Uncertainties and covariances in an analytical random walk

Andrew R. McCluskey[1,2,*]
[1] *European Spallation Source ERIC, P.O. Box 176, SE-221 00, Lund, Sweden*
[2] *Department of Chemistry, University of Bath, Claverton Down, Bath, BA2 7AY, UK*

I. VARIANCES IN AN ANALYTICAL RANDOM WALK

Here, we derive the variance on the mean-squared displacement (MSD) of random walk, clarifying some aspects from the work of Smith and Gillan [1]. We will consider a single particle, travelling in 1 dimension over time. The particle is displaced by $h = \pm d_1$ (where the subscript 1 is indicative of the dimensionality of the system) in a single hop, where the hops are proportional to the timestep that has elapsed. The MSD of this particle, after n hops, can be described with the following,

$$\begin{aligned}\langle \mathbf{r}_1^2(n)\rangle &= \left\langle \left[\sum_{i=1}^{n} h_i\right]^2 \right\rangle = \left\langle \sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j \right\rangle = \left\langle \sum_{i=1}^{n} h_i^2 \right\rangle + \left\langle \sum_{i=1}^{n}\sum_{j\neq i}^{n} h_i h_j \right\rangle \\ &= \sum_{i=1}^{n} \left\langle h_i^2 \right\rangle + \sum_{i=1}^{n}\sum_{j\neq i}^{n} \left\langle h_i h_j \right\rangle = n d_1^2.\end{aligned} \tag{1}$$

In the fourth line above, the cross term double summation (where $j \neq i$) is equal to zero, as the product of $h(i)h(j)$ is d_1^2 and therefore has equal probability of being 1 and -1 so the average must be zero. This shows the linear relationship between timestep and displacement.

The determination of the MSD allows for the derivation of the variance, $\sigma_1^2(n)$ of the MSD for each timestep. This variance can be found with the standard statistical formula,

$$\sigma_1^2\big(\mathbf{r}_1^2(n)\big) = \left\langle \Big[\mathbf{r}_1^2(n) - \langle \mathbf{r}_1^2(n)\rangle\Big]^2 \right\rangle, \tag{2}$$

which may be expanded and reformulated as,

$$\sigma_1^2\big(\mathbf{r}_1^2(n)\big) = \left\langle \big[\mathbf{r}_1^2(n)\big]^2 \right\rangle - 2\left\langle \mathbf{r}_1^2(n)\right\rangle\left\langle \mathbf{r}_1^2(n)\right\rangle + \left\langle \mathbf{r}_1^2(n)\right\rangle^2 = \left\langle \mathbf{r}_1^4(n)\right\rangle - \left\langle \mathbf{r}_1^2(n)\right\rangle^2, \tag{3}$$

where,

$$\left\langle \mathbf{r}_1^2(n)\right\rangle^2 = \left(n d_1^2\right)^2 \tag{4}$$

and,

$$\left\langle \mathbf{r}_1^4(n)\right\rangle = \left\langle \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} h_i h_j h_k h_l \right\rangle. \tag{5}$$

The term on the right-hand side of Equation 5 can be simplified substantially, as the four displacements are only uncorrelated when $i \neq j \neq k \neq l$. This leads to four possible conditions that will survive the averaging process,

(a) $i = j = k = l$,

(b) $(i = j) \neq (k = l)$,

(c) $(i = k) \neq (j = l)$,

(d) $(i = l) \neq (j = k)$,

---

* andrew.mccluskey@ess.eu; a.r.mccluskey@bath.ac.uk

these conditions will ensure a positive product for $h_i h_j h_k h_l$, additionally, the conditions (b), (c) and (d) are equivalent. The result from (a) will be,

$$\left\langle \sum_{i=1}^{n} h_i^4 \right\rangle = nd_1^4, \tag{6}$$

while from (b), (c), and (d) the following is obtained,

$$\left\langle \sum_{i=1}^{n} \sum_{j=1}^{n} h_i^2 h_j^2 \right\rangle = \left(nd_1^2\right)^2. \tag{7}$$

This allows Equation 5 to be rewritten as,

$$\left\langle \mathbf{r}_1^4(n) \right\rangle = nd^4 + 3\left[\left(nd^2\right)^2\right] = nd_1^4 + 3n^2 d_1^4 = (3n^2 + n)d_1^4, \tag{8}$$

as $n \to \infty$,

$$\left\langle \mathbf{r}_1^4(n) \right\rangle \to 3n^2 d_1^4. \tag{9}$$

This means that by combining this with Equation 4 the variance can be determined as,

$$\sigma_1^2\big(\mathbf{r}_1^2(n)\big) = 3n^2 d_1^4 - n^2 d_1^4 = 2n^2 d_1^4. \tag{10}$$

From this result, the variance, $\delta(n)^2$, over many *independent* particles, $N_i(n)$, can be found as,

$$\sigma_{1,i}^2\big(\mathbf{r}_{1,i}^2(n)\big) = \frac{2}{N_i(n)} n^2 d_1^4. \tag{11}$$

To convert these one-dimensional results into three-dimensional results (we note that this conversion is generalisable to ν-dimensions), we multiply both the MSD and its variance by 3. Then, we must consider d_3, which is the sum of three, orthogonal random walks in the x-, y-, and z-directions, so on average is $\sqrt{\langle d_x^2 \rangle + \langle d_y^2 \rangle + \langle d_z^2 \rangle}$ so $1/(\sqrt{3})d_3 = d_1$, therefore,

$$\langle \mathbf{r}_3^2(n) \rangle = 3nd_1^2 = 3n\left(\frac{1}{\sqrt{3}} d_3\right)^2 = nd_3^2, \tag{12}$$

and,

$$\sigma_3^2\big(\mathbf{r}_3^2(n)\big) = 3(2n^2 d_1^4) = 3\left(2n^2\left(\frac{1}{\sqrt{3}} d_3\right)^4\right) = 3\left(\frac{2n^2 d_3^4}{9}\right) = \frac{2n^2 d_3^4}{3}. \tag{13}$$

Therefore the result over N *independent* trajectories is,

$$\sigma_{3,i}^2\big(\mathbf{r}_{3,i}^2(n)\big) = \frac{2n^2 d_3^4}{3N_i(n)}. \tag{14}$$

where N_i is the sample size,

$$N_i(\Delta t) = N_{\text{atoms}}(\Delta t) \left\lfloor \frac{N_{\text{obs}}(0)}{N_{\text{obs}}(0) - N_{\text{obs}}(\Delta t) + 1} \right\rfloor, \tag{15}$$

II. EXTENSION TO COVARIANCE MATRIX

Above, it is has been shown how the variance of an MSD after a given number of steps can be defined. Now, we will extend this derivation to show how the covariance is defined similarly. Let's consider two MSD values, after i steps and $i + j$ steps,

$$\begin{aligned} \left\langle \mathbf{r}_1^2(n) \right\rangle &= nd_1^2 \\ \left\langle \mathbf{r}_1^2(n+m) \right\rangle &= (n+m)d_1^2 \end{aligned} \tag{16}$$

The covariance between these values is defined as,

$$\mathrm{cov}\Big(\big\langle\mathbf{r}_1^2(n)\big\rangle,\big\langle\mathbf{r}_1^2(n+m)\big\rangle\Big)=\Big\langle[\mathbf{r}_1^2(n)-\langle\mathbf{r}_1^2(n)\rangle][\mathbf{r}_1^2(n+m)-\langle\mathbf{r}_1^2(n+m)\rangle]\Big\rangle, \tag{17}$$

which can be expanded and reformulated,

$$\begin{aligned}\mathrm{cov}\Big(\big\langle\mathbf{r}_1^2(n)\big\rangle,\big\langle\mathbf{r}_1^2(n+m)\big\rangle\Big)&=\Big\langle\mathbf{r}_1^2(n)\mathbf{r}_1^2(n+m)-\mathbf{r}_1^2(n)\langle\mathbf{r}_1^2(n+m)\rangle-\mathbf{r}_1^2(n+m)\langle\mathbf{r}_1^2(n)\rangle+\langle\mathbf{r}_1^2(n)\rangle\langle\mathbf{r}_1^2(n+m)\rangle\Big\rangle\\&=\langle\mathbf{r}_1^2(n)\mathbf{r}_1^2(n+m)\rangle-\langle\mathbf{r}_1^2(n)\rangle\langle\mathbf{r}_1^2(n+m)\rangle,\end{aligned} \tag{18}$$

where,

$$\big\langle\mathbf{r}_1^2(n)\big\rangle\big\langle\mathbf{r}_1^2(n+m)\big\rangle=nd^2(n+m)d^2=n(n+m)d^4, \tag{19}$$

and by analogy to Equation 5,

$$\big\langle\mathbf{r}_1^2(n)\mathbf{r}_1^2(n+m)\big\rangle=\Bigg\langle\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n+m}\sum_{l=1}^{n+m}h_ih_jh_kh_l\Bigg\rangle, \tag{20}$$

which we will rewrite as,

$$\begin{aligned}\big\langle\mathbf{r}_1^2(n)\mathbf{r}_1^2(n+m)\big\rangle=\Bigg\langle&\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}h_ih_jh_kh_l+\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=n+1}^{n+m}h_ih_jh_kh_l\\&+\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=n+1}^{n+m}\sum_{l=1}^{n}h_ih_jh_kh_l+\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=n+1}^{n+m}\sum_{l=n+1}^{n+m}h_ih_jh_kh_l\Bigg\rangle.\end{aligned} \tag{21}$$

The second and third terms in Equation 21 tend to 0 due to the equal probability of positive and negative displacements, reducing to,

$$\big\langle\mathbf{r}_1^2(n)\mathbf{r}_1^2(n+m)\big\rangle=\Bigg\langle\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}h_ih_jh_kh_l\Bigg\rangle+\Bigg\langle\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=n+1}^{n+m}\sum_{l=n+1}^{n+m}h_ih_jh_kh_l\Bigg\rangle, \tag{22}$$

and using Equation 9 gives,

$$\big\langle\mathbf{r}_1^2(n)\mathbf{r}_1^2(n+m)\big\rangle=3n^2d_1^4+\Bigg\langle\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=n+1}^{n+m}\sum_{l=n+1}^{n+m}h_ih_jh_kh_l\Bigg\rangle. \tag{23}$$

Finally, we rewrite the above as,

$$\big\langle\mathbf{r}_1^2(n)\mathbf{r}_1^2(n+m)\big\rangle=3n^2d_1^4+\Bigg\langle\sum_{i=1}^{n}\sum_{j=1}^{n}h_ih_j\Bigg\rangle\Bigg\langle\sum_{k=n+1}^{n+m}\sum_{l=n+1}^{n+m}h_kh_l\Bigg\rangle, \tag{24}$$

where the following holds,

$$\big\langle\mathbf{r}_1^2(n)\mathbf{r}_1^2(n+m)\big\rangle=3n^2d_1^4+nd^2md^2=3n^2d_1^4+nmd_1^4. \tag{25}$$

Putting this result into Equation 18 allows the covariance to be written as,

$$\mathrm{cov}\Big(\big\langle\mathbf{r}_1^2(n)\big\rangle,\big\langle\mathbf{r}_1^2(n+m)\big\rangle\Big)=3n^2d_1^4+nmd_1^4-n(n+m)d_1^4=3n^2d_1^4-n^2d_1^4=2n^2d_1^4, \tag{26}$$

indicating that the covariance depends only on the number of overlapping points, n. This can be rationalised as the steps are random and therefore for any non-overlapping points the covariance must be 0. However, the difference between the variance and the covariance becomes clear when we consider the number of *independent* trajectories to consider, which for the covariance is,

$$\mathrm{cov}_{1,i}\Big(\big\langle\mathbf{r}_1^2(n)\big\rangle,\big\langle\mathbf{r}_1^2(n+m)\big\rangle\Big)=\frac{2}{N_i(n+m)}n^2d_1^4, \tag{27}$$

since $N_i(n+m)$ is the minimum number of *independent* trajectories in both n and $n+m$. This result extends to three-dimensions identically to the self-timestep uncertainty to give the following for *independent* trajectories,

$$\mathrm{cov}_{3,i}\Big(\big\langle \mathbf{r}_3^2(n)\big\rangle, \big\langle \mathbf{r}_3^2(n+m)\big\rangle\Big) = \frac{2n^2 d_3^4}{3N_i(n+m)} \tag{28}$$

[1] Smith, W. & Gillan, M. J. The Random Walk and the Mean Squared Displacement. *Inf. Q. Comput. Simul. Condens. Phases* 54–64 (1996).