UNIVERSITY OF GÖTEBORG

Department of Statistics

RESEARCH REPORT 1993:3
ISSN 0349-8034

RESAMPLING PROCEDURES IN LINEAR MODELS

by

Martin Gellerstedt

Statistiska institutionen
Göteborgs Universitet
Viktoriagatan 13
S-411 25 Göteborg
Sweden

RESAMPLING PROCEDURES IN LINEAR MODELS

by

Martin Gellerstedt

Department of Statistics
Viktoriagatan 13
S-411 25 Göteborg
SWEDEN

ABSTRACT

We will study here different resampling procedures for creating confidence sets in linear models. A special technique called abstract resampling makes it possible to use the true residuals and the true model for resampling. This may seem to be peculiar since the true residuals contains unknown parameters and thus are non observable; but for each specified parameter value the residuals are observable and can be used for resampling. Furthermore simulating the null distribution of some appropriate statistic gives the possibility to test the accuracy of a hypothetic parameter value. Finally a confidence set can be created by finding the parameter values which can not be rejected.

Bootstrapping the true residuals will be called abstract bootstrapping. We will show that the abstract bootstrap method is closely related to a permutation method.

A balanced abstract bootstrap method will also be presented, a method which treats the grand mean in linear models and can be applied in ordinary bootstrapping as well.

The resampling methods; bootstrap, abstract bootstrap and the permutation method are all closely related. Which method to use is discussed from a practical point of view.

CONTENTS

1. INTRODUCTION

During the last decade computer power has increased enormously and therefore the interest in computer intensive methods has grown. We will study here some of these methods especially for creating confidence sets. One of the great advantages with these methods is that they demand minimal assumptions about distributional forms.

For creating a confidence set we need information about the variability of the random variable studied. One way to get this information is to use resampling methods for example bootstrapping. Assume that X is a random variable with some unknown distribution function F, furthermore assume that θ is the parameter of interest and T(X) its estimate. The bootstrap method can be illustrated in step by step the following way:

θ is the parameter of interest and T(X) an estimate of θ
A sample X $\Rightarrow$ estimated distribution $\hat{F}$, observation T(X)
estimated distribution $\Rightarrow$ new samples X*
new samples X* $\Rightarrow$ new observations T(X*)
new observations T(X*) $\Rightarrow$ information about the variation of T(X)
information about the variation $\Rightarrow$ confidence set for θ

The crucial and most difficult step is how to use the information about the variation. The problem is that the new observations and the information about the variation are produced from the estimated model and not from the true model. It is difficult to understand and to calculate the relation between the variation in the estimated model and the variation in the true model.

In some situations it is possible to do the resampling in another way. For each resample find the parameter which forces the new resampled statistic T(X*) to be equal to the original one T(X). The parameters found are possible values of the true parameter θ. As we shall see this generated sequence of possible true parameter values estimates the likelihood function. This resampling technique will be called abstract resampling, because the statistic is abstract and non observable until we specify a parameter value.

2. ABSTRACT SIMULATION TECHNIQUE

2.1 Abstract samples

One technique for simulating outcomes of a random variable is to simulate $U_1, U_2, \ldots, U_n$ independently and uniformly distributed in the interval [0,1] and then transform these values according to the actual distribution; bearing in mind that if X is a random variable and F its c.d.f. then F(X) is uniformly distributed in the interval [0,1], in other words $F^{-1}(U_i)$ has the same distribution as X.

Example 1. Assume that X is binomially distributed with known parameters n and p. Simulating an outcome x is easily done by simulating n independent values uniformly distributed in the interval [0,1] and then count the number of values less than p.

Generally, assume that X is a random variable with a distribution possible to simulate by transforming U's. This means that each sequence of U's is a potential outcome for X. For a given U sequence the outcome depends on the value of the parameter(s) belonging to the actual distribution. The outcome is undecided until we specify the parameter value i.e. the outcome is abstract.

Example 2. Assume that X is binomially distributed with parameters n=10 and with an unknown p-value. A sequence U_1,U_2 , ... , U_{10} is a potential outcome of X, an abstract outcome, which will be undecided until we specify the parameter p. Observe that we have the possibility to get exactly the outcome that we want. Assume that we want x=3, then we just have to choose a p-value such that exactly three observations are less than this value, i.e. choose a p-value in the interval $[U_{(3)}$, $U_{(4)}]$. Observe that it is possible to get all of the outcomes in the sample space of X: for x=1,2,...,9 choose a p-value in the interval $[U_{(X)}$, $U_{(X+1)}]$, for x=0 choose a p-value$\in[0,U_{(1)}]$ and for x=10 choose a p-value $\in[U_{(10)},1]$, where $U_{(X)}$ is the X:th ordered value.

Example 3. Assume that X is normally distributed with parameters $\sigma=1$ and with an unknown μ value. If Z has standard normal distribution, then Z is a potential outcome of $X=\mu+Z$. The value of X is controllable, we can get exactly the value x that we want by choosing $\mu=x-Z$. For each simulated Z we get a possible value of μ to have produced the outcome x.

Assume that X is a random variable with some distribution depending on the parameter θ. Let x denote the original outcome and let $X^*(\theta)$ denote an abstract outcome. Furthermore assume that we repeatedly simulate U-sequences and for each sequence find the θ'-value that makes $X^*(\theta)=x$. The series of θ'-values are all possible true θ-values to have produced the outcome x. A confidence set is created by sorting these values and rejecting the most extreme ones .

2.2 Abstract simulation versus maximum likelihood

We will here examine the relation between the abstract simulation technique and the maximum likelihood theory. Suppose that $f(\theta,x)$ is the frequency or density function of the random variable X, where θ is the true parameter belonging to the set Θ. Consider the likelihood function $L(x,\theta)=f(\theta,x)$ as a function of θ for fixed x. Here x is thought of as an observation obtained in an experiment. In the discrete case $L(x,\theta)$ gives the probability of observing x. Thus we can regard $L(x,\theta)$ as a measure of how likely θ is to have produced the observation x.
The method of maximum likelihood consists of finding the value $\hat{\theta}$ which is most likely to have produced the observation x.

$$\hat{\theta}: \; L(x,\hat{\theta}) = f(\hat{\theta},x) \geq L(x,\theta) = f(\theta,x) \quad , \text{ for all } \theta\in\Theta .$$

The maximum likelihood estimate $\hat{\theta}$, the most likely parameter to have produced the observation, can also be found by keeping the x value fixed and simulate θ' values by abstract simulation. The most frequented θ' value gives us the maximum likelihood estimate $\hat{\theta}$.

Example 4. Assume that $X_1,X_2, \ldots, X_n$ are independent normally distributed random variables all with the parameters μ (unknown) and $\sigma=1$. As a statistic we will use $\bar{X}$ which is the maximum likelihood estimate of μ. Simulate n independently distributed normal standard variables $Z_1,Z_2, \ldots, Z_n$ and let

$$X_1^*=\mu'+Z_1 \, , \, X_2^*=\mu'+Z_2 \, , \ldots , \, X_n^*=\mu'+Z_n \quad , \quad \bar{X}^*=\frac{1}{n}\sum_{i=1}^{n} X_i^* \; .$$

The sample $X_1^*,X_2^*, \ldots, X_n^*$ is abstract and can be used for generating normally distributed samples. The outcome depends on which specific μ' values we choose, the most interesting μ' values are the values that make $\bar{X}^*=\bar{X} \Leftrightarrow \bar{X}=\mu+\bar{Z} \Leftrightarrow \mu'=\bar{X}-\bar{Z}$. Unconditionally the expectation $E[\mu']=\mu$, let $\bar{x}$ be an observation then conditionally $E[\mu' \mid \bar{X}=\bar{x}]=\bar{x}$. Also observe that $VAR[\mu' \mid \bar{X}=\bar{x}]=\frac{1}{\sqrt{n}}=VAR[\bar{X}]$.

Thus the expectation of our generating parameter μ' conditionally gives the maximum likelihood estimate and unconditionally the true parameter value μ. The variance result also indicates that if we can create a confidence interval by studying μ' we will get the same interval as the exact interval found by ordinary normal distribution theory $\bar{X} \pm z_{(1-\alpha/2)} \frac{1}{\sqrt{n}}$.

As we have pointed out the most frequented θ' value gives the maximum likelihood estimate $\hat{\theta}$. A more general result is that the generated sequence of possible parameters in fact estimates the whole likelihood function. This means that abstract simulation makes it possible to find the likelihood function with simulations instead of calculations. A confidence interval/set can be found by studying the integral of the likelihood function. When the likelihood function is found by abstract resampling, studying the integral means that we should study the generated sequence of possible parameters. A confidence set is found by rejecting the most extreme parameter values.

2.3 Creating confidence intervals by abstract simulation

Assume that X is a random variable with some distribution function F_θ , where θ is an unknown one dimensional parameter. Furthermore let T(X) be a statistic and assume that a large/small T(X) value indicates a large/small θ value.

For a given observation t(x):
The lower confidence limit θ_{low} , is found by

$$P_{\theta_{low}}(T(X) \geq t(x)) = \alpha/2$$

Observe that θ_{low} is the largest θ value that makes the outcome t(x) or more extreme (larger) outcomes unlikely. Analogously the upper confidence limit θ_{upp}, is found by

$$P_{\theta_{upp}}(T(X) \leq t(x)) = \alpha/2 .$$

Let $X^*(\theta)$ be an abstract variable and assume that $X^*(\theta)$ increases with θ. Define: $\theta_{inf} = \inf\{ \theta' ; t(X^*(\theta')) \geq t(x) \}$ and
$\theta_{sup} = \sup\{ \theta' ; t(X^*(\theta')) \leq t(x) \}$

Then $P(\theta_{inf} \leq \theta_{low}) = \alpha/2$
$P(\theta_{sup} \geq \theta_{upp}) = \alpha/2$

These results are easily motivated by the inequality relation
$\theta_{inf} \leq \theta_{low} \Leftrightarrow t(X^*(\theta_{low})) \geq t(x)$ and that
$P_{\theta_{low}}(T(X) \geq t(x)) = \alpha/2 \Leftrightarrow P(t(X^*(\theta_{low})) \geq t(x)) = \alpha/2.$

The practical use of this result is that we can create a confidence interval by abstract simulation. For each abstract simulation find the θ_{inf} and θ_{sup} value. Sort the θ_{inf} values in order and let the $(\alpha/2)100\%$ percentile be the lower confidence limit. Also sort the θ_{sup} values in order and let the $(1-\alpha/2)100\%$ percentile be the upper confidence limit. If the distribution F is continuous then there is a unique value θ' that makes $t(X^*(\theta'))=t(x)$. In this case $\theta_{inf}=\theta_{sup}$ and both percentiles are found from the same series.

2.4 Example: The binomial distribution

Assume that X is binomially distributed with parameters n and p. For a given outcome x the lower confidence limit p_{low} is found by

$$P_{p_{low}}(X \geq x) = \frac{\alpha}{2}$$

Analogously the upper confidence limit p_{up} is found by

$$P_{p_{up}}(X \leq x) = \frac{\alpha}{2}$$

The interval $[p_{low} , p_{up}]$ is a $(1-\alpha)100\%$ confidence interval.

We will now study the abstract binomial sample. Assume that $U_1, U_2 , \ldots , U_n$ are independent and uniformly distributed in the interval [0,1]. As we have seen this sample can be used for generating binomial samples. For each specified p'-value we get a binomial outcome by counting the number of U_i's less than p'. The interesting p'-values are the values that give an outcome which is equal to the basic outcome X i.e. the p'-values in the interval $[U_{(X)},U_{(X+1)}]$, where $U_{(X)}$ is the X:th ordered value. In this case the extreme values, inf and sup, are $U_{(X)}$ and $U_{(X+1)}$ respectively. These values should be studied in order to find the confidence limit, motivated by:

$$P(U_{(X)} \leq p_{low}) = \frac{\alpha}{2} \qquad \text{and}$$
$$P(U_{(X+1)} \geq p_{up}) = \frac{\alpha}{2}$$

In this case it is also rather easy to verify these results analytically. The binomial frequency function gives

$$P_{p_{low}}(X \geq x) = \sum_{i=x}^{n} \binom{n}{i} p_{low}^{\,i} (1-p_{low})^{n-i}$$

The frequency function for the ordered statistic $U_{(X)}$ is

$$f(v) = \frac{n!}{(X-1)!\ (n-X)!} v^{X-1} (1-v)^{n-X}$$

Furthermore

$$P(U_{(X)} \leq p_{low}) = \int_0^{p_{low}} \frac{n!}{(X-1)!\ (n-X)!}\ v^{X-1}\ (1-v)^{n-X}\ dv$$

and by using repeated partial integration this equals

$$\sum_{i=x}^{n} \binom{n}{i}\ p_{low}{}^{i}\ (1-p_{low})^{n-i}$$

which verifies the results.

This means that if we simulate $U_{(X)}$ a large number of times, and sort these values in order, we will find the lower confidence limit as the $(\alpha/2)100\%$ percentile. Analogously the upper confidence limit is found as the $(1-\alpha/2)100\%$ percentile in the ordered series of $U_{(X+1)}$ values.

2.5 A theorem for creating confidence sets by abstract simulation.

Let X denote a random variable with distribution function F depending on the parameter θ, where θ belongs to the set Θ. Furthermore let X* denote an abstract outcome, let $X^*(\theta')$ denote the value of the abstract outcome corresponding to the specific θ' value.

Theorem:
Suppose that there is a set $A(\theta)$ for all $\theta \in \Theta$ such that;
$P_\theta(X \in A(\theta))=1-\alpha$, and for each possible outcome x define
$S(x)=\{\theta\ ;\ x \in A(\theta)\}$. If we also define $S^*=\{\theta'\ ;\ X^*(\theta') \in A(\theta)\}$

then $P_\theta(\theta \in S(X))=P_\theta(\theta \in S^*)=1-\alpha$

Proof $\theta \in S^* \Leftrightarrow X^*(\theta) \in A(\theta)$ Observe that X* is an abstract sample of X which means that $X^*(\theta)$ has the same unconditional distribution as X. Thus the two events $X^*(\theta) \in A(\theta)$ and $X \in A(\theta)$ have the same probability which implicates that $\theta \in S(X)$ and $\theta \in S^*$ also have the same probability. Finally $\theta \in S(x) \Leftrightarrow x \in A(\theta) \Rightarrow P_\theta(\theta \in S(X)) = P_\theta(X \in A(\theta))=1-\alpha$.

The practical use of this theorem is that we can create confidence sets by simulating abstract outcomes X*. This is done by the following steps:

1. Simulate X* and find the parameter(s) θ' that makes $X^*(\theta')$ equal or less extreme than X. This means that $X \in A(\theta) \Rightarrow X^*(\theta') \in A(\theta)$.
Let $L=\{\theta' : X \in A(\theta) \Rightarrow X^*(\theta') \in A(\theta)\}$ (the set of likely parameters).

2. Save the most extreme parameters from L e.g for the one dimensional case save infL and supL.
Observe that $P_\theta(A(\theta) \text{ contains } L)=1-\alpha \Rightarrow$
$P_\theta(\text{the most extreme } \theta'\text{:s in } L \in A(\theta))=1-\alpha$

3. Simulate a large number of abstract samples and repeat step 1 & 2. All the parameters saved from step 2 should now be studied in order to create a confidence set. Which confidence limits that should be chosen depends on which shape of the confidence set that is wanted.

2.6 Minimum unlikelihood procedure

The abstract simulation technique generates possible true parameters, it generates the likelihood function. In the one dimensional case a confidence interval is easily created (as shown in 2.3), by sorting the possible parameters in order and then choosing the percentiles as confidence limits. The interval contains the q% most likely parameters, we have sorted out the (1-q)% most unlikely parameters. We could say that the confidence interval contains the q% least unlikely parameters. Because the procedure rejects the most unlikely parameters and saves the least unlikely ones, we will call this procedure the minimum unlikelihood procedure. Usually the number of resamples is recommended to be 999. This is so because the 999 observations is reasonably many and divide the real line into 1000 intervals with equal probability, different percentiles are now easily found. For instance the common percentiles 0.5%, 2.5% , 99.5% and 97.5% are found as the observation numbers 5, 25, 995 and 975.

Assume that the distribution is continuous, in this case it is rather easy to generalize the procedure to the multidimensional case with a parameter vector of size n. Each resample generates a possible parameter vector in the n-dimensional space. Repeating this resampling a large number of times gives a sequence of such parameter vectors. In the one dimensional case we sort the observations in order with the aim to divide the real line into intervals.

In this case we have to construct boxes in the space. The procedure is done by using following steps:

1. Create the largest box in the space including all parameter points.
2. Reject the used parameter vectors.
3. Among the remaining parameter vectors repeat step 1-2 until q100% of the parameter vectors are rejected.
4. Use the remaining (1-q)100% parameter vectors and create the largest box i.e. the confidence box.

Example 5. Assume that n=3. The first box (the largest box), is the box with the minimum and maximum values in each direction as limits, that is taking X_{min} and X_{max} as limits in X direction, Y_{min} and Y_{max} as limits in Y direction and Z_{min} and Z_{max} as limits in Z direction. The vectors containing these points are now used and the next box is found exactly in the same way studying the remaining parameter vectors. Each box uses a number of parameter vectors, in this case with n=3 a box can be constructed with 2,3,4,5 or 6 parameter vectors. The required number of parameter vectors is random. If the nominal level of confidence is 95% then we shall create new boxes until 5% of the parameter vectors are used. If for instance 1000 parameter vectors were generated, we shall create boxes until 50 (or as close as possible) parameter vectors are used. The largest box created from the remaining 950 parameter vectors is the 95% confidence box.

3. BOOTSTRAP

3.1 Distribution free models

The models examined this far are all models with some known distribution e.g. binomial or normal. We will now study the abstract simulation technique for distribution free models. The idea of bootstrap, Efron(1982), is to use the empirical distribution function not only for estimation but for resampling as well. In some situations it is possible to use these two techniques together. We estimate the distribution with the empirical one and then apply the abstract simulation technique. This mixture is called abstract bootstrap, Holm(1990,1993).

The model studied here is the linear model with explanatory variables possible for the experimenter to choose. A general procedure for creating confidence sets will be deduced. Also suggested is a special balanced method for treating the grand mean. The abstract bootstrapping is, as we shall see, closely related to some permutation methods, Maritz(1984).

3.2 Bootstrap confidence intervals.

Let $\mathbf{X}=(X_1, X_2,..., X_n)$ be an i.i.d sample with unknown distribution function and let θ be the parameter of interest and $\hat{\theta}=g(\mathbf{X})$ an estimate of θ. Furthermore let $\mathbf{X}^*=(X^*_1, X^*_2,..., X^*_n)$ be a bootstrap sample, independently drawn from $(X_1, X_2,..., X_n)$, with equal probability in each point and with replacement. The bootstrap sample gives the bootstrap estimate $\hat{\theta}^*=g(\mathbf{X}^*)$. Repeating this sampling a large number of times gives a sequence of bootstrap estimates which can be used to approximate the distribution of $\hat{\theta}$.

The original percentile method, Efron(1982), takes the percentiles, $[\hat{\theta}^*(\alpha/2) , \hat{\theta}^*(1-\alpha/2)]$, as a $(1-\alpha)100\%$ confidence interval for θ. There are suggested refinements of this method; the bias correction method, Efron(1982), and the accelerated bias correction method, Efron(1987).

Singh(1981), Bickel & Freedman(1981) and Beran(1987) use another method called the functional method, (root method, pivot method). The distribution of a pivot variable is approximated by the bootstrap distribution of the corresponding bootstrap pivot variable.

Example 6. Translation, parameter of interest the mean θ. Assume that $\hat{\theta}-\theta$ is our pivot variable and has a fixed distribution invariant with θ. The probability $P(a \leq \hat{\theta}-\theta \leq b)=1-\alpha$ corresponds to $P^*(a \leq \hat{\theta}^*-\hat{\theta} \leq b)=1-\alpha$, (* denotes bootstrap distr.). Thus $a+\hat{\theta}=\hat{\theta}^*_{low} \Rightarrow a=\hat{\theta}^*_{low}-\hat{\theta}$ and $b=\hat{\theta}^*_{up}-\hat{\theta}$, where $\hat{\theta}^*_{low}$ and $\hat{\theta}^*_{up}$ are the lower and upper percentiles in the bootstrap distribution. The approximation is:
$P(\hat{\theta}^*_{low}-\hat{\theta} \leq \hat{\theta}-\theta \leq \hat{\theta}^*_{up}-\hat{\theta}) \approx 1-\alpha$.
Thus the functional bootstrap method gives us the functional confidence interval, $[2\hat{\theta}-\hat{\theta}^*_{up}, 2\hat{\theta}-\hat{\theta}^*_{low}]$.

3.3 Bootstrap in linear models.

The method suggested by Efron (1982), resamples the empirical residuals. Model $Y=\alpha+X\beta+\varepsilon$, where Y is the observation vector $(n\times1)$, α is a $(n\times1)$ vector with all components equal to α, X is the design matrix $(n\times p)$, β the parameter vector $(p\times1)$ and ε the residual vector $(n\times1)$. The components of the residual vector are assumed to be i.i.d and to have expectation 0 and some variance σ^2. Let e be the empirical residual vector, $e_i=Y_i-\hat{\alpha}-X_i\hat{\beta}$. New observations are found by $Y_{new}=\hat{\alpha}+X\hat{\beta}+e^*$, where e* is a bootstrap sample drawn from e. It is now possible to use the Monte Carlo technique to find the estimated $\hat{\beta}$ distribution and create a confidence interval for β. However, the empirical residuals are neither independent nor equally distributed, thus they can at most serve as an approximation of a sample of true i.i.d residuals.

Example: Simple linear regression, $Y_i = \alpha + x_i\beta + \varepsilon_i$ $i=1,2,...,n$
where x= -2, -1, 0, 1, 2 gives $e_5=(2e_1-2e_3-4e_4+4e_5)1/10$ with variance $0.4\ \sigma^2$ and $e_3=(-2e_1-2e_2+8e_3-2e_4-2e_5)1/10$ with variance $0.8\ \sigma^2$.
The variance of e_3 is twice as high as that of e_5.
This means that we have a further approximation beside the bootstrap approximation itself.

Another method suggested by Holm(1990,1993), is to use the non observable abstract true residuals for resampling. This means that we depict the original experiment closer and that the only approximation is the pure bootstrap approximation. The method is a mixture of abstract simulation and bootstrap technique. The essential point in the paper by Holm is that although the true residuals are not observable we get an observable final result. The final result is a confidence set and can be calculated directly without knowledge of the true residuals. This is possible by using the theorem and applying the procedure for finding confidence limits. This procedure will now be studied in more detail.

4. ABSTRACT BOOTSTRAP

4.1 Abstract bootstrap in linear models.

The model $Y= \alpha + X\beta + \varepsilon$ is the same as in 3.3. Let ε^* denote the abstract bootstrap sample, drawn from ε (the true non observable residual vector). That is $\varepsilon^*= Y^* - \alpha - X^*\beta$, where Y^* consists of the randomly choosen observation and X^* consists of the corresponding rows.

The new (abstract!) observations are:

$$Y_{new} = \alpha + X\beta + \varepsilon^* = \alpha + X\beta + Y^* - \alpha - X^*\beta = X\beta + Y^* - X^*\beta$$

Having original observations and new abstract observations, we will now study the related estimates. The ordinary least square estimate is $\hat{\beta} = SY$, where S is the estimation matrix found by ordinary least square estimation. In the non full rank case S is found by using restrictions. The abstract bootstrap estimate is:

$$\hat{\beta}^* = SY_{new} = S\alpha + SX\beta + S\varepsilon^* .$$

As pivot variable we will use $\hat{\beta}-E[\hat{\beta}] = \hat{\beta} - S\alpha - SX\beta$. Furthermore we will assume that $S1=0$, where 1 is a $(n\times1)$ vector with all components equal to 1. This gives orthogonality between α and β. Observe that:

$$S1=0 \Rightarrow S\alpha=0 \Rightarrow \hat{\beta}-E[\hat{\beta}] = \hat{\beta} - SX\beta \ , \text{ (in the full rank case } SX\beta=\beta).$$

The corresponding abstract bootstrap pivot is:

$$\hat{\beta}^* - E^*[\hat{\beta}^*] =\hat{\beta}^* - E^*[S\alpha + SX\beta +S\varepsilon^*] = \hat{\beta}^* - S\alpha - SX\beta -SE^*[\varepsilon^*] =$$

$$\hat{\beta}^* - S\alpha - SX\beta -S\bar{\varepsilon} = \hat{\beta}^* - SX\beta \ , \text{ where } \bar{\varepsilon} \text{ is a } (n\times1) \text{ vector with all components equal to } \bar{\varepsilon} , (S1=0 \Rightarrow S\alpha=0 \text{ and } S\bar{\varepsilon}=0).$$

<u>The statistic, (the abstract bootstrap pivot variable) is non observable because β is unknown. But observe that for a hypothesis $\beta=\beta'$ the nulldistribution is possible to simulate. This notable feature and the relation between tests and confidence sets are just the facts that makes it possible to create a confidence set.</u>

According to the technique described in the theorem 2.5 we should study the parameters which gives an abstract outcome equal to the original one. That is, making the abstract bootstrap pivot equal to the original pivot variable:

$$\hat{\beta}^* - SX\beta = \hat{\beta} - SX\beta \Rightarrow \hat{\beta}^* = \hat{\beta} \Rightarrow S\alpha + SX\beta + S\varepsilon^* = SY \Rightarrow$$
$$S\alpha + SX\beta + SY^* - S\alpha - SX^*\beta = SY \Rightarrow S(X-X^*)\beta = S(Y-Y^*)$$

For each resample solve the equation above. This generates a sequence of β's. Sorting these β's and eliminating the most extreme ones leaves a confidence set, (or in the one dimensional case an interval).

The procedure can be summarised in the following steps:

1. Find the estimation matrix S
2. For each simulated bootstrap sample solve $S(X-X^*)\beta=S(Y-Y^*)$
3. Sort out the most extreme parameters, according to the minimum unlikelihood method.
4. The most extreme β's among the remaining ones are the limits which form the confidence set.

In the proceeding we will refer several times to this procedure and these four steps.

In the non full rank case the estimation matrix is found by using a restriction $K^T\beta=m$. The abstract bootstrap technique shall depict the original model as closely as possible and therefore this restriction must also be taken into account in the abstract bootstrap procedure. Either directly in the abstract bootstrap estimate or in step 2 of the procedure. That is, for each simulated abstract bootstrap sample find the solution to $S(X-X^*)\beta=S(Y-Y^*)$ complemented with $K^T\beta=m$. Both ways lead to the same final result as will be illustrated in 4.3, studying the one way analysis of variance.

4.2 Example: Simple linear regression

The model is $Y_i = \alpha + (X_i-\bar{X})\beta + \varepsilon_i$, i=1,2,...,n. The residuals are assumed to be i.i.d and to have expectation 0. The estimation matrix (vector), $S=(\mathbf{X}-\bar{\mathbf{X}})^T/Q_x$, where $(\mathbf{X}-\bar{\mathbf{X}})$ is a $(n\times 1)$ vector with i:th element equal to $X_i-\bar{X}$ and $Q_x=\sum_{i=1}^{n}(X_i-\bar{X})^2$. The second step in the procedure is to solve $S(X-X^*)\beta=S(Y-Y^*)$.

In this case $X=(\mathbf{X}-\bar{\mathbf{X}})$ and $X^*=(\mathbf{X}^*-\bar{\mathbf{X}}) \Rightarrow$

$(\mathbf{X}-\bar{\mathbf{X}})^T((\mathbf{X}-\bar{\mathbf{X}})-(\mathbf{X}^*-\bar{\mathbf{X}}))\beta/Q_x = (\mathbf{X}-\bar{\mathbf{X}})^T(Y-Y^*)/Q_x \Rightarrow$

$(1-W_x^*/Q_x)\beta = \hat{\beta} - W_y^*/Q_x$, where $W_x^*=\sum_{i=1}^{n}(X_i-\bar{X})X_i^*$ and

$W_y^*=\sum_{i=1}^{n}(X_i-\bar{X})Y_i^*$ which means that

$\beta=(\hat{\beta}\,Q_x-W_y^*)/(Q_x-W_x^*)$.

This is the generating variable which gives us the sequence of β's from which we create a confidence interval according to step 3. That is, just sorting the β's in order and choosing the $(1-\alpha/2)100\%$ and the $\alpha/2100\%$ percentiles as confidence limits.

A more detailed description of the simple linear regression case, asymptotic validity, comparisons between the abstract and the ordinary method and simulations, (which indicates good performance for the abstract method), is given in the report by S. Holm(1990,1993).

4.3 Example: One way analysis of variance

Model $Y_{ij}= \alpha + \beta_i + \varepsilon_{ij}$, i=1,...,k and j=1,...,m (replicate)

In matrix form $Y= \alpha + X\beta + \varepsilon$,this is a non full rank model and therefore we will use the restriction $\sum_{i=1}^{k} \beta_i =0$, in matrix form $1^T\beta=0$.

This gives the estimation matrix S (k×km) =

$$\begin{bmatrix} \frac{1}{m}-\frac{1}{km} \cdots \frac{1}{m}-\frac{1}{km} & -\frac{1}{km} \cdots -\frac{1}{km} & \cdots & -\frac{1}{km} \cdots -\frac{1}{km} \\ -\frac{1}{km} \cdots -\frac{1}{km} & \frac{1}{m}-\frac{1}{km} \cdots \frac{1}{m}-\frac{1}{km} & \cdots & -\frac{1}{km} \cdots -\frac{1}{km} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{km} \cdots -\frac{1}{km} & -\frac{1}{km} \cdots -\frac{1}{km} & \cdots & \frac{1}{m}-\frac{1}{km} \cdots \frac{1}{m}-\frac{1}{km} \end{bmatrix}$$

In row nr i the elements which affect the Y:series nr i equals $\frac{1}{m}-\frac{1}{km} \cdots \frac{1}{m}-\frac{1}{km}$ and the other elements equal $-\frac{1}{km} \cdots -\frac{1}{km}$.

$\Rightarrow \hat{\beta} = SY =(\bar{Y}_i - \bar{\bar{Y}})$, (k×1) vector.

Step 2. Solve $S(X-X^*)\beta=S(Y-Y^*)$ with the restriction $1^T\beta =0$

Observe that $1^T\beta =0 \Rightarrow SX\beta=\beta \Rightarrow S(X-X^*)\beta= (I_k-SX^*)\beta=S(Y-Y^*) \Rightarrow$

$\beta =(I_k - \frac{1}{m}[N_{i,j}^*] + \frac{1}{mk} 1_k [N_j^*]^T)^{-1} (\bar{Y}_i - \bar{\bar{Y}} - \bar{Y}_i^* + \bar{\bar{Y}}^*)$, where

I_k is a unit (k×k) matrix, 1_k is a (k×1) vector with all elements =1, $[N_{i,j}^*]$ denotes the number of Y:s choosen from series j in the bootstrap series i , N_j^* is the total number of Y:s chosen from series j , $\bar{Y}_i$ and $\bar{Y}_i^*$ is the series means respectively the bootstrapped series means , finally $\bar{\bar{Y}}$ and $\bar{\bar{Y}}^*$ are the two grand means. This is the variable which generates the sequence which we shall study in order to obtain a confidence set, according to step 3 and 4.

In the argumentation above the restriction was taken into account when solving the equation in step 2. It is also possible to use the restriction directly in the bootstrap estimate. We will now once more deduce the generating variable but use the restriction directly in the bootstrap estimate. Hopefully this will illustrate the idea of abstract bootstrap further.

With the restriction $1^T \hat{\beta} = 0$, the estimate $\hat{\beta}$ is found by solving:

$$X^TX \hat{\beta} + 1\lambda = X^T(Y-\alpha)$$

$$1^T \hat{\beta} = 0$$

Here $X^TX = mI_k$, where I_k is a unit matrix ($k \times k$) and thus

$$m1^T \hat{\beta} + k\lambda = 1^TX^T(Y-\alpha)$$

i.e.

$$m \hat{\beta} = X^T(Y-\alpha) - \frac{1}{k} 11^TX^T(Y-\alpha)$$

$$\hat{\beta} = \frac{1}{m} [X^T(Y-\alpha) - \frac{1}{k} 11^TX^T(Y-\alpha)]$$

Observe that $\frac{1}{m} [X^T\alpha - \frac{1}{k} 11^TX^T\alpha] = 0$ and thus

$$\hat{\beta} = \frac{1}{m} [X^TY - \frac{1}{k} 11^TX^TY]$$

These are the well-known differences between the mean in each group and the grand mean .

By resampling the true residuals new observations are found:

$$Y_{new} = \alpha + X\beta + \varepsilon^* = \alpha + X\beta + (Y^* - \alpha - X^*\beta) = Y^* + X\beta - X^*\beta$$

The abstract bootstrap estimate is found by using the same restriction $1^T\hat{\beta}^* = 0$. This gives:

$$\hat{\beta}^* = \frac{1}{m} [X^TY^* + X^TX \beta - X^TX^* \beta - \frac{1}{k} 11^TX^TY - \frac{1}{k} 11^TX^TX \beta + \frac{1}{k} 11^TX^TX^* \beta]$$

Using the restriction directly in the estimate gives:

$$\hat{\beta}^* = \frac{1}{m} [X^TY^* + X^TX \beta - X^TX^* \beta - \frac{1}{k} 11^TX^TY + \frac{1}{k} 11^TX^TX^* \beta]$$

For finding the β of interest, let $\hat{\beta}^*=\hat{\beta}$:

$$(I_k - \frac{1}{m} X^T X^* + \frac{1}{mk} 11^T X^T X^*)\beta = \frac{1}{m} X^T(Y-Y^*) - \frac{1}{mk} 11^T X^T(Y-Y^*)$$

Here $X^TX^*=[N_{i,j}^*]$ and $1^T[N_{i,j}^*]=[N_j^*]$ furthermore $\frac{1}{m} X^T(Y-Y^*)$ equals

$(\bar{Y}_i - \bar{\bar{Y}})$ and finally $1^TX^T(Y-Y^*)$ equals the number $n(\bar{\bar{Y}}-\bar{\bar{Y}}^*)$.
And thus we have reached the same result one more time.

The procedure in 4.1 is more general and may be more applicable for computer programming.

4.4 A numerical example

We will now put some figures into the one way analysis of variance.
Model $Y_{ij}= \alpha + \beta_i + \varepsilon_{ij}$, i=1,...,3 and j=1,2 (replicate)
Observation serie $Y=(Y_{11},Y_{12},Y_{21},Y_{22},Y_{31},Y_{32})=(10,12,16,18,20,20)$

, average $\bar{\bar{Y}}=16$.
Original estimate $\hat{\beta} =(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) =(-5,1,4)$
Assume that the resampled residual is equal to:
$\varepsilon^*=(\varepsilon_5 , \varepsilon_2 , \varepsilon_6 , \varepsilon_3 , \varepsilon_4 , \varepsilon_6)$,the resampled vector of observations:

$Y^*=(20,12,20,16,18,20)$ average $\bar{\bar{Y}}^*=17.666...$
A few calculations give the generating variable:
$\beta=(-6, 2, 4)$

Going backwards in the argumentation, this means that if the true parameter is equal to $\beta=(-6, 2, 4)$, and if the first residual had been equal to ε_5, the second residual had been equal to ε_2 and so on, then the estimate had been equal to $\hat{\beta}=(-5,1,4)$. We have used the exchangeability of the i.i.d. true residuals and have found a possible true parameter. Resampling the true residuals over and over again generates a sequence of possible parameters and by sorting out the q100% most extreme we have the (1-q)100% most likely parameter values left, i.e. a (1-q)100% confidence set. This is a typical mission for a powerful computer.

4.5 Simulation

A program for simulating the case of one way analysis of variance was constructed with the following steps:

1. Input are the model and seeds for the random number generation
2. Random number generation of residuals
3. Calculating the original estimates
4. Bootstrapping the residuals
5. Find the generated parameters (step 2)
   * By Gauss elimination with pivoting
6. Save the generated parameters
7. Repeat step 4-6 1000 times.
8. Find the confidence box:
   a, Create the largest box
   b, Count the number of parameter vectors needed for that box
   c, Repeat step a-b until 50 (or as close as possible) parameter vectors are used.
   d, The largest box created by using the remaining parameter vectors is the confidence box, (with level of confidence=95%).

Simulation results:

The model tested is: $Y_{ij}= \alpha + \beta_i + \varepsilon_{ij}$, i=1,...,3 and j=1,4
with $\alpha=0$, $\beta_1=-15$, $\beta_2=5$, $\beta_3=10$.
The residuals were generated from a normal distribution with $\mu=0$, $\sigma^2=4$. Repeating the steps 1-8 above 1000 times gave the following results:

Nominal level of confidence=95.23 %
Observed level of confidence=95.5 % (simultaneously)

	Observed level of confidence for the parameter:	mean length of interval for the parameter	mean values for the limits
β_1 :	98.1 %	4.5	-17.2 ,-12.7
β_2 :	98.1 %	4.6	2.7 , 7.3
β_3 :	98.6 %	4.5	7.8 , 12.3

4.6 Example: Two way analysis of variance

Two way analysis of variance

Model $Y_{ijk}= \alpha + \beta_i + \lambda_j + \varepsilon_{ijk}$, i=1,...,m j=1,...,n k=1,...,k (replicate) ,

in matrix form $Y = \alpha + X\beta + \varepsilon$, where Y is a (mnk×1) vector ,

α has the same form but with all elements equal to the grand mean α,

X is the design matrix and

β is a ((m+n)×1) vector containing the parameters β_i and λ_j .

As usual ε is the residual vector. This model is also a non full rank model, the parameters are not identifiable and therefore we will use the restrictions:

$\sum_{i=1}^{m}\beta_i=0$ and $\sum_{i=1}^{n}\lambda_i=0$. This gives the estimation matrix S ((m+n)×mnk).

S is not shown because of the large space it would need , anyhow it gives the ordinary estimates: $\hat{\beta}=SY \Rightarrow \quad \hat{\beta}_i=\bar{Y}_{i..}-\bar{\bar{Y}}_{...}$ and $\hat{\lambda}_j=\bar{Y}_{.j.}-\bar{\bar{Y}}_{...}$

In the next step solve: $S(X-X^*)\beta=S(Y-Y^*)$

with restrictions $\mathbf{1}^T\beta=0$ $\left(\sum_{i=1}^{m}\beta_i=0\right)$

and $\mathbf{1}^T\lambda=0$ $\left(\sum_{i=1}^{n}\lambda_i=0\right)$

Observe that the restrictions $\Rightarrow SX\beta=\beta \Rightarrow S(X-X^*)\beta = (I_{m+n}-SX^*)\beta =S(Y-Y^*)$.

The generating variable β equals:

$$\left(I_{m+n} - \frac{1}{k}\begin{bmatrix}\frac{1}{n}I_m & 0\\ 0 & \frac{1}{m}I_n\end{bmatrix}\begin{bmatrix}N_{i.,i.} & N_{i.,j}\\ N_{j,i.} & N_{j,j}\end{bmatrix} + \frac{1}{kmn}\begin{bmatrix}1_m & 0\\ 0 & 1_n\end{bmatrix}\begin{bmatrix}N_{..,i.} & N_{..,j}\\ N_{..,i.} & N_{..,j}\end{bmatrix}\right)^{-1}$$

$$\left(\begin{bmatrix}\bar{Y}_{i..}-\bar{\bar{Y}}_{...}\\ \bar{Y}_{.j.}-\bar{\bar{Y}}_{...}\end{bmatrix} - \begin{bmatrix}\bar{Y}^*_{i..}-\bar{\bar{Y}}^*_{...}\\ \bar{Y}^*_{.j.}-\bar{\bar{Y}}^*_{...}\end{bmatrix}\right)$$

, (where the different variables are of the same kind as in the one way lay out.

For example $N_{i.,j}$ is the number of Y:s in bootstrap series i , chosen from series j. Series nr i are the rows in X which have the element in column i equal to 1, i=1,...,m (for the β:s) , and equivalent series j ,(for the λ:s), are the rows in X which have the element in column m+j equal to 1, j=1,...,n). This variable generates the possible parameter vectors which should be studied for obtaining a confidence set. For instance by arranging boxes as described in 2.6.

4.7 Assuming normal distribution.

Model $Y= \alpha + X\beta + \varepsilon$, where we assume that the residuals are normally distributed with some known variance i.e. $\varepsilon_i \sim N(0,\sigma)$. The least square estimate $\hat{\beta}=SY$. Assume that the model is a full rank model, that $E[SY]=\beta$ and that $S\alpha=0$.

Simulate $\varepsilon_1^*, \varepsilon_2^*, ..., \varepsilon_n^*$ independently and $\sim N(0,\sigma)$.

New observations are found as $Y_{new} = \alpha + X\beta' + \varepsilon^*=$ and the new estimate is found as $\hat{\beta}^*=SY_{new} =S\alpha + SX\beta' + S\varepsilon^* = \beta' + S\varepsilon^*$.

Find the β' for which $\hat{\beta}^*=\hat{\beta} \Leftrightarrow \beta' + S\varepsilon^*= SY \Leftrightarrow \beta' = SY - S\varepsilon^* = S(Y-\varepsilon^*)$

This is the generating variable which can be used for simulating the likelihood function. Observe that β' is normally distributed with the mean SY and the variance σ^2. Confidence limits found by simulations are expected to be the same as found by ordinary Normaldistribution theory. The simulation error will be very small if we compute a large number of resamples, which is easily done with some computer power. Of course simulations are not necessary in this trivial case.

5. A BALANCED ABSTRACT BOOTSTRAP METHOD FOR TREATING THE GRAND MEAN

5.1 The balanced abstract bootstrap method

The model $Y = \alpha + X\beta + \varepsilon$ contains a grand mean which we have not treated yet. At first sight this seems to be impossible because of a twofold reason. Firstly, our new abstract observations ($Y_{new} = \alpha + X\beta + \varepsilon^* = \alpha + X\beta + Y^*-\alpha-X^*\beta = X\beta + Y^* - X^*\beta$) do not contain the α and secondly there is a β dependence ! But, by using a balanced resample it is possible to treat this problem as an ordinary translation problem.

The estimate we use is the average $\overline{Y}$, this means that our estimation matrix ,(vector) is $S= (\frac{1}{n} , \dots , \frac{1}{n})$, where n is the total number of Y:s. Assume that $SX\beta=0$, (orthogonality between α and β parts). As pivot statistic we will use $\hat{\alpha}-E[\hat{\alpha}] = \hat{\alpha}-\alpha$. Our bootstrap estimate is $\hat{\alpha}^*= SY_{new} =SX\beta + SY^* - SX^*\beta$, but $SX\beta=0$ and the term $SX^*\beta$ will be equal to 0 <u>if we balance our resampling and choose the same number of Y:s from each series .</u>

This is possible in replicate models, for example simple linear regression with replicates, one way and two way analysis of variance. <u>We balance the variation part out, by choosing equal numbers of observation from each level, and get the average without factor influence. In fact we have changed the problem to be an ordinary average study for treating the mean in a translation model, see example 6 in 3.2.</u> Observe that our chosen observations can be placed anywhere in the bootstrap series. This balance condition, that observations shall occur the same number of times from each level is easily achieved by a random permutation. Example: Assume that we have four different levels and that five observations shall be taken from each level. Write the string:
12341234123412341234 representing the series that we shall choose an observation from. Permutating this string gives a random order like:
21223443112341433214 which means that we shall start with choosing an observation randomly from series 2 and then choose an observation randomly from series 1, and so on.

In the balanced method $SX^*\beta = 0$ and the bootstrap estimate $\hat{\alpha}^* = SY^*$.

The corresponding pivot statistic is:

$$\hat{\alpha}^* - E^*[\hat{\alpha}] = \hat{\alpha}^* - SE^*[Y^*] = \hat{\alpha}^* - SE^*[\varepsilon^*] - S\alpha - SE[X^*\beta] =$$

$$\hat{\alpha}^* - S\bar{\varepsilon} - S\alpha = \hat{\alpha}^* - S\bar{\varepsilon} - S\alpha = \hat{\alpha}^* - \bar{\varepsilon} - \alpha$$

For finding the α of interest let the two pivots be equal:

$$\hat{\alpha} - \alpha = \hat{\alpha}^* - \bar{\varepsilon} - \alpha = SY^* - S1\left(\frac{1}{n} 1^T (Y - \alpha - X\beta)\right) - \alpha = SY^* - SY \Rightarrow$$

$$\alpha = SY - SY^* + \hat{\alpha} = 2SY - SY^*$$

This is the generating variable which should be studied in order to obtain a confidence interval, the limits are found by using the minimum unlikelihood method (2.6). This confidence interval is comparable with the confidence interval created by the ordinary bootstrap method in the translation example.

5.2 Balanced method, usual bootstrapping.

The balanced method described above is also possible to apply when using the usual bootstrap methods.

Example: One-way analysis of variance.

Model $Y_{ij} = \alpha + \beta_i + \varepsilon_{ij}$, $i=1,...,k$ and $j=1,...,m$ (replicate), with the restriction $\sum_{i=1}^{k} \beta_i = 0$. The grand mean is estimated by the average and if we balance the bootstrap sample, the bootstrap estimate will not be affected by the different factor levels. By balancing we can treat the grand mean as the mean in an ordinary translation problem, example 6 in 3.2.

5.3 Theoretical aspects of balancing.

Observe that in the bootstrap method there are two approximations involved, first the statistical approximation using the empirical distribution function as an estimate of the true distribution function and secondly the use of Monte Carlo simulation which is a numerical approximation.

Davidson *et al.* (1986) have shown that by using this type of balancing it is possible to reduce the simulation error, specially in bias estimation. This is so because many estimates have a large linear component. Furthermore Hinkley *et al.* (1990) extends the methodology to second order balance, which principally affects bootstrap estimation of variance.

The balanced method forces the resample to contain the same number of observations from each group, but note that there are no limitations in which observations we choose in a group, this means that we can choose the same observation in a group more than one time. A stronger balancing would be that an observation can be chosen just one time. If the bootstrap sample is of the same size as the original sample balance gives in fact a permutation. We will now show the close relation between permutation methods and bootstrap methods.

6. A PERMUTATION METHOD

6.1 The permutation method

The balanced methods described above are examples of bootstrapping with some kind of restriction. An even stronger restriction is to resample without replacement i.e. use permutations. We will study here a permutation method for linear regression presented in Maritz(1984). The basic idea is to use the exchangeability of i.i.d residuals. New observations are found by permutating the residuals:

Model $Y_j = \alpha + (x_j - \bar{x})\beta + \varepsilon_j \quad , j=1,2,...,n$

The residuals are assumed to be i.i.d and to have expectation 0. Thus analysing trends of the residuals , $\varepsilon_j(\beta) = Y_j - \alpha - (x_j - \bar{x})\beta$, tests accuracy of the parameter β. The general class of test statistics are

$$T(\beta,\Psi,H,Y,x) = \sum_{j=1}^{n} \Psi(x_j) H[\varepsilon_j(\beta)]$$

Example: Let $T(\beta,\Psi,H,Y,x) = \sum_{j=1}^{n} (x_j - \bar{x})\varepsilon_j = \sum_{j=1}^{n} (x_j - \bar{x})(Y_j - \alpha - (x_j - \bar{x})\beta)$

To test $H_0: \beta = \beta_0$ evaluate all n! permutations of the ε_i:s and for each permutation calculate the T value. Each of these T values have probability 1/n! in the null distribution. If our original observed T value is extreme in this distribution the hypothesis is rejected. A confidence interval is created by finding the set of β:s that we can not reject. Let

$$T_q(\beta) = \sum_{j=1}^{n} (x_j - \bar{x})\varepsilon_{nj} = \sum_{j=1}^{n} (x_j - \bar{x})(Y_{nj} - \alpha - (x_{nj} - \bar{x})\beta)$$

denote one of the n! possible permutations of $T(\beta)$. Let Q' be the set of all permutations excluding the observed one. Let $N(\beta)$ denote the number of $T_q(\beta)$; $q \in Q$ that are smaller than our observed $T(\beta)$. If $N(\beta) > r$ and $N(\beta) < n! - r$ then the β can not be rejected in a two-sided test with significance level r/n!.

Therefore a confidence interval is found by examining the values of $N(\beta)$ as β varies β varies from $-\infty$ to $+\infty$.

Represent $N(\beta)$ as:

$$N(\beta) = \sum_{q \in Q} I\left[\sum (x_j - \bar{x})(Y_j - \alpha - (x_j - \bar{x})\beta) - \sum_{q} (x_j - \bar{x})(Y_{nj} - \alpha - (x_{nj} - \bar{x})\beta) > 0\right]$$

$$= \sum_{q \in Q} I\left[\frac{\sum_{j=1}^{n} (x_j - \bar{x})(y_j - y_{nj})}{\sum_{j=1}^{n} (x_j - \bar{x})(x_j - x_{nj})} > \beta\right], \text{assuming that } \sum_{j=1}^{n} (x_j - \bar{x})(x_j - x_{nj}) > 0$$

$N(\beta)$ change value whenever β equals one of the slope estimates

$$\frac{\sum_{j=1}^{n} (x_j - \bar{x})(y_j - y_{nj})}{\sum_{j=1}^{n} (x_j - \bar{x})(x_j - x_{nj})}$$

Thus confidence limits can be found by calculating the slope estimate for each permutation, sorting these values in order and then choosing the upper and lower percentile as confidence limits.
Observe that the generating variable above is exactly the same as found with the abstract bootstrap method in 4.2. This is natural because the variable above is also deduced in an abstract manner. Each resample (permutation) gives a possible true parameter, namely the parameter which makes the resampled estimate equal to the original one. In linear regression the abstract bootstrap method and the permutation method are closely related, in fact the only difference between the two methods is the resampling procedure, with or without replacement.

The test statistic $T(\beta) = \sum_{j=1}^{n} (x_j - \bar{x})\varepsilon_j$ has in both methods conditional expectation equal to 0. Conditionally Var[T in the abstract method] = $\frac{n-1}{n}$ Var[T in the perm. meth.]. Thus the two methods have the same asymptotic properties. Further theoretical comparisons will not be considered in this paper, but we will discuss the use of different resampling methods from a practical point of view.

7. COMPARISONS

The ordinary bootstrap method, the abstract bootstrap method and the permutation method are all resampling methods. In many situations the choice of method depends on different practical problems. In some models there are non-existence problems in several points or even worse, a method does not work at all. These kinds of problems will now be illustrated in some examples.

Example 7. Translation, parameter of interest: the mean $\mu=E[X]$.
Model $Y_j= \mu + \varepsilon_j$, $j=1,2,...,n$
The abstract bootstrap method sometimes fails in step 2, that is finding the generating variable, the variable that makes the bootstrap estimate equal to the original one. Sometimes this variable does not exist, like in this very simple model:
$\varepsilon^*= Y^*- \mu^* = Y^* - \mu$, new observations: $Y_{new} = \mu + \varepsilon^* = \mu + Y^* - \mu = Y^*$
Observe that the new observations do not contain the parameter μ, and of course it is not possible to find any generating variable. Thus the abstract method does not work in this case. But observe that resampling the true residuals, in this case, gives new observable observations. These new observations are just the observations found by using the ordinary bootstrap method. Thus the ordinary method works but the abstract method does not. Using the ordinary bootstrap method usually means that we have to use the estimated residuals and the estimated model for resampling, but in this simple model it works anyway. The permutation method does not work either, for two reasons. Firstly all permutations give the same average and secondly the generating variable does not exist.

The three methods can be separated in two parts. On the first side the ordinary bootstrap method, which gives observable observations directly. On the other side the abstract bootstrap method and the permutation method, both methods generate non observable observations but by finding the generating variable a confidence set can be constructed. The problem is that in some resamples the generating variable does not exist. The most obvious resample is the one which is identical with the original one i.e. $\varepsilon^*= \varepsilon$.

If the generating variable, the variable that makes the bootstrap estimate equal to the original one, had been existing for this resample, it would have been too good to be true, because in that case the generating variable would be equal to the true parameter of the model! Unfortunately this is not always the only resample for which the generating variable does not exist, there may be others.

The number of resamples for which the generating variable does not exist may be the determing factor for which model to use. In the simple model above the generating variable did not exist for any resample at all and thus the abstract bootstrap- and permutation method did not work. It is also worth noting that in the resamples with non existing generating variable, we get an observable new observation and thus an ordinary bootstrap estimate.

In linear models (except the simple model in the example above), the disadvantage with the ordinary bootstrap method is that we have to use the estimated residuals (which are not i.i.d), and the estimated model for resampling, remembering that: $Y_{new} = \hat{\alpha} + X\hat{\beta} + e^*$, where e^* is a resample from the estimated residuals. The disadvantage with the abstract bootstrap- and the permutation method is the problem with non existing generating variables. But the number of resamples with this non existence is in most models relatively small even when the number of observations is moderate.

Example 8. Simple linear regression, $Y_i = \alpha + X_i\beta + \varepsilon_i$, $i=1,2,3$
and $X_1=-2 \quad X_2=-1 \quad X_3=3$.
To use the ordinary bootstrap method we have to use the estimated model for resampling. We will instead examine the two other methods:
The generating variable is equal to:

$$\beta = \left(\hat{\beta} \sum_{i=1}^{n} (X_i-\bar{X})^2 - \sum_{i=1}^{n} (X_i-\bar{X})Y_i^* \right) / \left(\sum_{i=1}^{n} (X_i-\bar{X})^2 - \sum_{i=1}^{n} (X_i-\bar{X})X_i^* \right)$$

The generating variable does not exist when:

$$\sum_{i=1}^{n} (X_i-\bar{X})^2 - \sum_{i=1}^{n} (X_i-\bar{X})X_i^* = 0 \Leftrightarrow 14 = -2X_1^* - X_2^* + 3X_3^*$$

There is only one resample which fulfils the equation above namely the obvious resample $\varepsilon^* = \varepsilon \Leftrightarrow X_i^* = X_i \ \forall\, i$. In the abstract bootstrap method this is one out of 27 resamples and in the permutation method this is one out of 6 resamples.
Suppose that we increase the model to the design:
$X_1=-3 \quad X_2=-2 \quad X_3=-1 \quad X_4=0 \quad X_5=1 \quad X_6=2 \quad X_7=3.$
Observe that this design includes the origin and thus the residual number four does not affect the estimate and thus just by changing residual number four to anyone of the others we will get a resample for which the generating variable does not exist. Totally there will be non existence for 742 resamples out of 823543 ($\approx$0.0009), in the abstract bootstrap method. In the permutation method there will be non existence for 7 out of 5040 ($\approx$0.0014), (results from a computer simulation).
Thus even if the number of observation is moderate the probability of getting a resample with a non existing generating variable is small.
Observe that the 742 resamples with non existing generating variable contains information, these resamples give observable new observations and new estimates possible to treat according to the ordinary bootstrap method.

8. DISCUSSION

Which method to use, the ordinary bootstrap, the abstract bootstrap, the balanced abstract bootstrap or the permutation method, is a rather difficult question, it depends on several circumstances, for instance:

1. The model: When studying the mean in the simple translation model the ordinary bootstrap method was the only applicable.
2. The parameter of interest: Studying the grand mean in a linear model, a balancation of the resamples was necessary to avoid influence from the variation part.
3. The design and the total number of observations: These factors affect the probability of the non existence problem in the abstract bootstrap- and the permutation method.

The abstract procedure for creating a confidence set is a useful technique for resampling methods. In linear models this technique makes it possible to resample the true residuals. The abstract bootstrap- and the permutation method use the true model for resampling. This is an advantage compared with the ordinary bootstrap method which use the estimated model for resampling and thus has more approximations involved.

The abstract bootstrap- and the permutation method are closely related, in the linear model the only difference is how to resample, with or without replacement. The two methods have the same asymptotic properties. To recommend one of these methods is difficult, because on one side the abstract bootstrap method seems to have a lower probability of the non existence problem, but on the other side balanced resampling may reduce the simulation error.

9. FURTHER DEVELOPMENT.

Even if the distribution is known there can be reasons for using abstract resampling, for instance if the statistic is complex and it is difficult to analytically calculate the likelihood function. This is the case of the interesting theory of generalized linear model, McCullagh & Nelder(1989).

Model $E[Y]=\mu$, $g(E[Y])= g(\mu)=X\beta$ for some function g. An estimate of β is found by iterating

$$X^TWX\hat{\beta}_k = WX^Tz \qquad \text{until} \quad \hat{\beta}_k-\hat{\beta}_{k-1} < 0.0001$$

Where W is a weight matrix $W^{-1}_{k-1}= [\ \frac{\delta g(\mu)}{\delta \mu}\ ]^2 V_{k-1}$, and V_{k-1} is the variance function evaluated at $\hat{\beta}_{k-1}$.

The variable $z= g(\mu_{k-1}) + (y-\mu_{k-1})(\ \frac{\delta g(\mu)}{\delta \mu}\ )_{k-1}$, is a linearized form of the link function applied to the data, $g(y)\approx g(\mu) + (y-\mu)g'(\mu)$.

Furthermore $\hat{\beta}_k-\hat{\beta}_{k-1} = (X^TWX)^{-1}WX^T(y-\mu_{k-1})(\ \frac{\delta g(\mu)}{\delta \mu}\ )_{k-1}$ is called the adjustment part.

It may be possible to create confidence set in this model as well. For a known distribution of Y, for instance binomial (logit link), simulate an abstract sample. Next step is to find the parameter which makes the resampled estimate equal to the original one. Let $\hat{\beta}_k$ be the start point and find the parameter which makes the adjustment part small i.e. so small that the iteration stops at once. Then the new estimate is equal to the original one.

There are, however, still some problems to solve, for instance if the known distribution is discrete the generating parameter will not be unique, this is a difficult problem especially if the parameter is multidimensional.

10. REFERENCES.

Beran R.J. (1987) Prepivoting to reduce level error of confidence set.
Biometrica, 74 p. 457-468.

Bickel P.J. & Freedman D.A (1981)
Some asymptotic theory for the bootstrap.
Ann. Stat. 9, p. 1196-1217

Davidsson A.C, Hinkley D.V & Schechtman E.(1986)
Efficient bootstrap simulation
Biometrika, 73 , p. 555-566

Efron B. (1982) The jackknife, the bootstrap and other resampling plans. Soc. Ind. Appl. Math. CBMS-NSF SIAM.

Efron B. (1987) Better bootstrap confidence intervals.
J. Amer. Stat. Assoc., 82, 171-200

Hinkley D.V., John P.W.M & Shi S. (1990)
Balanced Design of bootstrap simulations
J.R. Statist. Soc. B , 52, p. 185-202

Holm S. (1990) Abstract Bootstrap Confidence Intervals in Linear Models. Research Report, Stat. Dep. Univ. Göteborg.

Holm S. (1993) Abstract Bootstrap Confidence Intervals in Linear Models. Scand. Journal of Stat., 20, p. 157-170.

Maritz J.S. (1984) Distribution-Free Statistical Methods. Chapman and Hall

Mc Cullagh & Nelder (1989) Generalized Linear Models, 2nd edition. Chapman and Hall.

Singh K. (1981) On the asymptotic accuracy of Efron´s bootstrap. Ann. Stat., 9, 1187-1195.

ACKNOWLEDGEMENT.

My supervisor professor Sture Holm suggested the topic and has encouraged me in many ways during the work. He is truly a source of inspiration, always positive and happy.
For this I am deeply grateful.

I am also very grateful to my colleague Göran Åkermo for many enjoyable (and some constructive) discussions.

1992:1	Teräsvirta, T., Tjøstheim, D. & Granger, C.W.J.	Aspects of Modelling Nonlinear Time Series
1992:2	Palaszewski, B.	A conditional stepwise test for deviating parameters
1992::3	Guilbaud, O.	Exact Semiparametric Inference About the Within--Subject Variability in 2 x 2 Crossover Trails
1992:4	Svensson, E. & Holm, S.	Separation of systematic and random errors in ordinal rating scales
1993:1	Frisén, M & Åkermo, G.	Comparison between two methods of surveillance: exponentially weighted moving average vs cusum
1993:2	Jonsson, R.	Exact properties of McNemar's test in small samples.