Efficient Estimators of Abundance, for Fish and Plankton Surveys
Author(s): Michael Pennington
Source: *Biometrics*, Mar., 1983, Vol. 39, No. 1 (Mar., 1983), pp. 281-286
Published by: International Biometric Society
Stable URL: https://www.jstor.org/stable/2530830

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics*

This content downloaded from
86.177.242.126 on Fri, 14 Jul 2023 13:38:17 +00:00
All use subject to https://about.jstor.org/terms

Biometrics 39, 281–286
March 1983

Efficient Estimators of Abundance, for Fish and Plankton Surveys

Michael Pennington

National Marine Fisheries Service, National Oceanic and Atmospheric Administration,
Northeast Fisheries Center, Woods Hole, Massachusetts 02543, U.S.A.

Summary

The data from marine surveys often contain a large proportion of zeros. Treating the zeros separately can lead, in some cases, to more efficient estimators of abundance. To this end, the results of Aitchison (1955, *Journal of the American Statistical Association* **50,** 901–908) on the estimation of the mean and variance of a distribution with a discrete probability mass at zero are applied and extended to give an estimate of the variance associated with the estimate of the mean. It is shown that under some conditions these are minimum variance unbiased estimators. The case in which the nonzero values are lognormally distributed is examined in detail and applied to an ichthyoplankton survey.

1. Introduction

In multispecies fish and plankton surveys, large areas are sampled, and any particular species usually occupies only a part of the total survey area. In these circumstances the zero values can be taken to represent areas of unsuitable or unoccupied 'habitat'. The proportion of nonzero values in the sample estimates the proportion of the total survey area that is occupied by the species, and the nonzero values themselves can be used to estimate the abundance of the species in its area of occurrence.

The interpretation of the proportion of nonzeros in a sample as an estimate of habitat area may be vague in some situations, especially for mobile populations. A suitable habitat may change from time to time due to many factors including the timing of the survey, or an area may be unoccupied simply because of a low population level. However, keeping the zeros separate often enables one to fit a relatively simple distribution to the data, which leads, in some situations, to more efficient estimators of abundance.

Aitchison (1955) gave methods for estimating the mean and variance of a random variable which has a nonzero probability of being equal to zero and whose conditional distribution of nonzero values is some well-known distribution. In order to measure the precision of the abundance estimates, an estimate of the variance of the estimator for the mean is derived. It is shown that in some situations, these are minimum variance unbiased estimators. The case for which the nonzero values are lognormally distributed [what Aitchison and Brown (1957) called a 'Δ-distribution'] is examined in detail and applied to an ichthyoplankton survey. Due to the high variability of the log of the nonzero values for the plankton data, the estimator of the mean based on the Δ-distribution is much more efficient than the ordinary sample mean.

2. Estimation

Let X be a random variable defined on a population which contains a subset of elements on which X is equal to zero, and denote the mean and variance of the conditional distribution

Key words: Estimation of population abundance; Minimum variance unbiased estimator; Δ-distribution; Plankton survey.

This content downloaded from
86.177.242.126 on Fri, 14 Jul 2023 13:38:17 +00:00
All use subject to https://about.jstor.org/terms

of X for $X \neq 0$ by μ and σ^2, respectively. Then if $\mathrm{pr}(X \neq 0) = p$, the mean, α, and variance, β, of X are given by

$$\alpha = p\mu$$

and

$$\beta = p(1-p)\mu^2 + p\sigma^2.$$

Suppose $x_1, \ldots, x_m, \ldots, x_n$ is a random sample of size n, such that the first m values are nonzero. Aitchison (1955) showed that if, for $m > 0$, $a_{(m)}$, $e_{(m)}$ and $f_{(m)}$ are unbiased estimators of μ, μ^2 and σ^2, respectively, then

$$c = \begin{cases} \dfrac{m}{n} a_{(m)}, & m > 0, \\ 0, & m = 0, \end{cases}$$

and

$$d = \begin{cases} \dfrac{m}{n} f_{(m)}, + \dfrac{m}{n}\left(\dfrac{n-m}{n-1}\right) e_{(m)}, & m > 0, \\ 0, & m = 0, \end{cases}$$

are unbiased estimators of α and β, respectively. He further showed that if $a_{(m)}$ is also a sufficient statistic for μ, then m/n and $a_{(m)}$ are jointly sufficient and hence c has minimum attainable variance. A similar result holds for the estimator d.

It can also be shown that if $a_{(m)}$ is a complete sufficient statistic for μ, then m/n and $a_{(m)}$ are joint complete sufficient statistics for the parameters p and μ. For this case, c will be the minimum variance unbiased estimator of α. Again, an analogous result holds for d.

What is needed in most applications is an estimate of $\mathrm{var}(c)$. If $a_{(m)}$ is the arithmetic mean of the m values, then c is the arithmetic mean of the entire sample. Hence, $\mathrm{var}(c) = \beta/n$ and can be efficiently estimated by d/n.

It can be proved in general that

$$\mathrm{var}(c) = \frac{1}{n^2} \mathrm{E}\{m^2 \mathrm{var}(a_{(m)})\} + \frac{p(1-p)}{n} \mu^2,$$

and if $g_{(m)}$ is an unbiased estimator of $\mathrm{var}(a_{(m)})$ for $m > 0$, then

$$\mathrm{var}_{\mathrm{est}}(c) = \begin{cases} \dfrac{m}{n}\left(\dfrac{m-1}{n-1}\right) g_{(m)} + \dfrac{m}{n^2}\left(\dfrac{n-m}{n-1}\right) a^2_{(m)}, & m > 0, \\ 0, & m = 0, \end{cases} \tag{1}$$

is an unbiased estimator of $\mathrm{var}(c)$. As before, if $g_{(m)}$ and $a_{(m)}$ are joint complete sufficient statistics, then $\mathrm{var}_{\mathrm{est}}(c)$ is the minimum variance unbiased estimator of $\mathrm{var}(c)$.

3. The Δ-Distribution

A useful special case occurs when the conditional distribution of X for $X \neq 0$ is lognormal. This was called the Δ-distribution by Aitchison and Brown (1957). If $\mathrm{pr}(X \neq 0) = 1$, then X is lognormally distributed. Unbiased estimators of the mean and variance of the Δ-distribution

This content downloaded from
86.177.242.126 on Fri, 14 Jul 2023 13:38:17 +00:00
All use subject to https://about.jstor.org/terms

are given by (Aitchison, 1955)

$$c = \begin{cases} \dfrac{m}{n} \exp(\bar{y}) G_m(\tfrac{1}{2}\, s^2), & m > 1, \\ \dfrac{x_1}{n}, & m = 1, \\ 0, & m = 0, \end{cases} \tag{2}$$

and

$$d = \begin{cases} \dfrac{m}{n} \exp(2\bar{y}) \left\{ G_m(2s^2) - \left(\dfrac{m-1}{n-1}\right) G_m\left(\dfrac{m-2}{m-1}\, s^2\right) \right\}, & m > 1, \\ \dfrac{x_1^2}{n}, & m = 1, \\ 0, & m = 0, \end{cases}$$

where $\bar{y}$ and s^2 are the sample mean and sample variance, respectively, of the log of the nonzero values and

$$G_n(t) = 1 + \frac{n-1}{n} t + \sum_{j=2}^{\infty} \frac{(n-1)^{2j-1}}{n^j (n+1)(n+3) \cdots (n+2j-3)} \times \frac{t^j}{j!}.$$

Hoyle (1968) showed that an unbiased estimate of the variance of $\exp(\bar{y}) G_n(\frac{1}{2} s^2)$ is, in the same notation as above,

$$\exp(2\bar{y}) \left\{ G_n^2(\tfrac{1}{2}\, s^2) - G_n\left(\frac{n-2}{n-1}\, s^2\right) \right\}. \tag{3}$$

Denoting (3) by $g_{(m)}$ and substituting it and $a_{(m)} = \exp(\bar{y}) G_m(\frac{1}{2} s^2)$ into (1) yields the unbiased estimator

$$\mathrm{var}_{\mathrm{est}}(c) = \begin{cases} \dfrac{m}{n} \exp(2\bar{y}) \left\{ \dfrac{m}{n}\, G_m^2(\tfrac{1}{2}\, s^2) - \left(\dfrac{m-1}{n-1}\right) G_m\left(\dfrac{m-2}{m-1}\, s^2\right) \right\}, & m > 1, \\ \left(\dfrac{x_1}{n}\right)^2, & m = 1, \\ 0, & m = 0. \end{cases} \tag{4}$$

Since $\bar{y}$ and s^2 are joint complete sufficient statistics for the parameters of the lognormal distribution, it follows that m/n, $\bar{y}$ and s^2 are joint complete sufficient statistics for the Δ-distribution, and hence, c, d and $\mathrm{var}_{\mathrm{est}}(c)$ are the minimum variance unbiased estimators. If $\mathrm{var}\{\ln(X) | X \neq 0\}$ is large, as is often the case in fish and plankton surveys, then c is considerably more efficient than the ordinary sample mean.

4. An Example

In 1977 a series of ichthyoplankton surveys was conducted off the northeast coast of the United States. From these surveys, it was desired to estimate the total egg production of Atlantic mackerel, *Scomber scombrus*, in this region. With this estimate of egg production, the abundance of spawning stock could be estimated for comparison with other independent estimates of mackerel abundance. The surveys were not designed solely to measure the abundance of mackerel eggs but also to monitor the distribution and abundance of the entire zooplankton community. Thus the design embraced the requirements of many investigations and was not optimal for estimating mackerel egg production. Nevertheless, the surveys provided estimates of egg production at several times over the season, and these were integrated over time to produce an estimate of total seasonal production. For a more detailed discussion of procedures and methodology, see Berrien, Naplin and Pennington (1981).

This content downloaded from
86.177.242.126 on Fri, 14 Jul 2023 13:38:17 +00:00
All use subject to https://about.jstor.org/terms

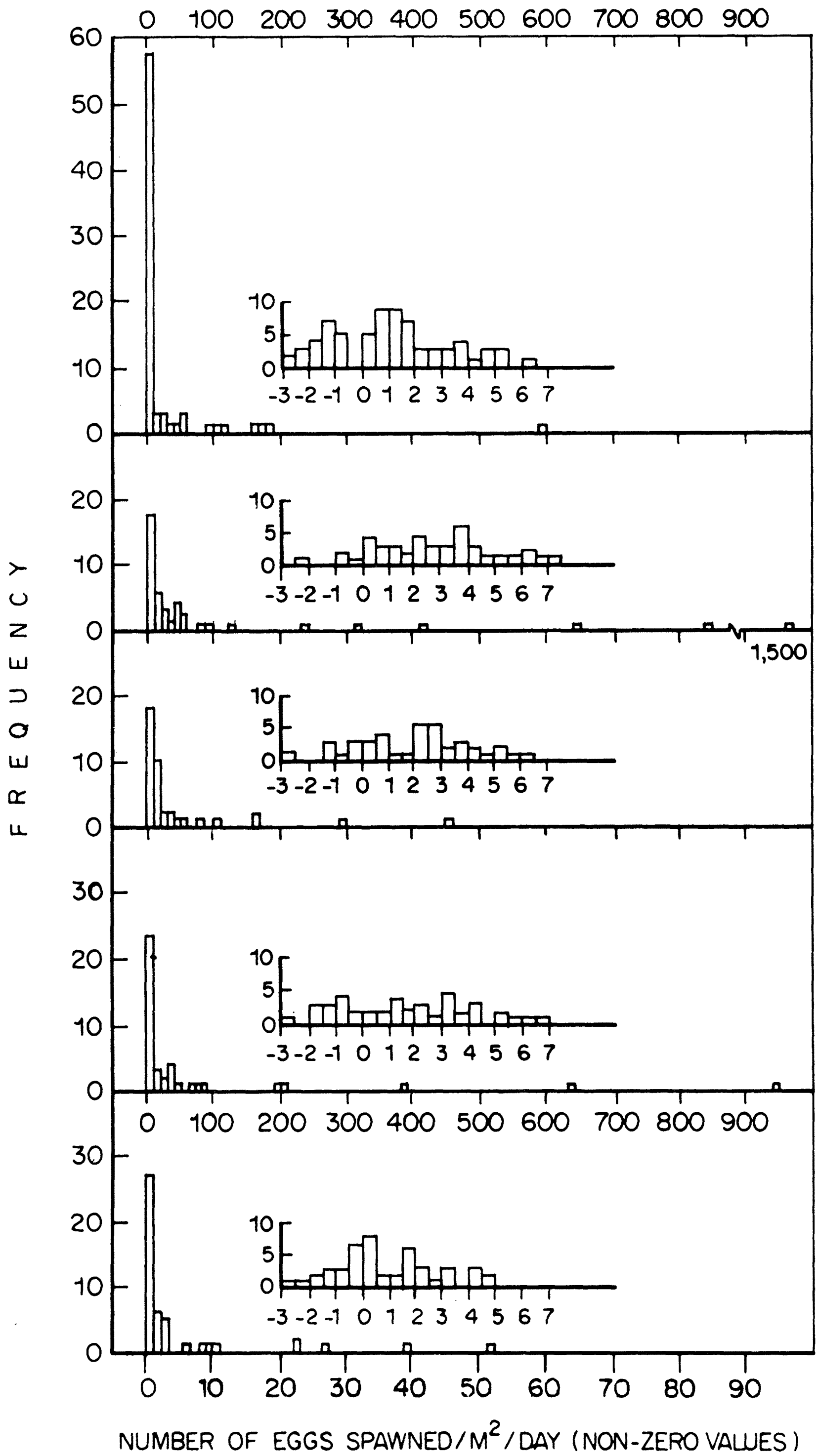

Figure 1. Frequency distributions of the nonzero egg production densities for five cruises. Insets are the distributions of $\ln(x)$, $x \neq 0$.

This content downloaded from
86.177.242.126 on Fri, 14 Jul 2023 13:38:17 +00:00
All use subject to https://about.jstor.org/terms

The data from an individual data-gathering cruise contained many zero values and were highly skewed to the right (Fig. 1). Various tests of normality did not reject the assumption that the log of the nonzero values is normally distributed. So estimators based on the Δ-distribution theory were used.

A typical example of the calculations made to estimate the egg abundance at a particular time is given below. The sample statistics needed are

$$n = 91, \qquad m = 41,$$
$$\bar{y} = 2.09, \qquad s^2 = 4.41.$$

The middle graph in Fig. 1 shows the frequency distribution of the nonzero values.

Estimates of the mean and its variance were calculated from (2) and (4), respectively:

$$\begin{aligned} c &= \tfrac{41}{91} \exp(2.09) G_{41}(2.21) \\ &= 28.4 \end{aligned}$$

and

$$\begin{aligned} \mathrm{var}_{\mathrm{est}}(c) &= \tfrac{41}{91} \exp(4.18)\{\tfrac{41}{91}\, G_{41}^2(2.21) - \tfrac{40}{90} G_{41}(\tfrac{39}{40} \times 4.41)\} \\ &= 187. \end{aligned}$$

These values compare with a sample mean, $\bar{x}$, of 17.6 and estimates of $\mathrm{var}(\bar{x})$ of 39 from the sample variance, s_x^2/n, and of 521 from d/n.

The appropriate weighted sum of the individual estimates of c and $\mathrm{var}_{\mathrm{est}}(c)$ were then calculated in order to estimate total egg production and its standard error. The final estimate of spawning stock size (1.225×10^9 fish), based on the total egg production, compares favorably with an estimate (1.379×10^9 fish) made independently by other techniques (Berrien *et al.*, 1981).

5. Discussion

The example above indicates another problem associated with the use of ordinary sample statistics for egg data. The distribution of egg densities is highly skewed. In fact 10% of the sample contains 85% of the eggs. For moderate sample sizes the distribution of s_x^2/n will also be highly skewed to the right. Thus if s_x^2/n is used to measure the precision of $\bar{x}$, then because of the skewness of its distribution, it will, much more often than not, underestimate the true variability of $\bar{x}$. Hence, s_x^2/n will frequently give a too optimistic impression of the precision obtained by a particular survey. Furthermore, it should be noted that $\mathrm{var}_{\mathrm{est}}(c)$ can also be larger than s_x^2/n. But again, this seeming loss of precision is the result of using s_x^2, which is extremely inefficient in comparison to d, to estimate the variance.

Acknowledgement

The data in the example were collected as part of the National Marine Fisheries Service's MARMAP program.

Résumé

Les données d'enquêtes marines contiennent souvent une forte proportion de zéros. Traiter les zéros séparément peut permettre dans certains cas d'obtenir des estimateurs de l'abondance plus efficaces. Dans ce but, les résultats d'Aitchison (1955, *Journal of the American Statistical Association* **50**, 901–908) sur l'estimation de l'espérance et de la variance d'une distribution avec une masse de probabilité discrète au point zéro sont appliqués et étendus pour donner un estimateur de la variance associé à l'estimateur de l'espérance. On montre que sous certaines conditions ce sont des estimateurs sans biais de variance

This content downloaded from
86.177.242.126 on Fri, 14 Jul 2023 13:38:17 +00:00
All use subject to https://about.jstor.org/terms

minimum. On examine en détail la cas où les valeurs différentes de zéro ont une distribution lognormale et on l'applique à une enquête sur l'ichtyoplancton.

References

Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association* **50,** 901–908.

Aitchison, J. and Brown, J. A. C. (1957). *The Lognormal Distribution.* Cambridge: Cambridge University Press.

Berrien, P. L., Naplin, N. A. and Pennington, M. (1981). Atlantic mackerel, *Scomber scombrus*, egg production and spawning population estimates for 1977 in the Gulf of Maine, Georges Bank, and Middle Atlantic Bight. *Rapports et Procès-Verbaux des Réunions, Conseil International pour l'Exploration de la Mer* **178,** 279–288.

Hoyle, M. H. (1968). The estimation of variance after using a Gaussianating transformation. *Annals of Mathematical Statistics* **39,** 1125–1143.

Received January 1981; *revised November* 1981

This content downloaded from
86.177.242.126 on Fri, 14 Jul 2023 13:38:17 +00:00
All use subject to https://about.jstor.org/terms