Abdelhak M. Zoubir and D. Robert Iskander

Bootstrap Methods in Signal Processing

Your first reaction when you read the title of this issue of *IEEE Signal Processing Magazine* could have been: "Why a special section on the bootstrap?" Or maybe "What is the bootstrap?" Maybe you did not ask yourself any questions and you are eager to discover what the authors of this special section report on. Regardless of which of the above groups you belong to, we will attempt to briefly answer these questions and in particular what led us to propose a special section on the bootstrap. More importantly, we would like to give you a little taste for what promises to be a collection of exciting tutorial articles with a wide range of real-life applications and conclude with some suggestions for further reading.

THE RATIONALE

The use of more accurate models has become a fundamental requirement in signal processing applications such as wireless communications, radar, sonar, biomedical engineering, machine and car engine monitoring, and speech and image processing. With improved accuracy, the models have become more complex and inferential statistical signal processing is often intractable. The signal processing practitioner requires a simple but accurate method for assessing errors of estimates and answering inferential questions. Asymptotic approximations are useful only when a sufficient amount of data is available, which is not always possible due to time constraints, the nature of the signal or the measurement setting. Also, many of the theoretical derivations for assessing errors make assumptions relating to the probability distribution of the noise process. In today's applications, the Gaussian noise assumption is not always valid. It may hold for sensor noise, for example, but not necessarily for the environmental noise encountered in array signal processing. The signal processing practitioner thus requires methods that are simple and work under minimal assumptions. At the same time these techniques should be robust against changes in environmental conditions and perform equally well under conventional assumptions, such as Gaussian noise.

[IN ESSENCE, THE BOOTSTRAP DOES WITH A COMPUTER WHAT THE EXPERIMENTER WOULD DO IN PRACTICE IF IT WERE POSSIBLE: THEY WOULD REPEAT THE EXPERIMENT.]

What should we do if in a practical situation we cannot invoke procedures based on analytical mathematics and asymptotic approximations? One may think of Monte Carlo simulations, but how often do we face a situation where we cannot repeat the experiment or, even if we could, it would be unjustified due to changes in experimental conditions or due to the unaffordable cost? A real alternative to these problems is the *bootstrap*. It is the statistician's choice for answering inferential questions such as determining confidence limits for parameters of interest and testing hypotheses. In signal processing terms, the bootstrap could provide the tools for the design of a detector for signals buried in noise/interference of unknown distribution or for model selection. The aim of this special section is to stimulate more interest among signal processing practitioners to discover the power of bootstrap methods and to advocate their systematic use in inferential signal processing.

WHAT IS THE BOOTSTRAP?

The bootstrap is a computer-intensive tool for answering inferential questions. In essence, the bootstrap does with a computer what the experimenter would do in practice if it were possible: they would repeat the experiment. The underlying idea of the bootstrap is based on the substitution and simulation principles. If you cannot rerun your experiment, use the already acquired data and create "new" data and recalculate your estimate. The "new" data is obtained through random sampling from the original data. Since random sampling is performed on the original data (empirical distribution), you can draw as many data as you wanted and recalculate your estimate each time in the same way you obtained it with the original data. These many estimates of the parameter of interest, originating from resampling your original data, can be used to approximate measures of your estimator such as its distribution function. The cost of resampling and recalculating estimates is being constantly reduced by the ever-increasing computational power of today's computers.

This intuitive bootstrap approach is backed up by a vast literature in mathematical statistics, demonstrating its theoretical validity. Its applicability is far reaching as one can infer from the articles of this special section, which are all backed up with real data analysis.

Much statistical work is now performed with computers in ways that are too complicated for realistic analytical treatment. One then refers to computer-intensive statistical methods and the bootstrap is only one of them. Particle filters, for example, have become a popular tool for many signal processing applications such as target tracking. Other methods include the jackknife, cross-validation, and Markov chain Monte Carlo (MCMC) methods. Some readers may want to know more about the differences between these techniques, their advantages, and their shortfalls. This is beyond the scope of this special section, but the reader is referred to [1] and [4] for some answers.

CONTENT OF THIS SPECIAL SECTION

The articles in this special section are ordered based on complexity. All articles are written in a tutorial style. The first three contributions are introductory tutorials with real-life applications while the following four articles focus on the use of the bootstrap in advanced signal processing applications. We start with our tutorial article, which is written primarily for a bootstrap novice. At the same time, we believe that it will prove to be useful for all signal processing practitioners as it promotes a pragmatic approach to the use of bootstrap methods. This is followed by an article by Thomson who, for many years, has worked on and advocated another resampling technique, *the jackknife,* which is closely related to the bootstrap. In his article, the author considers the application of the jackknife to multitaper spectra, coherences, and frequency transfer functions in a range of diverse applications ranging from dropped-call rates in cellular phones to 663-year level records of the river Nile. Franke and Halim continue with a tutorial on bootstrap hypothesis testing and in particular on the so-called *wild* bootstrap method. They support the theory with applications to signal and image analysis, which include testing defects in texture analysis.

Wendt, Abry, and Jaffard use the bootstrap in multifractal analysis of hydrodynamic turbulence. They highlight the limitations of practical scaling analysis and show how, with the bootstrap, they can overcome this problem. The article covers the many intricacies that need to be taken into account when even basic bootstrap techniques for hypothesis testing are used in a practical application. Next, Foster and Żychaluk consider bootstrap techniques in the nonparametric modeling of biological transducer functions. Their article provides a tutorial to local linear fitting with bootstrap bandwidth estimation. The theory conveniently interleaves with practical applications of biological transducer function modeling in animals, ranging from turtles, through goldfish to guinea-pigs as well as humans. Further, Polikar reports on bootstrap variants such as *bagging* and *boosting* and their application in computational intelligence. His article will prove to be useful to all interested in emerging areas of machine learning. Finally, Candy addresses the concept of bootstrap particle filtering and considers the application of tracking in synthetic aperture sonar.

The articles of this special section have been chosen with extreme care. The reviewers were very enthusiastic about each of them and praised the excellent technical content disseminated in a very accessible tutorial style. We are confident that you will share this view.

SUGGESTED FURTHER READING

Some readers will be inspired by the collection of articles in this special section and will want to apply bootstrap techniques to their own signal processing problems. Those readers are encouraged to consult statistical textbooks such as the one by Efron and Tibshirani [2] or the one specially dedicated to signal processing practitioners [6] for further reading. We also hope that the readers will find inspirations in other tutorials on the bootstrap published earlier in this magazine [3], [5].

ACKNOWLEDGMENTS

We would like to thank all contributors, including those who submitted excellent articles, that we could not consider for this special section. Space has been critical, and we needed to make many difficult decisions, but we are confident that those articles will find their place in other journals. We are grateful to all reviewers who did excellent work in a timely fashion. Our thanks also go to Doug Williams, area editor for special sections, for his continued support and encouragement and Geri Krolin-Taylor, senior managing editor of *IEEE Signal Processing Magazine,* for her timely coordination of this special section.

REFERENCES

[1] B. Efron, "The bootstrap and modern statistics," *J. Amer. Statist. Assoc.*, vol. 95, no. 452, pp. 1293–1296, 2000.

[2] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

[3] D.N. Politis, "Computer-intensive methods in statistical analysis," *IEEE Signal Processing Mag.*, vol. 15, no. 1, pp. 39–55, 1998.

[4] J. Shao and D. Tu, *The Jackknife and Bootstrap*. New York: Springer Verlag, 1995.

[5] A.M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Processing Mag.*, vol. 15, no. 1, pp. 56–76, 1998.

[6] A.M. Zoubir and D.R. Iskander, *Bootstrap Techniques for Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2004.