MeanandVar

Paolo Luchini

CPLcode.net

February 27, 2021

Abstract

Program meanandvar estimates the running mean of a statistically stationary time series simultaneously with a variance of the estimate itself (not to be confused with the variance of the time series). Its use is to produce a realistic error bar (standard deviation, or square root of the variance) taking into account the statistical error associated with the finite time sampling of a process with nonzero autocorrelation.

Introduction

For an uncorrelated stochastic process the statistical error in an estimate of its mean is the standard deviation of the process divided by the square root of the number of samples. When the time series is extracted from a stationary stochastic process with unknown autocorrelation, for instance from sampling a statistically stationary continuous physical process or the numerical simulation of one, this simple relationship fails; although the asymptotic dependence with the inverse square root of the number of samples generally remains, the constant in front of it (the integral of the correlation function) is empirically unknown. Here the estimate of the mean is classically obtained from the arithmetic mean of the sample, and is correct for both correlated and uncorrelated input. The estimate of the variance of this estimate is obtained by an original unbiased method of batch means which internally calculates an adaptive batch size. For a description of the algorithm, see [1].

Usage

meanandvar.cpl can be used either as a standalone program or as a library. To compile as a standalone program, use option -Dstandalone, as in "makecpl meanand-

var -Dstandalone". Input is accepted from stdin in a multicolumn format, possibly preceded by comment lines starting with #. Each column is assumed to be a separate time series. The first command line parameter, if present, specifies the column, or range of columns, to be operated upon (default: 1). The second parameter, if present, is a number of lines to be skipped at the beginning of the file (default: 0). The third parameter, if present, is a specified fixed batch size (default: adaptive). Command line prototype:

meanandvar [-h] [col[-tocol]] [skip] [bsize]

Outputs are the estimate of the mean and the standard deviation of this estimate for each selected column, one per line. A final line provides the total number of samples used for the estimate and the automatically selected batch size (which can provide a rough estimate of correlation time). The number of samples used may differ from the number of samples received by the size of the last partially filled batch.

To compile meanandvar as a library, just include "USE meanandvar" in your program. For each mean and variance you want to accumulate define a variable of type MEANANDVAR, as "MEANANDVAR accum" does in the enclosed example program, and initialize it with mvinit(accum). Then to accumulate a sample value x, call meanandvar(accum,x). At any time in your program you can access the current mean as accum.mean, variance as accum.var, standard deviation as accum.rms and number of samples as accum.nt. To reset the counters and start afresh, just repeat mvinit.

Example

The provided example.cpl program simulates a first-order dynamical system driven by white noise, thus generating colored noise with a simple lorentzian spectrum, and plots the running estimate of the standard error of the estimate of the mean as a function of sample length, together with its theoretical exact value.

References

[1] S. Russo and P. Luchini. A fast algorithm for the estimation of statistical error in DNS (or experimental) time averages. *Journal of Computational Physics*, 347:328–340, 2017. doi:10.1016/j.jcp.2017.07.005.