Error in Ensemble Mean and Covariance Follows Counting Statistics

Bill Menke, January 28, 2024

I consider a 2D Normal pdf with true mean and covariance

$$\bar{\mathbf{d}} = \begin{bmatrix} ½ \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{C}_d = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

I use the Metropolis-Hastings algorithm to generate an ensemble drawn from the pdf of size N (in excess of a burn-in of length 1000). The initial member of the Markov chain is set to $\mathbf{d}^{(1)} = [0,0]^T$ and successors are drawn from an uncorrelated Normal pdf with a variance of $\sigma^2 = 1$. This value generated successors that were accepted about 70% of the time (Figure 1, bottom).

I computed the sample mean and the sample covariance of each ensemble, and compared them with the true mean and covariance, quantifying the relative (fractional) error using the L_2 norm. Tests with $10^2 \leq N \leq 10^6$ indicate that the relative error in the mean tends to be about half of the relative error of the covariance, and that both are proportional to $N^{-½}$ (Figure 1, top). This is the same fall-off as is encountered in counting statistics.

An ensemble size of $N = 10^6$ is sufficient to determine mean and covariance to about a percent.

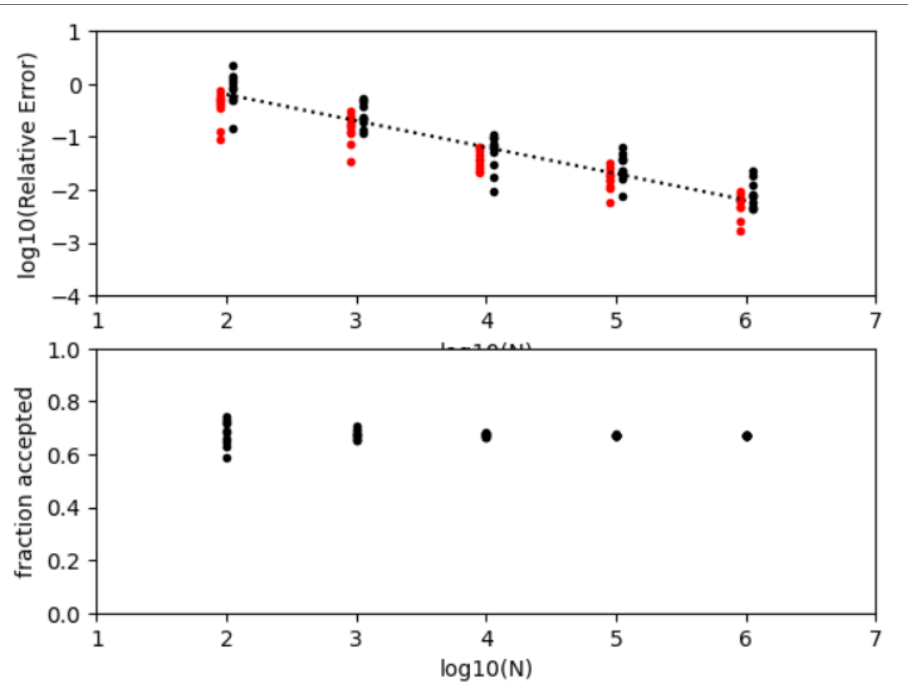

Fig. 1. Results of numerical tests. (Top) Relative error in the estimated mean (black) and estimated covariance (red) as a function of ensemble size N, for $\log_{10} N = 2,3,\cdots 6$. Ten repetitions are performed for each value of N. An $N^{-½}$ trend line is shown for comparison. (Bottom) Fraction of successors accepted during the production of the Markov chain during Metropolis-Hastings algorithm.