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, November 10, 2020

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):

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

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

methods

5 Simulation Settings

simulation setting

6 Empirical Analysis

empirical

7 Conclusion

conclusion

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.

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

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.

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.

Appendix A

appendix