

Review of statistical bootstrap methods for estimating random uncertainty in turbulent flows

Jacob Bryan

Mechanical and Aerospace Engineering, Utah State University

E-mail: `jacob.bryan@aggiemail.usu.edu`

Barton Smith

Mechanical and Aerospace Engineering, Utah State University

E-mail: `barton.smith@usu.edu`

Douglas Neil

LaVision Inc.

E-mail:

Geordie Richards

Mechanical and Aerospace Engineering, Utah State University

E-mail: `geordie.richards@usu.edu`

Abstract. Estimating the uncertainty of the sample mean is of critical importance to experimentalists across disciplines. While this is a trivial task for independent samples, time-resolved measurement methods, such as time-resolved particle image velocimetry for turbulent flows, typically produce samples which are no longer independent of each other. This significantly complicates the task of estimating the random uncertainty of the mean. An assortment of bootstrap methods that accomplish this task have been presented in the statistics literature, with several being applied previously to experimental data in fluid dynamics. However, ...

Keywords:

1. Introduction

Estimating the standard deviation of a finite sample mean computed from evenly spaced observations in a stationary stochastic process is a basic problem in statistics of fundamental significance to experimental science. The answer is elementary and familiar in the setting of independent (or nearly independent) samples, provided by the standard error of the mean, but when stronger correlations are present this simple problem has a surprising amount of literature devoted to various strategies for resolution. This problem is of particular interest in experimental fluid mechanics, where increasingly high resolution Particle Image Velocimetry (PIV) and related measurement techniques produce data with substantial temporal and spatial correlations, particularly in the context of turbulent flows, and there is some need for clarity on best practices.

Among the strategies for estimating the standard deviation of the sample mean in the presence of correlations, block bootstrap methods, and related techniques such as the jackknife, and the method of batch means, have the virtue of simplicity in application. Indeed, in contrast to modeling by time series and Bayesian estimation approaches [1, 2], the block bootstrap methods with automatic block length selection can be applied by simply replacing the standard error formula with a modified equation. There are other strategies advocated in the fluids literature which retain this simplicity, such as averaging a truncated sum of autocorrelation coefficients, but these alternate approaches can be fraught with sources of error that are avoided by bootstrap techniques (see [5] for discussion). This manuscript will provide a quick reference for applying block bootstrap methods, with a specific focus on estimating the standard deviation of the sample mean computed from a stationary process with correlations, and summarize a comparison of these methods with the more standard techniques in the fluids literature. The methods will be compared and contrasted using a large collection of data sets from three experiments performed at the Experimental Fluid Dynamics Laboratory at Utah State University. We aim to demonstrate that block bootstrap methods, particularly the tapered block bootstrap (and equivalently, the tapered jackknife), which is previously unused in the fluids literature, can ameliorate problems with more standard methods while retaining ease of application.

2. Theory

Suppose that X_1, \dots, X_N are samples from a stationary real-valued sequence $\{X_n\}_{n \in \mathbb{Z}}$ with mean $\mu = \mathbb{E}X_t$, and autocovariance sequence $R(s) = \mathbb{E}((X_t - \mu)(X_{t+s} - \mu))$. The variance of the sample mean $\bar{X}_N = \frac{1}{N}(X_1 + \dots + X_N)$ can be computed as $\text{Var}(\bar{X}_N) = \sigma_N^2/N$, where $\sigma_N^2 := R(0) + 2 \sum_{k=1}^N (1 - k/N)R(k)$ (see e.g. [4]). We are interested in estimating the standard deviation of the sample mean, $\sigma_{\bar{X}_N} := \sqrt{\text{Var}(\bar{X}_N)}$, for N large, particularly as \bar{X}_N is asymptotically normal under appropriate conditions. Indeed, under sufficient moment and mixing conditions for a central limit theorem to be applicable [3], the limit $\sigma_\infty^2 := \lim_{N \rightarrow \infty} \sigma_N^2 = \sum_{k=-\infty}^{\infty} R(k)$ exists, and $\sqrt{N}(\bar{X}_N - \mu)$ converges in distribution to a normal random variable with mean zero and variance σ_∞^2 . Thus, for N large, we may focus on estimators of σ_∞^2 in an effort to approximate

the standard deviation of the sample mean as $\sigma_{\bar{X}_N} \approx \sigma_\infty / \sqrt{N}$. Note that, in the context of negligible correlations, this reduces to the standard error of the mean, $\sigma_{\bar{X}_N} \approx \sigma / \sqrt{N}$, where $\sigma^2 := R(0) = \text{Var}(X_t)$.

This approach is closely related to an interpretation in the fluids literature, where, for example, the measurements X_1, \dots, X_N are drawn from a stationary stochastic process $\{Z_t\}_{t \in \mathbb{R}}$ at evenly spaced time points, i.e. $X_k = Z_{k\Delta t}$ for each $k = 1, \dots, N$, with some fixed temporal spacing Δt . In this context, the formula $\sigma_{\bar{X}_N} \approx \sigma_\infty / \sqrt{N}$ is re-expressed as $\sigma_{\bar{X}_N} \approx \sigma / \sqrt{N_{\text{eff}}}$, typically referred to as the random uncertainty of the mean, where $N_{\text{eff}} := \frac{N}{2T_u/\Delta t}$ is the “effective number of samples”, itself defined in terms of the integral time scale $T_u = \int_0^\infty \rho(\tau) d\tau \approx \frac{1}{2} (\sum_{k=-\infty}^\infty \rho(k\Delta t)) \Delta t$. Here $\rho(\tau)$ is the temporal autocorrelation of the process $\{Z_t\}_{t \in \mathbb{R}}$, related to the interpretation above by the formula $R(k) = \sigma^2 \rho(k\Delta t)$, for $k = 1, \dots, N$. That is, chasing through these definitions,

$$\sigma_{\bar{X}_N} \approx \frac{\sigma}{\sqrt{N_{\text{eff}}}} = \sigma \sqrt{\frac{2T_u/\Delta t}{N}} = \sigma \sqrt{\frac{\sum_{k=-\infty}^\infty \rho(k\Delta t)}{N}} = \sqrt{\frac{\sum_{k=-\infty}^\infty R(k)}{N}} = \frac{\sigma_\infty}{\sqrt{N}},$$

and the formulas coincide. Thus, in this signal processing context, asymptotic approximation of σ_∞ , the integral time scale T_u , or the effective number of samples N_{eff} , are all equivalent. We proceed to discuss some techniques for estimating T_u and N_{eff} that are presented in the fluids literature, and then review bootstrapping techniques advocated for estimation of σ_∞ , and will highlight some previously clarified overlap.

2.1. Estimating the Integral Time Scale

This section should focus on methods for estimating the integral time scale already existent in the fluids literature.

2.2. Bootstrap Methods

This section should focus on what a bootstrap method is and how it works. This should be fairly short. This sets the conceptual foundation for outlining the specifics of each method in subsequent subsections.

2.3. Setup

IDK what Geordie was intending with this subsection... This requires some more thought. Maybe this is where we can talk about using bootstrap statistics and how to calculate $\text{Var}^*(\sqrt{N}Y^*)$ for the various bootstrap methods.

2.4. Moving and Circular Block Bootstrap

This section should explain why the MBB and CBB methods are basically equivalent. We use the CBB here. Any reasons at all why CBB is better than MBB?

2.5. Stationary Bootstrap

This section should show the details of how the stationary bootstrap uses a geometric distribution to weight the autocorrelation coefficient summation and explain that the stationary bootstrap pseudo time series is also stationary, a potential benefit over MBB/CBB in some applications. Biggest things to note here are that CB and SB have similar bias but SB has higher variance due to the additional randomness introduced when drawing random block sizes. However, SB generates stationary pseudoseries and is less sensitive to the choice of block length compared to CB. This can be useful given that the optimal block length is never known in practice (see Politis and White).

2.6. Tapered Block Bootstrap

Make a big deal about this section. Should include a note that TBB has better asymptotic performance than MBB/CBB/SB and that it has not been seen in the fluids literature thus far. Include the formula to calculate the variance included in the appendix of the original paper.

$$\hat{\sigma}_{b,\text{TBB}} = \frac{1}{N-b+1} \frac{1}{\|w_b\|_2^2} \sum_{i=0}^{N-b} \left\{ \sum_{j=1}^b w_b(j) X_{i+j} - \sum_{t=1}^N \sum_{j=0}^{N-b} w_b(t-j) \frac{X_t}{N-b+1} \right\}^2 \approx \sigma_\infty \quad (1)$$

2.7. Optimal Block Length Selection

This section should probably look like Section 3 of Theunissen, noting the differences in optimal block length selection for each method. However this optimal block size is *never* known in practice, so an It may also be useful to note that the proofs of asymptotic performance of each method make the assumption that the MSE-optimal block length is used in each method, though for finite-sample data, this is an unknown quantity. A derivation of the MSE-optimal block length for each method involves infinite sums which must be approximated for finitely many samples. This is done using the flat-top lag window method of Politis and Romano, though that detail likely isn't important for this paper.

All of the bootstrap methods presented here require the specification of a block length b , with MSE-optimal estimates of σ_∞ being dependent on selecting the MSE-optimal block length b_{opt} . However, b_{opt} is *never* known in practice. We then have the additional problem of trying to calculate an approximate optimal block length, \hat{b}_{opt} . A method for doing this is outlined in [?], which proceeds as follows:

3. Experimental Data Sets

- Description of each data set, experimental setup. Go to Smith et al (2018) and the conference paper for this info.
- Include ensemble-averaged autocorrelation function plot, a la Theunissen, Fig. 3a.

- Note that stochastic structure influences estimates of the integral time scale (see Thiebaux and Zwiers) and that each data set has a different autocorrelative structure

4. Results

4.1. Confidence Interval Accuracy

NOTES: To be inserted

1. 3 pairs of plots (2 for each experiment) showing coverage for fixed NPIN across sampling rates, parents stats vs. record stats.

4.2. Mean Squared Error

NOTES: To be inserted

1. 3 Box plots showing MSE at the highest sampling rates.
2. Let's get those heat maps going again? I think they are useful, just need adequate discussion.
Questions: (i) can we remove the lowest sampling rates in the arrays? (ii) Can these be performed plotting NPIN (rather than N) vs. sampling rate?

Acknowledgement

References

- [1] David Barber, A Taylan Cemgil, and Silvia Chiappa. *Bayesian time series models*. Cambridge University Press, 2011.
- [2] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [3] Murray Rosenblatt. *Stationary sequences and random fields*. Springer Science & Business Media, 2012.
- [4] Barton L Smith, Douglas R Neal, Mark A Feero, and Geordie Richards. Assessing the limitations of effective number of samples for finding the uncertainty of the mean of correlated data. *Measurement Science and Technology*, 29(12):125304, 2018.
- [5] R Theunissen, A Di Sante, ML Riethmuller, and RA Van den Braembussche. Confidence estimation using dependent circular block bootstrapping: application to the statistical analysis of piv measurements. *Experiments in Fluids*, 44(4):591–596, 2008.
- [6] HJ Thiebaux and FW Zwiers. The interpretation and estimation of effective sample size.