How to generate exponentially correlated Gaussian random numbers

Markus Deserno
Department of Chemistry and Biochemistry, UCLA, USA
(Dated: August 26, 2002)

An algorithm is described which generates a sequence of random numbers $r_1, r_2, \ldots$ with the following two properties: (*i*) each individual r_i is a Gaussian deviate with zero mean and unit variance; (*ii*) the autocorrelation function of the sequence decays exponentially with a predetermined decay time τ. A correlated random walk is discussed as a simple application.

I. THE ALGORITHM

Let g_n be a sequence of independent Gaussian deviates with zero mean and unit variance, *i. e.*

$$\mathrm{prob}(g_n = x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2} \qquad \forall n \in \mathbb{N}_0, \tag{1}$$

and in particular $\langle g_n \rangle = 0$ and $\langle g_n^2 \rangle = 1$. Such numbers can be generated quite easily on a computer [1, 2], and we will not discuss this issue any further.

Let $\tau > 0$ be a real number and introduce the correlation coefficient f as

$$f \;:=\; \mathrm{e}^{-1/\tau}. \tag{2}$$

Now define the sequence of numbers r_n recursively via

$$r_0 \;:=\; g_0 \quad ; \quad r_{n+1} \;:=\; f\,r_n + \sqrt{1-f^2}\,g_{n+1}. \tag{3}$$

This can also be written in a closed expression as

$$r_n \;=\; f^n g_0 + \sqrt{1-f^2}\sum_{i=1}^{n} g_i f^{n-i}. \tag{4}$$

Theorem 1 *The random numbers r_n are Gaussian deviates with zero mean and unit variance.*

Proof: Since each r_n is the sum of Gaussian deviates, it is also a Gaussian deviate. By construction r_0 has zero mean and unit variance, and by induction we have

$$\langle r_{n+1}\rangle \;=\; f\underbrace{\langle r_n\rangle}_{=0} + \sqrt{1-f^2}\underbrace{\langle g_{n+1}\rangle}_{=0} \;=\; 0.$$

and

$$\langle r_{n+1}^2\rangle \;=\; f^2\underbrace{\langle r_n^2\rangle}_{=1} + \big(1-f^2\big)\underbrace{\langle g_{n+1}^2\rangle}_{=1} \;=\; 1,$$

where the cross-term $\langle r_n g_{n+1}\rangle$ vanishes because r_n and g_{n+1} are independent and thus uncorrelated. We also used the result for the variance of a linear combination of independent random variables [4]. □

We now look at the autocorrelation coefficient $c(n;m)$ of the sequence r_n, which we shall define by

$$c(n;m) \;:=\; \frac{\langle r_m r_{m+n}\rangle - \langle r_m\rangle\langle r_{m+n}\rangle}{\langle (r_m - \langle r_m\rangle)^2\rangle^{1/2}\langle (r_{m+n} - \langle r_{m+n}\rangle)^2\rangle^{1/2}}. \tag{5}$$

Theorem 2 *The autocorrelation coefficient $c(n;m)$ is independent of m (i. e., the corresponding stochastic process is* stationary*) and is given by*

$$c(n;m) \;\equiv\; c(n) \;=\; f^n \;=\; \mathrm{e}^{-n/\tau}. \tag{6}$$

Proof: The r_n have zero mean and unit variance, therefore $c(n;m)$ reduces to $\langle r_m r_{m+n}\rangle$. We now simply have to calculate:

$$\begin{aligned}\langle r_m r_{m+n}\rangle \;&=\; \Big\langle r_m\Big(f^n r_m + \sqrt{1-f^2}\sum_{i=m+1}^{m+n} g_i f^{m+n-i}\Big)\Big\rangle \\ &\overset{*}{=}\; f^n\,\langle r_m^2\rangle \;=\; \mathrm{e}^{-n/\tau}.\end{aligned}$$

At the key step $*$ we used the fact that the Gaussian deviates g_i are *not* correlated with the number r_m, since $i > m$. □

Since correlation coefficients are invariant under (affine) linear transformations of random variables, $X \to a + bX$, we have the

Corrolary 1 *The random numbers*

$$\tilde{r}_n \;:=\; \mu + \sigma r_n \tag{7}$$

are Gaussian with mean μ and variance σ^2, and their correlation time is also τ.

Remarks:

1. The sequence r_n of random numbers is a Markov process [3]. Eqn. (3) specifies precisely the two necessary ingredients: (*i*) the initial state and (*ii*) the transition probability. The latter is obviously given by a Gaussian with mean $(f-1)r_n$ and variance $1-f^2$.

2. The above prescription for getting correlated random numbers is closely related to the following method of getting *two* correlated Gaussian random numbers. Let a be a Gaussian random variable with mean μ_a and variance σ_a^2. Let g be a Gaussian random variable with zero mean and unit variance. Then it is easily checked that

$$b \;:=\; f\frac{\sigma_b}{\sigma_a}\big(a-\mu_a\big) + \sqrt{1-f^2}\sigma_b g + \mu_b \tag{8}$$

is a Gaussian random variable with mean μ_b and variance σ_b^2, which has a correlation coefficient f with the random variable a.

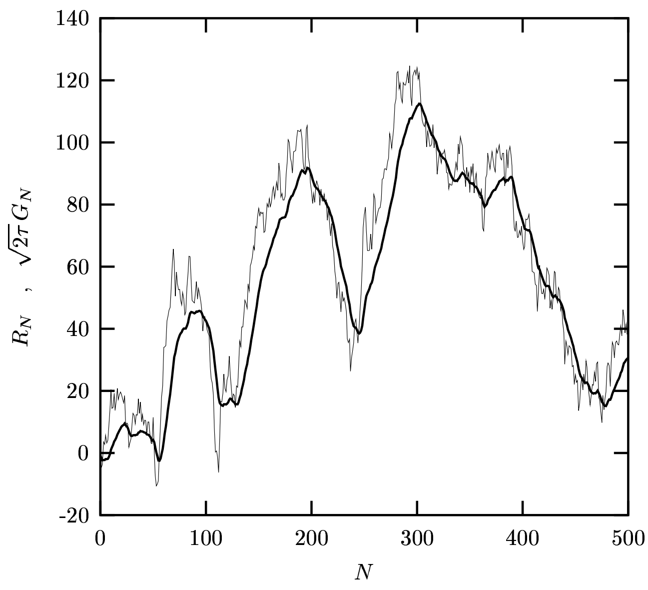

FIG. 1: Comparison between a (rescaled!) Gaussian random walk G_N (fine solid line) and a correlated random walk R_N (bold solid line) for a correlation time of $\tau = 10$.

II. APPLICATION: CORRELATED RANDOM WALK

By adding up the Gaussian deviates g_n or the correlated deviates r_n, we can create so called "random walks":

$$G_N := \sum_{n=0}^{N} g_n \qquad , \qquad R_N := \sum_{n=0}^{N} r_n. \tag{9}$$

All increments for G_N are independent, so this is again Markov process. But the increments in R_N are strongly backwards correlated and this is no Markov process any more. However, it can be *made* into a Markov process by suitably *extending the "phase space"* and including the increments r_n, since we then have

$$\begin{pmatrix} R_{n+1} \\ r_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & f \end{pmatrix} \begin{pmatrix} R_n \\ r_n \end{pmatrix} + \sqrt{1-f^2} \begin{pmatrix} 0 \\ g_{n+1} \end{pmatrix}. \tag{10}$$

In this form the process evidently possesses the Markov property. This just illustrates that Markov processes can also have some sort of memory – even though the kind of memory is limited to special forms.

We will now investigate how these two random walks are related to each other. A straightforward calculation gives

$$\begin{aligned} R_N &= \sum_{n=0}^{N} \left[f^n g_0 + \sqrt{1-f^2} \sum_{i=1}^{n} g_i f^{n-i} \right] \\ &= g_0 \frac{1-f^{N+1}}{1-f} + \sqrt{1-f^2} \sum_{i=1}^{N} g_i \sum_{n=i}^{N} f^{n-i} \\ &= g_0 \frac{1-f^{N+1}}{1-f} + \frac{\sqrt{1-f^2}}{1-f} \left[\sum_{i=1}^{N} g_i - f \sum_{i=1}^{N} g_i f^{N-i} \right]. \end{aligned}$$

The first sum in the bracket is $G_N - g_0$, and using Eqn. (4) the second sum can be re-expressed by r_N. We thus get

$$R_N = \sqrt{\frac{1+f}{1-f}}\, G_N - \frac{f}{1-f}\, r_N + \frac{1-\sqrt{1-f^2}}{1-f}\, g_0. \tag{11}$$

The first term is a Gaussian random walk with a prefactor, it will thus on average grow with time. The second and third term do not grow, since they are just Gaussian deviates with prefactors (the third is a constant anyway). For sufficiently large N the first term hence dominates the expression. Rewriting f in terms of the correlation time τ we then get the approximate relation

$$R_N \sim \sqrt{\coth \frac{1}{2\tau}}\, G_N \approx \begin{cases} G_N & : \tau \ll 1 \\ \sqrt{2\tau}\, G_N & : \tau \gg 1 \end{cases} \tag{12}$$

It can hence be seen that the correlated Gaussian random walk is "faster" than the uncorrelated random walk, even though in both cases the increments are Gaussian deviates with zero mean and unit variance! More precisely, the mean square displacement of the correlated walk will grow stronger – in the case $\tau \gg 1$ by a factor of 2τ. Persistence gives distance!

As an illustration, Fig.1 shows a comparison between G_N and R_N for a correlation time $\tau = 10$. It is clearly visible that the correlated random walk is smoother on time scales smaller or equal to τ, but for longer times it follows the same *large-scale motion* as the Gaussian random walk (only the amplitude is larger by a factor of about $\sqrt{2\tau}$). The difference in short-term "wiggliness" between both random walks is due to the second term in Eqn. (11). Another way of looking at this is the following: The correlated random walk can be generated from the uncorrelated one by a specific "filtering process", which suitably combines the previous values (over essentially a distance τ). This also "explains" why the coarse grained random walk lags behind.

[1] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed., Cambridge University Press, Cambridge (1994).

[2] D. E. Knuth, *The Art of Computer Programming*, Vol. 2, Seminumerical Algorithms, Addison-Wesley (1969).

[3] S. Karlin and H. M. Taylor, *A first course in stochastic processes*, 2nd ed.. Academic Press, New York (1975).

[4] Let X and Y be random variables with variances σ_X^2 and σ_Y^2, respectively. Then the linear combination $Z := aX + bY$ has variance $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$.