Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER 499744
UNCLASSIFIED	

TITLE

SO WHAT IS IT ABOUT THE BOOTSTRAP?

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to:

DEPARTMENT OF NATIONAL DEFENCE
CANADA

OPERATIONAL RESEARCH DIVISION

DIRECTORATE OF OPERATIONAL RESEARCH (JOINT & LAND)

LOGORT RESEARCH NOTE RN9606

SO WHAT IS IT ABOUT THE BOOTSTRAP?

by

M. Provencher

OCTOBER 1996

OTTAWA, CANADA

National Defence Défense nationale

OPERATIONAL RESEARCH DIVISION

CATEGORIES OF PUBLICATION

ORD Reports are the most authoritative and most carefully considered publications of the DGOR scientific community. They normally embody the results of major research activities or are significant works of lasting value or provide a comprehensive view on major defence research initiatives. ORD Reports are approved personally by DGOR, and are subject to peer review.

ORD Project Reports record the analysis and results of studies conducted for specific sponsors. This Category is the main vehicle to report completed research to the sponsors and may also describe a significant milestone in ongoing work. They are approved by DGOR and are subject to peer review. They are released initially to sponsors and may, with sponsor approval, be released to other agencies having an interest in the material.

Directorate Research Notes are issued by directorates. They are intended to outline, develop or document proposals, ideas, analysis or models which do not warrant more formal publication. They may record development work done in support of sponsored projects which could be applied elsewhere in the future. As such they help serve as the corporate scientific memory of the directorates.

ORD Journal Reprints provide readily available copies of articles published with DGOR approval, by OR researchers in learned journals, open technical publications, proceedings, etc.

ORD Contractor Reports document research done under contract of DGOR agencies by industrial concerns, universities, consultants, other government departments or agencies, etc. The scientific content is the responsibility of the originator but has been reviewed by the scientific authority for the contract and approved for release by DGOR.

DEPARTMENT OF NATIONAL DEFENCE

CANADA

OPERATIONAL RESEARCH DIVISION

DIRECTORATE OF OPERATIONAL RESEARCH (JOINT & LAND)

LOGORT RESEARCH NOTE RN 9606

SO WHAT IS IT ABOUT THE BOOTSTRAP?

by

M. Provencher

Recommended by: P.A. Vincent, LOGORT

Approved by: G. Lafond, DOR(J&L)

Directorate Research Notes are written to document material which does not warrant or require more formal publication. The contents do not necessarily reflect the views of ORD or the Canadian Department of National Defence.

OTTAWA, ONTARIO

OCTOBER 1996

ABSTRACT

The bootstrap method of re-using a sample is introduced. Its advantages and disadvantages are discussed. It is proposed as a sound method to use particularly when little is known about the type of distribution under study.

RÉSUMÉ

La méthode "bootstrap" de réutilisation d'un échantillon est introduite. Ses avantages et inconvénients sont discutés. La méthode est proposée comme étant valable particulièrement lorsqu'on connaît peu le type de distribution à l'étude.

TABLE OF CONTENTS

SO WHAT IS IT ABOUT THE BOOTSTRAP?

"Mathematics is the most exact science, and its conclusions are capable of absolute proof. But this is so only because mathematics does not attempt to draw absolute conclusions. All mathematical truths are relative, conditional"

Charles Proteus Steinmetz (1923)

I. INTRODUCTION

1. Statistical techniques are used in many areas. Surveys are designed to collect opinions in order to forecast the outcome of an election. Companies sample consumers to predict a product's appeal. In each case, the large body of data that is the target of interest is called the population while the subset selected from it is called the sample.

2. Sometimes the sample is all that you can capture from the population. The measurement of a guidance system for an aircraft may have been taken from only three existing systems in order to estimate the same measurement for other systems that might be manufactured sometime in the future. In this case, the population is conceptual.

3. The purpose of statistics is to make inferences on the population on the basis of the sample (see Figure 1).

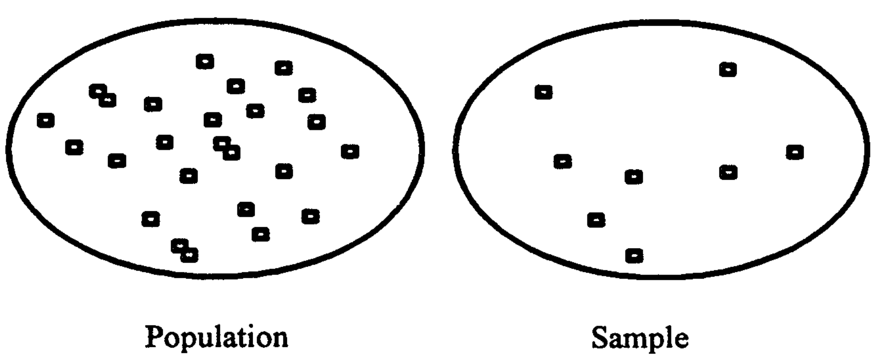

Figure 1: Population versus Sample

4. Statistical techniques have evolved over time always with the purpose of drawing as much information as possible from the sample. Financial and practical constraints, which are at the source of the development of this branch of mathematics, will always be the determinant factor in statistics. The reality is that we cannot sample as much of a population as we would like because it is either too costly or too difficult. In some cases sampling is destructive.

5. However some practical constraints which were present in the past no longer exist. This is true particularly when it comes to computation. Statistical techniques that are computationally intensive have been developed in the last few decades thanks to the power of computers. Re-sampling techniques fall under this category. The author used one such resampling technique, the bootstrap, in a study on the Scheduled Road Freight Service [1]. The current paper seeks to address the good and the bad points about the bootstrap.

II. CLASSICAL STATISTICAL TECHNIQUES

6. Classical statistical techniques have been used for a very long time. Regression analysis, for example, dates back to about the 1870s [2]. In those days most computations were performed by hand and consequently the techniques adopted reflected the existing limitations in terms of calculations. The formulas used were kept as simple as possible.

7. In almost every data analysis, the statistician observes some data (attributes or measurements) $\boldsymbol{x} = (x_1, x_2, x_3, \ldots, x_n)$ that are independent observations sampled from an unknown probability distribution function f (the population distribution). From $\boldsymbol{x}$ the statistician constructs an estimate $\hat{\theta} = t(x)$ for a parameter of interest θ. In the most familiar case, the parameter of interest θ is the true mean μ of f and the statistic $t(x)$ is the sample mean $\bar{x} = \frac{1}{n}\Sigma_{i=1}^{n} x_i$.

8. Having selected $\hat{\theta}$, a second and more difficult task arises: to assess the accuracy of $\hat{\theta}$ as an estimator of the true value θ. The standard error of $\hat{\theta}$, the square root of its variance

$$se\ (\hat{\theta}\ ; f\ ) = [var_f\ t(x)\ ]^{1/2} \tag{1}$$

is the most common measure of accuracy for estimators $\hat{\theta}$ that are unbiased or nearly so.

9. Elementary statistics focus on formulas for the standard error of $\hat{\mu} = \bar{x}$, the mean. A simple but powerful formula relates *se* $(\bar{x}\ ; f)$ to the variance of f, σ^2,

$$se\ (\hat{\mu}\ ; f) = \left[\sigma^2/n\right]^{1/2} \tag{2}$$

This looks useless for practical purposes since σ^2 is itself a function of the unknown distribution function f. However a simple unbiased estimate exists for σ^2

$$\hat{\sigma}^2 = \frac{\Sigma_{i=1}^{n}\ (x_i - \bar{x})^2}{(n-1)}$$

thus

$$\widehat{se}\ (\hat{\mu}) = \left[\frac{\Sigma_{i=1}^{n}\ (x_i - \bar{x})^2}{n(n-1)}\right]^{1/2} \tag{3}$$

10. Having estimated the parameter of interest, and assessed the standard error of the estimator, it is often desirable to build a confidence interval around the true value of the parameter. To do that, one tries to know more about the distribution function f.

11. Classical statistical techniques depend for their power on assumptions made about the data under analysis. The most common assumption is that the data follows a normal or Gaussian distribution (named after the mathematician Karl Friedrich Gauss, 1777-1855 [3]) as shown in Figure 2. We generally either assume that such an assumption is realistic for the type of data under study or we verify that the sampled data follows a relatively bell-shaped distribution. In practice, as long as, the observed distribution has one bump, we consider the assumption to be met.

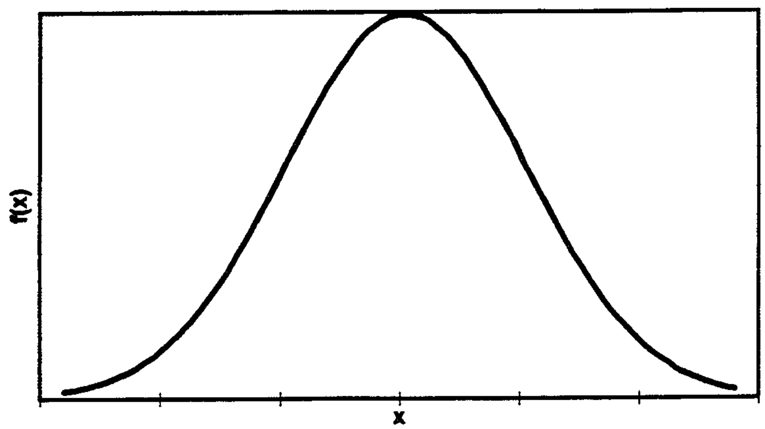

Figure 2: Normal distribution function of variable x

III. SAMPLE RE-USE TECHNIQUES AND THE BOOTSTRAP

12. Resampling techniques aim at re-using an initial sample by drawing more samples from it.

The Jackknife

13. Sample re-use methods such as the bootstrap are not entirely new, and perhaps the most well-known predecessor is the jackknife. In 1958 John Tukey revolutionized error estimation with his "jackknife" method, built upon Quenouille's older technique for estimating biases, [4]. The jackknife aims at easily obtaining trustworthy variance estimates. Tukey's method avoided formula (2) entirely, going directly to a generalization of (3).

14. Let $x_{(-i)}$ be the data sample with the i[th] observation removed,

$$x_{(-i)} = (x_1, x_2, x_3, \ldots, x_{(i-1)}, x_{(i+1)}, \ldots, x_n)$$

and let $\hat{\theta}_{(-i)} = t(x_{(-i)})$, the statistic $\hat{\theta}$ re-evaluated for the deleted-point data set $x_{(-i)}$. The jackknife estimate of standard error is

$$se_{Jack}(\hat{\theta}) = \left[\frac{n-1}{n} \Sigma_{i=1}^{n} \left(\hat{\theta}_{(-i)} - \overline{\theta_{(\cdot)}}\right)^2\right]^{1/2} \tag{4}$$

where $\overline{\theta_{(.)}} = \Sigma_{i=1}^{n} \hat{\theta}_{(-i)}/n$.

15. It is easy to verify that formula (4) reduces to formula (3) when $\hat{\theta} = \bar{x}$ (see Annex A for the details).

16. The beauty of Tukey's jackknife is that it automatically produces a standard error estimate for even the most complicated estimator $\hat{\theta}$. Its weakness is that there is no theoritical justification for the $\frac{n-1}{n}$ factor in (4). Thus the method brings "perhaps" better variance estimates. The jackknife turned out to work poorly on very un-smooth estimators like the sample median, but otherwise it seems to give generally trustworthy results [4].

The Bootstrap

17. Efron's bootstrap (1979) began as an attempt to better understand the jackknife. This involved re-examining formula (2) which had been bypassed in going to (3). Suppose that in (2) the empirical distribution function $\hat{f}$ of the observed data is such that each observation x_i $i=1, 2, ..., n$ has probability $1/n$. And that $\hat{f}$ is substitued for f. Then

$$\hat{se}(\hat{\mu}) = \left[\frac{1}{n^2} \Sigma_{i=1}^{n} (x_i - \bar{x})^2\right]^{1/2} \tag{5}$$

almost the same as the traditional estimate in (3).

18. Applying that same reasoning for the general case in formula (1), then

$$se_{Boot}(\hat{\theta}) = se\ (\hat{\theta}\ ; \hat{f}\ ) = \left[var_{\hat{f}}\ t(x)\right]^{1/2}. \tag{6}$$

This is the theoritical avenue to find the bootstrap variance (based on the initial sample alone). The second avenue and the most popular, relies on simulation. Although it is not possible to get many samples from the population with distribution function f, under the bootstrap simulation method, a multitude of samples are drawn with replacement from the observed data. One draws repeated samples of the same size as the observed sample, computes the estimator and then calculates the variance of this set of estimators.

19. The bootstrap algorithm for the estimator $\hat{\theta}$ is:

a. Create B bootstrap samples of size n by sampling with replacement from the observations. The bth bootstrap sample is denoted $X^{(b)} = \left(x_1^{(b)}, x_2^{(b)}, \ldots, x_n^{(b)}\right)$ $b = 1, 2, \ldots, B$ where each $x_i^{(b)}$ is a random selection from the original sample.

b. Compute the estimator $\hat{\theta}^{(b)}$ from each of the bootstrap samples

$$\hat{\theta}^{(b)} = t\left(x^{(b)}\right).$$

c. Use the B bootstrap estimators $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots, \hat{\theta}^{(B)}$ to calculate the simulated bootstrap standard error estimate

$$se_{Boot}(\hat{\theta}) = \left[\frac{1}{(B-1)} \Sigma_{b=1}^{B} \left(\hat{\theta}^{(b)} - \overline{\theta^{(\cdot)}}\right)^2\right]^{1/2}$$

where $\overline{\theta^{(\cdot)}} = \Sigma_{b=1}^{B} \hat{\theta}^{(b)} \big/ B.$

20. The strengths of this method are: it is easy to understand intuitively; it makes no assumption on the underlying population distribution; it is theoritically simple no matter what estimator is used; and not only does it provide a standard error estimate, but it also provides an estimate for the distribution of the estimated parameter and in particular confidence intervals around that parameter.

IV. WHEN TO USE THE BOOTSTRAP

21. The term "bootstrap" which Efron gave to his procedure is unfortunate. It has tended to convey the notion of the statistician striving to obtain something for nothing. The thinking should rather be that of a procedure by which sound statistical conclusions can often be reached, being entirely conditioned by the sample data [5].

22. The results of sample surveys are always subject to some uncertainty because only part of the population has been measured and because of errors of measurement. This uncertainty can be reduced by taking larger samples and by using superior instruments of measurement. But this usually costs time and money. Consequently this specification of the

degree of precision wanted in the results is an important step. It may present difficulties though. Formulas determining the sample size n for a relative error r in the estimated population mean exist [6]. They rely on the assumption that the mean is normally distributed, which is a reasonable assumption for a relatively large n (Central Limit Theorem). However, formulas giving the sample size n for a relative error r do not exist for all types of estimates.

23. The bootstrap technique can at times provide a less-than-accurate picture of the real data. Since we are essentially trying to construct a picture of a complete data set from only a small fragment or sample of the data set, we may be misled from time to time. The bootstrap, then, does not guarantee accuracy in our estimates, but it will give a better picture of the original data set than classical statistical techniques most of the time. The one best reason for using the bootstrap technique remains that it involves no simplying assumptions about the underlying probability distribution.

24. The key step of the bootstrap approach is to complement sampling from the population defined by f with bootstrap resampling from the data, the population defined by $\hat{f}$. When is this a reasonable thing to do? Clearly, it depends on $\hat{f}$ being a good estimator of f. Without making other assumptions about the nature of the population, such as symmetry, $\hat{f}$ is about the best we can do [7].

25. So there is no clear answer for the size of the sample. The other question is: how many bootstrap replications should we make? It is suggested in [4] that typical problems require 50 to 200 bootstrap replications to estimate a standard error, and 1000 to 2000 replications to compute a bootstrap confidence interval. The algorithm presented in para 19 has been programmed for the mean (see Annex B) and applied to the Scheduled Road Freight Service weight data of 1994 for the Toronto depot [1]. In this case, the reason for using the bootstrap was the lack of knowledge of the population. Statistical packages such as "Resampling Stats" are available to do that type of analysis. Table I presents the sample data. Figure 3 illustrates frequency distributions of the bootstrap means for various replications. In this case Figure 3 tends to confirm the number of replications suggested earlier in this paragraph. The idea is to stop replicating when the distribution is sufficiently smooth.

TABLE I
SRFS-TORONTO: SAMPLE DATA

Month (Fiscal Year 94-95)	Weight (Tons)	Month (Fiscal Year 94-95)	Weight (Tons)
April	372.27	October	327.43
May	306.36	November	279.00
June	530.32	December	187.13
July	287.55	January	264.77
August	369.20	February	234.72
September	118.86	March	182.46

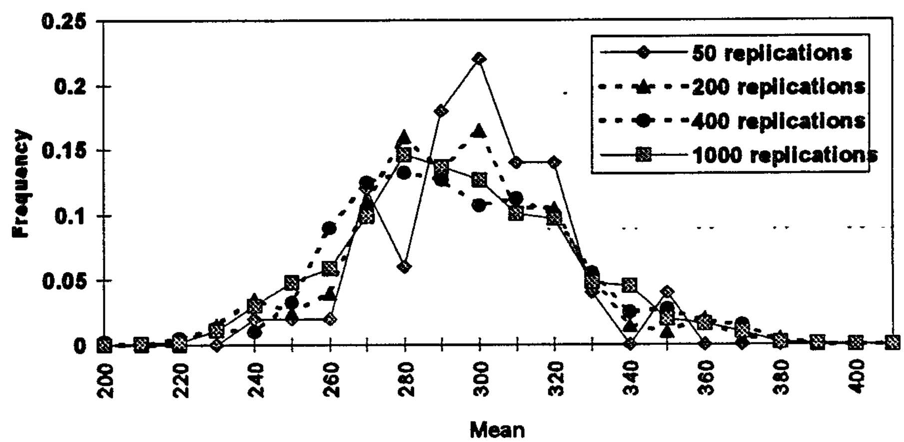

Figure 3: Frequency Distributions for different numbers of replications

26. In this case the assumption of Normality seems relatively well met but we had no prior knowledge of this. The 90% bootstrap confidence interval for the frequency distribution of the 1000 replications is: [241.85, 339.96]. The 90% traditional confidence interval using the assumption of Normality is: [237.33, 339.35], not all that different.

V. SO WHAT

27. A sample re-use technique, the boostrap, was presented. The advantage of the method is that no assumption on the nature of the sampled distribution is made. It requires a little more computational effort because of the replications made but this is negligeable in this day and age. The only reserve is the one common to any statistical analysis: is the sample representative enough of the population?

28. The bootstrap is a sound method to use (theoritically or practically) in general when full sampling is not possible but more so if little is known about the type of distribution under study.

REFERENCES

1. Provencher, M. & Ng, K.Y.K. *A Scheduled Road Freight Service Review*. ORA Project Report PR9519, September 1995.

2. Wilson, J.H. & Keating, B. *Business Forecasting*, Irwin, 1990.

3. Bell, E.T. *Men of Mathematics*, Simon and Schuster, 1937.

4. Lepage R. & Billard, L. *Exploring the Limits of Bootstrap*, John Wiley & Son, 1992.

5. Young, G.A. *Bootstrap: More than a Stab in the Dark?* Statistical Science, 1994, vol. 9, no. 3, 382-415.

6. Cochran, W.G. *Sampling Techniques*, John Wiley & Sons, 1977.

7. Stine, R. *An Introduction to Bootstrap Methods*, Sociological Methods & Research, vol. 18, nos 2 & 3, November 1989/February 1990, 243-291.

THE JACKKNIFE STANDARD ERROR ESTIMATE OF $\bar{X}$ REDUCES TO THE TRADITIONAL ESTIMATE USED

$$\bar{x}_{(.)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \neq i} \frac{x_j}{(n-1)}$$

$$= \frac{1}{n(n-1)} \left[\sum_{\substack{j=1 \\ j \neq 1}}^{n} x_j + \sum_{\substack{j=1 \\ j \neq 2}}^{n} x_j + \ldots + \sum_{\substack{j=1 \\ j \neq n}}^{n} x_j \right]$$

$$= \frac{1}{n(n-1)} \left[(n-1)\, x_1 + (n-1)\, x_2 + \ldots + (n-1)\, x_n \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

$$\frac{(n-1)}{n}\sum_{i=1}^{n}\left(\bar{x}_{(-i)}-\bar{x}_{(\cdot)}\right)^2=\frac{n-1}{n}\left[\left[\frac{1}{(n-1)}\sum_{j\neq 1}x_j-\bar{x}\right]^2+\left[\frac{1}{(n-1)}\sum_{j\neq 2}x_j-\bar{x}\right]^2+\ldots\right.$$

$$\left.+\left[\frac{1}{(n-1)}\sum_{j\neq n}x_j-\bar{x}\right]^2\right]$$

$$=\frac{(n-1)}{n}\,\frac{1}{(n-1)^2}\left[\left[\frac{\sum_{j\neq 1}x_j-(n-1)\,x_1}{n}\right]^2+\right.$$

$$\left.\left[\frac{\sum_{j\neq 2}x_j-(n-1)\,x_2}{n}\right]^2+\ldots+\left[\frac{\sum_{j\neq n}x_j-(n-1)\,x_n}{n}\right]^2\right]$$

$$=\frac{1}{n(n-1)}\left[\left[\frac{n\bar{x}-nx_1}{n}\right]^2+\left[\frac{n\bar{x}-nx_2}{n}\right]^2+\ldots\right.$$

$$\left.+\left[\frac{n\bar{x}-nx_n}{n}\right]^2\right]$$

$$=\frac{1}{n(n-1)}\left[\left(\bar{x}-x_1\right)^2+\left(\bar{x}-x_2\right)^2+\ldots\right.$$

$$\left.+\left(\bar{x}-x_n\right)^2\right]$$

$$=\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\bar{x}-x_i\right)^2$$

FORTRAN PROGRAM FOR BOOTSTRAP ESTIMATE

```
C*************************************************************************
C       Bqotstrap program developed to                                  *
C       calculate a confidence interval around the mean                 *
C       M. Provencher July 1996                                         *
C*************************************************************************

        DIMENSION X(100),XB(100),XBARB(1000)
        CHARACTER*30 DATFIL,OUTFIL
        WRITE(*,*) 'Bootstrap Analysis'
        WRITE(*,*) 'Limitations:  max 100 data, 1000 replicates'
        WRITE(*,*) 'The datafile must start with the number of data'
        WRITE(*,*) ' '
        WRITE(*,*) 'Give the name of the data file: '
        READ(*,99) DATFIL
        WRITE(*,*) 'Give the name of the output file: '
        READ(*,99) OUTFIL
99      FORMAT(A)
        OPEN(21,FILE=DATFIL)
        OPEN(22,FILE=OUTFIL)
        READ(21,*) N
        READ(21,*) (X(I),I=1,N)
        WRITE(*,*) 'Give the number of boot-replicates, max 1000: '
        READ(*,*) NB
        R=RRAND()

        DO 300 IB=1,NB

C       Make bootstrap samples

           XBARB(IB)=0.0

           DO 210 J=1,N
              R=RND()
              INDEX=INT(R*FLOAT(N))+1
              IF(INDEX.GT.N) INDEX=N
              XB(J)=X(INDEX)
              XBARB(IB)=XBARB(IB)+XB(J)
210        CONTINUE
           XBARB(IB)=XBARB(IB)/FLOAT(N)
300     CONTINUE

        AVE=0.0
        SUM2=0.0

        DO 400 IB=1,NB
           AVE=AVE+XBARB(IB)
           SUM2=SUM2+XBARB(IB)*XBARB(IB)
400     CONTINUE

        AVE=AVE/FLOAT(NB)
        VAR=(SUM2-FLOAT(NB)*AVE*AVE)/FLOAT(NB-1)
        WRITE(*,*) ' '
        WRITE(*,*) 'Bootstrap mean: ',AVE
        WRITE(*,*) 'Bootstrap variance of the mean: ',VAR
        WRITE(*,*) ' '
        WRITE(22,*) 'Bootstrap Analysis'
        WRITE(22,*) '=================='
        WRITE(22,*) ' '
        WRITE(22,*) 'Name of input file: ',DATFIL
        WRITE(22,*) 'Number of boot-replicates: ',NB
        WRITE(22,*) ' '
        WRITE(22,*) 'Bootstrap mean: ',AVE
        WRITE(22,*) 'Bootstrap variance of the mean: ',VAR
```

```
      WRITE(22,*) ' '

C       Selection sort on XBARB
C       (see Algorithms by Robert Sedgewick (1988))

      CALL SORT(XBARB,NB)
      WRITE(*,*) '   VALUE    PERCENTILE'
      WRITE(*,*) '   =====    =========='
      WRITE(22,*) '   VALUE    PERCENTILE'
      WRITE(22,*) '   =====    =========='
      DO 450 J=1,99

C       The jth percentile, j=1,2, ... ,99 is given
C       by the j(n+1)/100 th value.

         T=J*(NB+1)/100.
         II=T

C       Interpolation

         IF((T<1).OR.(T>NB)) GO TO 450
         X(J)=XBARB(II)*(II+1-T)+XBARB(II+1)*(T-II)
         WRITE(*,425) X(J),J
         WRITE(22,425) X(J),J
425       FORMAT(1X,F8.2,6X,I2)
450   CONTINUE
      WRITE(22,*) ' '
      WRITE(22,*) ' Raw Data:  Bootstrap Means '
      WRITE(22,*) ' ========================== '
      WRITE(22,*) XBARB

      STOP
      END

      SUBROUTINE SORT(X,N)
      DIMENSION X(1000)
      DO 30 I=1,(N-1)
         DO 20 J=(I+1),N
            IF(X(J)<X(I)) THEN
            T=X(J)
            X(J)=X(I)
            X(I)=T
            ENDIF
20       CONTINUE
30    CONTINUE
      RETURN
      END
```

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM
(highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared e.g. Establishment Sponsoring a contractor's report, or tasking agency, are entered in Section 8). Operational Research Division Department of National Defence Ottawa, Ontario K1A 0K2	2. SECURITY CLASSIFICATION (overall security classification of the document, including special warning terms if applicable) UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title)

So What is it About the Bootstrap?

4. AUTHORS (last name, first name, middle initial)

Provencher, M.

5. DATE OF PUBLICATION (month Year of Publication of document) October 1996	6a. NO OF PAGES (total containing information. Include Annexes, Appendices, etc.) 19	6b. NO OF REFS (total cited in document) 7

7. DESCRIPTIVE NOTES (the category of document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Research Note

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address).

LOGORT

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.) 3561-MOD-G	9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written.) ---
10a. ORIGINATOR's document number (the official document number by which the document is identified by the originating activity. This number must be unique to this document.) LOGORT Research Note RN 9606	10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor.) ---

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification.)

(X) Unlimited distribution
( ) Distribution limited to defence departments and defence contractors: further distribution only as approved
( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
( ) Distribution limited to government departments and agencies; further distribution only as approved
( ) Distribution limited to defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM

13. **ABSTRACT** (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the test is bilingual).

The bootstrap method of re-using a sample is introduced. Its advantages and disadvantages are discussed. It is proposed as a sound method to use particularly when little is known about the type of distribution under study.

14. **KEYWORDS, DESCRIPTORS or IDENTIFIERS** (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified . If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

bootstrap
jackknife
standard error
estimate

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM

NO. OF COPIES NOMBRE DE COPIES	COPY NO. COPIE Nº	INFORMATION SCIENTIST'S INITIALS INITIALES DE L'AGENT D'INFORMATION SCIENTIFIQUE
1	1	BA

AQUISITION ROUTE FOURNI PAR ▶	ORD
DATE ▶	25 Oct 96
DSIS ACCESSION NO. NUMÉRO DSIS ▶	

DND 1158 (6-87)

#499744

National Defence / Défense nationale

PLEASE RETURN THIS DOCUMENT TO THE FOLLOWING ADDRESS:

DIRECTOR
SCIENTIFIC INFORMATION SERVICES
NATIONAL DEFENCE
HEADQUARTERS
OTTAWA, ONT. - CANADA K1A 0K2

PRIÈRE DE RETOURNER CE DOCUMENT À L'ADRESSE SUIVANTE:

DIRECTEUR
SERVICES D'INFORMATION SCIENTIFIQUES
QUARTIER GÉNÉRAL
DE LA DÉFENSE NATIONALE
OTTAWA, ONT. - CANADA K1A 0K2

Canada