# The Bootstrap - Numerical Procedures and Applications

Seminar Paper submitted

to

**Prof. Dr. Brenda López-Cabrera**

Humboldt-Universität zu Berlin
School of Business and Economics
Climate, Weather and Energy Analysis

by

**Erin Sprünken**
(581608)

in partial fulfillment of the requirements
for the Seminar
**Numerical Introductory Course**
Berlin, January 12, 2021

# Contents

# List of Tables

# List of Figures

# 1 Motivation

motivation

## 2 Related Literature

In this section we provide an overview of related literature to the topic of the bootstrap. Important to mention is the article of Efron (1979) who generalised the jackknife principle to what is the classical bootstrap today. Since then, many advancements, generalizations and applications were conducted by numerous authors. Kerns (2010) provides a general overview of resampling techniques in chapter 13 and Hastie et al. (2008) describe and relate bootstrap methods to Maximum Likelihood, Bayesian Inference and Machine Learning techniques in chapter 8.

Konietschke and Pauly (2014) find that bootstrap and permutation methods for matched pairs give valid asymptotic results even for different distributions which have no treatment effect under the null hypothesis. Their paper relates to estimation of t-type test statistics for paired samples and overcomes the problems of small sample sizes, heterogeneity and different distributions. Friedrich et al. (2017) find valid results applying the so-called wild bootstrap method to problems of repeated measurements. The nonparametric bootstrap approach delivers better results than WALD or ANOVA statistics, which are especially useful for small sample sizes. Umlauft et al. (2019) find that wild-bootstrapping ranked-based procedures overcome assumptions about normality or homogeneity in general factorial measure designs and deliver asymptotically correct multiple contrast tests. Furthermore, Bathke et al. (2018) find that parametric bootstrap can, under minimal assumptions, overcome the problems of normality assumptions or equal covariance matrices in multivariate factorial designs.

Prášková (2002) and Politis and Romano (1994) describe bootstrap methods for timeseries, which again are used in applied research. For example, Petukhina et al. (2018) refer to the latter one cominbing portfolios. Petukhina and Sprünken (2020) use a parametric bootstrap approach to numerically compute optimal CVaR (Conditional Value-at-Risk) portfolios.

# 3 Applications

As already mentioned before, various applications for the bootstrap method exist. For example, early clinical trials usually suffer from small sample sizes. For example, consider a very early stage for a new drug. Since it's effects are only considered theoretically, ethical standards would forbid testing such a drug on a large sample of patients. On the other hand, small samples are not reliable when it comes to statistical inference. However, valid inference is necessary to judge about approving or declining the public usage of this treatment. The solution lies in bootstrapping a small sample. For example, the clinical trial consists of 16 patients receiving the new treatment or the placebo, this leads to a two-sample of matched pairs design. Here, bootstrap solves this small sample problem and leads to reliable results, see Konietschke and Pauly (2014) and Friedrich et al. (2017).

In finance, especially portfolio management, parametric bootstrap and resampling in general can provide large samples when analytic solutions are not possible. Consider the following minimization problem (Condition Value-at-Risk):

$$\min_{x} \quad -\frac{1}{1-\alpha}\int_{x^\intercal\mu\leq -VaR_\alpha(x)} x^\intercal\mu f\left(x^\intercal\mu \mid x\right)\mathrm{d}x^\intercal\mu. \tag{1}$$

This has no analytic solutions. However, a portfolio manager wants to obtain the portfolio weights $x$. Petukhina and Sprünken (2020) solve this problem by sampling a vector of weights $x$ from the uniform distribution $n_b$ times and compute $n_b$ (empirical) CVaRs and choose the vector $x$ of weights which corresponds to the smallest CVaR of all.

Bagging, as mentioned and described in Hastie et al. (2008), uses bootstrapping to improve Machine Learning techniques. By bootstrapping and it's property of drawing with replacement, one creates $n_b$ samples on which an algorithm can be trained and estimated. The $n_b$ models are then aggregated into one (for example by majority vote), see Breiman (1996). This parallel design is distinct to sequential designs such as Boosting. Random Forests are a special case of this, since it grows $n_b$ trees instead of only one and thus significantly reduces the variance of the tree model, see Ho (1995) and Breiman (2001).

Generally, bootstrapping can be applied to many fields of statistical and applied sciences.

# 4 Methodology

This section will cover formal aspects of the bootstrapping and is divided into two parts. The first one is a short overview of advantages and disadvantages. The second part will briefly describe mathematical aspects of the different applications covered within this seminar paper.

## 4.1 Advantages and Disadvantages

One of the main advantages given by bootstrapping is the sample size. Usually, especially for frequentist statistics, small sample sizes are a huge problem since many applications or theorems rely on large samples or asymptotic behaviour. However, there are many situations where the researcher has to deal with a small sample size. Classical statistics fastly become unreliable in such a setting. Here, resampling provides a solution to overcome this issue.

However, if the original sample from which resampling is conducted is a bad sample, the resampling techniques will not help any further. For example, consider a sample size of four and two observations of this are outliers, then the probability of resampling the outliers is quite high, although they might be unlikely in general. So, the resampling (as the name implies) relies on the sample as well and this is the main disadvantage.

## 4.2 Mathematical Aspects

This subsection deals with the formal aspects of the algorithm, specifically how and why it works. The general idea of bootstrapping is the following (for $nboot$ number of resampling iterations):

1. Fix the data x
2. FOR $i$ IN $1$ TO $nboot$ DO
3. (a) Sample $x\star$ from x
   (b) Compute the statistic of interest for $x\star$
   (c) Save the respective statistic at the $i$-th position of a vector $T\star$
4. Compute Mean (or Median, Confidence Intervals, or any other statistic) of the vector $T\star$.

Mathematically, the bootstrapping algorithm makes use of the strong low of large numbers and creates an empirical distribution of the statistic of interest to mimic the true but unknown distribution.

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i \quad \overset{a.s.}{\to} \quad \mathbb{E}[X]$$
$$\lim_{n\to\infty} F_n(x) \quad \overset{a.s.}{\to} F^X(x).$$

Note, that the first one (law of large numbers) relies on pairwise independence if $X \in \mathcal{L}^1$ or at least uncorrelatedness if $X \in \mathcal{L}^2$. The second equation is the convergence of the empirical distribution function $F_n(x)$ to the true distribution function $F^X(x)$. This follows immediately from the law of large numbers.

There are two versions of bootstrap incorporated in this paper: The nonparametric bootstrap and the wild bootstrap (Rademacher-version). The nonparametric bootstrap procedes pretty much in the way described above, where in the resampling step the data is resampled with replacement and each observation has an equal probability of getting into the bootstrap sample. On the other hand, the wild bootstrap proceeds differently. The following Pseudocode demonstrates the Rademacher wild bootstrap:

1. Fix the data x
2. FOR $i$ IN $1$ TO $nboot$ DO
3. (a) Sample vector $z$ from $\{-1, 1\}$ with replacement and of same length as x
   (b) Create bootstrap sample $x\star = z \odot (x - \bar{x})$ (Hadamard product)
   (c) Compute the statistic of interest for $x\star$
   (d) Save the respective statistic at the $i$-th position of a vector $T\star$
4. Compute Mean (or Median, Confidence Intervals, or any other statistic) of the vector $T\star$.

### Summary Statistics

As mentioned before, there are three types of statistics we are going to analyze in the empirical section of this paper. The first type are summary statistics, computing the following metrics: Minimum, $25\%$-quantile, Median, Mean, $75\%$-quantile, Maximum and Standard Deviation. Of course, the Minimum and the Maximum are not really able to be resampled, since the absolute Minimum (Maximum) that could be obtained in the resampling is the actual Minimum (Maximum) of the original sample. The quantiles (and the Median is the $50\%$-quantile) are computed in the default way of the respective R-function *quantile*, see R Core Team (2020).

The Standard Deviation is the squareroot of the corrected sample Variance. Since the original data is centered for the wild bootstrap, finally the original's sample mean has to be added to the bootstrapped location parameters. Sloppy said, this procedure of estimating parameters can be thought of as a nonparametric maximum likelihood approach, although this is not quite precisely formulated.

### t-Type Test Statistics

Although simple, the t-type tests are a nice example for bootstrapping. However, this procedure can be extended to all possible test procedures one can think of (Lilliefors, Wilcoxon-Signed-Rank, Friedman, etc.). The idea is basically the same as before and the algorithm works in the same way. However, the statistic of interest is the T-statistic. For the one-sample case that is the following equation, see Toutenburg and Heumann (2008):

$$T(X) \quad = \quad \frac{\bar{X} - \mathbb{E}[\bar{X}]}{\hat{\sigma}} \sqrt{n}.$$

When bootstrapping this statistic, we aim to mimic it's distribution and finally refer to the quantiles of the bootstrap distribution to make a decision whether to reject or not reject the null hypothesis. However, in the resampling iteration we need to compute a slightly different statistic, since there is a setting of a bootstrap sample given the fixed data:

$$T(X\star \mid X) \quad = \quad \frac{\bar{X\star} - \mathbb{E}[\bar{X\star} \mid X]}{\hat{\sigma}} \sqrt{n},$$

where $\mathbb{E}[X\star \mid X] = \bar{X}$, so the conditional expectation of the bootstrap sample is the sample mean of the original data. For the two-sample t-Test the statistic looks different as well, since not only the difference of means has to be taken into account, but the pooled standard deviation, too:

$$\begin{aligned} X'\star \quad &:= \quad X_2\star - X_1\star \\ T(X'\star \mid X) \quad &= \quad \frac{\bar{X'\star} - \mathbb{E}[\bar{X'\star}]}{\sqrt{\frac{\hat{\sigma_1^2}\star}{n_1} + \frac{\hat{\sigma_2^2}\star}{n_2}}}, \end{aligned}$$

where $n_1$ and $n_2$ refer to the sample sizes. Also, under the null hypothesis when testing for equality of means the expectation of mean difference is zero (i.e. $\mathbb{E}[X\bar{*}\star] = 0$). The resampling procedure for nonparametric and wild stays the same. However, for the two-sample problem we introduce the groupwise nonparametric bootstrap as well. The nonparametric bootstrap samples from the whole sample $\{\{x \in X_1\}, \{x \in X_2\}\}$, whereas the groupwise nonparametric bootstrap respects the different samples observations came from. That means, that $X_1\star$ is directly sampled from $\{x \in X_1\}$ (and the same for $X_2\star$).

**Regression Coefficients**

Bootstrapping the regression is also a matter where bootstrap can provide significant improvements to estimation. Especially the wild bootstrap has it's origins in the field of regression analysis, see Wu (1986). Furthermore, it is continuously studied, see Davidson and Flachaire (2008). As before, bootstrap grants the possibility to obtain point- and set-estimates for the coefficients. According to Fox and Weisberg (2018), if the matrix of covariates $X$ is random (or at least not fixed), one has to resample the whole pairs of $(Y_i, X_{i\cdot})$ and estimate the linear model for the new data. We apply such an approach in our nonparametric bootstrap version of the linear regression. However, if X is assumed to be deterministic/fixed, then it's enough to resample the residuals of the regression. This is, where we apply the wild bootstrap in the following way:

1. Fix the data $(Y, X)$
2. Estimate linear model for $(Y, X)$ and predict fitted values $\hat{Y}$
3. Save residuals $\epsilon = Y - \hat{Y}$
4. FOR $i$ IN $1$ TO $nboot$ DO
5. (a) Sample vector $z$ from $\{-1, 1\}$ with replacement and of same length as $\epsilon$
   (b) Create bootstrap sample $\epsilon\star = z \odot \epsilon$ (Hadamard product)
   (c) Create bootstrap sample $Y\star = \hat{Y} + \epsilon\star$
   (d) Estimate linear model $(Y\star, X)$ and save coefficients $\beta_i\star$ of $i$-th iteration
6. Compute Mean (or Median, Confidence Intervals, or any other statistic) of the array $\beta\star$.

# 5 Simulation Settings

simulation setting

# 6 Empirical Analysis

This section covers the empirical results using the methods mentioned earlier. Our interest lies in the aspects of accuracy, computation time and complexity, and for tests the errors of type I and II.

## 6.1 Accuracy

## 6.2 Type I and II Errors

## 6.3 Computational Complexity

# 7 Conclusion

conclusion outlook: * different wild bootstraps

* parametric bootstraps

* multiple contrast tests

* other tests (F-Test, Anderson-Darling, ...)

* effect statistics

## References


Bathke, A., Friedrich, S., Pauly, M., Konietschke, F., Staffen, W., Strobl, N., and Höller, Y. (2018). Testing Mean Differences among Groups: Multivariate and Repeated Measures Analysis with Minimal Assumptions. *Multivariate Behavioral Research* 53, pp. 384 –359.

Breiman, L. (1996). Bagging Predictors. *Machine Learning* 24, pp. 123 –140.

— (2001). Random Forests. *Machine Learning* 45, pp. 5 –32.

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics* 146, pp. 162 –169.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 7, pp. 1 –26.

Fox, J. and Weisberg, S. (2018). An R Companion to Applied Regression.

Friedrich, S., Konietschke, F., and Pauly, M. (2017). A wild bootstrap approach for nonparametric repeated measurements. *Computational Statistics & Data Analysis* 113, pp. 38 –52.

Hastie, T., Tibshirani, R., and Friedman, J. (2008). The Elements of Statistical Learning.

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1, pp. 278 –282.

Kerns, G. J. (2010). Introduction to Probability and Statistics Using R.

Konietschke, F. and Pauly, M. (2014). Bootstrapping and permuting paired t-test type statistics. *Statistics and Computing* 24, pp. 283 –296.

Petukhina, A. and Sprünken, E. (2020). Evaluation of Multi-Asset Investment Strategies with Digital Assets. Available at SSRN: https://ssrn.com/abstract=3664219.

Petukhina, A., Trimborn, S., Härdle, W. K., and Elendner, H. (2018). Investing with Cryptocurrencies - evaluating their potential for portfolio allocation strategies. Available at SSRN: https://ssrn.com/abstract=3274193.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on Subsamples under minimal assumptions. *The Annals of Statistics* 22, pp. 2031 –2050.

Prášková, Z. (2002). Bootstrap in Nonstationary Autoregression. *Kybernetika* 38, pp. 389 –404.

R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Toutenburg, H. and Heumann, C. (2008). Induktive Statistik - Eine Einführung mit R und SPSS.

Umlauft, M., Placzek, M., Konietschke, F., and Pauly, M. (2019). Wild bootstrapping rank-based procedures: Multiple testing in nonparametric factorial repeated measures designs. *Journal of Multivariate Analysis* 171, pp. 176 –192.

Wu, C. F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *Annals of Statistics* 14, pp. 1261–1295.

# Appendix A

appendix

* proof $\hat{F}(x) \stackrel{a.s.}{\to} F(x)$
* proof $\mathbb{E}[\bar{X}_{.} \mid \mathbf{X}] = \bar{X}$