STATISTICAL PROBLEMS IN CALIBRATING RADIOCARBON DATES

HEROLD DEHLING

Mathematics Department, University of Groningen, Blauwborgje 3, 9747 AG, The Netherlands

and

JOHANNES VAN DER PLICHT

Centre for Isotope Research, University of Groningen, Nijenborgh 4, 9747 AG, The Netherlands

INTRODUCTION

The transformation of radiocarbon years to calendar years (cal AD/BC) is not straightforward because of past variations in atmospheric ^{14}C content (de Vries 1958; Suess 1970). A calibration curve, $y = f(x)$, transforms each dendrochronologically dated calendar age (x) to a ^{14}C date (y). By inverting this relationship, one can determine the calibrated calendar age of a given sample. In some time intervals, the calibration curve is problematic in that $f(x)$ is not uniquely invertible (Fig. 1); even an exact measurement of y cannot be converted to a single calendar age (see examples in van der Plicht & Mook (1987)).

The most widely distributed calibration programs are those of Seattle (Stuiver & Reimer 1986) and Groningen (van der Plicht & Mook 1989). Calibration procedures in both programs are essentially the same (Aitchison *et al.* 1989). Questions have been raised about the validity of mathematical procedures underlying the calibrations. Here we discuss problems in the statistical theory to derive correct expressions of the error limits of a calibrated ^{14}C date.

An important purpose of statistical estimation theory is to provide error limits for estimates given. A ^{14}C age, the result of repeated measurements, can be assigned a sample standard deviation, σ, centered around the arithmetic mean. The key question is how one should transform these error limits in the calibration process. The computer program developed in Groningen (van der Plicht & Mook 1989) uses the error function

$$\varepsilon(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\hat{y})^2}{2\sigma^2}} \tag{1}$$

where y is the true ^{14}C date, $\hat{y}$ is the measurement outcome (the mean value), σ is the standard deviation. The error curve for the ^{14}C age is transformed to an error curve for the calibrated age by $\varepsilon^*(x) = \varepsilon(f(x))$. This procedure differs from "classical probability" density transformations according to $\varepsilon^*(x) = |\,dy/dx\,|\,\varepsilon f(x)$.

Figure 2 illustrates problematic consequences of obtaining probability densities under a monotone transformation $y = f(x)$. Here, consider a piecewise monotone function, with a change in slope at $\hat{y}$ (the outcome of the measurement), as transformation function. Figure 2A shows the result of transforming $y \pm \sigma$ along the x-axis, using the Groningen program. Figure 2B shows the result of the classical transformation. The area under the Gaussian curve is symmetric with respect to the measurement outcome $\hat{y}$. For the calibrated distribution, this is also true for the example shown in Figure 2B, whereas this is not true for the example shown in Figure 2A. Although the two resulting distributions are different, one should not conclude that the program is wrong. This is illustrated
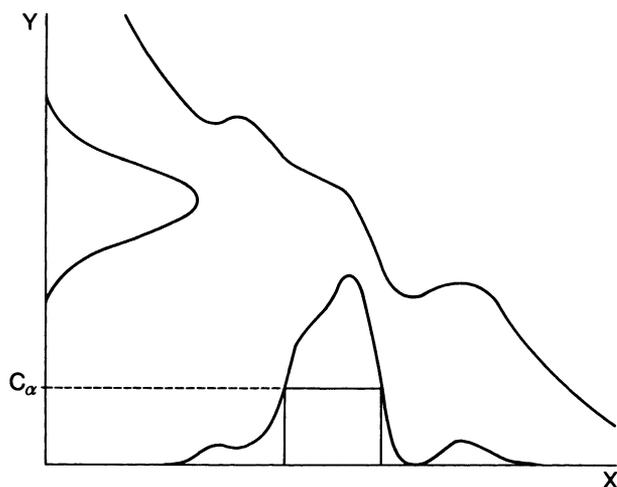
Fig. 1. Example of the ^{14}C calibration procedure: Transformation of a ^{14}C date into a calibrated calendar age probability distribution. The graph shows the calibration curve, $y = f(x)$, the Gaussian probability distribution corresponding to the ^{14}C age along the y-axis, and the calibrated calendar age probability distribution along the x-axis. The confidence limit, c_α, is defined in the text.

in Figure 3, where the transformation function has a horizontal stretch. Figure 3A shows the Groningen result, Figure 3B, the classical result. The latter is counterintuitive; why should the calibrated interval corresponding to the horizontal part of the curve have a near-zero probability, whereas the measurement agrees perfectly with any age in this interval? We aim to resolve this ambiguity by offering a clear interpretation of the error curve being transformed, and shall do this within both classical statistical and Bayesian frameworks.
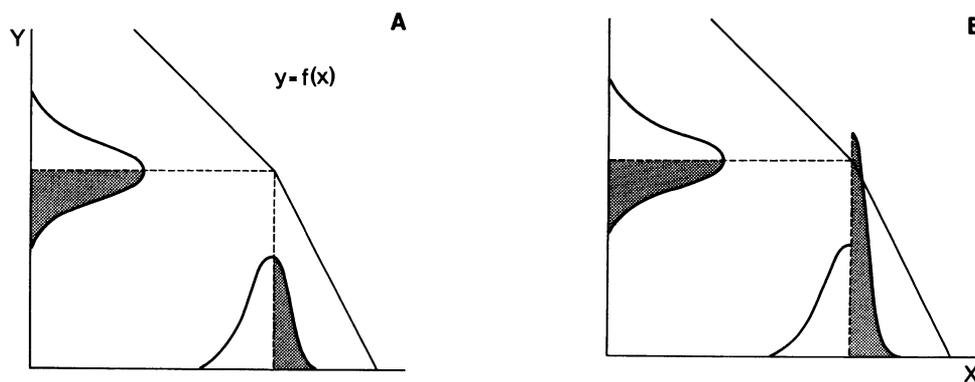
Fig. 2A,B. The calibration paradox (1): Two ways of obtaining a calibrated probability distribution, using an artificial calibration curve

Bayesian methods and classical statistics differ in their approach to statistical inferences. Significance tests, point estimates and confidence limits are essential parts of classical inference, whereas Bayesian inference is expressed as a posterior probability distribution. The fundamental difference between these methods concerns the meaning of probability. The classical objective approach assigns a probability only for events in repeated experiments. In contrast, the Bayesian subjective approach interprets a probability distribution as a degree of belief, in which probabilities can be assigned in situations where the objective approach does not apply.

A Bayesian approach, though subjective, has wider applicability. For instance, two archaeologists examining the same object might have different opinions about the probability that it dates from
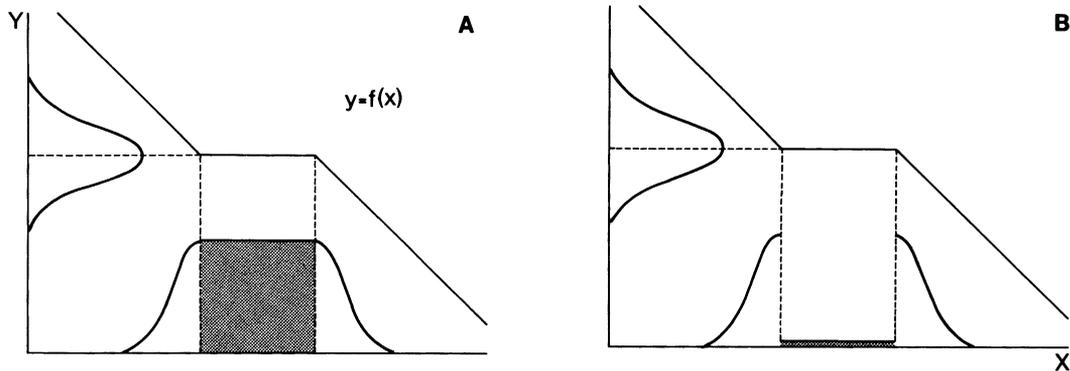
Fig. 3A,B. The calibration paradox (2): Two ways of obtaining a calibrated probability distribution, using an artificial calibration curve

a certain period. If there is no prior information, a common approach would be to assume a uniform distribution. However, the uniform distribution is not invariant under coordinate transformations, so that the choice of coordinate system is somewhat arbitrary. In our calibration example, the uniform distribution along the cal AD/BC axis is different from the uniform distribution along the yr BP axis. This leads to the Bayesian explanation of the aforementioned ambiguity.

A STATISTICAL MODEL FOR ^{14}C MEASUREMENTS

A ^{14}C age measurement is subject to random disturbances. We model this by assuming that the actual measurement, y, is the real value of a random variable with a distribution around the true, but unknown ^{14}C age η. We employ the Gaussian error model, where y has a normal probability density with mean η and variance σ^2:

$$p(y|\eta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\eta)^2}{2\sigma^2}} \quad . \tag{2}$$

So far, we have viewed the density $p(y|\eta)$ as a function of y for a fixed η. Now, if y is the actual outcome of the measurement, we can study $p(y|\eta)$ as a function of η, given as the so-called likelihood function L. The maximum value of L is the widely used maximum likelihood estimate.

In the Gaussian measurement error model, the likelihood is given by

$$L(\eta) = p(y|\eta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\eta-y)^2}{2\sigma^2}} \quad . \tag{3}$$

Note that because of symmetry regarding y and η, the likelihood has exactly the same shape as the density. The Gaussian error curve, drawn along the yr BP axis around y, is the likelihood function.

Using the calibration curve, we reparametrize the model in order to derive calibrated probabilities, relevant to archaeologists. If ξ denotes the true calendar age, the ^{14}C age is given by $\eta = f(\xi)$, and hence, the measured ^{14}C age follows the density:

$$p^*(y|\xi) = p(y|f(\xi)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-f(\xi))^2}{2\sigma^2}} \quad . \tag{4}$$

If we take the calendar age as the model parameter, the likelihood is given by

$$L^*(\xi) = p^*(y|\xi) = p(y|f(\xi)) = L(f(\xi)) \quad . \tag{5}$$

Thus, the likelihood function $L(\eta)$ on the ^{14}C age scale (y-axis) is transformed into the likelihood function $L^*(\xi)$ on the calendar age scale (x-axis).

AN OBJECTIVE APPROACH: CONFIDENCE REGIONS

A point estimate, $\hat{\eta}$ of η, is inadequate in statistical application, because it gives no information about the uncertainty involved. The objective approach expresses this uncertainty as a confidence region R. This function associates, to each outcome y, a set of parameters $C(y) \subset R$. We say that C is a $(1 - \alpha)$ confidence region for η, with $0 < \alpha < 1$, if

$$P_\eta(\{\eta \in C(Y)\}) = \int_{\{y:\eta\in C(y)\}} p(y|\eta)\,dy = 1-\alpha \tag{6}$$

for all η. This specifies that in a long series of measurements where the true parameter is η, the confidence regions $C(y_1)$, $C(y_2)$, . . . obtained will contain η in about $100(1 - \alpha)\%$ of all cases. For example, for a normal distribution $p(y|\eta)$ with known variance σ^2, the (random) interval $[y - \sigma, y + \sigma]$ is a 68% confidence region, because $C(y)$ will cover the parameter η if the observed value y has a distance less than or equal to σ from η. For any value η of the parameter, the random confidence region, $C(y)$, will cover η with probability $(1 - \alpha)$. From a classical viewpoint, the parameter η either lies in the confidence region C_o or not, but we cannot determine which. Neither could a Bayesian approach compute the probability that $\eta \in C_o$, because this depends essentially on the prior probability.

One approach to obtain confidence limits, based on the likelihood principle, suggests confidence regions given by level sets $C(y) = \{\eta : L_\eta(y) \le c_\alpha\}$ of the likelihood function. The constant c_α must be chosen in such a way that $C(y)$ is a $(1 - \alpha)$ confidence region.

As observed, error curves drawn on the yr BP and cal AD/BC axes are likelihood functions for ξ and η, respectively. Thus, the 1 σ (2 σ) interval on the BP axis, along with the level set on the AD/BC axis, are likelihood-based confidence regions with confidence level 68 (95)%. Now the confidence region for ξ covers the true age ξ if and only if $|f(\xi) - y| \le \sigma$ (2 σ). But this has again a probability of 68 (95)%, since y follows a normal distribution with mean ξ and variance σ^2.

A BAYESIAN APPROACH

The Bayesian calibration solution provides a basis for interpretation of the error curve (drawn in Figures 2 and 3) as a probability density on the age space. It requires specification of a prior probability density on the parameter space, representing parameter information prior to any measurement. Lindley (1965) and Naylor and Smith (1988) give a thorough treatment of the Bayesian approach. In our situation, either η or ξ must be regarded as the fundamental parameter about which prior assumptions are made. For instance, an archaeologist might, on the basis of

previous samples from a single site, be convinced that this material is from a certain age period. This could be expressed by choosing an appropriate density, either $\pi(\eta)$ for η or $\pi^*(\xi)$ for ξ.

Beginning with $\pi(\eta)$, note that we have a conditional density $p(y|\eta)$, where η is now considered random. If y denotes the observation actually made, we can compute the density of η, given y, using a continuous analog of Bayes' theorem on inverse probabilities

$$\pi(\eta|y) = \frac{p(y|\eta)\pi(\eta)}{\int p(y|\eta)\pi(\eta)\,d\eta} \quad . \tag{7}$$

$\pi(\eta|y)$ is the posterior density, and it incorporates both our prior information and the measurement result. Alternatively, we can express our prior information with a prior density $\pi^*(\xi)$ on the calendar age. Then the posterior density is given by

$$\pi^*(\xi|y) = \frac{p(y|f(\xi))\pi^*(\xi)}{\int p(y|f(\xi))\pi^*(\xi)\,d\xi} \quad . \tag{8}$$

A special "improper" case of this is the uniform prior distribution $\pi(\eta) = 1$ ($\pi^*(\xi) = 1$) on the real line, also called the non-informative prior distribution. (Note that this is not a probability density, since $\int_{-\infty}^{\infty} 1\,d\eta = \infty$.) In this case, the posterior probability is formally given by the normalized likelihood function

$$\pi(\eta|y) = \frac{p(y|\eta)}{\int p(y|\eta)\,d\eta} \qquad \left(\pi^*(\xi|y) = \frac{p(y|f(\xi))}{\int p(y|f(\xi))\,d\xi}\right) \quad . \tag{9}$$

In this way, we justify the interpretation of the Gaussian error curve drawn in Figures 2 and 3 as a posterior density on ^{14}C age space, corresponding to a non-informative prior on this space.

As $\pi(\eta|y)$ is a Bayesian density, it must be transformed according to the classical formula for densities

$$\pi^*(\xi|y) = \pi(f(\xi)|y) \cdot |f'(\xi)| \quad . \tag{10}$$

How can we now explain the examples discussed in Figures 2 and 3? As mentioned above, a Bayesian interpretation of the Gaussian error curve (the likelihood function) as a posterior density implicitly assumes a uniform prior distribution $\pi(\eta)$ on the ^{14}C age, and corresponds to a non-uniform prior distribution

$$\pi^*(\xi) = \pi(f(\xi))|f'(\xi)|$$

on the calendar age BC/AD. This accounts for the unusual transformations shown, *e.g.*, in Figure 3B, where the prior $\pi^*(\xi)$ places no mass in the interval where f is constant. Of course, if we choose a prior distribution specifying that the probability of an observation in this range is zero, the posterior density will also place mass zero here. In this light, it seems unreasonable to express the lack of prior information on the age of the sample by a uniform prior distribution on the yr BP axis. However, a uniform prior distribution on the calendar axis would correspond to a non-uniform

prior distribution on yr BP, so that the Gaussian error curve on age BP is no longer the posterior density. But, the curve produced by the calibration programs, being the correct likelihood function, is then, with respect to normalization, the posterior distribution on calendar age, corresponding to the uniform prior on calendar age.

We observe that both the Groningen and classical error curves are posterior densities, the first corresponding to a uniform distribution on cal AD/BC years, the latter to a uniform distribution on yr BP as prior densities. That the latter is an unreasonable prior distribution is evident from Figures 2 and 3.

CONCLUSION

Calibration of radiocarbon dates involves the transformation of a measured ^{14}C age (BP $\pm$ σ) into a calibrated age distribution (cal AD/BC range). Because of the wiggly nature of the calibration curve, the correct procedure to obtain calibrated age ranges and confidence intervals is not straightforward. Mathematical pitfalls can cause calibration procedures to contradict classical formulas.

We show that these ambiguities can be understood in terms of classical and Bayesian approaches to statistical theory. The classical formulas correspond to a uniform prior distribution along the BP axis, the calibration procedure to a uniform prior distribution along the calendar axis. We argue that the latter is the correct choice, *i.e.*, the computer programs used for radiocarbon calibration are correct.

REFERENCES

Aitchison, T. C., Leese, M., Michczynska, D. J., Mook, W. G., Otlet, R. L., Ottaway, B. S., Pazdur, M. F., van der Plicht, J., Reimer, P. J., Robinson, S. W., Scott, E. M., Stuiver M. and Weninger, B. 1989 A comparison of methods used for the calibration of radiocarbon dates. *In* Long, A. and Kra, R. S., eds., Proceedings of the 13th International ^{14}C Conference. *Radiocarbon* 31(3): 846–864.

de Vries, H. 1958 Variation in concentration of radiocarbon with time and location on earth. *Kon. Ned. Acad. Wet. Proc. Ser. B* 61: 1–9.

Lindley, D. V. 1965 *Introduction to Probability Theory and Statistics from a Bayesian Viewpoint.* Cambridge, Cambridge University Press: 2 vols.

Naylor, J. C. and Smith, A. F. M. 1988 An archaeological inference problem. *Journal of the American Statistical Association* 403: 588–595.

Stuiver, M. and Reimer, P. J. 1986 A computer program for radiocarbon age calibration. *In* Stuiver, M. and Kra, R. S., eds., Proceedings of the 12th International ^{14}C Conference. *Radiocarbon* 28(2B): 1022–1030.

Suess, H. E. 1970 The three causes of secular ^{14}C fluctuations, their amplitudes and time constants. In Olsson, I. U., ed., *Radiocarbon Variations and Absolute Chronology.* Proceedings of the 12th Nobel Symposium. Stockholm, Almqvist and Viksell: 595–605.

van der Plicht, J. and Mook, W. G. 1987 Automatic radiocarbon calibration: Illustrative examples. *Palaeohistoria* 29: 173–182.

_____1989 Calibration of radiocarbon ages by computer. *In* Long, A. and Kra, R. S., eds., Proceedings of the 13th International ^{14}C Conference. *Radiocarbon* 31(3): 805–816.