, Y, the dung density (estimated by line transect), r, the dung decay rate (estimated through experimentation) and D, the defecation rate (estimated through field observations).

The 95% CI has been calculated in different ways. The most commonly used method (e.g. Sale *et al.* 1990) is,

$$95\%\ \text{CI(of } E\text{), lower limit} = \frac{\{Y - (1.96 \times se[Y])\}\ r}{D}$$

$$95\%\ \text{CI(of } E\text{), upper limit} = \frac{\{Y + (1.96 \times se[Y])\}\ r}{D} \quad \ldots\ldots \text{(Eqn. 4).}$$

The problems with using Eqn. 4 are the same as above; that of a misleadingly narrow CI resulting from inadequate incorporation of parameter variances.

A conservative suggestion made at the 1992 workshop was to use extreme values of parameter distributions to arrive at the 95% CI. The problem with this, again, is that this would give us a very wide CI.

Dawson and Dekker (1992; Pg. 34-35) have provided a formula (based on Goodman 1960) which incorporates variances of the three parameters. However, the term *covariance* in their formula should actually read as *coefficient of variation* for it to be meaningful (N.V. Joshi, pers. comm.). Even then the problem of the underlying distribution of parameters still exists. We have examined a vast body of field data from many places (for both the Asian and the African elephant) and have found that the distributions are far from symmetrical. It is theoretically possible to modify one of the known distributions to fit the data at hand and to then derive complex equations for the estimation of variance.

Another problem that we face in the indirect method is in calculating the variance of the dung decay rate. Data for decay rates are collected through dung decay experiments on the field (Barnes and Jensen 1987). Data are collected at intervals of time that can at times be quite large, because of constraints on the field. Because of time intervals involved while noting the number of dung piles that have disappeared, the midpoint of each class interval is taken to be the life span of the dung pile. This leads to a slight under-estimation of the variation in the life spans of the dung piles. This can be avoided during the experiment itself if the time interval between two observations is reduced. But this may not be practically possible because of field constraints.

There is another minor problem concerned with the estimation of the mean decay rate. Currently, the mean decay rate is calculated by a fit between the number of days since the experiment began and the proportions of dung piles surviving up to that day. Some authors assume a distribution *a priori*. For example, Sale *et al.* (1990) assume an exponential decay and use the slope of the fit line as the decay rate. Other equations can be used to predict life spans of dung piles, and then the reciprocal of mean life span is used as the mean decay rate (e.g., Barnes 1992). There is no problem with using the reciprocal of the mean life span as an estimate of the mean decay rate. The problem here is that since these methods use functions to fit data, the data has to be continuous, which is not so. For example, if 20 dung piles have disappeared between days 10 and 20, it is assumed that all 20 dung piles have disappeared at the midpoint of the class interval, i.e., at day 15. This makes the distribution discrete making it not very suitable for analysis requiring continuous distributions.

Solutions

Our approach to incorporate parameter variances into the final elephant density estimate involves computer-aided stochastic simulations.

We use Monté Carlo simulations to repeatedly (1000 times) sample from appropriate distributions of parameters to generate distributions of elephant densities. The 95% CI is then directly obtained from this distribution. Distributions of the different parameters used for sampling during simulation are generated using statistics from field or experimental data.

For the direct count we sample from, (a) distributions of group density generated using the mean and standard error obtained from the line transect data and (b) group size distributions actually observed on the transect. In the indirect count, sampling is from (a) distributions of dung density generated using mean and standard error from line transect data, (b) distributions of decay rate generated from mean and standard error from experimental data, and (c) distributions of defecation rate generated using mean and standard error from field data. A series of elephant densities is computed from each combination of the sampled parameters. For both the direct and the indirect method, the resulting elephant densities are then simply arranged in ascending order and 2.5% of the lowest and highest values are cut off to give us the 95% CI. Distributions were generated for group density, dung density, decay rate and defecation rate by transforming the standard normal distribution ($\mu = 0, \sigma = 1$) using estimated means and standard errors of the parameters. The procedure is as follows.

$$x = \text{random}(0, 1)$$

$$\text{Standard normal variate, } N(0, 1) = 6 - \sum_{i=1}^{12} x_i$$

$$\text{Normal variate of parameter, } y(\mu, \sigma) = \mu + \sigma(N)$$

For the problem of loss of variance in life spans of dung piles (in the indirect count), we recommend the following method. Since dung piles would have disappeared anywhere between two observation time-steps, the day of disappearance is randomized between the two intervals instead of assuming that all piles disappeared at the midpoint of the interval.

For example, if it was found that 22 dung piles disappeared between day 10 and day 18, the days of actual disappearance would be randomly scattered between 11 and 18 (e.g., 11, 14, 17, 13, 14, etc.). The life-spans of the dung-piles would thus be this number if the starting day was taken as 0. The mean decay rate is calculated as the reciprocal of the mean life-span of the dung piles.

Mean Decay Rate = $1/x$,

where x is the mean life-span of the dung-piles.

The variance of $(1/x)$ is estimated using the following formula :

$$\mathrm{Var}(1/x) = \frac{\mathrm{var}(x)/n}{x^4}$$

where n is the number of dung piles.

DISCUSSION

We have attempted to address some problems persistent in the analytical procedures in estimation of elephant density. Although deterministic solutions exist, they are not without problems (very complex derivations and assumptions of distributions). The simple solutions we have offered here to deal with some of the flaws in analytical techniques have been found satisfactory, both theoretically and practically.

The use of Monté Carlo simulations deals with the problem of incorporating the variances of parameters into the final elephant density estimate satisfactorily. The method does not assume any symmetry in the distributions and the CI is obtained by simply cutting of 2.5% of outliers on either side of the mean to yield non-equidistant confidence limits around the mean.

We provide the results of sample simulations for the direct and indirect count in figures 1a to 7. Figures 1a to 3 are distributions involving the direct count simulation. Figures 4a to 7 are distributions involving the indirect count. Notice the highly skewed distribution for elephant density in both, the direct (Fig. 3) and the indirect (Fig. 7) count. The source for parameters has been data collected at different points of time and may not correspond to each other. This is only to illustrate our analytical methods and elephant density values should not be taken as realistic.

We are aware that the solutions offered here are highly computer intensive. To make these solutions accessible to all elephant (and other vertebrate) researchers, we have prepared a menu driven, user-friendly software for DOS called GAJAHA (with which our simulations were carried out). This is available on request from the Asian Elephant Conservation Centre.

ACKNOWLEDGEMENTS

We thank N. V. Joshi, A.P. Gore, Mark Boyce and Richard Barnes for intensive discussions on the theoretical background to this paper. Surendra Varman, Uma Ramakrishnan and Milind Watve provided much of the field data used for our illustration.

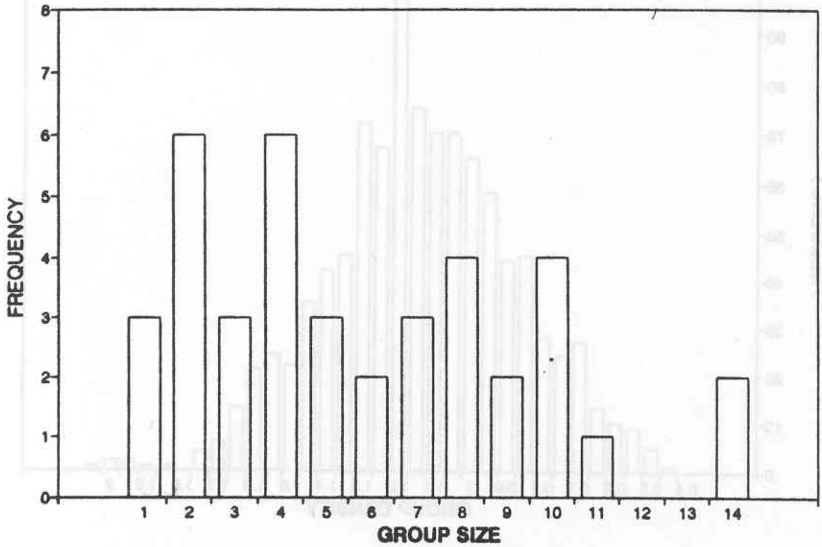

Fig. 1a. Distribution of groups collected on line transects at Mudumalai Wildlife Sanctuary, southern India. Group size distributions often have no apparent pattern and may be characterized by a large variance. Thus it may not be advisable to use the mean as a good representative of central tendency. Data distributions like these, form the pool from which we recommend sampling for the simulation. In this example, n = 39, and transect length = 409 km.

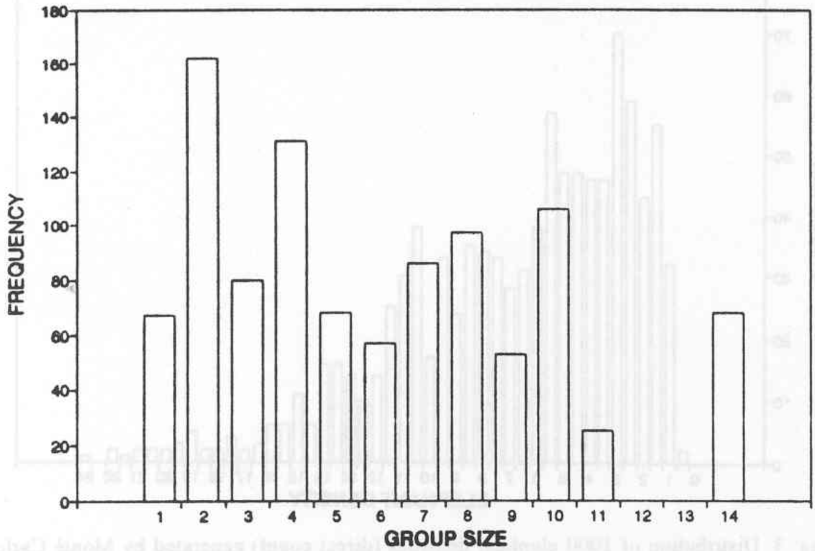

Fig. 1b. Distribution of 1000 group sizes used in a Montè Carlo simulation. This distribution was arrived at by sampling randomly from the distribution of field data displayed by Fig. 1a. The relative shapes of the distribution are not drastically altered as in the case when we use a mean and standard error in a prescribed formula that attempts to incorporate parameter variances.

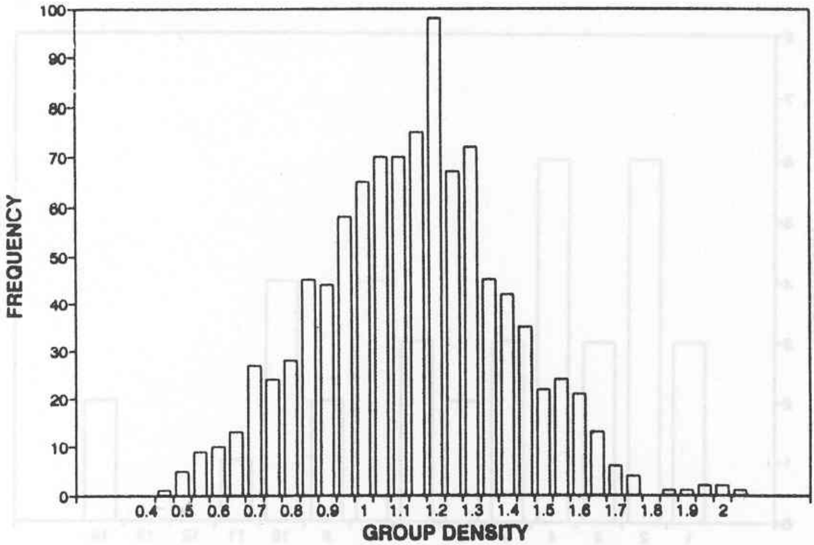

Fig. 2. Distribution of 1000 group densities (groups/sq km) used in a Montè Carlo simulation. The parameters to generate this distribution ($\mu = 1.1084$, $\sigma = 0.2608$) were obtained using a Fourier series estimator for line transect data from Mudumalai Wildlife Sanctuary. The distribution was generated by transforming a standard normal distribution (see text).

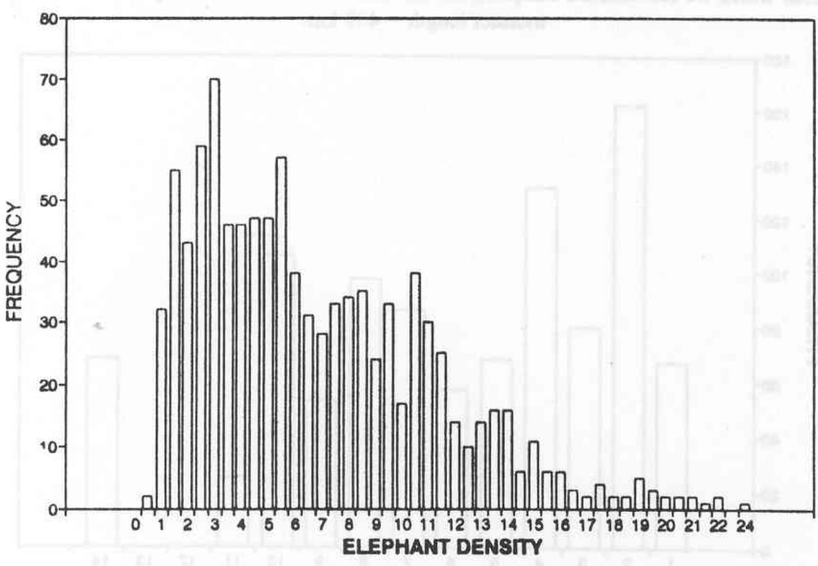

Fig. 3. Distribution of 1000 elephant densities (direct count) generated by Montè Carlo simulations that sample from field data distributions of group sizes (Fig. 1b.) and group densities (Fig. 2). The distribution is heavily skewed to the left making it inappropriate to use formulae that assume symmetry while estimating the CI. The mean elephant density, in this example, is 6.28.sq km and the 95% CI = 3.89 to 9.17 (note that the confidence limits are not equidistant from mean).

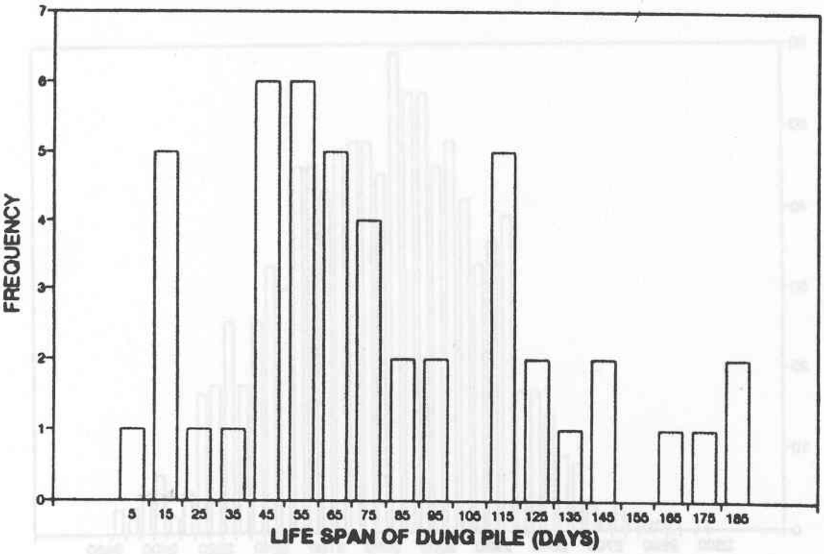

Fig. 4a. Distribution of life spans of dung piles obtained by randomizing days between two observations during the decay monitoring experiments. These are life spans of dung piles monitored at Mudumalai wildlife Sanctuary. The mean life span of dung piles for this data was found to be 73.36 days.

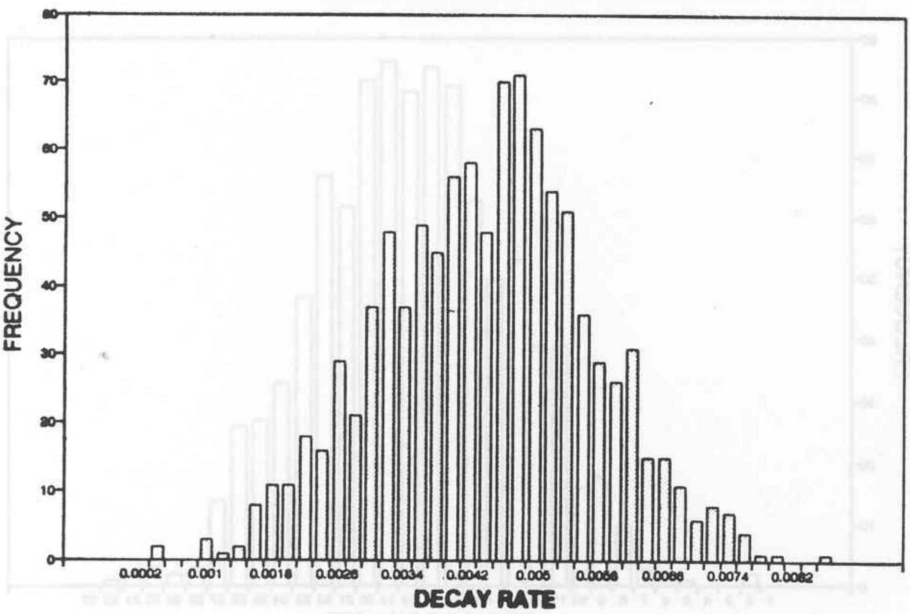

Fig. 4b. Distribution of 1000 decay rates used in Montè Carlo simulations. This distribution was obtained using the reciprocal of the mean life span from Fig. 4a. as the mean decay rate. The formula from variance of the decay rate is in the text. The parameters for this distribution are, $\mu = 0.0136$ and $\sigma = 0.0013$.

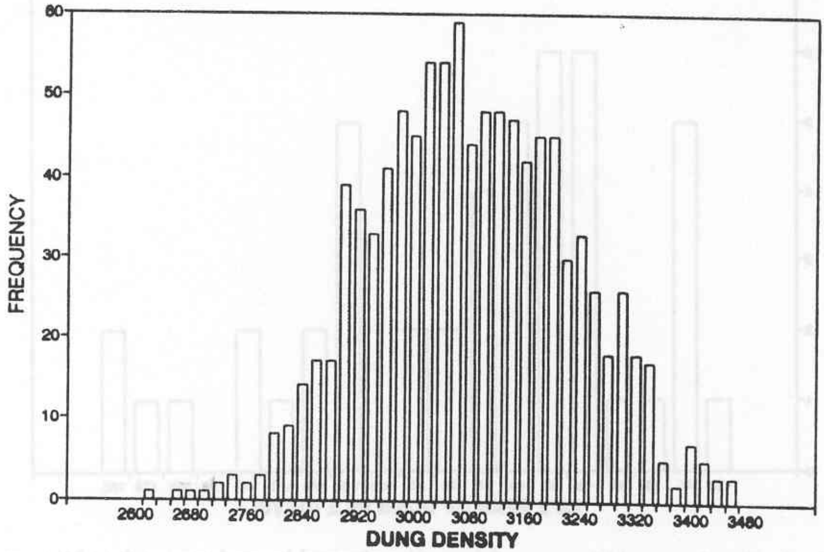

Fig. 5. Distribution of 1000 dung densities (dung piles/sq km) used in Montè Carlo simulations. The parameters used (μ = 3069, σ = 148.9) were obtained from a Fourier Series Estimator using data from line transects at Mudumalai Wildlife Sanctuary.

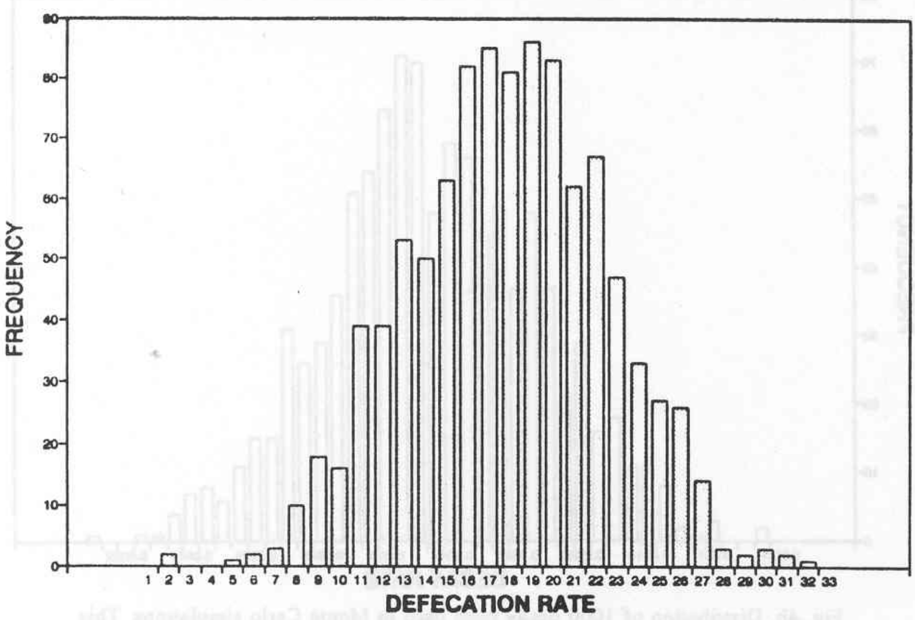

Fig. 6. Distribution of 1000 daily defecation rates used in Montè Carlo simulations. The distribution is based on more than 600 hours of observation of elephants in Mudumalai Wildlife Sanctuary. The distribution was obtained by transforming a standard normal distribution (see text).

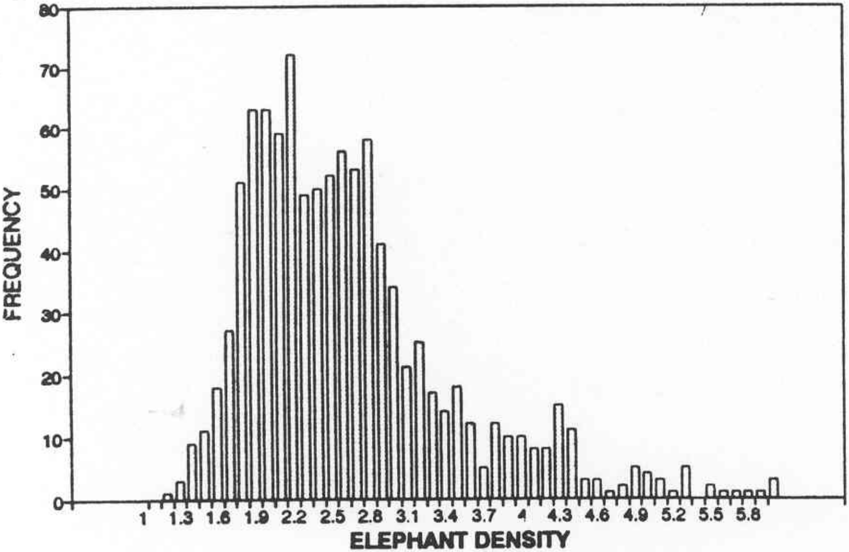

Fig. 7. Distribution of 1000 elephant densities (indirect count) generated by Montè Carlo simulations, sampling from parameter distribution displayed in Figs. 4b., 5 and 6. The mean elephant density is 2.68/sq km and 95% CI = 1.51 to 5.03 (note that the confidence limits are not equidistant from the mean).

REFERENCES

BARNES, R.F.W. AND K.L. JENSEN, (1987): How to count elephants in forests. IUCN African Elephant & Rhino Specialist Group Technical Bulletin *1*:1-6.

DAWSON, S., AND A.J.F.M. DEKKER (1992): Counting elephants in forests : a techniques manual. FAO, Bangkok.

DRUMMER, T.D. (1987): Program documentation and user's guide for SIZETRAN. Michigan Technological University, Houghton. 26 pp.

KARANTH, K.U., AND M.E. SUNQUIST (1992): Population structure, density and biomass of large herbivores in the tropical forests of Nagarhole, India. *J. Trop. Ecol. 8*: 21-35.

RAMAKRISHNAN, U., J.A. SANTOSH,, AND R. SUKUMAR (eds.) (1991): Censusing Elephants in Forests: Proceedings of an International Workshop. Technical Report No. 2, IUCN/SSC Asian Elephant Conservation Centre, Bangalore, India.

SALE, J.B., A.J.T. JOHNSINGH AND S. DAWSON (1990): Preliminary trials with an indirect method of estimating Asian elephant numbers. Report presented to the IUCN/SSC Asian Elephant Specialist Group.