

Radiocarbon

1995

IMPROVED RADIOCARBON AGE ESTIMATION USING THE BOOTSTRAP

AMIR D. ACZEL

Department of Mathematical Sciences, Bentley College, Waltham, Massachusetts 02154 USA

ABSTRACT. This paper proposes the use of the statistical bootstrap technique as an aid in combining radiocarbon date estimates. The rationale for the use of the bootstrap is the theoretical result that, even if individual date estimates are normally distributed, their combination by the usual formula results in a random quantity that is not normal but rather a mixture of distributions. The bootstrap is a non-parametric, computer-intensive technique. This technique can better estimate the actual distribution of the combined age, leading to more precise confidence intervals. While the bootstrap cannot solve the multiple-intercepts problem in calibration, it can nonetheless lead to better estimates. The benefits of using the bootstrap are especially noticeable when sample sizes are small (as is the case in other applications of this technique).

INTRODUCTION

The current approach to the combination of radiocarbon age estimates is described in Ward and Wilson (1978). The “Case I” situation discussed in that article involves separate estimates of the age of the same item. This provides the authors with the justification for suggesting that (once a chi-square criterion is satisfied) estimates and their standard errors be combined into a single age estimate.

The methodology assumes n age estimates and standard errors, $A_i \pm E_i$ ($i = 1, \dots, n$), which are combined as follows:

$$A_p = (\sum A_i / E_i^2) / (\sum 1 / E_i^2) \quad (1)$$

$$V(A_p) = (\sum 1 / E_i^2)^{-1} \quad (2)$$

where A_p is the pooled age estimate and $V(A_p)$ is its variance.

Even if all the separate A_i are exactly normally distributed, the statistic A_p is not. The reason for this is that the standard errors, E_i , are themselves random variables. The quantity A_p in Equation (1) is distributed as a mixture. Therefore, using a normal distribution in setting confidence limits for the ^{14}C age as suggested by Ward and Wilson (1978) is inaccurate. The statistic in (1) does, however, enjoy desirable properties of estimators. It is derived as a combination of estimates that is optimal in some sense—it is the result of solving a particular integral equation. This makes the solution, A_p , an M-estimate (see Serfling 1980: 243).

Arcones and Gine (1992) showed that the distribution of M-estimates may be estimated well by the bootstrap method (Efron 1979). They proved the convergence of the bootstrap distribution of an M-estimate to the actual, unknown distribution of the statistic. The bootstrap does not require the assumption of normality, or any other parametric assumption, and can be estimated with high accuracy using a computer. This provides an excellent alternative method for combining ^{14}C age estimates. I form the combination of age estimates from various sources as indicated by Equations (1) and (2). Then, instead of assuming a normal distribution for A_p , I bootstrap the statistic. As I show here, the results can be promising. I describe the bootstrap method below and apply it to ^{14}C dating in two well-known cases, showing that the bootstrap-derived confidence intervals are narrower than the ones obtained by incorrectly assuming a normal distribution.

THE BOOTSTRAP METHOD

The bootstrap (Efron 1979), makes use of the tremendous advances in computing power that have taken place in recent years. The idea is to resample a large number of times from our limited data, each time recomputing the value of the statistic of interest. Then a histogram may be plotted, which allows one to actually "see" a close estimate of the sampling distribution of the statistic. The procedure is non-parametric, and its validity does not rely on special assumptions such as normality. The bootstrap-estimated distribution can then be used in constructing confidence intervals for the estimated parameter, or in other statistical inference. Since its inception, the popularity of the bootstrap has been increasing dramatically (Kempthorne *et al.* (1991) counted over 360 bootstrap-related articles in the statistical literature in the last few years). Reports of surprising results on the method's accuracy and versatility have been published (Bickel and Freedman 1981; Singh 1981; Efron 1982; Efron and Tibshirani 1986; Hall 1988). The bootstrap is a relatively new technique and questions have been raised about its use. Many of these questions, however, have been answered in a positive way as new theory becomes available confirming the method's usefulness. An excellent reference to bootstrap applications and theory is the book by Efron and Tibshirani (1993). Issues of bootstrap theory are addressed in the book by Hall (1992).

Given a set of n data points, x_1, x_2, \dots, x_n , one first forms the empirical distribution of these data. The only assumption one needs is that the data are independently and randomly chosen from some population with an unknown distribution $F(x)$. One attempts to estimate this common population distribution by the empirical distribution of the data, $\hat{F}(x)$. The empirical distribution is defined as that probability distribution that assigns a probability of $1/n$ to each one of the n data points. It can be shown that $\hat{F}(x)$ is the maximum likelihood estimate of the actual, unknown population distribution $F(x)$. In this sense, $\hat{F}(x)$ is the best estimate of $F(x)$ and sampling from the distribution $\hat{F}(x)$ should closely resemble sampling from $F(x)$. Sampling from $\hat{F}(x)$ means resampling, with replacement, from our data set x_1, x_2, \dots, x_n .

The resampling is done many times, B . Typically, B is from 200 to 1000, even more. Each time I resample from the original sample of n observations, I select a random sample of size n with replacement. If $n = 4$ and our original observations are 8, 5, 13, 9, a bootstrap sample may consist of 8, 13, 9, 9, or 8, 5, 5, 13, or a similar set of four numbers randomly chosen with replacement from the original data. A bootstrap sample is indicated by *. Given a statistic of interest, T , one computes the value of the statistic from each of the B bootstrap samples. This leads to the bootstrap distribution of the statistic, the distribution of T^* .

The cumulative bootstrap distribution of a statistic, T , is:

$$G(t) = P(T^* \leq t) = \#(\text{bootstraps} \leq t) / B. \quad (3)$$

The bootstrap distribution of T can be shown to be an excellent estimate of the actual sampling distribution of the statistic. The bootstrap distribution may be used in inference. If T is an estimator of a population parameter, Θ , then a $(1-2\alpha)100\%$ confidence interval for Θ can be obtained using the percentile method (Efron 1982). This is done following the usual non-parametric procedure for setting confidence intervals by cutting off an accumulated probability α at each end of the distribution. Using the cumulative bootstrap distribution defined in Equation (3), a $(1-2\alpha)100\%$ confidence interval for Θ is

$$[G^{-1}(\alpha), G^{-1}(1-\alpha)]. \quad (4)$$

The results of Arcones and Gine (1992) on \sqrt{n} -consistency of the bootstrap for M-estimates allows for using the bootstrap in estimating the distribution of the combined ^{14}C age statistic, A_p . In forming the combined statistic, as given in Equation (1), one must use pairs of estimates: (A_i, E_i) . The n pairs are resampled $B = 1000$ times, each time computing the value of A_p from the n new pairs. This provides 1000 values of the combined statistic. I then produce a histogram of these values, obtaining the bootstrap distribution of A_p^* . Finally, I use Equation (4) to compute confidence intervals for the ^{14}C age.

The accuracy of bootstrap computations can be improved. Efron (1982) offers a step in this direction by introducing the idea of bias correction. Efron (1987) offered an even finer advance in which he discusses confidence intervals that account for "acceleration". Another area of extension of the bootstrap approach may be a Bayesian bootstrap, where age estimates are resampled with probability inversely proportional to their standard deviations and computed with a simple (rather than weighted) average of these estimates.

DATING THE THERA ERUPTION

^{14}C dating the eruption of the Thera volcano in the Aegean has received much attention recently. At the heart of the debate is the apparent disagreement of ^{14}C results with the accepted Late Minoan chronology, which places the eruption about a century earlier than the ^{14}C dates (Aitken 1988). Michael and Betancourt (1988: 172) present a table of nine calibrated ^{14}C dates. The dates, in years BC, are 1521, 1606, 1628, 1615, 1622, 1681, 1615, 1642 and 1642. The standard errors reported in the table are very uniform, at *ca.* 50 yr each.

I use these data in constructing the statistic A_p and carrying out a bootstrap estimation of its sampling distribution. Note that the bootstrap replaces both the normal distribution and the estimated variance as given in Equation (2). Because in this case the standard errors are roughly equal, we may ignore them in bootstrapping the statistic in Equation (1). Using the standard errors as weights or not using them leads to the same results. The distribution of 1000 bootstrap replicates of the statistic is shown in Figure 1. The resulting 95% confidence interval for the actual date of the eruption using the percentile method is only slightly wider than the one-sigma ($1-\sigma$) interval reported in Michael and Betancourt (1988); a $2-\sigma$ interval (required for 95% under the normal assumption) would be much wider than the result obtained by the bootstrap. Note the skewness of the bootstrap distribution, which provides another argument against using the normal (or any symmetric) distribution in this case.

DATING THE SHROUD OF TURIN

Damon *et al.* (1989) reported the results of ^{14}C dating of the Shroud of Turin. Samples of the Shroud, along with three control samples, were analyzed by the laboratories of Arizona, Oxford and Zürich. I treat the Sample 1 (Shroud) results reported by the three laboratories equally. This gives the following 12 pairs of observations (mean age BP, standard error), from Table 1 of Damon *et al.* (1989): (591,30), (690,35), (606,41), (701,33), (795,65), (730,45), (745,55), (733,61), (722,56), (635,57), (639,45), (679,51). These data were transformed using the Pearson and Stuiver (1986) curve, the error in the calibration curve incorporated in the standard errors. Wherever multiple intercepts occurred in the calibration curve, the earliest date was chosen (thus giving the null hypothesis of authenticity the "benefit of the doubt"). I recognize, of course, that the bootstrap cannot solve the serious difficulties caused by multiple intercepts in the calibration curve.

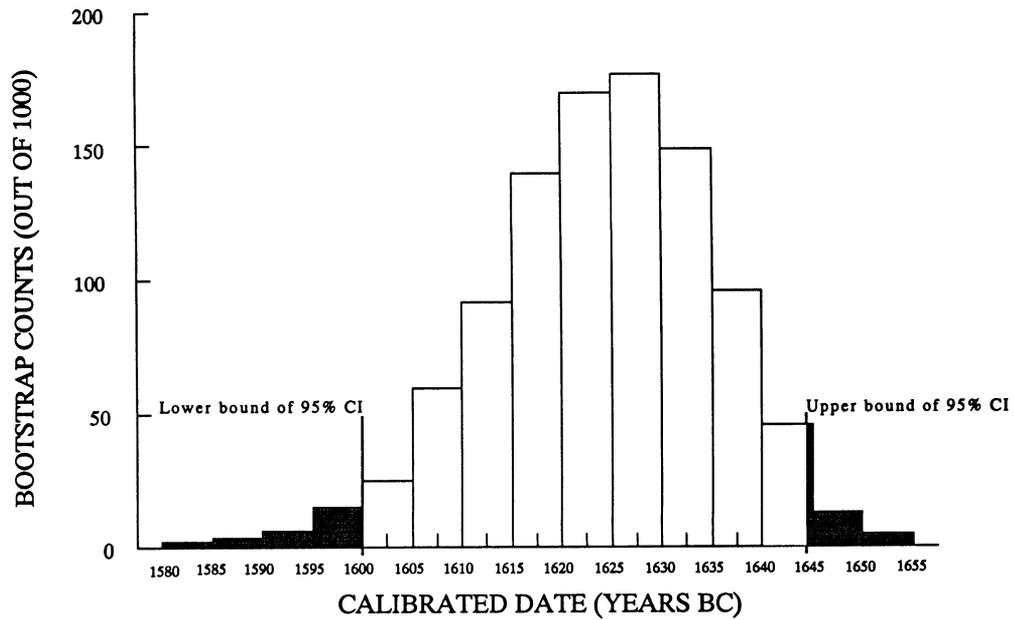


Fig. 1. Thera eruption date bootstrap distribution

The transformed data pairs were bootstrapped 2000 times (for improved precision) using the method described above. Figure 2 shows the resulting bootstrap distribution histogram. Each column in the histogram represents a single year. A 95% confidence interval for the date of the Shroud can be read

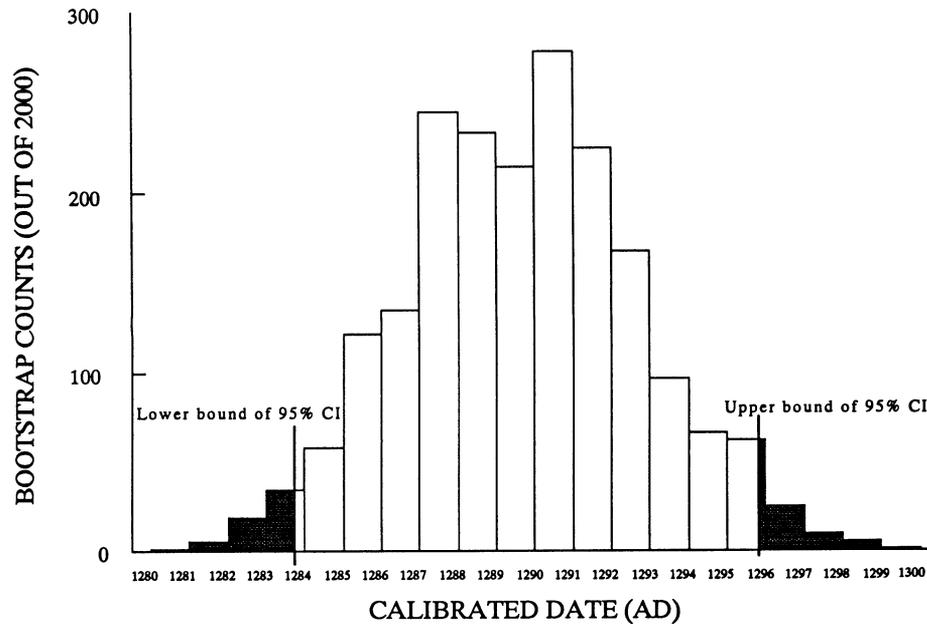


Fig. 2. Turin Shroud date bootstrap distribution

from the histogram as AD 1283–AD 1295. The interval is much narrower than that reported by Damon *et al.* (1989: Table 3) of AD 1262–1312, 1353–1384. Although the right endpoint may be closer to the center due to ignoring multiple intercepts to the right, the much more important left endpoint of the confidence interval is also tighter—by >20 yr. From the histogram, one can also observe a zero empirical (bootstrap) probability that the Shroud was created earlier than AD 1280.

CONCLUSION

The bootstrap offers many advantages as an alternative method for combining ^{14}C dates. Two examples were given in this article; hopefully other researchers will try the method and provide further evidence. For the calculations, a simple computer will suffice, even a hand-held programmable calculator may perform the required computations; thus, some quick estimates could be performed in the field. The code required by the bootstrap is minimal (<50 commands were necessary to perform the analyses discussed here).

ACKNOWLEDGMENTS

The author gratefully acknowledges the help of Professors Evarist Gine and Vladimir Kolcinskii of the University of Connecticut, and of Professor Norman Josephy of Bentley College.

REFERENCES

- Aitken, M. J. 1988 The Thera eruption: Continuing discussion of the dating. *Archaeometry* 30: 165–169.
- Arcones, M. and Gine, E. 1992 The bootstrap of M-estimates. In LePage, R. and Billard, L., eds., *Exploring the Limits of Bootstrap*. New York, John Wiley & Sons: 426 p.
- Bickel, P. J. and Freedman, D. A. 1981 Some asymptotic theory for the bootstrap. *Annals of Statistics* 9: 1196–1217.
- Damon, P. E., Donahue, D. J., Gore, B. H., Hatheway, A. L., Jull, A. J. T., Linick, T. W., Sercel, P. J., Toolin, L. J., Bronk, C. R., Hall, E. T., Hedges, R. E. M., Housley, R., Law, I. A., Perry, C., Bonani, G., Trumbore, S., Woelfli, W., Ambers, J. C., Bowman, S. G. E., Leese, M. N. and Tite, M. S. 1989 Radiocarbon dating of the Shroud of Turin. *Nature* 337: 611–615.
- Efron, B. 1979 Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7: 1–26.
- _____. 1982 *The Jackknife, the Bootstrap, and Other Resampling Schemes*. Philadelphia, Pennsylvania, Society of Industrial and Applied Mathematics: 92 p.
- _____. 1987 Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82: 171–185.
- Efron, B. and Tibshirani, R. 1986 Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1: 54–77.
- _____. 1993 *An Introduction to the Bootstrap*. New York, Chapman and Hall: 436 p.
- Hall, P. 1988 Theoretical comparison of bootstrap confidence intervals. *Annals of Statistics* 16: 927–952.
- _____. 1992 *The Bootstrap and Edgeworth Expansion*. New York, Springer-Verlag: 352 p.
- Kempthorne, P., Mukhopadhyay, N., Sen, P. K. and Zacks, S. 1991 Research, a panel discussion. *Statistical Science* 6: 149–163.
- Michael, H. N. and Betancourt, P. 1988 The Thera eruption: Further arguments for an early date. *Archaeometry* 30: 169–174.
- Serfling, R. J. 1980 *Approximation Theorems of Mathematical Statistics*. New York, John Wiley & Sons: 371 p.
- Singh, K. 1981 On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics* 9: 1187–1195.
- Stuiver, M. and Pearson, G. W. 1986 High-precision calibration of the radiocarbon time scale, AD 1950–500 BC. In Stuiver, M. and Kra, R., eds., *Proceedings of the 12th International ^{14}C Conference*. *Radiocarbon* 28 (2B): 805–838.
- Ward, G. K. and Wilson, S. R. 1978 Procedures for comparing and combining radiocarbon age determinations: A critique. *Archaeometry* 20: 19–31.