A Generalized Bootstrap Technique for Dependent Observations

[1]Roberto N. Padua [2]Mark S. Borres and [3]Randy K. Salazar

Abstract

The bootstrap method for re-sampling essentially obtains the re-sampled observations from the empirical distribution function of the original data. The method relies heavily on the assumption of independence of the observations (iid). When the original data are correlated, then the usual bootstrap technique may fail to give appropriate re-sampled data. The present study proposes a new method for generating bootstrap observations from dependent observations knowing the original correlation structure of the data. Independent and identically distributed initial bootstrap samples are obtained from the empirical cumulative distribution function of the data. The bootstrap re-samples from the original data are obtained from the space generated by the initial bootstrap subsamples. It is shown that the correlation structure of the bootstrap samples obtained is the same as the original data. Simulations show that the relative error and the mean-squared error decrease with increasing sample size. However, both types of error increase with increasing dimensionality of a multivariate normal distribution

Keywords and Phrases: bootstrap, dependent observations, subspace, Cholesky's method, LU-Decomposition

1.0 Introduction

Jackknife and bootstrap re-sampling procedures have gained widespread popularity since the seminal paper of Efron (1987) as methods for generating random samples "without getting fresh observations". Essentially, the methods obtain a "new" set of observations from a given random sample x1, x2, … ,xn by re-sampling from these observations. In the Jackknife method, sample of size k < n can be generated by obtaining all possible sub-sample of size k from n i.e. (nk) possible jackknife sub-samples. In the bootstrap method, one sets up the empirical cumulative distribution function Fn(x) of the sample x1,…,xn and uses this empirical cdf to generate "new" samples of any size.These methods are particularly useful when the available information is limited to the fixed random sample x1,…,xn and it is desired to study the behavior of an estimator Tn of a parameter θ, viz mean, variance and bias of Tn for a fixed sample size k<n. However, both methods apply only when x1, x2,…,xn are assumed iid (independent and identically) distributed F(x). When either the assumption of independence or identical distribution is violated, these re-sampling procedures will fail.

When the observations are correlated, a direct application of the bootstrapping technique destroys this correlation structure, i.e. the bootstrap sample will become independent. Several modifications to the bootstrap procedure had been proposed in the past to deal with various types of structured correlations among the observations. Chen (2003) proposed a bootstrap

[1]Consultant,
[2, 3]University of San Jose Recoletos

procedure for - mixing processes; Lomangcaya (2010) studied the situation when the observations obey an AR (1) model.

The dependence structure of the data could, however, be quite general and not necessarily following an AR model nor a - mixing process. More specifically, suppose that the data X = (x1, x2…xn) have:

(3) cov (x) = Σ

where Σ is an nxn positive- definite matrix. We are asked to obtain bootstrap samples Y = (y1, y2…yn) such that the covariance structure is preserved:

(4) cov (Y) = cov (X) = Σ.

The problem is complex and there is no obvious way of approaching it.

At the outset, we see that there are $\left(\frac{n(n+1)}{2}\right)$ unknown parameters from the covariance matrix Σ, and with only n observations, these parameters are clearly not estimable. This paper is concerned with the more general problem of obtaining a bootstrap sample from a set of observations that have a general covariance structure Σ.

2.0 Literature Review

The empirical cdf $F_n(x)$ is used as a natural estimator of F(x) for each x. When the observations are independent, the quantity $nF_n(x)$ is a binomial random variable with parameters n and p = F(x). Also, $\sigma^2 = F(x)(1 - F(x), \quad \mu = n\,F(x)$.

For large n, the Strong Law of Large Numbers (SLLN) tells us that:

(5) $$F_n(x) \xrightarrow{p} F(x) \text{ as } n \to \infty.$$

Not only is this true, the Glivenko-Cantelli theorem also shows that:

(6) $$\underset{x}{sup}|F_n(x) - F(x)| \to 0 \; as\; n \to \infty.$$

An estimate of just how close Fn(x) is to F(x) was in fact, provided by Dvoretsky – Kiefer – Wolfwitz (1956):

(7) $$\Pr\left(\underset{x}{sup}|F_n(x) - F(x)| > \varepsilon\right) \leq \; C$$

The unspecified constant C was later obtained by Pascal Massart (1990) to be:

(8) $$\Pr\left(\underset{x}{sup}|F_n(x) - F(x)| > \varepsilon\right) \leq \; 2e^{-2n\varepsilon^2} \qquad \forall\varepsilon > 0.$$

This strengthens the Glivenko-Cantelli Theorem by quantifying the rate of convergence as n→ ∞ .

The Massart (1990) result is useful when, for example, we wish to determine just how large a sample n is required to estimate F(x) to with $\varepsilon = \frac{1}{10}$ in with at least 90% confidence. This turn out to be n by using the Massart (1990) bound.

Dependent Observations

Suppose now that $x_1, x_2, \ldots, x_n$ are identically distributed dependent observations with $\sigma_{ij} = cov(x_i x_j) \neq 0$. for as long as $\rho_{ij} \neq 1 \; or \; \rho_{ij} \neq -1$, the observations $x_1, x_2, \ldots, x_n$ can still be used to construct the estimator $F_n(x)$. However , the sample size n will be larger than the sample size needed to achieve the same level of accuracy when they are independent. To see this, treat the empirical cdf$F_n(x)$ as an average of the observations $I(x_i),\ldots, I(x_n)$. Then, by the Chebychev's inequality:

(9) $$\Pr\left(|F_n(x) - F(x)| > \varepsilon\right) \leq 1 - \; Var(Fn(x))/\varepsilon^2 \qquad \forall\varepsilon > 0.$$

However, $Var(F_n(x)) = \sum_1^n var(I(xi))/n^2 + 2\sum_{i \neq j} cov(I(xi), I(xj))/n^2$

Thus,

(10) $1/n\,[F(x)(1-F(x)] + 2/n^2 \sum_{i,j} \sigma ij \leq \alpha\varepsilon^2$, which we can use to solve for n. Obviously, this sample size n will be larger than when the second term on the left hand side of the inequality (10) were not present.

Bootstrapping Based on Autoregressive Models

We review Lomangcaya's (2010) procedure for generating bootstrap samples for an autoregressive process. Consider a first-order autoregressive process:

(7) $X_t = \varphi_1 X_{t-1} + \varepsilon_t$, where ε_t are iid with zero mean and constant variance σ^2 .

We can , of course, consider a Gaussian process regression and then generate replicates of the process. We shall, however, use a simpler bootstrap regression approach as follows:

We compute the variance and covariance of the process as follows:

(8) $\mathrm{var}(X_{t)} = \nu(0) = E(\varphi_1 X_{t-1} + \varepsilon t)(\varphi_1 X_{t-1} + \varepsilon_{t'}) = \varphi_1 \nu(0) + \sigma^2$,

Hence: $(1-\varphi_1)\nu(0) = \sigma^2$ or $\nu(0) = \sigma^2/\,(1-\varphi_1)$.

Similarly,

(9) $\nu(1) = E(X_t X_{t-1}) = E(\varphi_1 X_{t-1} + \varepsilon_t)(X_{t-1}) = \varphi_1 \nu(0)$,

Hence: $\nu(1) = \varphi_1 \sigma^2/\,(1-\varphi_1)$ and more generally:

(10) $\nu(k) = \varphi_1^{\,k} \sigma^2/\,(1-\varphi 1)$.

The autocorrelation function can be easily derived from (10):

(11) $\rho_k = \varphi_1^{\,k}$ for $k = 1,2,\ldots$

We now attempt to fit a regression model using $Y = X_t$ and $X = X_{t-1}$.

In order to avoid the regression constant, center the data so that $Y_t = X_t - \mu_t$. The following steps will be followed:

1. Fit the model and retain the fitted values and the residuals $\hat{\epsilon}_i = y_i - \hat{y}_i, (i = 1, \ldots, n)$.

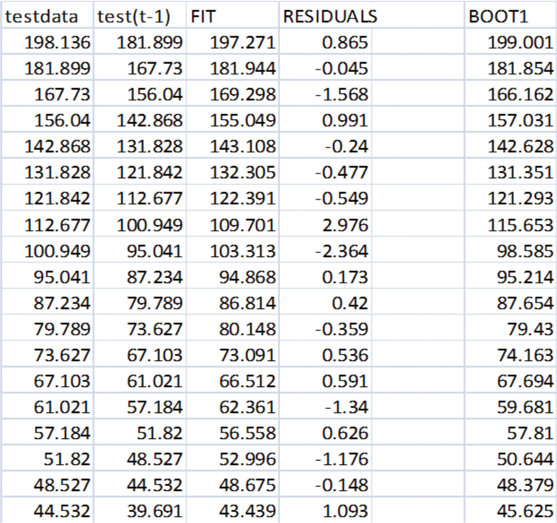
2. For each pair, (xi,yi), in which xi is the (possibly multivariate) explanatory variable, add a randomly resampled residual, $\hat{\epsilon}_j$, to the response variable yi. To resample from the residuals (which are now serially correlated), use the bootstrap algorithm developed in Section 3. In other words create synthetic response variables $y_i^* = y_i + \hat{\epsilon}_j$where j is selected randomly from the list $(1, \ldots, n)$for every i.

3. Refit the model using the fictitious response variables y_i^*, and retain the quantities of interest (often the parameters, $\hat{\mu}_i^*$, estimated from the synthetic y_i^*).

Repeat steps 2 and 3 as many times as bootstrapped subsamples are desired.

Sample Computations:

The following data were generated using Minitab. We performed the usual regression analysis on Xt and $X_{t-1.}$ Table 1 shows the data (X_t), data at lag 1 (X_{t-1}), the fitted values of X_t, the residuals and one bootstrap sample.



testdata	test(t-1)	FIT	RESIDUALS	BOOT1
198.136	181.899	197.271	0.865	199.001
181.899	167.73	181.944	-0.045	181.854
167.73	156.04	169.298	-1.568	166.162
156.04	142.868	155.049	0.991	157.031
142.868	131.828	143.108	-0.24	142.628
131.828	121.842	132.305	-0.477	131.351
121.842	112.677	122.391	-0.549	121.293
112.677	100.949	109.701	2.976	115.653
100.949	95.041	103.313	-2.364	98.585
95.041	87.234	94.868	0.173	95.214
87.234	79.789	86.814	0.42	87.654
79.789	73.627	80.148	-0.359	79.43
73.627	67.103	73.091	0.536	74.163
67.103	61.021	66.512	0.591	67.694
61.021	57.184	62.361	-1.34	59.681
57.184	51.82	56.558	0.626	57.81
51.82	48.527	52.996	-1.176	50.644
48.527	44.532	48.675	-0.148	48.379
44.532	39.691	43.439	1.093	45.625

3. The Proposed Bootstrap Method for General Data Dependentce Structure

Let (x1, x2,…,xn)be identically distributed, F(x), but dependent observations with an unknown covariance structure Σ such that ρij≠ 1 or -1. Without loss of generality, we assume that the observations are standardized so that $E(x_i) = 0 \; and \; Var(x_i) = 1$ Form the empirical distribution function Fn(x) as in (1) and generate the independent bootstrap sample $x_1^*, x_2^*, \ldots, x_n^*$. .

Let:

(9) B = $\{(x_1^*, x_2^*, \ldots, x_n^*) \,/\, x_i^* \; are \; bootstrap \; samples\}$

be the set of n- dimensional bootstrap samples generated from Fn(x). Consider the subspace D of B where:

(10) D = $\{(y_1, y_2, \ldots ,y_n)/ \, y_i = \sum a_i x_i^*, \; a_i \in R\}$

of constant multiples of each element of the vector $(x_1^*, \ldots, x_n^*)$. We attempt to construct bootstrap samples that preserve Σ based on D by finding appropriate values . More succinctly, we want to find replicates of the original observations that preserve the covariance structure.

In order to fix the ideas, consider the problem of finding replicates of (x_1,x_2) where the correlation ρ12 = corr(x_1,x_2). We proceed as follows: Generate two independent bootstrap samples x_1* and x_2* (note that in this case, the values of x_1* and x_2* may be x_1 or x_2 only).

Form:

(11) $y_1 = a_1x_1^* + a_2x_2^*$ and
$y2 = b_1x_1^* + b_2x_2^*$.

We can, without loss of generality, assume that $var(x_1) = var(x_2) = 1$ and $E(x_1) = E(x_2) = 0$.

From (11), we find that:

(12) $corr(y_1,y_2) = a_1b_1 + a_2b_2 = corr(x_1,x_2) = \rho_{12}$
$var(y_1) = a_1^2 + a_2^2 = 1 = var(x_1)$, and
$var(y_2) = b_1^2 + b_2^2 = 1 = var(x_t)$.

These now mean that we need to solve the system of non-linear equations:

(13) $a_1b_1 + a_2b_2 = \rho_{12}$
$a_1^2 + {}_{a2}^{2} = 1$
$b_1^2 + b_2^2 = 1$

From the second equation, let a2 =$\sqrt{1\text{-}a_1^2}$ and arbitrarily set b1= 0 (so that b2 = 1), then:

(14) $a_1 = \sqrt{1-\rho_{12}^2}$, $a_2 = \rho_{12}$, $b_1 = 0$, $b_2 = 1$

Hence,

(15) $y_1 = \sqrt{1-\rho_{12}^2}\, x_1^* + \rho_{12}\, x_2^*$
$y_2 = \rho_{12}x_2^*$

In order to generalize to the n-dimensional case, we need to realize that there are n2 unknown constants while there are n(n-1)/2 equations relating covariance of the y's to the original x's plus n equations for the variances or a total of n(n+1)/2 equations. Thus, we have an over-determined system. We can therefore set the values of some constants arbitrarily to zero so that the number of equations will equal to the number of unknowns. To this end, we set:

(16) $y_1 = a_{11}x_1^* + a_{12}x_2^* + a_{13}x_3^* \ \dots + a_{1n}x_n^*$

$y_2 = a_{22}x_2^* + a_{23}x_3^* + \dots + a_{2n}x_n^*$

$y_3 = a_{33}x_3^* + \dots + a_{3n}x_n^*$

$y_n = a_{nn}x_n^*$

which can be conveniently written in matrix form as:

(17) Y = AX* where A = $\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{bmatrix}$ is an upper triangular matrix.

It follows that:

(18) ∑ = cov(Y) = cov(X) = Acov(X*)A′ = AA′ since cov(X*) = I.

If cov (X) = ∑, then equation (18) simplifies to: ∑ = AA′. We also note in passing that since the original observations had been standardized, cov(X) = corr(X) = ρ. Equation (18) can now be solved for the entries of the upper triangular matrix A. We go back to this computational problem later. For now, we state the following main theorem for our generalized bootstrap for dependent observations which is proved by Equation (18).

Theorem 1.Let $(x_{1,} x_{2,} \dots, x_n)$ be identically distributed but dependent observations from F(x). Let $(x_1^*, x_2^*, \dots, x_n^*)$be n independent bootstrap samples from $F_n(x)$, the empirical cdfof $(x_{1,} \dots, x_n)$. Then, the bootstrap sample $(y_1,\ y_2,\ \dots\ , y_n)$ where $y_i = \sum_1^n a_{ij}x_j{}^*$ preserves the correlation structure Σ of the original observations where aij are solutions to Equation(18).

We now go back to the computational linear algebra problem of decomposing a positive-definite matrix into a lower triangular matrix (A) times an upper triangular matrix (A′). In linear algebra, the LU decomposition attempts to write a matrix as the product of a lower triangular matrix and an upper triangular matrix. The decomposition ∑ = AA′ is called the Cholesky decomposition. The Cholesky decomposition will always exists and is unique. Furthermore, it is known that computing the Cholesky decomposition is far more efficient and numerically more stable than computing the general LU decomposition problem. The method was discovered by André-Louis Cholesky(1948) for real matrices and gives an example of a square root of a matrix. When it is applicable, the Cholesky decomposition is shown to be roughly twice as efficient as the LU decomposition for solving systems of linear equations.

The Cholesky-Banachiewicz and Cholesky-Crout algorithms

If we write out the equation ∑= A =LL*,

The Cholesky-Banachiewicz and Cholesky-

$$\mathbf{A} = \mathbf{L}\mathbf{L}^{\mathbf{T}} = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix} = \begin{pmatrix} L_{11}^2 & & (symmetric) \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$L_{i,j} = \frac{1}{L_{j,j}} \left(A_{i,j} - \sum_{k=1}^{j-1} L_{i,k}L_{j,k} \right), \qquad \text{for } i > j.$$

Crout algorithms

If we write out the equation ∑= A =LL*,

we obtain the following formula for the entries of L:

$$L_{i,i} = \sqrt{A_{i,i} - \sum_{k=1}^{i-1} L_{i,k}^2}.$$

So we can compute the (i, j) entry if we know the entries to the left and above. The Cholesky-Banachiewicz algorithm starts from the upper left corner of the matrix L and proceeds to calculate the matrix row by row.

MATLAB Syntax for Cholesky Decomposition

cholcov-Cholesky covariance decomposition

Syntax

```
T=cholcov(SIGMA)
[T,num] = cholcov(SIGMA)
[T,num] = cholcov(SIGMA,0)
```

Description

T = cholcov(SIGMA) computes T such that SIGMA = T'*T. SIGMA must be square, symmetric, and positive semi-definite. If SIGMA is positive definite, then T is the square, upper triangular Cholesky factor. If SIGMA is not positive definite, T is computed from an eigenvalue decomposition of SIGMA. T is not necessarily triangular or square in this case. Any eigenvectors whose corresponding eigenvalue is close to zero (within a small tolerance) are omitted. If any remaining eigenvalues are negative, T is empty.

[T,num] = cholcov(SIGMA) returns the number num of negative eigenvalues of SIGMA, and T is empty if num is positive. If num is zero, SIGMA is positive semi-definite. If SIGMA is not square and symmetric, num is NaN and T is empty.

[T,num] = cholcov(SIGMA,0) returns num equal to zero if SIGMA is positive definite, and T is the Cholesky factor. If SIGMA is not positive definite, num is a positive integer and T is empty. [...] = cholcov(SIGMA,1) is equivalent to [...] = cholcov(SIGMA).

Measure of "Goodness"

The test for determining whether or not the yi's forms a "good enough" bootstrap replacement of the xi's is to assess how well Σx has been reproduced by the yi's. Let Σx be the covariance structure of the x's and let Σy denote the covariance of the bootstrap sample yi's. We look at

$$(17)\quad \left\|\Sigma_x - \Sigma_y\right\|_p = d_p \quad , p = 1{,}2\ and\ \infty$$

as a measure of goodness of the sub-sampling procedure.

The usual definitions of the matrix norms $\|A\|_p \quad , \mathrm{p} = 1, 2, \infty$ are:

$$(18)\ \|A\|_1 = max \sum_{1\leq i\leq n} |a_{ij}|\ \text{(Column sum maximum)}$$

$$\|A\|_\infty = max \sum_{1\leq j\leq n} |a_{ij}|\ \text{(Row sum maximum)}$$

$$\|A\|_2 = \sqrt{\lambda_{max}}\ \ where\ \ \lambda_{max} = largest\ eigvalue\ of\ A$$

We shall adopt the following measure of accuracy:

$$(19)\quad \mathrm{RE} = \frac{\left\|\Sigma_x - \Sigma_y\right\|}{\|\Sigma_{\mathrm{x}}\|} \mathrm{x}\ 100\%$$

Since the matrix norms obey the following relation:

$$(20)\ .\quad \|A\|_2 = \sqrt{\|A\|_1 \|A\|_\infty}$$

It follows that we can adopt. as our norm.

4.0 Simulation Results

We generated observations from a 3x3 multivariate normal distribution with an equicorrelated structure using different sample sizes n = 10, 20, 30. For each sample size, we did 1000 runs and computed the relative error (RE) and Mahalanobis distance for each. The results are shown in Table 1.

Table 1.

Sample Size	Relative Error	Mean-Squared Error (MSE)
10	12.58%	0.441
20	3.38%	0.122
30	1.11%	0.045

We note that both the relative error and the mean-squared error monotonically decrease with increasing sample size.

Next, we generated observations from a 4 x 4 and a 5 x 5 multivariate normal distribution with the same sample sizes and observed the relative errors and mean-squared errors for comparison purposes. The results are shown in Table 2.

	3-Variate Normal		4-Variate Normal		5-Variate Normal	
Sample Size	Relative Error	Mean-Squared Error	Relative Error	Mean-Squared Error	Relative Error	Mean-Squared Error
10	12.58%	0.441	14.67%	0.667	21.23%	0.922
20	3.38%	0.122	4.87%	0.334	5.38%	0.444
30	1.11%	0.045	1.95%	0.158	2.25%	0.201

For one type of multivariate normal distribution, both types of errors decrease monotonically with increasing sample size. However, when these figures are compared across the different distributions, we see that the errors increase with increasing dimensionality of the multivariate normal random variables.

5.0 Conclusions

The decomposition of the covariance matrix of dependent random observations using the LU-Decomposition technique or the Cholesky method yields a practical approach for obtaining bootstrap samples from a given set of dependent observations.

References

Davis, T. A., UMFPACK Version 4.6 User Guide (Dept. of Computer and Information Science and Engineering, Univ. of Florida, Gainesville, FL, 2002.

Kolman, Bernard. Elementary Linear Algebra (New York: McGrawhill, 1998)

Johnson, D. and Wichern, W. Applied Multivariate Statistics (New York: McMillan, 1987)

Lomangaya, Wilma. Sampling from a Univariate Time Series. (Unpublished Doctoral Dissertation, 2010, University of Rizal System)