Statistics Formulas

(last modified 9/7/99)

Table 1: Names and formulas for statistical quantities. The first column lists the exact definitions. The second column indicates how the formulas are used in practice, when only a finite data set is available. Capital letters represent the exact quantities, small letters are estimates. Note that σ is used as general notation for standard deviation of data, average, etc., so be careful for what it stands.

Name	Formula	Estimator
Data ("Trace")	$A_k;\ -\infty < k < +\infty$	$A_k;\ 1 \le k \le N$
Equilibriated Data Range	$-\infty < k < +\infty$	$k_1 \le k \le k_2;\ N_{eq} = k_2 - k_1 + 1$
Mean of A	$\langle A \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_k A_k$	$a = \frac{1}{N_{eq}} \sum_{k=k_1}^{k_2} A_k \approx \langle A \rangle$
Variance of A	$V = \langle (A - \langle A \rangle)^2 \rangle$	$v = \frac{1}{N_{eq}-1} \sum_{k=k_1}^{k_2} (A_k - a)^2$
Standard Deviation of Data	$\sigma = \sqrt{V}$	
Autocorrelation of A	$C_A(i) = \frac{1}{V} \sum_{k=1}^{N-i} \langle (A_k - \langle A \rangle)(A_{k+i} - \langle A \rangle) \rangle$	$c_A(i) = \frac{1}{v} \frac{1}{N_{eq}-i} \sum_{k=k_1}^{k_2-i} (A_k - a)(A_{k+i} - a)$
Correlation Time of A	$\kappa = 1 + 2 \sum_{i=1}^{\infty} C_A(i)$	$\kappa = 1 + 2 \sum_{i=1}^{i_{cutoff}} c_A(i)$
Effective Number of Points	$N_{eff} = \lim_{N\to\infty} \frac{N}{\kappa} \to \infty$	$N_{eff} = \frac{N_{eq}}{\kappa}$
Error of Mean	$\sigma = \sqrt{\frac{V_A}{N_{eff}}} \to 0$	$\sigma = \sqrt{\frac{v_A}{N_{eff}}} = \sqrt{\frac{v_A \kappa}{N_{eq}}} \propto \frac{1}{\sqrt{N_{eq}}}$

The block transformation transform function can be used to reduce the autocorrelation of a set of data. For a blocking size of m data points per block, the new blocked data A' take the form:

$$\begin{aligned} A'_k &= \frac{1}{m} \sum_{i=0}^{m-1} A_{m(k-1)+i};\ 1 \le k \le \frac{N_{eq}}{m} \\ N' &= N/m \\ \kappa' &\approx \kappa/m \\ a' &= a \end{aligned}$$

Blocking should be used until $N' \approx N_{eff}$, or equivalently $\kappa' \approx 1$. If you know the approximate correlation time, blocking your data can be an easy way to remove autocorrelation in your programs.

Please email any questions or corrections to ceperley@uiuc.edu