A STATISTICAL TECHNIQUE FOR INTEGRATING C-14 DATES WITH OTHER FORMS OF DATING EVIDENCE

Clive Orton

Institute of Archaeology, 31-34 Gordon Square, London, W.C.1.

Introduction

Although the C-14 dating method is widely used in archaeology, little theoretical work has been done on the problem of using dates obtained by this method with other sorts of dating evidence, e.g. pottery or stratigraphy. There is a problem because C-14 dates are absolute, usually expressed as a point estimate and a standard deviation, while many other forms of dating are relative. For example, stratigraphic relationships can be expressed in terms of a lattice and pottery chronologies may be established by formal or informal seriation techniques. When date ranges can be attached to pottery types, they usually appear as (implicitly) Uniform distributions, often with rather 'fuzzy' end points. We need to be able to relate these different sorts of evidence to each other in a coherent way in order to gain the greatest amount of information from each, and to be able to attach absolute dates to relative sequences.

In looking at this problem, the question of calibration will not be discussed here, for reasons of space. However, it is discussed in both the excavation report for which this work was done (Cunliffe, forthcoming) and the archival report (copies of which may be obtained from the author on request at cost of reproduction).

The data

This work is based on 70 C-14 dates from the 1969-78 excavations at the Iron Age hillfort of Danebury, Hants. (Cunliffe, forthcoming), of which 51 are from wood charcoal, nine from grain and ten from animal bone. Five samples appeared to be contaminated and were not used, reducing the number of samples to 65. Statistical considerations reduced the number finally used to 54 (see below).

The dates obtained from these samples were examined in relationship to the pottery phases with which they were associated. It was hoped that the pottery styles observed would correspond to chronological divisions of activity on the site ('ceramic phases'), and that the C-14 dates would support these divisions. Because of the large variation in the numbers of samples relating to different ceramic phases, the phases were grouped together as cp 1-3, cp 4-5, cp 6, and cp 7. The statistical problem is to assess the chronological distinctness of these supposed phases and to put absolute dates to them if the evidence warrants it.

Statistical model

It was felt to be useful to develop a model which could have general application. In this model there are k archaeological (e.g. ceramic) phases; the kth runs from date a_k to a_{k+1}. The question of whether these dates should be calendar or C-14 dates is discussed elsewhere (see above). The number of dates available from the jth phase is n_j, and the cumulative number from the first j phases is m_j. The dates are represented by x_i, where $i = 1, 2, \ldots, n = m_k$, with those from the first phase listed first ($i = 1, 2, \ldots, m_1$), those from the second phase next ($i = m_1 + 1, \ldots, m_2$) and so on. We assume that each x_i has a Normal distribution $N(\mu_i, \sigma_i)$, where μ_i is unknown and σ_i is the quoted standard deviation, and that the μ_i have Uniform distributions on the appropriate date ranges $(a_j, a_{j+1}]$. This last assumption is arbitrary and needs to be critically examined in any application. In particular, it is disturbed by non-linearity in the calibration function if we are working in C-14 years. However, it provides a useful starting point for the development of a statistical approach to the problem.

Development

The approach is to estimate the 'frontier' dates a_j between phases by the method of maximum likelihood estimation (mle), and to use likelihood ratio (lr) tests to examine the validity of these frontiers against a more general hypothesis of overlapping frontiers.

If the distribution of x is given by $f_i(x_i | a_j, a_{j+1}, \sigma_i)$, where j is such that $m_{j-i} < i \leq m_j$, then the overall likelihood function is

$$L(a) = \prod_{j=1}^{k} g_j,$$

where $g_j = \prod_i f_i(x_i | a_j, a_{j+i}, \sigma_i),\ i = m_{j-1} + 1, \ldots, m_j$ (1)

The ml estimates of the a_j are the values which maximise L or $\log L$,

i.e. solutions to $\partial L / \partial a_j = 0$,

or $\partial \log L / \partial a_j = 0,\ j = 1, 1, \ldots, k + 1$ (2)

To find $L(a)$, we write $x_i = \mu_i + z_i$,

where z_i has a distribution $N(0, \sigma_i)$

and μ_i has a Uniform distribution on $(a_j, a_{j+1}]$.

Then the d.f. of x_i is given by

$$f_i(x_i) = \int_{-\infty}^{\infty} f_z(z)\, f_\mu(x_i - z)\,dz$$ (Lindgren, 1960, 137),

which becomes on substitution

$$f_i(x_i) = (\Phi((x_i - a_j)/\sigma_i) - \Phi((x_i - a_{j+1})/\sigma_i))/(a_{j+1} - a_j),$$

where $\Phi(y) = \int_{-\infty}^{y} (1/\sqrt{2\pi}) \exp(-t^2/2)\,dt$ (3)

Substituting this expression in equation (1) gives

$$\log L = -\sum_{j=1}^{k} n_j \log (a_{j+1} - a_j) + \sum_j \sum_i \log (\Phi((x_i - a_j)/\sigma_i) - \Phi((x_i - a_{j+1})/\sigma_i)). \quad (4)$$

The problem becomes that of maximising $\log L$ as given in equation (4) over the space defined by $(a_1, a_2, \ldots, a_{k+1})$, where $a_1 < a_2 < \ldots < a_{k+1}$.

Equation (4) is an example of the classical optimisation problem in $k + 1$ dimensions, with constraints. Here it is solved not directly, by use of a standard technique (e.g. steepest-ascent, Newton's method, Davidon's method) but indirectly using two simpler approaches, one to reduce the dimensionality of the problem and the other a new technique.

Reduction of dimension

The idea is to work in only one or two dimensions, first by estimating a_1 and a_{k+1} from all the data, and secondly by estimating each a_j $(1<j\leq k)$ in turn by putting all the data from the first j phases into a dummy 'phase 1' and those from the last $k - j$ phases into a dummy 'phase 2'. Both require a further critical look at the assumption of Uniform distributions because they change the definition of 'phase'. We proceed as follows:

set $k = 1$ in equation (4), leading to

$$\log L = -n \log (a_2 - a_1) + \sum_{i=1}^{n} \log (\Phi((x_i - a_1)/\sigma_i) - \Phi((x_i - a_2)/\sigma_i)), \quad (5)$$

which on differentiation leads to

$$n/(a_2 - a_1) - \sum_{i=1}^{n} \frac{1/(\sigma_i\sqrt{2\pi})\exp(-(x_i - a_1)^2/2\sigma_i^2)}{\Phi((x_i - a_1)/\sigma_i) - \Phi((x_i - a_2)/\sigma_i)} = 0$$

and

$$n\ (a_2 - a_1) - \sum^{n} \frac{1/(\sigma_i\sqrt{2\pi})\exp(-(x_i - a_1)^2/2\sigma_i^2)}{\Phi((x_i - a_2)/\sigma_i) - \Phi((x_i - a_2)/\sigma_i)} = 0, \quad (6)$$

the optimisation problem in two dimensions, which can readily be solved by one of the standard techniques, e.g. steepest-ascent.

Substituting the values of a_1 and a_{k+1} so obtained into equation (4), we can estimate a_{r+1} $(r = 1, 2, \ldots, k-1)$ by writing

$$\log L = - m_r \log (a_{r+1} - a_1) - (n - m_j) \log (a_{k+1} - a_{r+1})$$

$$+ \sum_{i=1}^{m_r} \log \left(\Phi((x_i - a_1)/\sigma_i) - \Phi((x_i - a_{r+1})/\sigma_i) \right)$$
$$+ \sum_{i>mr} \log \left(\Phi((x_i - a_{r+1})/\sigma_i) - \Phi((x_i - a_{k+1})/\sigma_i) \right). \quad (7)$$

Differentiating and substituting a for a_{r+1} gives

$$((n - m_r)/(a_{k+1} - a)) - (mr/(a - a_1))$$
$$+ \sum_{i=1}^{m_r} 1/(\sigma_i \sqrt{2\pi}) \exp(-(x_i - a)^2/2\sigma_i^2)/(\Phi((x_i - a_1)/\sigma_i) - \Phi((x_i - a)/\sigma_i))$$
$$- \sum_{i>mr} 1/(\sigma_i \sqrt{2\pi}) \exp(-(x_i - a)^2/2\sigma_i^2)/(\Phi((x_i - a)/\sigma_i) - ((x_i - a_{k+1})/\sigma_i)) = 0, \quad (8)$$

which can be solved by a standard technique, e.g. the secant method.

The validity of the archaeological phases can be tested by the following procedure:

(i) estimate the opening and closing dates of each phase independently, using equation (6), and calculate the likelihood L_u, which has $2k$ degrees of freedom,

(ii) estimate the overall opening and closing dates, again using equation (6),

(iii) estimate the 'frontier' dates between each successive pair of phases, using equation (8), and calculate the likelihood L_c, which has $k + 1$ degrees of freedom,

(iv) calculate the likelihood ratio statistic $l = L_c/L_u$.

Under the null hypothesis that the phases do not overlap, $-2\log l$ has approximately a chi-squared distribution with $2k - (k + 1) = k-1$ degrees of freedom, so the test statistic is

$$-2\log(L_c/L_u) = 2(\log L_u - \log L_c).$$

Alternative method

A question raised in discussion - 'given the length of a phase, a typical standard deviation and an assumed Uniform distribution, what proportion of observed dates can be expected to lie outside the true date range of a phase?' - led to the exploration of an alternative technique for estimating opening and closing dates of phases.

The dates are first standardised so that $x = \mu + z$ (omitting subscripts), where μ has a Uniform distribution on $(0, l]$ and z has the distribution N (0, 1), i.e. l is the standardised length of the phase.

We require the probability $p = \text{pr}(x < 0 \text{ or} > l)$,

which by integration can be shown to be

$$p = 2/l \int_{-l}^{0} \Phi(b)\,db. \quad (9)$$

Approximate values of p for various values of l are shown in Table 1. It can be seen that for relatively short phases (1-2 standard deviations in length), about half of the observed dates can be expected to lie outside the true range of their phase.

This approach gives a second way of estimating the opening and closing dates of a phase. Opening and closing dates are initially guessed, and the standardised length and expected proportion of dates outside their range are calculated. If too many dates lie outside their range it is lengthened, and if too few, it is shortened. Successive approximation usually leads to a satisfactory outcome within four or five iterations.

Unfortunately, this technique does not always give a unique solution, and must therefore be used with caution. However, it can provide useful starting points for an iterative solution to equation (6), and a useful check on such a solution. It is here called the poor (proportion outside of range) method of estimation.

The program

The program DANEBURY was written in PASCAL (as defined by Grogono, 1980) for use on the Cambridge University Computing Service's IBM 370/165, and has since been transferred to the Service's new IBM 3081. It can operate in three 'modes' - A (arbitrary), R (restricted) or U (unrestricted), which by invoking different combinations of procedures can carry out all the tasks described in Development above. Input may consist of up to 100 dates and their standard deviations, assigned to up to eight phases: these limits could easily be increased if need be.

The program is small and, on the Cambridge machines, reasonably fast. The compiled program code occupies about 9K of storage and about 13K of working space is needed. The CPU time needed is roughly proportional to the product of the number of phases and the number of dates, but varies according to the mode used. As an illustration, the CPU times for the standard Danebury dataset (4 phases, 54 dates) on the 370/165 average 0.2s ('A' mode), 0.5s ('R' mode) and 2.2s ('U' mode), of which perhaps 0.1s should be allowed for overheads.

A listing of the program, and instructions for its use, are available (at cost of reproduction) from the author. The program does have weaknesses: in particular, the optimisation procedure is very crude and may encounter difficulties with local extrema. It would benefit greatly from a proper multi-dimensional constrained optimisation procedure.

Initial examination of the data

The 65 dates supplied were critically examined before use with the program DANEBURY, and the number of usable dates reduced to 54, as follows:

(i) two pairs of check measurements were made on the same charcoal samples: since both were consistent they have been replaced by their respective weighted means, and their standard deviations recalculated. The sample was thus reduced to 63 dates,

(ii) nine pairs of measurements were made on bone and charcoal, one on grain and bone, and two on charcoal and grain, taken from the same archaeological contexts. Statistical comparisons suggest that two of the measurements on bone are unreliable, and these have been omitted,

(iii) nineteen samples were obtained from a sequence of five stratigraphic phases (phases l, k, i, h and a-e). examination of the consistency of dating both within and between phases leads to the rejection of all four dates from phase k, and their replacement by a single interpolated date,

(iv) four further dates could not be related to a ceramic phase.

Thus comparisons between materials, and stratigraphic considerations (details are given in the archival report) reduce the usable sample from 65 to 54 dates: 9 each in cp 1-3 and cp 4-5, 12 in cp 6 and 24 in cp 7. These dates are summarised in Table 2.

Comparison of C-14 dates and ceramic phases

Before the opening and closing dates of the phases were estimated, a check was made on the chronological distinctness of the four ceramic phases. The k-sample generalisation of the Wilcoxon-Mann-Whitney test (Kendall and Stuart, 1973, 522-4) gave a value of z of about 5.2, confirming both the distinctness and the order of the phases.

Straightforward application of the mle technique described above gave the results presented in Table 3. All results are rounded to the nearest five years. One date (630 bc) in cp 4-5 appeared to be an outlier, and analyses were repeated with this value omitted.

Comparison of Tables 2 and 3 shows that the overall sequence is dominated by the high density of dates in the later part, and that the assumption of uniformity over the whole range does not hold. It might therefore be better to use cp 1-3 by itself to give an opening date for the whole sequence (i.e. 550 bc as against 470 or 440 bc).

Use of the poor method of estimation leads to the results shown in Table 4. The solution for cp 7 is not unique. The method supports an early (6th century bc) opening date, while agreeing broadly with the ml estimate of the closing date. Fortunately, variations in the overall opening date make relatively little difference to the frontier dates between the phases.

The 'preferred' dates shown in Table 5 are based on an assessment of the outcomes of both methods. They give a very good fit to the 'poor' model without diverging greatly from the mle results. The estimated frontiers of the phases have been added to Table 2, to show which dates lie within their respective ranges.

There is an apparent discrepancy between the two methods, in that the poor method shows a good fit between the suggested phases and the poor model, while the mle method gives unacceptably high values for chi-squared (10.6, or 8.6 if the outlier is omitted, on three degrees of freedom). In other words, although about the expected numbers of dates lie outside their ranges, they tend to lie further outside than one would expect (best seen on Table 2). An archaeological explanation can be offered: the long tails of late dates for pits of cp1-3 and cp 4-5 date might be explained by their including only residual earlier ceramic material (i.e. these pits are contemporary with later ceramic phases but contain no contemporaneous ceramic material). The opposite effect, a tailback of early dates in demonstrably later contexts, is to be expected in the later ceramic phases in which old timbers were probably being extensively re-used.

The distinctness of cp 6 and cp 7 was questioned, because the pottery assemblage of cp 7 differs from that of cp 6 only in that it includes decorated vessels. Thus if an assemblage of cp 7 date was without decorated types it would be classified as cp 6. It could be argued that the difference between cp 6 and cp 7 is not chronological, but perhaps represents function or even size of assemblage (some assemblages might lack decorated types simply because they are small). To test this, the frontiers were re-estimated allowing cp 6 and cp 7 to overlap. There was a marginal increase in likelihood, but the lr statistic was not significant (chi-squared = 0.6 on one degree of freedom), suggesting that the phases are chronologically distinct, although the late (2nd century bc) cp 6 assemblages might really belong to cp 7.

Bearing in mind the limitations imposed by the nature of archaeological material, the C-14 dates do seem to confirm the excavator's assignment of the pottery to chronological phases. The two techniques described, when used together, give a useful way of estimating the frontier dates between such phases.

Discussion

Despite the degree of overlap between the dates from the different ceramic phases, it is possible to disentangle them using a technique based on mle. It is more difficult to estimate overall opening and closing dates, but a combination of mle and a new technique, poor estimation, gives reasonable results. The amount of overlapping between the phases can be theoretically predicted as a function of the ratio of range to standard deviation. The high proportions of dates that can be expected to lie outside the range of their phase (Table 1) have serious implications for sites with few dates from each phase.

It is not clear whether it is better to carry out these calculations on uncalibrated dates and to calibrate the results, or calibrate each date first and then perform the calculations. Where the calibration curve is reasonably straight the difference will be negligible, but where it has 'kinks' both approaches run into difficulty. Non-linearity disrupts the assumption of a Uniform distribution of C-14 dates (on the a priori more reasonable assumption of a Uniform distribution of calendar dates) and so undermines the theoretical basis of the methods used. The robustness of the techniques to such disturbances needs to be examined. On the other hand, calibrating first gives us an arbitrary choice of three calendar dates for each of a subset of the C-14 dates. Also, calibration at this stage can introduce serious errors due to 'near misses' at turning points in the calibration curve. On balance, it seems preferable to carry out the calculations on uncalibrated dates.

Given a generally agreed calibration curve, it would be possible to replace the assumption of a Uniform distribution of dates in a phase with a distribution reflecting the 'bunching' caused by kinks in the curve. The mathematics would be correspondingly more difficult, but should not be insuperable. There is clearly much scope for improvement in the techniques presented here, which nevertheless seem to be a useful first step towards solving an important problem.

Acknowledgements

Thanks are due to Professor B. W. Cunliffe for presenting the data and encouraging me to undertake this project, which was mainly done while working on S.E.R.C. Grant GR/B 09377. The presentation of this report has benefitted from comments made at seminars at the British Museum Research Laboratory and the Institute of Archaeology, London.

References

Cunliffe, B. W. forthcoming — AN IRON AGE HILLFORT IN HAMPSHIRE. C.B.A. Res. Rep.

Grogono, P. 1980 — PROGRAMMING IN PASCAL, Addison-Wesley.

Kendall, M. G. and Stuart, A. 1973 — THE ADVANCED THEORY OF STATISTICS, Vol. 2 (third edition), Griffin.

Lindgren, B. W. 1960 — STATISTICAL THEORY, Macmillan.

	range/s.d.	expected percentage of dates outside range
calculated values	1 2 3	60 40 25
extrapolated values	4 6 8	20 12½ 10

Table 1: expected percentages of observed dates lying outside the true date range of a phase, for selected values of the ratio of range to standard deviation of dates.

years bc	cp 1-3	cp 4-5	cp 6	cp 7	all
639-620		1			1
-600					
-580	1				1
-560					
-540					
-520					
-500	2				2
-480					-
-460	2				2
-440					-
-420			1	1	2
-400					-
-380	1	2	1		4
-360					-
-340		2	1	2	5
-320	1	1			2
-300				1	1
-280			1		1
-260	1		4		5
-240				1	1
-220		1		2	3
-200	1			2	3
-180		1		2	3
-160		1	1	3	5
-140			2	2	4
-120				1	1
-100			1	2	3
-80				2	2
-60					
-40				2	2
-20				1	1
total	9	9	12	24	54

Table 2: numbers of dates lying in 20-year spans, showing proposed frontiers between phases.

all dates	restricted	unrestricted	χ^2	d.f.
cp 1-3 cp 4-5 cp 6 cp 7	470-370 370-295 295-220 220-80	550-240 460-240 370-140 300-75	10.6	3
outlier removed				
cp 1-3 cp 4-5 cp 6 cp 7	440-350 350-300 300-225 225-80	550-240 350-225 370-140 300-75	8.6	3

Table 3: restricted and unrestricted mle dates for each phase, and lr tests, with and without the 'outlying' date (all dates bc).

phase	dates bc
cp 1-3 cp 4-5 cp 6 cp 7	520-370 390-310 390-140 350-50 or 240-100
overall	520-100 or 50

Table 4: unrestricted dates for each phase, using the 'poor' technique.

phase	date range (bc)	range	s.d.	number of dates	no. outside of range	
					observed	expected
cp 1-3	520-360	160	75	9	4	3.4
cp 4-5	360-290	70	80	9	6	5.7
cp 6	290-210	80	75	12	7	7.1
cp 7	210-80	130	70	24	10.5	10.3
total	-	-	-	54	27.5	26.5

Table 5: poor estimate of opening date of overall sequence, combined with mle dates for end of sequence and frontiers; also showing expected and observed numbers of dates lying outside each date range.