

## ESTIMATING SETTLEMENT TIME

ANDREW R. SOLOW

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 USA

**ABSTRACT.** This note describes point and interval estimation of a settlement time from a collection of radiocarbon dates. An application concerning the settlement time of New Zealand is presented.

### INTRODUCTION

Radiocarbon dating is sometimes used to draw inferences about a settlement or other origination time for a human population. For example, Higham and Hogg (1997) cited as evidence of a recent settlement time for New Zealand the fact that no calibrated  $^{14}\text{C}$  date among a collection of archaeological samples predated AD 1250. In general, the information that a collection of  $^{14}\text{C}$  dates provides about a settlement time depends on factors including their number and precision. The purpose of this note is to describe and illustrate some statistical methods for this problem.

### MODEL AND METHODS

Let the random variables  $Y_1, Y_2, \dots, Y_n$  be the calibrated  $^{14}\text{C}$  ages of a collection of  $n$  independent objects, with  $0 \leq Y_i \leq \infty$ . The basic model considered in this note is

$$Y_i = X_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad (1)$$

where the random variable  $X_i$  is the true age of the object and  $\varepsilon_i$  is a normally distributed dating error independent of  $X_i$  with mean 0 and known or well-estimated variance  $\sigma_i^2$ . The randomness in  $X_i$  reflects the process by which objects are sampled for dating. Strictly speaking, it is possible under this model for  $Y_i < 0$ . However, provided  $X_i$  is large in relation to  $\sigma_i$ , this possibility can be discounted.

This note focuses on the special case in which  $X_i$  is uniformly distributed over the interval  $(\beta, \gamma)$ , with  $0 \leq \beta < \gamma < \infty$ . Although both bounds  $\beta$  and  $\gamma$  are unknown, interest centers here on inference about  $\gamma$ , which is assumed to correspond to the settlement time. The uniform model is appropriate when the sampling rate of objects is independent of age (an alternative model is discussed briefly in the final section). It is straightforward to show that, under the uniform model, the probability density function of  $Y_i$  is

$$g(y; \beta, \gamma) = (\Phi((y-\beta) / \sigma_i) - \Phi((y-\gamma) / \sigma_i)) / (\gamma-\beta) \quad (2)$$

where  $\Phi$  is the standard normal distribution function.

A point estimate of  $\gamma$  can be found by the method of maximum likelihood (*e.g.*, Silvey 1975). For the collection of  $n$  independent objects, the log likelihood function is

$$L(\beta, \gamma) = \sum_{i=1}^n \log g(y_i; \beta, \gamma) \quad (3)$$

where  $y_i$  is the observed value of  $Y_i$ . The maximum likelihood estimates of  $\beta$  and  $\gamma$  are found by maximizing  $L(\beta, \gamma)$  over  $\beta$  and  $\gamma$  jointly. This maximization must be done numerically.

In many situations, it is important to go beyond point estimation to provide a confidence interval for  $\gamma$ . A simple way to do this is through the profile likelihood function (Barndorff-Nielsen 1991). The profile log likelihood function for  $\gamma$  is defined as

$$L(\gamma) = L(\gamma, \beta^*(\gamma)) \quad (4)$$

where  $\beta^*(\gamma)$  is the maximum likelihood estimate of  $\beta$  for  $\gamma$  fixed. The profile log likelihood function shares many of the features of the ordinary likelihood function. For example, the estimate of  $\gamma$  found by maximizing  $L(\gamma)$  is also the ordinary maximum likelihood estimate of  $\gamma$ . Also, an approximate 1- $\alpha$  confidence interval for  $\gamma$  consists of the set of values of  $\gamma_0$  that satisfy

$$L(\gamma^*) - L(\gamma_0) < \chi_1^2(\alpha) / 2 \quad (5)$$

where  $\gamma^*$  is the maximum likelihood estimate of  $\gamma$  and  $\chi_1^2(\alpha)$  is the upper  $\alpha$ -quantile of the chi-squared distribution with one degree of freedom. For example, to construct a 0.95 confidence interval for  $\gamma$ , the value  $\chi_1^2(0.05) = 3.84$  is used.

Interestingly, the standard likelihood theory does not apply under the uniform model when age is measured without error (David 1981). The problem arises because the support of  $Y_i$  depends on the unknown parameters. The introduction of normal dating error eliminates this problem.

TABLE 1. Approximate calibrated age (Y) and error standard deviation ( $\sigma$ ) of shell samples from New Zealand (extracted from Figure 3 of Higham and Hogg (1997)).

Sample	Y (yr BP)	$\sigma$ (yr)
Wk-2282	670	30
Wk-2768	670	30
Wk-2770	620	40
Wk-1083	530	25
Wk-2502	565	55
Wk-1515	595	50
Wk-2515	610	50
Wk-1167	580	35
Wk-1737	610	40
Wk-2751	565	45
Wk-2410	580	50
Wk-2411	595	50
Wk-2412	635	30
Wk-2362	580	45
Wk-2508	640	25
Wk-2752	565	45
Wk-2856	575	40
Wk-2632	575	40
Wk-2857	560	45
Wk-2440	625	30
Wk-2441	645	25
Wk-2851	570	40

## APPLICATION

In this section, the methods outlined in the previous section are applied to some data extracted from Higham and Hogg (1997). These data represent calibrated ages  $y_i$  and dating error standard deviations  $\sigma_i$  of 22 samples of shell found at archaeological sites in and around New Zealand. The data, which are given in Table 1, were estimated from Figure 3 of the original paper and therefore are approximate. The profile log likelihood function  $L(\gamma)$ , (Fig. 1), reaches a maximum of  $-113.26$  at  $\gamma^* = 651$  yr BP. The corresponding maximum likelihood estimate of  $\beta$  is 553 yr BP. From (5), an approximate 0.95 confidence interval is 617–687 yr BP. As shown in Figure 1, this interval contains all values of  $\gamma$  for which  $L(\gamma) > -115.18$ . These results confirm the conclusion of Higham and Hogg (1997) that the data support the “late settlement” model with a settlement time sometime after 1250 AD.

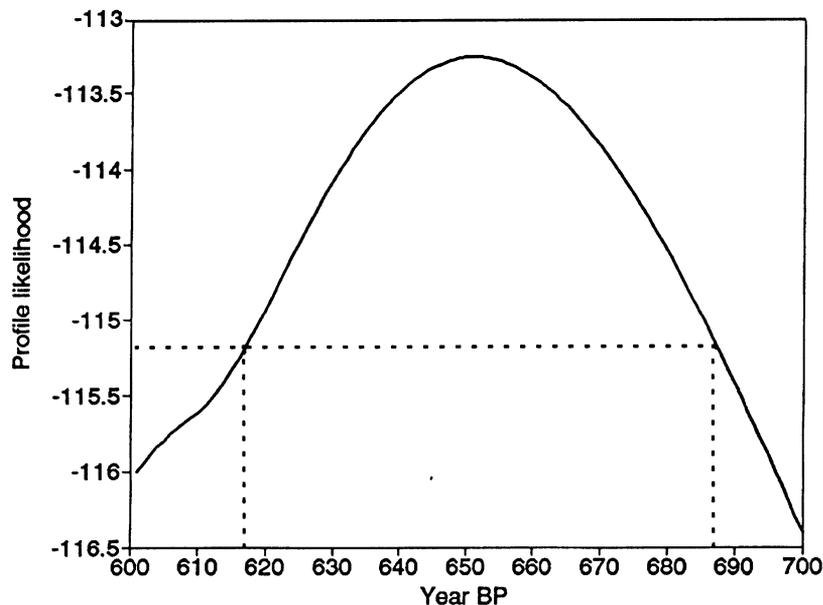


Fig. 1. The profile log-likelihood function for  $\gamma$ . The 0.95 confidence interval for  $\gamma$  consists of the values contained within the vertical projections of the dotted lines onto the horizontal axis.

## DISCUSSION

One key assumption underlying the methods described in this note is that the samples are independent of each other. This assumption may be violated when the collection contains clusters of samples from a few sites. For example, of the 22 samples in Table 1, 9 were collected at a site at Shag Mouth. In this case, the variability in the estimated ages among the Shag Mouth samples is not noticeably smaller than the variability between estimated ages from different sites, so the evidence for dependence is not strong. In general, this source of dependence could be eliminated by retaining only a single sample from each cluster.

A second key assumption is that the true ages are uniformly distributed over an interval whose upper bound corresponds to settlement time. This assumption would be violated if either the production rate or preservation rate of objects varies with age. If the density of true ages declines with age, then adopting the uniform model will lead to underestimation of  $\gamma$ . This problem may be avoided by a judicious screening of objects. For example, Higham and Hogg (1997) selected only shell samples

from archaeological sites predating 930 BP. It is also possible to adopt a nonuniform model for true age. A flexible model of this type uses in place of (2) the probability density function

$$f(x; \beta, \gamma, \lambda_1, \lambda_2) \propto x^{\lambda_1} (1-x)^{\lambda_2} \quad (6)$$

with  $\lambda_1, \lambda_2 > -1$  and where  $x' = (x-\beta)/(\gamma-\beta)$ . While inference about  $\gamma$  under this model, which can exhibit a wide range of behavior, can again be based on the profile likelihood,  $g(y_i; \beta, \gamma, \lambda_1, \lambda_2)$  would have to be evaluated numerically.

#### ACKNOWLEDGMENTS

Helpful comments from T. Higham, D. Sewell, and an anonymous reviewer are acknowledged with gratitude.

#### REFERENCES

- Barndorff-Nielsen, O. E. 1991 Likelihood theory. In Hinkley, D. V., Reid, N. and Snell, E. J., eds., *Statistical Theory and Modelling*. London, Chapman and Hall: 232–264.
- David, H. A. 1981 *Order Statistics*. New York, Wiley: 360 p.
- Higham, T. F. G. and Hogg, A. G. 1997 Evidence for late Polynesian colonization of New Zealand: University of Waikato radiocarbon measurements. *Radiocarbon* 39(2): 149–192.
- Silvey, S. D. 1975 *Statistical Inference*. London, Chapman and Hall: 158 p.