RLE

Comments on Ulltang's "A note on the confidence limits of the egg production estimates from the mackerel egg surveys in the Western area"

M. R. Pennington

National Marine Fisheries Service
Northeast Fisheries Center
Woods Hole Laboratory
Woods Hole, Massachusetts 02543

Laboratory Reference No. 78-23
April 1978

Comments on Ulltang's "A note on the confidence limits of the egg production estimates from the mackerel egg surveys in the Western area"

M. R. Pennington

INTRODUCTION

Ulltang notes that the empirical distribution of $\ln(x + 2)$ for the egg production survey appears to be far from normally distributed, and hence any confidence intervals using a normality assumption may be falacious. He further observes that though $\ln(x + a)$ is not normal for any value of a, the distribution of $\ln(x)$, $x \neq 0$, appears to be so distributed. From these observations, it is suggested that the distribution of egg counts is not lognormal but follows a closely related distribution, the Δ-distribution. Confidence intervals derived under this assumption appear to be more realistic than the ones obtained under the assumption that egg counts are lognormally distributed. The following note represents a preliminary examination of this problem.

THE Δ-DISTRIBUTION

If a population is such that there is a proportion of zeros, and the distribution of non-zero values is lognormal, then the resulting distribution is called a Δ-distribution (see Aitchison and Brown [1957] for a detailed description of the properties of this distribution). If x is a variable so distributed, then its mean is:

$$\mu_x = (1 - q)e^{\mu' + \frac{1}{2}\sigma^{2'}}$$

where μ' and $\sigma^{2'}$ are the mean and variance respectively of $\ln x$, $x \neq 0$, and the variance of x is:

$$\sigma^2 = (1 - q)e^{2\mu' + \sigma^{2'}}(e^{\sigma^{2'}} - 1 + q).$$

The following estimators for μ and σ^2 are unbiased and efficient:

$$m = \begin{cases} \frac{n_1}{n} e^{\bar{y}} g_n(\tfrac{1}{2}s_y^2) & n_1 > 1 \\ \frac{x_1}{n} & n_1 = 1 \\ 0 & n_1 = 0 \end{cases}$$

$$r^2 = \begin{cases} \frac{n_1}{n} e^{2\bar{y}}\left\{g_{n_1}(2s_y^2) - \frac{n_1 - 1}{n - 1} g_{n_1}\left(\frac{n_1 - 2}{n_1 - 1} s_y^2\right)\right\} & n_1 > 1 \\ \frac{x_1^2}{n} & n_1 = 1 \\ 0 & n_1 = 0 \end{cases}$$

where n_1 is the number of non-zero sample values, and g_n is the usual adjustment function for the retransformation of the lognormal distribution which can be approximated by e^x for larger values of n.

It may be noted that there is no intrinsic relationship between the zero counts and the non-zero counts other than that a fixed proportion is zero. This may be contrasted to another possible model of the distribution of egg counts, the truncated lognormal distribution (Thompson 1951). For this distribution the frequency of x counts per sampling unit is given by:

$$\int_{\ln x}^{\ln(x+1)} dN(\mu,\sigma^2)$$

In particular, the frequency of zeros is:

$$\int_{-\infty}^{0} dN(\mu,\sigma^2),$$

and hence the zero and non-zero counts are part of a unified distribution. It may be noted that the frequency distribution of $\ln(x + 1)$ for the truncated normal model will look similar to those obtained by Ulltang for the egg data. Thompson (1951) gives an easy method for estimating the parameters of a truncated lognormal distribution. Which of the two models will most closely fit the distribution at hand can only be determined from a closer examination of the data; however from Ulltang's plot of $\ln(x)$ $x \neq 0$, the Δ-distribution appears to be the most plausible.

<u>A simple method for deriving approximate confidence intervals for the Δ-distribution.</u>

From a quick perusal of the literature, no method has been found for calculating confidence intervals for the Δ-distribution. The major difficulty is that one would like a confidence interval for the product of the parameter p, the proportion of non-zeros and μ, the mean of x, $x \neq 0$, i.e., a confidence interval for $p\mu$, the mean of x.

An approximate confidence interval may be obtained as follows:

(1) Calculate separately confidence intervals for p (which has a binomial distribution) and μ, say:

$$p' \leq p \leq p''$$

and

$$\mu' \leq \mu \leq \mu''$$

are 95% confidence intervals (where the fact that $\ln x$, $x \neq 0$, is normal is used to calculate the interval for μ).

(2) Let the confidence interval for $p\mu$ be defined by the minimum and the maximum of $p\mu$ on the above rectangular region. Hence the confidence interval will be:

$$p'\mu' \leq p\mu \leq p''\mu''.$$

The level of confidence will clearly be greater than 90%, and probably at least 95%. (See Halperin and Mantel [1963], and Halperin [1964] for an elaboration of the problems involved in determining confidence intervals for non-linear functions of parameters.)

For the CIROLANA 5/77 survey, it was estimated from Ulltang's Figure 1 that there were 9 zeros in a sample of size 75. Hence a 95% confidence interval for p is approximately:

$$.79 \leq p \leq .94.$$

From Ulltang's Table 1, a 95% confidence interval for μ is:

$$49 \leq \mu \leq 121.$$

Hence, an approximate 95% confidence interval for $p\mu$ is:

$$39 \leq p\mu \leq 114.$$

This compares favorably to the results of Ulltang's simulation in which he found that if the actual distribution of the CIROLANA 5/77 data is taken as the true distribution of the underlying population, then 95% of the time the mean of a sample of size 75 will be in the interval (41, 120). Furthermore, from Ulltang's Table 1, the retransformed mean of $\ln x$, $x \neq 0$, reduced according to the number of 0 hauls (which is equivalent to the estimator of the mean of the Δ-distribution shown in this note on page 2) gives values which are consistently close to the arithmetic mean, giving support to the hypothesis that the underlying distribution is a Δ-distribution.

REFERENCES

Aitchison, J., and J. A. C. Brown. 1957. The Lognormal Distribution. Cambridge University Press, London.

Halperin, M. 1964. Interval estimation of nonlinear functions. Parametric function II. J. Amer. Stat. Assoc. 59(1964):168.

_____, and W. Mantel. 1963. Interval estimation of nonlinear parametric functions. J. Amer. Stat. Assoc. 58(1963):611.

Thompson, M. R. 1951. Truncated lognormal distribution. Biometrika 38:414.

Ulltang, Ø. 1978. A note on the confidence limits of egg production estimates from the mackerel egg surveys in the Western area. Int. Comm. Northw. Atlant. Fish., Mackerel Working Group, Working Doc., February 1978.