

TREATMENT OF DISCREPANCIES IN RADIOCARBON DATING

C H VINCENT

Building A70.2, Ministry of Defence, AWRE, Aldermaston
Berkshire, England

ABSTRACT. A method is given for modifying the estimated radiocarbon dates of two samples when they are known to be in the wrong order.

A radiocarbon date expressed in the form $1720 \text{ BP} \pm 80$, for example, implies that the best estimate of the date is 1720 BP and that the uncertainty in the measurement is estimated to be given by a Gaussian error distribution with a standard deviation of 80 years (Orton, 1980). Orton has pointed out, however, that a particular difficulty arises when two samples have been dated in this way and one has been found to be earlier than the other, when there are reliable archaeological grounds, such as stratigraphy, for taking it to be the later. Such an anomaly can, of course, be clearly detected only when the times concerned lie within a monotonic region of the radiocarbon calibration curve. This paper provides a method of dealing with the situation, when the order of the radiocarbon dates is taken to be incorrect. Similar difficulties could arise with other physical methods of dating with results expressed in the same way. All the times referred to in the calculations are radiocarbon dates (BP). The problem is closely related to that of the estimation of differential counting measurements of positive quantities discussed by Vincent (1982, 1984; Vincent & Vincent, 1984). However, it is now required to estimate each date, rather than merely the difference. Other relevant publications are (Katz, 1963; Blumenthal & Cohen, 1968).

The method used is based on Bayes' (1763) principle. Since its strong advocacy by Jeffreys (1948), Bayes' method has been the subject of much controversy, but its soundness has received increasing recognition by many statisticians in recent years (eg, Lindley, 1965; Berger, 1980). As the method has been widely described in the literature, familiarity with its basic principles and terminology will be assumed here. Let t_1 and t_2 be the two dates and let the stratigraphic information be that $t_2 > t_1$. The theory that follows will also apply and will modify the estimated dates, when the radiocarbon order agrees with this but there is an appreciable overlap between the two Gaussian distributions.

Since t_1 and t_2 must both be positive, we assume an initial prior distribution that is uniform over the region $0 < t_1 < \infty$, $0 < t_2 < \infty$. When the information obtained from the radiocarbon measurements is applied by Bayes' method, the Gaussian error distributions assumed provide a likelihood proportional to

$$\exp \left\{ -\frac{(t_1 - t'_1)^2}{2\sigma_1^2} - \frac{(t_2 - t'_2)^2}{2\sigma_2^2} \right\} \quad (1)$$

where t'_1 and t'_2 are the measured values for t_1 and t_2 , respectively, and σ_1 and σ_2 are the corresponding standard deviations. t'_2 may be greater or less than t'_1 .

For the uniform prior distribution assumed, the resulting posterior

distribution is also proportional to this expression. This posterior distribution then becomes the prior distribution needed to apply the stratigraphic information. This information is of a particularly simple type in which all (t_1, t_2) values belong to one or other of two regions, having likelihoods of 0 and 1, respectively. To take such information into account, it is only necessary to omit the region of likelihood 0 from the integrals required for calculating the posterior expected value and variance, with the uniform likelihood of 1 in the other region (Vincent, 1973; Berger, 1980).

The expected value of t_1 is then

$$\bar{t}_1 = \frac{\int_0^\infty \int_0^{t_2} t_1 \exp \left\{ -\frac{(t_1 - t'_1)^2}{2\sigma_1^2} - \frac{(t_2 - t'_2)^2}{2\sigma_2^2} \right\} dt_1 dt_2}{\int_0^\infty \int_0^{t_2} \exp \left\{ -\frac{(t_1 - t'_1)^2}{2\sigma_1^2} - \frac{(t_2 - t'_2)^2}{2\sigma_2^2} \right\} dt_1 dt_2}. \quad (2)$$

Approximating on the assumptions that $t'_1 \gg \sigma_1$ and $t'_2 \gg \sigma_2$ the lower limits of integration may be extended to $-\infty$, as a very good approximation, and we then have

$$\bar{t}_1 = t'_1 - \frac{\frac{\sigma_1 \sigma_0}{\sqrt{2\pi} \sigma_2} \exp \left\{ -\frac{(t'_2 - t'_1)^2}{2(\sigma_1^2 + \sigma_2^2)} \right\}}{\mathbf{I} \left[\frac{t'_2 - t'_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right]} \quad (3)$$

where

$$\sigma_0^{-2} = \sigma_1^{-2} + \sigma_2^{-2} \quad (4)$$

and

$$\mathbf{I}(y) = \frac{\int_{-\infty}^y \exp(-x^2/2) dx}{\sqrt{2\pi}}. \quad (5)$$

The denominator in Eq (3) is the probability, based on the radiocarbon measurements, that $t_2 > t_1$. Those measurements give a Gaussian distribution for the difference $t_2 - t_1$ with mean $t'_2 - t'_1$ and standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2}$.

The value of \bar{t}_2 can be obtained similarly, by interchanging t_1, t'_1 and σ_1 with t_2, t'_2 and σ_2 , respectively, in Eq (2). The integral $\int_0^{t_2}$ must be replaced by $\int_{t_1}^\infty$. This gives

$$\bar{t}_2 = t'_2 + \frac{\frac{\sigma_2 \sigma_0}{\sqrt{2\pi} \sigma_1} \exp \left\{ -\frac{(t'_2 - t'_1)^2}{2(\sigma_1^2 + \sigma_2^2)} \right\}}{\mathbf{I} \left[\frac{t'_2 - t'_1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right]}. \quad (6)$$

We also have

$$\overline{t_2 - t_1} = t'_2 - t'_1 + \frac{\frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sqrt{2\pi}} \exp\left\{-\frac{(t'_2 - t'_1)^2}{2(\sigma_1^2 + \sigma_2^2)}\right\}}{I\left\{\frac{t'_2 - t'_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right\}} \quad (7)$$

which is in accordance with Eq (9) of Vincent (1982). It is shown in that paper that this value is always positive, as is to be expected. The function $I(y)$ can be obtained from any table of the normal error integral, such as the Biometrika Tables for Statisticians (Pearson & Hartley, 1966), or by numerical integration.

The modified distributions are no longer Gaussian, but standard deviations σ_{1m} and σ_{2m} can be calculated for them, if desired, as follows:

$$\sigma_{1m}^2 = \frac{\int_{-\infty}^{\infty} \exp\left\{-\frac{(t_2 - t'_2)^2}{2\sigma_2^2}\right\} \int_{-\infty}^{t_2} (t_1 - t'_1)^2 \exp\left\{-\frac{(t_1 - t'_1)^2}{2\sigma_1^2}\right\} dt_1 dt_2}{2\pi\sigma_1\sigma_2 I\left\{\frac{t'_2 - t'_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right\} - (\Delta t_1)^2 = \sigma_1^2 \left\{ \cos^2 \theta + V\left(\frac{t'_2 - t'_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \sin^2 \theta \right\}} \quad (8)$$

$$\text{where } \Delta t_1 = t'_1 - \bar{t}_1 \quad (9)$$

$$\theta = \tan^{-1}(\sigma_1/\sigma_2) \quad (10)$$

and $V(y)$ is the variance of the truncated Gaussian distribution $\exp(-x^2/2)/\sqrt{2\pi}$, given by

$$V(y) = \frac{\int_{-\infty}^y x^2 \exp(-x^2/2) dx}{\sqrt{2\pi}I(y)} - \left\{ \frac{\int_{-\infty}^y x \exp(-x^2/2) dx}{\sqrt{2\pi}I(y)} \right\}^2 = \frac{\int_{-\infty}^y x^2 \exp(-x^2/2) dx}{\sqrt{2\pi}I(y)} - \left\{ \frac{\exp(-y^2/2)}{\sqrt{2\pi}I(y)} \right\}^2 \quad (11)$$

which can readily be obtained by numerical integration. Similarly,

$$\sigma_{2m}^2 = \sigma_2^2 \left\{ \sin^2 \theta + V\left(\frac{t'_2 - t'_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \cos^2 \theta \right\}. \quad (12)$$

The following examples were calculated using a simple computer program for the functions $I(y)$ and $V(y)$ and show typical results. The first is an example with a discrepancy in the order, while the second has agreement in the order but an appreciable overlap between the two Gaussian distributions. The same time interval and standard deviations have been chosen to facilitate comparison.

Radiocarbon measurement 1	$t'_1 \pm \sigma_1$	1900 BP \pm 85	1800 BP \pm 85
Radiocarbon measurement 2	$t'_2 \pm \sigma_2$	1800 BP \pm 70	1900 BP \pm 70
Stratigraphic indication	$t_2 > t_1$		
Expected value of t_1	\bar{t}_1	1805 BP	1779 BP
Expected value of t_2	\bar{t}_2	1865 BP	1914 BP
Modified standard deviation 1	σ_{1m}	62	74
Modified standard deviation 2	σ_{2m}	58	64

It can be seen from Eqs (3) and (6) that the changes in the two dates are proportional to their respective radiocarbon variances and are greater in the discrepancy case, for given σ_1 , σ_2 and $|t'_2 - t'_1|$. If the latter value is large compared with both standard deviations and there is agreement with the stratigraphic order, the changes in the values are small.

Because $V(y)$ is a monotonic increasing function of y with an upper limit of 1 at $y = \infty$, it follows that $\sigma_{1m} < \sigma_1$ and $\sigma_{2m} < \sigma_2$. It also follows that the modified standard deviations are smaller for the discrepancy case, for given σ_1 , σ_2 and $|t'_2 - t'_1|$.

The validity of all these conclusions depends, of course, on the correctness of the assumptions on which they are based. On these assumptions, large discrepancies with the probability $I(y) \ll 1$ for

$$y = (t'_2 - t'_1) / \sqrt{\sigma_1^2 + \sigma_2^2} \quad (13)$$

must be rare. When such a discrepancy occurs, therefore, it is necessary to keep in mind the possibility that some systematic error may have occurred (in addition to the random Gaussian errors on which this analysis is based) thus invalidating the results. This situation must be judged in relation to both the value of $I(y)$ and the nature and circumstances of the two samples measured. The smaller the value of $I(y)$, the greater is the confidence that is needed in the suitability of the samples and in the freedom of the measurements from any sources of systematic error, such as contamination of a sample.

REFERENCES

- Bayes, T, 1763, An essay towards solving a problem in the doctrine of chances: *Philos Trans*, v 53, p 370.
- Berger, J O, 1980, *Statistical decision theory*: New York, Springer, p 101.
- Blumenthal, S and Cohen, A, 1968, Estimation of the larger translation parameter: *Annals Mathematical Statistics*, v 39, p 502–516.
- Jeffreys, H, 1948, *Theory of probability*, 2nd ed: Oxford, Oxford Univ Press.
- Katz, M W, 1963, Estimating ordered probabilities: *Annals Mathematical Statistics*, v 34, p 967–972.
- Lindley, D V, 1965, *Introduction to probability and statistics*: Cambridge, Cambridge Univ Press.
- Orton, C, 1980, *Mathematics in archaeology*: London, Collins, p 90–95.
- Pearson, E S and Hartley, H O, 1966, *Biometrika tables for statisticians*, vol 1, 3rd ed: Cambridge, Cambridge Univ Press.
- Vincent, C H, 1973, Random pulse trains—their measurement and statistical properties: London, Peter Peregrinus, p 13–23, 30–33.
- 1982, The estimation of differential counting measurements of positive quantities with relatively large statistical errors: *Nuclear Instruments & Methods*, v 203, p 307–315.
- 1984, Prior distributions for Bayesian estimations of positive quantities: *Nuclear Instruments & Methods*, v 219, p 259–261.
- Vincent, C H and Vincent, R C, 1984, Practical estimation methods for differential counting: *Nuclear Instruments & Methods*, v 223, p 147–149.