Theor. Appl. Climatol. 45, 229–233 (1992)

Theoretical and Applied Climatology

551.58

[1] Climatology Research Group, University of the Witwatersrand, Johannesburg, South Africa
[2] Department of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, South Africa

The Use of Bootstrap Confidence Intervals for the Correlation Coefficient in Climatology

S. J. Mason[1] and **G. M. Mimmack**[2]

Received October 14, 1991
Revised March 6, 1992

Summary

The importance of defining confidence intervals for sample statistics that are used to estimate characteristics of the parent population(s) is emphasised. Not all sample statistics are unbiased estimators or have normally distributed sampling distributions and so it is not always easy to reflect the reliability of the estimator. In such cases, Efron's "bias corrected percentile method", which uses bootstrap samples to estimate the bias and makes no assumptions about the distribution of the sample statistic can be used to define confidence limits for the population parameter. The method is explained and the procedure for calculating the confidence limits is outlined.

As an example, bootstrap confidence limits calculated for the maximum correlation between the Southern Oscillation Index and rainfall at South African stations over the period 1935–1983 suggest that the sample correlation is an unreliable measure of the true association. One possible reason for this is that the association is thought to have broken down during the 1940s. However, the reliability of the estimator does not seem to improve when confidence limits are calculated for the 30-year period 1954–1983. It is possible that the width of the confidence interval is an indication of more than one distinct statistical population.

1. Introduction

Despite the hazards of using the correlation coefficient as a statistical tool in climatology (Ramage, 1983; Brown and Katz, 1991), it is still widely used within the discipline since it provides a very useful and simple procedure for identifying associations between variables. The correlation coefficient can be used to quantify the strength of the observed association by means of r^2, the coefficient of determination, which is used to define the precentage of variance observed in the dependent variable that can be "explained" or modelled by a linear association with the independent variable. In calculating the correlation coefficient we are implicitly fitting a linear regression line to the data and quantifying the extent to which the total variability observed in the dependent variable can be accounted for by the variability of the estimated values.

Since it is possible to find an apparent association between two sets of random numbers, it is standard practice to assess the significance of the correlation by calculating the probability of exceeding the observed correlation by accident. The significance test of the correlation coefficient is an analysis of variance test to assess whether the variance modelled by the linear regression is significantly greater than the variance not modelled (the residual variance). The need for testing the collective significance of individual significance tests when calculating a correlation field, for example, has been detailed (Livezey and Chen, 1983; Brown and Katz, 1991) and so is not considered further here. The present concern is to comment on the use of the correlation coefficient in a descriptive capacity, in the form of r^2, for example, once it has "passed" a statistical test. Although the

paper concentrates on the correlation coefficient, it should be emphasised that the principles can be applied to any statistic.

2. Theory and Derivation

If a statistic is found to be "significant" it may be of interest to obtain an estimate of the corresponding population parameter. For example, the sample correlation coefficient is often used as an estimate of the population correlation in order to define the percentage of variance which can be modelled by a simple linear statistical relationship between two variables. Frequently, once a sample statistic is found to be significant, it is treated as an accurate estimate of the relevant population quantity. As the Neyman–Pearson method of hypothesis testing yields only the probability of achieving the observed result purely by chance it does not give any indication of the reliability of the sample statistic as an estimator of the population parameter. Although r can be seen as an estimator of ρ, it is necessary to indicate its reliability in this role.

One way to incorporate the reliability of an estimator is to give the estimate in the form of a confidence interval, thus defining a range of estimates for the population parameter. The reliability of any estimator is reflected in part by the estimated standard deviation, or standard error, of the estimator. The standard errors of some statistics can be obtained conveniently from other estimates. For example, consider estimating the standard deviation of the sample mean, $\bar{X}$. Since the standard deviation of $\bar{X}$ is $\sigma/\sqrt{n}$ and since s estimates σ, the standard error of $\bar{x}$, $SE_{\bar{x}}$, is given by $s/\sqrt{n}$. Another consideration in assessing the reliability of an estimator is the form of its sampling distribution. In the case of $\bar{X}$, a large sample is sufficient to guarantee that $\bar{X}$ has (at least approximately) a normal distribution. Since the distribution of $\bar{X}$ is centred at the population mean, μ,

$$X \dot{\sim} N(\mu, \sigma^2/n). \tag{1}$$

Consequently, $\bar{X} - \mu$ has a normal distribution centred at zero. If the sample size is large, it is appropriate to estimate σ by s and use percentage points of the standard normal distribution to define confidence intervals for μ. For example, $(\bar{x} - 2SE_{\bar{x}}, \bar{x} + 2SE_{\bar{x}})$ is a 95% confidence interval for μ because

$$P(\bar{X} - 2SE_{\bar{X}} < \mu < \bar{X} + 2SE_{\bar{X}}) \cong 0.95. \tag{2}$$

In the case of the correlation coefficient the problem is to identify the limits denoted r_l and r_u that will enclose ρ with probability $1 - 2\alpha$

$$P(r_l < \rho < r_u) = 1 - 2\alpha. \tag{3}$$

Since r is used to estimate ρ, the length and location of the interval (r_l, r_u) is determined by the standard error and the sampling distribution of r. Unfortunately, r does not have a normal distribution. Except in the case of $\rho = 0$ (and in the trivial case of $\rho = \pm 1$), the sampling distribution of r is negatively skewed. Consequently, even if the sampling distribution of r can be established, it is inappropriate to define confidence intervals for ρ using percentiles from a set of sample correlations obtained by repeated sampling. A solution to this problem is to transform r into a random variable that does have a normal distribution. A monotone increasing transformation function, g, can be defined whose values, $g(r)$, have a normal distribution with unit variance irrespective of the original sampling distribution of r. The confidence limits are $g(r_l)$ and $g(r_u)$ and Eq. 3 becomes

$$P(g(r_l) < g(\rho) < g(r_u)) = 1 - 2\alpha. \tag{4}$$

A further complication is that r is not an unbiased estimator of ρ so the expected value of r should be expressed as $\rho + a$, say, where a is non-zero. Consequently, a confidence interval for ρ should not be centred at r. The bias can be carried over to the transformed variable so that the distribution of $g(r)$ is centred at $g(\rho) + b$, say, where b is non-zero. Thus,

$$g(r) \sim N(g(\rho) + b, 1). \tag{5}$$

The confidence limits for $g(\rho)$ can now be defined as

$$g(r_l) = g(r) - b - \Delta \tag{6}$$

$$g(r_u) = g(r) - b + \Delta \tag{7}$$

and for ρ as

$$r_l = g^{-1}(g(r) - b - \Delta) \tag{8}$$

$$r_u = g^{-1}(g(r) - b + \Delta) \tag{9}$$

where b is a measure of the bias of $g(r)$ and 2Δ is the length of the confidence interval.

If ρ is known and if the sampling distribution of r can be established, it is possible to calculate the probability of r being less than ρ, using

$$CDF(\rho) = P(r < \rho) \tag{10}$$

where $CDF(\rho)$ represents the proportion of sample correlation coefficients that are less than the population correlation coefficient. Since g, being a monotone increasing function, does not affect the rank order of the values of r, it follows that

$$CDF(\rho) = CDG(g(\rho)) = P(g(r) < g(\rho)) \tag{11}$$

where $CDG(g(\rho))$ represents the proportion of the transformed sample correlation coefficients that are less than $g(\rho)$. From Eq. 11 it follows that

$$CDF(\rho) = P(g(r) - g(\rho) - b < -b) \tag{12}$$

Equation 5 can be re-arranged to give

$$g(r) - g(\rho) - b \sim N(0, 1). \tag{13}$$

Therefore Eq. 12 becomes

$$CDF(\rho) = P(Z < -b) \tag{14}$$

where $Z \sim N(0, 1)$. Hence

$$b = -\Phi^{-1}(CDF(\rho)) \tag{15}$$

where Φ is the cumulative standard normal distribution function.

The length Δ can be calculated from the distribution of $g(r)$ as follows. The confidence limits for $g(\rho)$, which have been defined in Eqs. 6 and 7, yield the $100(1-2\alpha)\%$ confidence interval as

$$P(g(r) - b - \Delta < g(\rho) < g(r) - b + \Delta) = 1 - 2\alpha \tag{16}$$

which can be simplified to

$$P(-\Delta < g(r) - g(\rho) - b < \Delta) = 1 - 2\alpha. \tag{17}$$

The middle part of Eq. 17 has been identified already, in Eq. 13, as a standard normal random variable and so Eq. 17 becomes

$$P(-\Delta < Z < \Delta) = 1 - 2\alpha. \tag{18}$$

Hence,

$$\Delta = \Phi^{-1}(1-\alpha). \tag{19}$$

Although the bias and confidence interval have now been defined, their calculation assumes that ρ and the sampling distribution of r are known. In climatology, such assumptions cannot usually be met. However, the bootstrap resampling procedure of Efron (1982) can be used to obtain information about the distribution of r. The procedure entails obtaining many sub-samples, called bootstrap samples, from the sample that is used to calculate r. Each of these bootstrap samples yields a bootstrap correlation coefficient, denoted $\hat{r}$. It is necessary to take a sufficient number of bootstrap samples in order to produce a stable estimate of the sampling distribution of $\hat{r}$. Efron (1982) recommends 400. The distribution of the bootstrap sample correlation coefficients is then used to provide information about the distribution of r.

Just as r is a biased estimator of ρ, so also the distribution of the bootstrap estimates is not centred at r. Since the sampling distribution of r is estimated by the empirical distribution of the bootstrap estimates, the bias can be re-defined from Eq. 15 as

$$b = -\Phi^{-1}(\widehat{CDF}(r)) \tag{20}$$

where $\widehat{CDF}(r)$ is the proportion of bootstrap sample correlation coefficients that are less than r. Also

$$g(\hat{r}) - g(r) \sim N(b, 1). \tag{21}$$

Thus the end-points of the interval defined in Eq. 16 satisfy

$$P(g(\hat{r}) < g(r) - b \pm \Delta) = P(g(\hat{r}) - g(r) - b < -2b \pm \Delta). \tag{22}$$

That is

$$P(g(\hat{r}) < g(r) - b \pm \Delta) = \Phi(-2b \pm \Delta). \tag{23}$$

If $\widehat{CDF}(k)$ is the observed proportion of bootstrap sample correlation coefficients that are less than k and $\widehat{CDG}(k)$ is the observed proportion of transformed bootstrap estimates that are less than k, the empirical counterpart of Eq. 23 is

$$\widehat{CDG}(g(\hat{r}) < g(r) - b \pm \Delta) = \Phi(-2b \pm \Delta). \tag{24}$$

Equivalently,

$$(g(\hat{r}) < g(r) - b \pm \Delta) = \widehat{CDG}^{-1}(\Phi(-2b \pm \Delta)). \tag{25}$$

The left hand side of Eq. 25 defines the confidence limits for $g(\rho)$ given in Eqs. 6 and 7, and so, from Eqs. 8 and 9, the confidence limits for ρ are defined from

$$g^{-1}(g(\hat{r}) < g(r) - b \pm \Delta) = g^{-1}(\widehat{CDG}^{-1}(\Phi(-2b \pm \Delta))). \tag{26}$$

From Eq. 11 it is evident that $CDG(g(k)) = CDF(k)$ and so the $100(1-2\alpha)\%$ bootstrap confidence limits for ρ are

$$r_l = \widehat{CDF}^{-1}(\Phi(-2b - \Delta)) \tag{27}$$

$$r_u = \widehat{CDF}^{-1}(\Phi(-2b + \Delta)) \quad (28)$$

where b has been defined in Eq. 20 and Δ in Eq. 19.

3. Operational Procedure

The procedure for the numerical calculation of the $100(1-2\alpha)\%$ confidence interval for ρ using Efron's bias corrected percentile method is as follows:

1) calculate r from the sample of n data points;
2) generate 400 bootstrap samples of size n^*;
3) calculate the bootstrap estimates, $\hat{r}$;
4) sort the bootstrap estimates into ascending order of magnitude;
5) obtain $\widehat{CDF}(r)$, the proportion of bootstrap estimates that are less than r;
6) calculate b, the standard normal score that satisfies

   $b = -\Phi^{-1}(CDF(r))$;

7) obtain

   $\Delta = \Phi^{-1}(1-\alpha)$;

8) calculate the proportions $\Phi(-2b-\Delta)$ and $\Phi(-2b+\Delta)$;
9) find, from the sorted list of bootstrap estimates, the two values of $\hat{r}$ that have ranks $400\Phi(-2b-\Delta)$ and $400\Phi(-2b+\Delta)$.

4. Example

Rainfall over large parts of central South Africa is known to be modulated with the phase of the Southern Oscillation (Lindesay, 1988; van Heerden et al., 1988). The association is most apparent during January–March when above (below) normal rainfall occurs concurrently with high (low) phases of the Southern Oscillation. It has been estimated that up to 25% of the rainfall variance observed during January–March over central South Africa can be "explained" by reference to the Southern Oscillation (Lindesay, 1988). However, here is a case of using the correlation coefficient in a descriptive manner after it has been declared to be significant (using $\alpha = 0.05$). Since it has been suggested that the sample correlation coefficient is a biased estimator of the population coefficient, the quoted modelled variance is applicable only to the particular sample used and it is not immediately possible to give an estimate of the percentage of variance modelled in the population. The estimator is biased and because it is not possible to calculate the percentage exactly, it is preferable to calculate confidence limits. Bootstrap confidence limits for the correlation between the Southern Oscillation Index and rainfall at 59 stations in South Africa over the 49-year period 1935–1983 have been calculated at $\alpha = 0.05$ using a sub-sample size of 30 years. The 90% bootstrap confidence limits of the maximum modelled variance at any of the analysed rainfall stations in South Africa are 11% and 36%. Estimates near the lower end of the confidence interval indicate that the association between the Southern Oscillation and South African rainfall is negligible, while estimates near the upper end of the interval indicate that the association is strong over most of the country. It is evident that the confidence interval is very wide, suggesting that, in this case, the sample correlation coefficient is an unstable estimator of the population correlation.

One of the possible reasons for the observed unreliability of the sample correlation is that over the period analysed the association between the Southern Oscillation and South African rainfall is known to have been unstable, virtually breaking down completely in the 1940s (Lindesay, 1989). The analysis was therefore repeated for the period 1954–1983 taking bootstrap samples of 20 years. During this period the association is thought to have stabilised. If the maximum modelled variance is estimated from the sample correlation coefficient it increases to 44% over the shorter 30-year period and the 90% confidence limits are 16% and 54%. This represents a small improvement in the lower confidence limit and a large improvement in the upper confidence limit. Even though the association between the Southern Oscillation and South African rainfall is supposed to have stabilised during the shorter 30-year period, as is partly reflected in that the sample correlation and upper confidence limit are larger, the sample coefficient remains a very unreliable estimate. The confidence interval may be so wide because of the possibility that the association between the Southern Oscillation and South African rainfall during the January–March season is apparent only during years in which the Quasi-Biennial Oscillation is in its westerly phase (Mason and Lindesay, 1992). The poor lower confidence limit may therefore be the result of randomly selecting predominantly easterly years and the high upper limit predominantly westerly years. Whatever the case, a wide confidence interval

emphasises the unreliability of the sample correlation coefficient and highlights the need to incorporate the stability of the sample statistic in estimation. In the case of the Southern Oscillation-South African rainfall association, the lack of reliability suggests that there are more than one distinct statistical populations.

5. Conclusions

It is often statistically incorrect to use the values of a sample statistic as a direct estimate of the population parameter because the standard method of testing an hypothesis tells us only whether the observed value is significantly different from some pre-defined score. An indication of the reliability of the sample statistic as an estimator is required. Just as it is standard practice to quote confidence limits when estimating the population mean, for example, so also it is advisable to quote confidence limits when trying to estimate any other population parameter. This is especially true when estimating the coefficient of determination since the sample correlation coefficient, from which it is derived, is usually a biased estimator of the population correlation. In such cases, Efron's "bias corrected percentile method", which uses bootstrap samples to estimate the bias and makes no assumptions about the distribution of the sample statistic, should be used to define confidence limits for the population parameter.

Bootstrap confidence limits calculated for the maximum correlation between the Southern Oscillation Index and rainfall at South African stations over the period 1935–1983 suggest that the sample correlation is an unreliable estimator of the true association. One possible reason for this is that the association is thought to have broken down during the 1940s. However, the reliability of the estimator does not seem to improve when confidence limits are calculated for the 30-year period 1954–1983. It is possible that the wide confidence interval is an indication of more than one distinct statistical populations.

This research forms part of a Special Programme on South African Climatic Change: Analysis, Interpretation and Modelling (SACCAIM) funded by the Foundation for Research Development.

References

Brown, B. G., Katz, R. W., 1991: Use of statistical methods in the search for teleconnections: past, present, and future. In: Glantz, M. H., Katz, R. W., Nicholls, N. (eds.) *Teleconnections Linking Worldwide Climate Anomalies: Scientific Basis and Societal Impact*. Cambridge: Cambridge University Press, 371–400.

Efron, B., 1982: The jackknife, the bootstrap and other resampling plans. *Soc. Industr. Appl. Math.*, J.W. Arrowsmith, 92 pp.

Lindesay, J. A., 1988: Southern African rainfall, the Southern oscillation and a Southern Hemisphere semi-annual cycle. *J. Climatol.*, **8**, 17–30.

Lindesay, J. A., 1992: Temporal variations in Southern Oscillation-Southern African rainfall associations. *J. Climate* (submitted).

Livezey, R. E., Chen, W. Y., 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59.

Mason, S. J., Lindesay, J. A., 1992: A note on the modulation of Southern Oscillation-South African rainfall associations with the Quasi-Biennial Oscillation. *J. Geophys. Res.* (in press).

Ramage, C. S., 1983: Teleconnections and the seige of time. *J. Climatol.*, **3**, 223–231.

van Heerden, J., Terblanche, D. E., Schultze, G. C., 1988: The Southern oscillation and South African summer rainfall. *J. Climatol.*, **8**, 577–597.

Authors' addresses: S. J. Mason, Climatology Research Group, University of Witwatersrand, Johannesburg 2050, South Africa and G. M. Mimmack, Department of Statistics and Actuarial Science, University of Witwatersrand, Johannesburg 2050, South Africa.