MEANANDVAR

Paolo Luchini
CPLcode.net

February 15, 2021

Abstract

Program meanandvar estimates the mean of a statistically stationary time series simultaneously with a variance of the estimate itself (a different concept from the variance of the time series). Its use is to produce a realistic error bar taking into account the statistical error associated with the finite time sampling of a correlated process.

Introduction

For an uncorrelated stochastic process the statistical error in an estimate of the 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 a statistically stationary continuous physical process or from the numerical simulation of one, this simple relationship fails. 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 its variance 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, and in "makecpl meanandvar -Dstandalone". Input is accepted from stdin in a multicolumn format, possibly prefixed 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 on (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]

Output is its estimate of the mean and standard deviation for each selected column, one per line. A final line provides the total number of samples received and the automatic batch size (a rough estimate of correlation time).

To compile 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 again, just repeat mvinit.

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.