READER REACTION

Comment on Cowling's "Spatial Methods for Line Transect Surveys"

Magne Aldrin,[1,*] **Marit Holden,**[1,**] **and Tore Schweder**[1,2,***]

[1]Norwegian Computing Center, P.O. Box 114 Blindern, N-0314 Oslo, Norway
[2]Department of Economics, University of Oslo, P.O. Box 1095 Blindern, N-0317 Oslo, Norway
[*]*email:* magne.aldrin@nr.no
[**]*email:* marit.holden@nr.no
[***]*email:* tore.schweder@econ.uio.no

SUMMARY. We consider the problem of estimating the parameters of a two-dimensional Neyman-Scott process, from data collected through a line transect survey. Cowling (1998, *Biometrics* **54,** 828–839) suggested an estimation method based on a one-dimensional K-function along the transect line. However, her expression for the theoretical K-function is wrong. In this article, we correct her K-function.

KEY WORDS: K-function; Line transect survey; Neyman-Scott process.

1. Introduction

In an article in *Biometrics,* Cowling (1998) (henceforth, AC) considered the problem of estimating the parameters of a Neyman-Scott process, from data collected through a line transect survey. The actual Neyman-Scott process is a two-dimensional spatial point process, suitable for modeling the spatial distribution of clustered populations. When points (or individuals) in the population are observed by a line transect survey, the probability of detecting points decreases rapidly, as a function of the perpendicular distance x from the transect line, so only points close to the transect are detected.

Hagen and Schweder (1995) proposed an estimation method for use in line transect surveys for minke whales. They projected the detected points onto the transect line (the y-coordinates), and ignored the x-coordinates. The rationale for ignoring the x-coordinates is that the distance to the transect line is usually considerably shorter than the transect length, and that most of the information about the spatial pattern is contained in the coordinate along the transect. However, even if they used only one dimension of the data, they used a two-dimensional K-function (Ripley, 1977) for parameter estimation. AC constructed a more whole-hearted one-dimensional estimation method, based on a real one-dimensional K-function along the transect line. However, her expression for the theoretical K-function is wrong. In this article, we follow the idea of AC, and correct her K-function.

2. The Neyman-Scott Process and the Detection Function

The definition of a Neyman-Scott process is found in, for instance, Cressie (1993, p. 662). We will consider the following special version:

- Invisible parent events are Poisson distributed with intensity λ (per unit area).
- Each parent independently produces a Poisson (μ) number of offspring.
- The positions of the offspring relative to their parents are independent, and have an isotropic bivariate normal distribution with variance ρ^2 in all directions.

The detection function $g(x)$ is the probability of detecting an offspring at a distance x from the transect line. Following AC, assume a normal detection function,

$$g(x) = g_0 \cdot \exp(-x^2/2\sigma^2), \qquad (1)$$

where $g_0 = g(0)$ is the detection probability at $x = 0$. AC assumes $g_0 = 1$, but this assumption is unnecessary. For North Atlantic minke whales, g_0 is around 0.35—due to diving, etc. (Schweder et al., 1999). The two parameters g_0 and σ are typically estimated from external data, and are assumed known in the present context.

Suppose that a line transect survey of infinite length is carried out along $x = 0$, and consider a parent event located at $x = c$. Let T denote the number of detected offspring in that cluster. AC shows that, conditioned on c, the expected number of detected offspring is

$$E(T \mid c) = g_0 \mu p_c, \qquad (2)$$

where $p_c = \sigma(\sigma^2 + \rho^2)^{-1/2} \exp\{-c^2/2(\sigma^2 + \rho^2)\}$. Further, the conditional distribution of T is Poisson, i.e., $T \mid c \sim$ Poisson$(g_0 \mu p_c)$.

3. Parameter Estimation

The plan is now to project the detected points onto the transect line, calculate the theoretical K-function for the resulting one-dimensional process, and fit it to the empirical one-dimensional K-function.

The K-function (Ripley, 1977) of a stationary spatial point process, with intensity τ, is defined as

$$K(h) = \tau^{-1}\mathrm{E}\left(\begin{array}{c}\text{number of extra events within distance } h \\ \text{of a randomly chosen event}\end{array}\right).$$
(3)

Cressie (1993, p. 665) gives the K-function for a d-dimensional Neyman-Scott process. In one dimension, this becomes

$$K_1(h) = 2h + \frac{\mathrm{E}\{T(T-1)\}}{\lambda_1\{\mathrm{E}(T)\}^2}F(h),$$
(4)

where $F(h)$ is the distribution function of the distance between two points in the same cluster, λ_1 is the intensity of the one-dimensional parent process, and T is the number of offspring in a cluster.

The one-dimensional process will still fulfill the conditions for a (stationary and isotropic) Neyman-Scott process. The offspring will be normally distributed around the cluster centers, and

$$F(h) = 2\Phi(h/\rho\sqrt{2}) - 1,$$
(5)

where Φ is the distribution function of the standard normal distribution (AC).

However, even if T is Poisson conditional on the x-coordinate of the cluster ($x = c$), the marginal distribution of T is *not* Poisson. AC's expressions for $\mathrm{E}\{T(T-1)\}/(\lambda_1\{\mathrm{E}(T)\}^2)$ and K_1 are, unfortunately, wrong, since T was assumed Poisson distributed. The correct K-function for the detected points projected onto the transect line is

$$K_1 h = 2h + \{2\Phi(h/\rho\sqrt{2}) - 1\}/(2\sqrt{\pi}\lambda\sqrt{\sigma^2 + \rho^2});$$
(6)

see the Appendix for a proof.

The empirical K-function is given by

$$\widehat{K}_1(h) = 2n^{-2}\,\mathrm{L}\sum_j\sum_{i<j}\mathrm{I}(|y_i - y_j| < h),$$
(7)

where L is the length of the transect, n is the number of detected points, and y_i and y_j are the positions along the transect line. Now, the parameters λ and ρ can be estimated by fitting the theoretical K-function to its empirical counterpart; see AC. From (6), we see that $K_1(h)$ depends on σ in addition to λ and ρ, so it should be possible to estimate σ as well, in situations where it is not known from other sources.

We further have that

$$\mu = [\{E(n)\}/(2\sqrt{\pi}\lambda\sigma g_0 L)]$$
(8)

(see Appendix for proof)—allowing μ to be estimated by substituting $\mathrm{E}(n)$ by the observed n, and λ and σ by their estimates.

In a simulation study similar to that of AC, we have shown that Cowling's corrected method clearly outperforms the method of Hagen and Schweder (1995) (see Aldrin, Holden, and Schweder, 2001).

References

Aldrin, M., Holden, M., and Schweder, T. (2001). *Estimation of a Neyman-Scott process from line transect data by a one-dimensional K-function.* Technical Report SAMBA/18/01, Norwegian Computing Center, Oslo. Available at http://publications.nr.no/NS.estbyK.Biometrics.pdf.

Cowling, A. (1998). Spatial methods for line transect surveys. *Biometrics* **54,** 828–839.

Cressie, N. A. C. (1993). *Statistics for Spatial Data,* Revised edition. New York: Wiley.

Hagen, G. and Schweder, T. (1995). Point clustering of minke whales in the northeastern Atlantic. In *Whales, Seals, Fish and Man*, A. S. Blix, L. Walløe, and Ø. Ulltang (eds), 27–33. Amsterdam: Elsevier Science B.V.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, Series B* **39,** 172–212.

Schweder, T., Skaug, H. J., Langaas, M., and Dimakos, X. K. (1999). Simulated likelihood methods for complex double-platform line transect surveys. *Biometrics* **55,** 678–687.

Appendix

Proof of equations (6) *and* (8)

To prove equation (6), we will need the following results

$$\int_{-\infty}^{\infty} p_c\, dc = \sigma\sqrt{2\pi},$$
(A.1)

$$\int_{-\infty}^{\infty} p_c^2\, dc = \sigma^2\sqrt{\pi}/\sqrt{\sigma^2 + \rho^2},$$
(A.2)

which are easily shown by using the fact that the normal distribution density integrates to 1. Further, if Z is Poisson distributed with mean m,

$$\mathrm{E}\{Z(Z - 1)\} = m^2.$$
(A.3)

Consider first clusters with centers at $x = c$ within the range $(-C, C)$. Then the intensity of the one-dimensional parent process is $\lambda_1 = 2C\lambda$.

Parent events follow a Poisson process, so within $(-C, C)$, c is uniformly distributed. Thus, by conditioning on c,

$$\mathrm{E}(T) = \mathrm{E}(g_0\mu p_c) = g_0\mu \int_{-C}^{C} \frac{1}{2C}p_c\, dc,$$
(A.4)

and from (A.4)

$$\mathrm{E}\{T(T-1)\} = \mathrm{E}\big(g_0^2\mu^2 p_c^2\big) = g_0^2\mu^2 \int_{-C}^{C} \frac{1}{2C} p_c^2 \, dc. \quad \text{(A.5)}$$

This gives

$$\frac{\mathrm{E}\{T(T-1)\}}{\lambda_1\{\mathrm{E}(T)\}^2} = \frac{\displaystyle\int_{-C}^{C} p_c^2 \, dc}{\lambda \left(\displaystyle\int_{-C}^{C} p_c \, dc\right)^2}. \quad \text{(A.6)}$$

Letting $C \to \infty$, (6) is then obtained from (4), (5) (A.1) and (A.2).

To prove (8), first restrict attention to clusters within $(-C, C)$. Then the expected number of detected points on a transect line of length L

$$\mathrm{E}(n) = L2C\lambda\mathrm{E}(T) = L2C\lambda g_0\mu \int_{-C}^{C} \frac{1}{2C} p_c \, dc. \quad \text{(A.7)}$$

Letting C go to infinity and solving for μ gives (8).

The author replied as follows:

I would like to thank the authors for correcting the error in my paper.

Ann Cowling
Statistical Consulting Unit
Australian National University
John Dedman Building
ACT 0200, Australia
email: ann.cowling@an.edu.au