Environment Canada
Environnement Canada

National Water Research Institute
Institut National de Recherche sur les Eaux

THE BOOTSTRAP AND THE JACKKNIFE IN ECOTOXICOLOGY OR NONPARAMETRIC ESTIMATES OF STANDARD ERROR

by

Efraim Halfon

Inland Waters Directorate
Direction Générale des Eaux Intérieures

THE BOOTSTRAP AND THE JACKKNIFE IN ECOTOXICOLOGY OR NONPARAMETRIC ESTIMATES OF STANDARD ERROR

by

Efraim Halfon

National Water Research Institute
Canada Centre for Inland Waters
Burlington, Ontario
Canada L7R 4A6

ABSTRACT

A recently invented statistical method, the bootstrap, is used to verify whether a hypothesis, developed from a limited data set, would be valid if all possible data would have been available, i.e. this statistical method allows generalization to chemicals of the same class not included in the original analysis. The validity of the relation, hypothesized by Neely, between the water solubility of an organic chemical and the ratio of the acute fish LC50 at two different time periods has been tested. The hypothesis has been shown correct by first fitting a linear model with a geometric mean (GM) functional regression, which takes into account errors in both the independent and dependent variables, to compare observed and predicted ratios; the generality of the model has been tested by computing the confidence limits of the correlation coefficient, of the slope and intercept of the GM regression model using the bootstrap. The results show that the correlation between predicted and observed data is statistically significant within one standard deviation, but sometimes it may not be significant at the 95% confidence limit. Neely's model is probably correct but it might have a systematic bias which makes the theoretical ratio somewhat higher than the observed ratio.

EXECUTIVE SUMMARY

Laboratory tests for determining chemical and environmental properties of toxic contaminants are time consuming and expensive. Over the years much research has been performed to predict the relation between certain easily measurable properties and others more difficult to obtain. Statistical relations, such as linear regression models, have been developed for this purpose. The problem is, however, that no test has been developed to assure that the predictions of these statistical models are valid also for new chemicals. The bootstrap is a new statistical procedure that allows generalization of the results to chemicals not used in the development of the original model; the bootstrap is used to verify whether a hypothesis developed from a limited data set would be valid, if all possible data would have been available. Thus, the method is very useful to reduce the amount of data to be collected from laboratory experiments to evaluate the toxicity and environmental hazard of toxic contaminants. When the statistical models are used for prediction of new chemical properties, the bootstrap allows an estimate of the probability and range of the chemical property, such as toxicity, bioconcentration, etc.

RESUME POUR LA DIRECTION

Les etudes en laboratoire visant a determiner les proprietes chimiques et environnementales des polluants toxiques sont longues et couteuses. Au fil des ans, beaucoup de recherches ont ete effectuees pour prevoir la relation entre certaines proprietes facilement mesurables et d'autres proprietes plu difficiles a evaluer. C'est pourquoi des relations statistiques, comme des modeles de regression lineaire, ont ete mises au point. Toutefois, aucun test ne permet de verifier que les previsions de ces modeles a des substances chimiques qui n'ont pas ete considerees lorsque ceux-ci ont ete mis au point; il sert a verifier si une hypothese elaboree a partir d'un ensemble de donnees limite serait egalement valable si l'on avait dispose de toutes les donnees. Cette methode est donc tres utile en ce qu'elle permet de reduire les quantites de donnees necessaires a l'evaluation en laboratoire de la toxicite et des dangers pour l'environnement des polluants. Lorsque les modeles statistiques servent a la prevision des proprietes de nouvelles substances chimiques, la methode "bootstrap" permet d'estimer la probabilite et l'importance de ces proprietes, telles la toxicite, la bioconcentration, etc.

INTRODUCTION

In the late 1940's Quenouille (7) invented a nonparametric estimate of bias, later named the jackknife. The method is based on computing some statistic of interest by sequentially deleting points X_i one at a time and then recomputing the statistic to assess the bias or the relevance of each data point. Later in 1958, Tukey (9) suggested how to use the jackknife to provide a nonparametric estimate of variance. In 1977, Efron (2-4) invented a new statistical test, the bootstrap, which generalizes the jackknife and uses information from the given data to estimate the statistic, for example the correlation with its confidence limits, if all possible data from a population, or distribution F, would have been available. As an interview in Science in Science discloses (5) "Efron's most recent work has been on a general question that often plagues scientists: What would be seen if there were a lot more data?" The method is simple to describe and even simpler to program, but it is so dependent on the computer, even if based on strong theoretical foundations, that it would have been unthinkable 30 years ago (1).

Efron (5) calls his method the bootstrap because "you use the data to estimate probabilities and then you pick yourself up by your bootstraps and see how variable the data are in that framework". The frequency distribution of the statistic of interest is computed using Monte Carlo simulations on the assumption that all independent identical samplings come from a distribution F, which is unknown but is taken equal to the observed distribution F. For example, in exotoxicology, regression models are developed using data from a limited number of chemicals. The jackknife analyzes the data by eliminating one chemical at a time; the bootstrap generalizes the results to infer the conclusions we could expect if all chemicals of the same class, i.e. the complete distribution F, would have been included in the statistical analysis. Efron showed that the jackknife is only a particular case of the bootstrap and presented the

theoretical fundamentals in several papers (2,3,4) and most clearly in (1). The boostrap method can be used to estimate the uncertainty in the relationships as well as confidence limits and has been widely used not only with linear models but also with other statistics such as factor analysis, principal component analysis and others (1,5). The rest of the paper will show an application to ecotoxicological data and explain how easily the method can be applied.

THE DATA

Neely (6) recently compared 96h and 24h acute LC50 fish data with aquatic toxicity data estimated by a theoretical model based on water solubility of 24 chemicals. His results seem to prove the validity of the hypothesis that prediction of acute LC50 fish data is possible using solubility data. To compare predicted and observed results he used correlation and linear regression analysis. However, Neely used only data from 24 chemicals with solubilities ranging over several orders of magnitude and he did not prove statistically the generality of his hypothesis, since tens of other chemicals should be tested to validate the model. This effort might be time consuming and expensive. The bootstrap can be used for this purpose.

Another fact the Neely did not take into consideration is that both sets of data contained errors. The observed values have errors of observations and natural variability, the theoretical values are also uncertain since they were derived from another regression analysis, which in itself has some variability and therefore uncertainty. The bootstrap method will therefore be applied to a modified linear model which takes into account errors in both variables.

THE LINEAR MODEL

Neely (6) used regression analysis to predict the observed (96h LC50/24h LC50) ratio (R_o) from a theoretical ratio (R), which was

obtained from the simulation of a mathematical model of contaminant dynamics in fish. He expected that a one to one correspondence would yield a slope of 1.0 and an intercept of zero between the two ratios, i.e.

$$R_o = 0.0 + 1.0\ R \qquad (1)$$

The slope computed from Neely's published data using standard regression analysis is 0.845 with an intercept of 0.033; the correlation factor is 0.847. Since the slope is significantly different from 1.0, is Neely's hypothesis wrong or is there a bias in the model? The problem is that Neely used an incorrect procedure to compute the parameters of the regression model $Y = a + bX$, by assuming the X's, or independent variables, known without error. Instead, the X's, or the theoretical ratios, are estimated with error since they are computed from solubilities, measured with error, by a mathematical model, which is uncertain by definition.

The correct procedure is to use the geometric mean (GM) linear regression model (8), which takes into account that both the X's and the Y's are measured or estimated with error. The slope b is computed from the formula,

$$b = \text{sign}\ (r)\ \ (SY\ /SX\ ) \qquad (2)$$

where

$$Sx = X - (X)\ /N, \qquad (3)$$

and

$$Sy = X - (X)\ /N, \qquad (4)$$

and sign (r) is the sign, + or −, of the correlation coefficient r. N is the number of paired observations.

The intercept a is estimated as usual as

$$a = \bar{Y} - b\ \bar{X}, \qquad (5)$$

where $\bar{Y}$ and $\bar{X}$ are the averages of the Y's and X's, respectively.

Using the GM linear regression model with the data published by Neely produces a slope of 0.997 with an intercept of -0.092; the computed slope is very close to the one to one correspondence hypothesized by Neely, or

$$R_o = -0.092 + 0.997\ R \qquad r = 0.847,\ n = 24. \qquad (6)$$

THE BOOTSTRAP

The fact that the GM model has a slope very similar to the theoretical one does not imply that the model is generally valid for all chemical classes represented by the 24 chemicals used by Neely. The bootstrap can be used to seek this generalization by estimating the unknown distribution F; with the bootstrap, the frequency distribution F does not have to be assumed normal, which is very useful since chemicals with different structures and chemical properties are used in this model. The bootstrap, as mentioned before, assumes that the unknown distribution F can be estimated from the observed distribution F, i.e., we can infer from the observed data the validity of Neely's hypothesis for all other chemicals with similar properties without having to perform more experiments; the generality of the hypothesis can be inferred from the standard errors associated with the correlation and with the linear model parameters.

To perform the bootstrap test, each of the 24 data points is replicated a very large number of times, e.g., one billion times, and then this large amount of data is sampled 100 to 1000 times, the bootstrap samples. From a practical point of view the data are not really replicated a billion times, but a random number generator is used. The statistics of interest, in this case the standard errors and the confidence limits of the slope of the intercept and of the correlation coefficient, are computed for each such bootstrap sample. Given the fact that the assumption of normality has been abandoned, the

confidence limits may not be symmetrical around the mean, if the probability density function is skewed.

Results from the bootstrap calculations are shown in Table 1. The bootstrap average estimate of the correlation coefficient is 0.835 (Fig. 1) with a standard error of 0.096, the standard error is 0.062 if the assumption of normality is maintained. The one standard deviation (68%) confidence limits of the correlation are 0.76 and 0.90 while the 95% confidence limits are .48 and .98. As can be seen from these results, the distribution is skewed to the right and the limits are not symmetrical. For the GM linear model the one standard deviation confidence limits for the slope (Fig. 2) are 0.990 and 1.148 with a bootstrap average of 1.018, and for the intercept (Fig. 3) -0.170, -0.039 and -0.112, respectively.

DISCUSSION

The validity of the relation, hypothesized by Neely (6), between the water solubility of an organic chemical and the ratio of the acute fish LC50 at two different time periods has been tested using the bootstrap method. The hypothesis has been shown correct by first fitting a linear model with the GM functional regression to compare observed and predicted ratios, and the generality of the model has been tested by computing the confidence limits of the correlation coefficient, of the slope and intercept of the regression model using the bootstrap. This statistical method allows generalization from a limited set of data to chemicals not included in the original analysis (10). The results show that the correlation between predicted and observed data is statistically significant within one standard

Table I. **Bootstrap estimates and confidence limits of the correlation coefficient, r, the slope, b, and the intercept, a, of the GM functional regression model based on 1000 replicates.**

	Correlation r	Slope b	Intercept a
expected values	-	1.0	0.0
-95%	.48	.877	-.720
-1 SD	.76	.990	-.170
bootstrap average (stand. dev.)	.835 (.096)	1.018 (.151)	-.112 (.146)
+1 SD	.90	1.148	-.039
+95%	.98	1.622	+.092
GM linear model	.847 (.062)	.997	-.092
Neely's model	.847	.845	.033

Figure 1. Frequency distribution of the correlation coefficient between the predicted (R) and observed (R_o) ratios. The distribution is based on 1000 bootstrap samples. The bootstrap average of the correlation coefficient is 0.835 (95% confidence limits .48; .98).

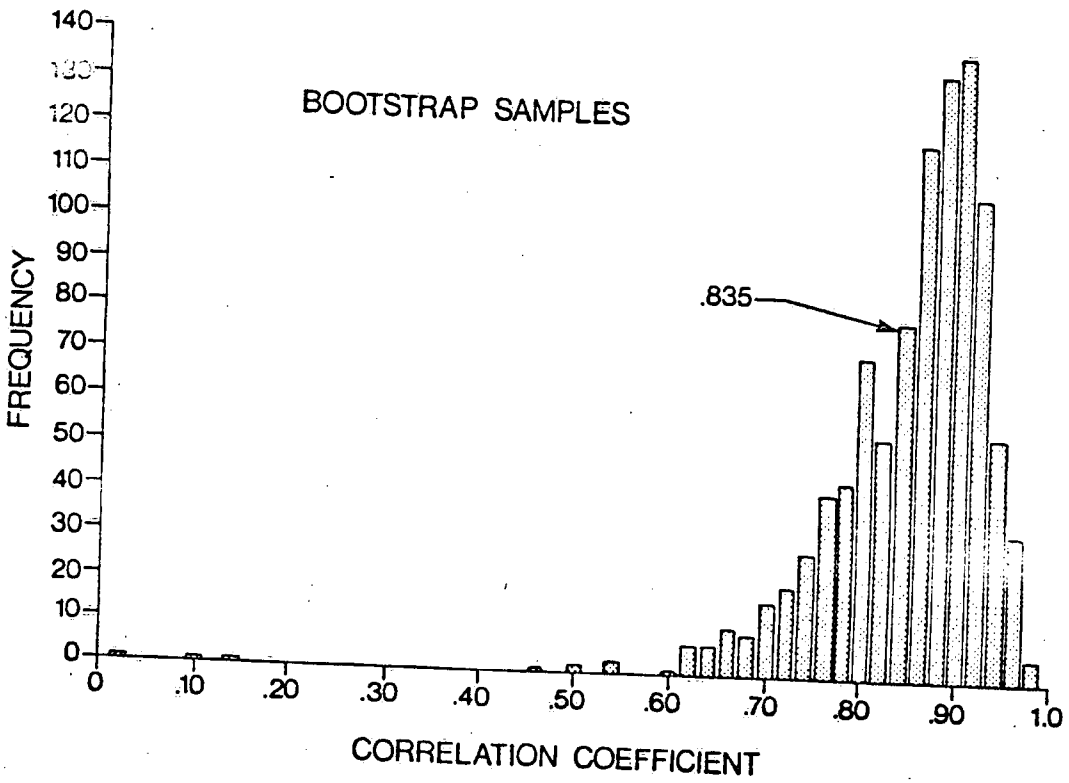


Figure 2. Frequency distribution of the slope b of the linear model $R_o=a+bR$. The distribution is based on 1000 bootstrap samples. The bootstrap average of the slope of the GM linear model is 1.018 (95% confidence limits .877; 1.622) statistically equal to the theoretical value of 1.

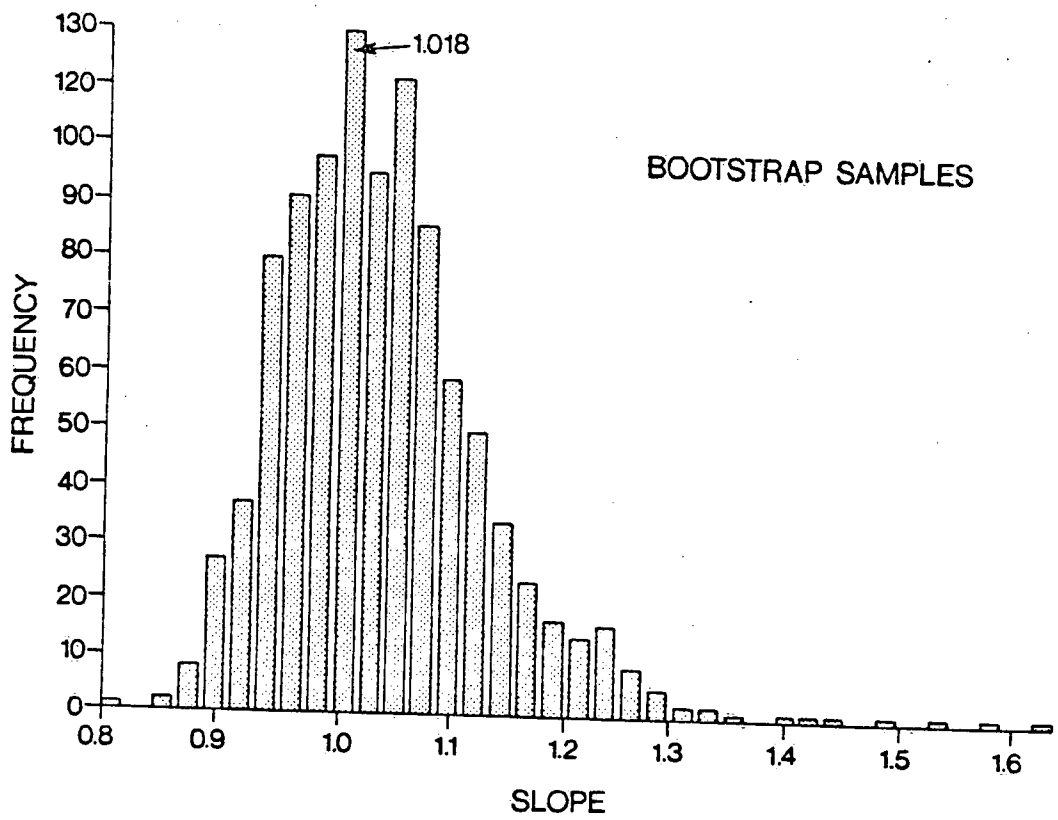
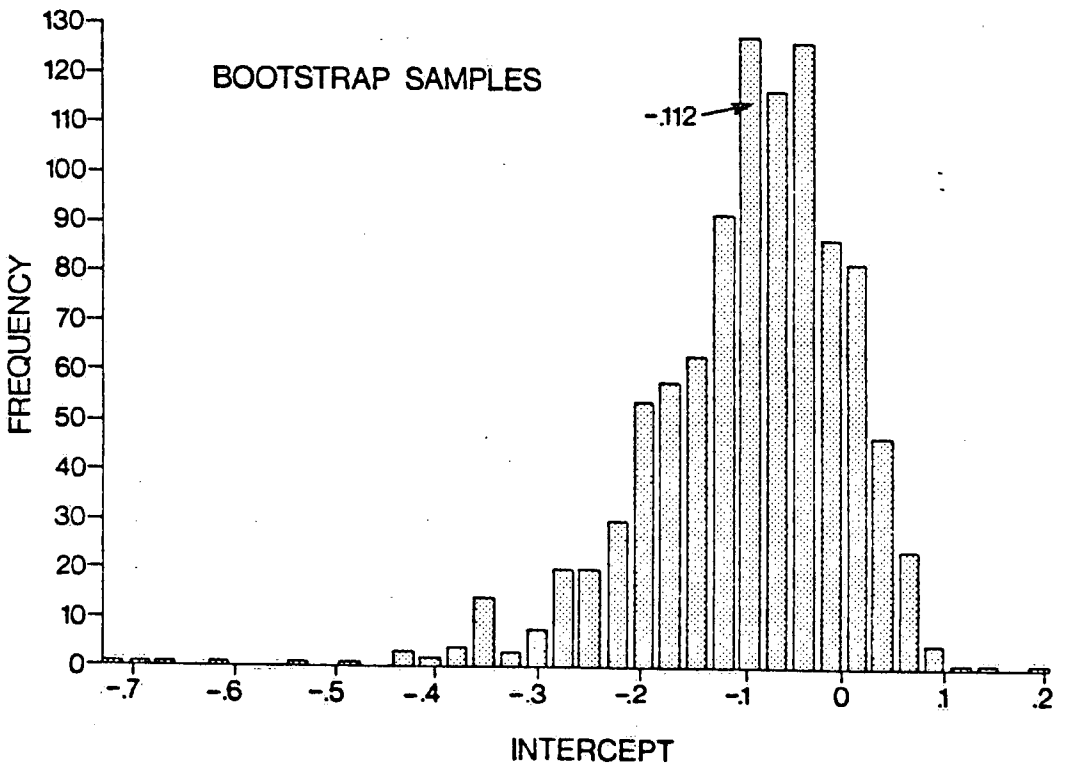


Figure 3. Frequency distribution of the intercept a of the linear model $R_o=a+bR$. The distribution is based on 1000 bootstrap samples. The boostrap average of the intercept of the GM linear model is -.112 (95% confidence limits -720; +.092) significantly different from the theoretical value of 0.0.



deviation, but sometimes it may not be significant at the 95% confidence limit.

However, since the observed distribution F is skewed to the right with an average correlation of 0.835 and an upper limit of 0.98, the theory is probably correct but for a few chemicals (the lower 95% confidence interval is 0.48). For all chemicals the predicted average ratio is 1.018 with the hypothesized 1.0 falling within one standard deviation; however, the average intercept is -0.112 and the probability of the intercept being 0.0 or positive is only about 12%. Therefore the conclusion is that the model is probably correct but it might have a systematic bias which makes the theoretical ratio somewhat higher than the observed ratio.

The GM functional regression method should always be used to compute the coefficients of a linear model when measurement errors or natural uncertainty is expected in the independent variables X as well as in the dependent variables Y. The theoretical ratios that Neely used as independent variables in the linear model were clearly uncertain; the standard linear regression method that he used produced a large underestimate of the slope, thus undermining his hypothesis whereas the correct statistical procedure showed his hypothesis right, even if the linear model had a positive systematic bias.

The bootstrap is a computer intensive statistical method that uses Monte Carlo simulations to provide information when theoretical analytical solutions are not possible, for example if the original frequency distribution is not normal. The present analysis was performed on a CDC Cyber 171 computer and it took 27 CPU seconds for 1000 replications or bootstrap samples; Efron (2,3) suggests 128 to 512 replications since the method converges asymptotically. The method is simple enough numerically that it can be programmed on a microcomputer such as an Atari 400 or a Commodore 64. The application to ecotoxicological problems is intriguing since very often correlation and regression models are published in the literature based on few data and the reliability of the results is usually difficult to establish given the diversity of chemicals. The bootstrap is an interesting

method that should be often used to establish the uncertainty of the proposed hypothesis.

REFERENCES

1. Diaconis, P. and B. Efron. Computer-intensive methods in statistics. Scientific American, 248: 116-130 (1983).
2. Efron, B. The jackknife, the bootstrap and other resampling methods. SIAM, 92 pp, (1982).
3. Efron, B. Nonparametric standard errors and confidence intervals. Can. J. Statistics 9: 139-172 (1981).
4. Efron, B. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika, 68: 589-599 (1981).
5. Kolata, G. The art of learning from experience. Science, 225: 156-158 (1984).
6. Neely, B.C. An analysis of aquatic toxicity data: water solubility and acute LC50 data. Chemosphere, 13, 813-819 (1984).
7. Quenouille, M. Approximate tests of correlation in time series. J. Royal Statist. Soc. Ser. B, 11: 18-84 (1949).
8. Teisser, G. La relation d'allometrie: sa significance statistique et biologique. Biometrics, 4: 14-48 (1948).
9. Tukey, J. Bias and confidence in not quite large samples. Abstract, Ann. Math. Statist., 29: 614 (1958).
10. A footnote. In the interviews reported by Gina Kolata (5), David Freedman states that some people were skeptical of the bootstrap since they "were afraid it was all done with mirrors". Mosteller remarked that "the bootstrap seems incestuous, since you are trying to learn about the sample error by sampling the sample and statisticians are not ordinarily involved with something as anti-intuitive as this". However, Efron articles (3,4) and book (2) present all the basic theoretical background for a complete understanding of the method and he is now working on additional theoretical statistics proofs.

9734