Note on the variance of the random variables used in the Global Consciousness Project

Chris Phillips
7 March 2023

[A new final paragraph has been added to the Conclusions section. The original version of this note can be seen here.]

1. Introduction

This note examines the variance of the random variables used in the Global Consciousness Project (GCP), which was a long-term study of the behaviour of a worldwide network of random number generators and its possible connection to the timing of events arousing widespread human interest. In a 17-year series of more than 500 formal pre-specified statistical hypotheses for particular events (1998-2015), an increase in correlations between the random number generators was observed, to a degree that was extremely statistically significant, amounting cumulatively to more than 7 standard deviations above chance expectation (Nelson, 1999-2020).

Technical details of the random number generators are given in Bancel (2016), on which the following description is based. Most of the random number generators used in the project were either Mindsong or Orion devices. Both of these work on essentially the same principles, with a random stream of bits (0s and 1s) initially being generated by the hardware device[1], a bitmask (again of 0s and 1s) then being applied to the bitstream using the XOR (exclusive or) operation, and finally a sample of the resulting bits being taken and added together to produce a random variable. In the XOR operation, at positions in the stream where the value of the mask is 0 the random bit is left unchanged, but at positions where the value of the mask is 1 the random bit is inverted. Ideally the bits generated by the hardware should be unbiased between 0s and 1s and the values of different bits should be statistically independent of one another. The mask is intended to mitigate the effects of any departure from this ideal behaviour in the hardware-generated bitstreams. In particular, if the part of the mask corresponding to the sample contains equal numbers of 0s and 1s, either for each individual sample or on average over all the samples, then its application will eliminate any overall bias between 0s and 1s.

In the GCP, the sample size for both the devices is 200 bits. But there are significant differences between the operation of the two devices:
(1) In the Mindsong device, the hardware produces a single random bitstream and then an internal 560-bit mask is applied to it. The mask consists of all 70 of the possible 8-bit sequences that contain equal numbers of 0s and 1s. Evidently the mask is intended not only to eliminate the effects of overall bias between 0s and 1s, but also to reduce the effects of autocorrelation between different bits within the stream.
(2) In the Orion device, the hardware generates two separate random bitstreams and then combines them using the XOR operation. In the Global Consciousness Project, a mask consisting simply of alternating 0s and 1s is then applied to the output of the device.

While it was found that the variance of the random variables produced by the Orion devices was generally close to the ideal value, for the Mindsong devices the variance was nearly always significantly larger than the ideal value, in many cases by as much as 0.1% (see Bancel, 2016, Figure 1). For an event lasting several hours, with several dozen random number generators being

1 Details of the hardware operation, which in both cases relies on the quantum tunnelling phenomenon, can be found on the GCP website (Nelson, 1999-2020).

sampled once a second, hundreds of thousands of values are involved, and even relatively small departures from ideality may have a large cumulative effect. For this reason, it was found necessary to normalise the data by using the measured variance for each device.

The purpose of this note is to try to understand how this non-ideal behaviour may arise - particularly for the Mindsong device - by examining the relationship between the variance and the statistical properties of the bitstreams from which the random variables are produced.

2. General result

First, a general result for the variance is derived, by considering a sequence X_n of *N* random binary variables (for $n=1$ to *N*), to which a binary mask is applied. For computational convenience, it will be assumed that the variables take the values 1 and -1. The *N* corresponding elements of the mask, which will be denoted by I_n , can also be represented by values 1 and -1, so that the application of the mask is equivalent to multiplying X_n by I_n , for each value of *n*. (If the I_n represents a mask composed of bits, this means that $I_n = 1$ represents a 0 and $I_n = -1$ represents a 1.)

The sum of the *N* random variables, after the application of the mask, is

$$Y = \sum_{n=1}^{N} I_n X_n \tag{1}$$

At this stage, the only assumption that needs to be made is that the statistical properties of the X_n do not vary with time. This implies that the mean of X_n is independent of *n*, and the covariance of X_n and X_{n+m} depends only on the difference *m*. Therefore, denoting the expectation by an overbar, the mean and covariance can be written as:

$$\mu = \overline{X_n} \tag{2}$$

$$\gamma_m = \overline{(X_n - \mu)(X_{n+m} - \mu)} \qquad \text{for } m \geq 1 \tag{3}$$

Note that because X_n takes only the values 1 and -1, the mean square of X_n is 1, so that $\gamma_0 = 1 - \mu^2$.

In addition to averaging over different values of the random variables X_n , it is also necessary to average over different positions of the mask relative to the *N* elements of the sequence that are sampled. Performing these two averages in sequence and denoting the second by $\langle ... \rangle$, it is straightforward to show that the mean and mean square of *Y* are given by

$$\langle \overline{Y} \rangle = \mu \sum_{n=1}^{N} \langle I_n \rangle \tag{4}$$

$$\langle \overline{Y^2} \rangle = N + 2 \sum_{m=1}^{N-1} \gamma_m \sum_{n=1}^{N-m} \langle I_n I_{n+m} \rangle + \mu^2 [J(N) - N] \tag{5}$$

in which

$$J(N) = \left\langle \left(\sum_{n=1}^{N} I_n \right)^2 \right\rangle \tag{6}$$

Note that this general result requires no assumption about the statistical distribution of mask positions. The position of the mask relative to the elements of the sequence sampled is not assumed to be uniformly distributed. It may be non-uniformly distributed, or may even take a single fixed value.

In equation (5) for the mean square value of *Y*, the second term on the right-hand side represents the effect of autocorrelation within the sequence X_n, and the third represents the effect of bias.

3. Application to the Mindsong random number generator

In order to use equations (4) and (5) to calculate the variance of *Y* when the Mindsong mask is used, it is necessary to evaluate the terms containing the I_n, and this requires an assumption to be made about the statistical distribution of the position of the mask relative to the *N* elements of the sequence X_n that are sampled. It is assumed that this position can be taken to be randomly and uniformly distributed.[2] This means that there is an equal probability that the first variable sampled, X_1, will correspond to each of the 560 elements of the mask.

Because the whole mask contains equal numbers of 0s and 1s, for each value of *n*, I_n takes the values 1 and -1 with equal probability, so $\langle I_n \rangle = 0$, and therefore from equation (4) the mean value $\langle \overline{Y} \rangle = 0$. For the Mindsong mask, for $m = 1$, the average over mask positions in equation (5) is found by direct calculation to be:

$$\langle I_n I_{n+1} \rangle = -\tfrac{1}{7} \tag{7}$$

It remains to consider the influence of autocorrelation and bias in the sequence X_n, which determine μ and the γ_m in equation (5). Although it has been assumed that the statistical properties of the sequence do not change with time, in general the distribution of each X_n may still depend on the values of all the other elements. But for illustration, a simple model will be considered, in which the distribution of X_n depends only on the value of the variable immediately before it, X_{n-1}. It will also be assumed that the distribution differs by only a small amount from the ideal distribution, in which X_n has mean 0 and variance 1. With these assumptions, the dependence can be expressed as

$$\text{Probability}(X_n = 1) = \begin{cases} \tfrac{1}{2} + (\alpha+\beta)\epsilon & \text{if } X_{n-1} = 1 \\ \tfrac{1}{2} + (\alpha-\beta)\epsilon & \text{if } X_{n-1} = -1 \end{cases} \tag{8}$$

In these equations, the parameter ϵ reflects the overall size of the deviation from ideality, which is assumed to be small, and the coefficients α and β determine the bias and autocorrelation respectively in the bitstream.

2 As discussed by Bancel (2016), the timing of the 200-bit sample is determined by the computer's clock, but the application of the Mindsong device's internal mask is not. Synchronisation between the sample and the mask position would require both these timings to agree to within about a third of a millisecond. In contrast, for most of the computers in the network, the clocks were found to be inaccurate by at least a second, and in most tests this inaccuracy drifted by at least a second in one hour.

Recalling the assumption that μ, the mean value of X_n, is independent of *n*, it is straightforward to show from equation (8) that consistency requires

$$\mu = \frac{2\alpha\epsilon}{1-2\beta\epsilon} \tag{9}$$

Also, defining p_m to be the probability that the value of X_{n+m} is equal to that of X_n, it is found that

$$p_1 = \tfrac{1}{2} + \beta\epsilon + \alpha\mu\epsilon \tag{10}$$

$$p_{m+1} = p_1\, p_m + (1-p_1)\,(1-p_m) \tag{11}$$

Equations (9-11) are exact, but considering the limit in which ϵ, the deviation from ideality, becomes small, the quantities that appear in equation (5) are obtained approximately as

$$\mu = 2\alpha\epsilon + O(\epsilon^2) \tag{12}$$

$$\gamma_1 = 2\beta\epsilon - \mu^2 + O(\epsilon^2) \tag{13}$$

$$\gamma_m = -\mu^2 + O(\epsilon^2) \qquad \text{when } m > 1 \tag{14}$$

where in each equation the notation means that the final error term is asymptotically proportional to ϵ^2 when ϵ is small.

Finally, recalling that $\langle Y \rangle = 0$, equation (5) gives the variance of *Y*. This can be expressed as a fraction of its ideal value, $\mathrm{Var}(Y_0) = N$. This eliminates the effect of the rescaling of X_n used above, so that the following results are directly applicable to the bit-based variables used in the GCP.

Two cases must be considered. First, when $\beta \neq 0$ and autocorrelation is present, the leading-order effect on the variance comes from the γ_1 term of the sum and is proportional to ϵ, while the contributions from the other γ_m and from μ^2 are smaller, and proportional to ϵ^2. Using equation (7), this gives

$$\frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y_0)} = 1 - \tfrac{4}{7}(1-N^{-1})\beta\epsilon + O(\epsilon^2) \tag{15}$$

In the second case, when $\beta = 0$, autocorrelation is absent. This means all the γ_m are zero, and the effect on the variance comes entirely from the final term in equation (5), proportional to μ^2. This gives

$$\frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y_0)} = 1 - 4\left[1-N^{-1}J(N)\right]\alpha^2\epsilon^2 + O(\epsilon^3) \tag{16}$$

in which the value of $J(N)$ for the value of *N* used in the GCP is

$$J(200) = \tfrac{22}{7} \tag{17}$$

For the Mindsong device, in this model and for the sample size used, when autocorrelation is absent the variance is always lower than the ideal value. When autocorrelation is present, the variance may be either higher or lower than the ideal value, but it will be higher if each of the X_n is *negatively* correlated with its immediate predecessor (so that $\beta < 0$).

4. Application to the Orion random number generator

The Orion random number generator, as used in the GCP, differs from the Mindsong in two respects. Firstly, two random bitstreams are combined using XOR before a mask is applied. Secondly, in contrast to the elaborate mask incorporated in the Mindsong device, the mask applied here consists simply of 0s alternating with 1s.

Because of the simple alternating nature of the mask, it is not necessary to make the same assumption as for the Mindsong device - namely that the mask position can be taken to be randomly and uniformly distributed relative to the *N* elements sampled. It is sufficient just to assume that *N* is an even number.[3] With this assumption, it follows that

$$I_n I_{n+1} = -1 \qquad \sum_{n=1}^{N} I_n = 0 \tag{18}$$

Using the same formalism as above, based on random variables X_n taking the values 1 and -1, the combination of the two bitstreams by XOR is equivalent to simply multiplying together the variables representing the two sequences, and then reversing the sign of the product. If the same simple model as above is used for each of the sequences, this gives straightforwardly

$$\mu = -4\alpha_1\alpha_2\epsilon^2 + O(\epsilon^3) \tag{19}$$

$$\gamma_1 = 4\beta_1\beta_2\epsilon^2 + O(\epsilon^3) \tag{20}$$

$$\gamma_m = O(\epsilon^4) \qquad \text{when } m > 1 \tag{21}$$

in which the subscripts of the α and β coefficients indicate the two bitstreams that are combined.

As before, equations (4) and (18) show that $\langle \overline{Y} \rangle = 0$, and the variance of *Y* can be obtained directly from equation (5). Again, the result depends on whether autocorrelation is present in the sequences of random variables. If it is present in both sequences, then $\beta_1\beta_2 \neq 0$ and the leading-order effect comes from the γ_1 term of the sum, with other contributions asymptotically smaller:

$$\frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y_0)} = 1 - 8(1-N^{-1})\beta_1\beta_2\epsilon^2 + O(\epsilon^3) \tag{22}$$

In contrast, if autocorrelation is absent from both sequences, then all the γ_m are zero and the variance is found to be

3 This property of the mask is important because, while in the Mindsong device the mask is internal and not synchronised with the computer's clock, for Orion in the GCP the mask is applied by the computer to the output of the device. Therefore the position of the mask relative to the sample cannot be taken to have a uniform random distribution. In this situation, in order to avoid bias it is necessary not just that the mask as a whole is balanced between 0s and 1s, but that the part of the mask corresponding to the sample is balanced.

$$\frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y_0)} = 1 - 16\alpha_1^2\alpha_2^2\epsilon^4 + O(\epsilon^5) \tag{23}$$

For the Orion device as used in the GCP, in this model when autocorrelation is absent the variance is always lower than its ideal value (in the same way as for the Mindsong device). When autocorrelation is present in both the bitstreams that are combined, the variance may be either higher or lower than its ideal value, but it will be higher if the combined effect of the two autocorrelations is such that each of the X_n, obtained by combining the bitstreams, is *negatively* correlated with its immediate predecessor (again in the same way as for the Mindsong device).

The difference from the variance produced by the Mindsong device is that here it is much smaller - of order ϵ^2 instead of ϵ in the presence of autocorrelation, and of order ϵ^4 instead of ϵ^2 in its absence. This is the result of combining two separate hardware-produced bitstreams.

5. Conclusions

The results for the variance of the random variables produced by the two devices are equations (15-17), (22) and (23), based on the simplified model equation (8) of the statistical properties of the bitstreams. In these equations, *N* is the number of bits sampled, ϵ reflects the overall size of the deviations of the bitstreams from ideality (assumed to be small), and the coefficients α and β determine the bias and autocorrelation of the individual bitstreams respectively.

In qualitative terms, the behaviour is similar for both devices. Bias of the bitstreams alone, in the absence of autocorrelation, will always produce a decrease in the variance below its ideal value. But if autocorrelation is also present, there may be either an increase or a decrease in the variance, depending on the direction of the autocorrelation. For both devices, the variance will be increased if the direction of the autocorrelation is such as to produce a negative correlation between successive bits.

But quantitatively, the difference between the devices is that the changes in the variance from its ideal value, whether positive or negative, are much larger for the Mindsong than for the Orion device. In terms of the small parameter ϵ, in the leading-order approximation for Mindsong they are proportional to ϵ in the presence of autocorrelation and ϵ^2 in its absence, but for Orion they are proportional to only ϵ^2 in the presence of autocorrelation and ϵ^4 in its absence.

These scalings are consistent with the plot in Figure 1 of Bancel (2016), which shows that nearly all the measured Mindsong variances are significantly larger than the ideal value, but that the measured Orion variances are clustered tightly around the ideal value, with very few differing from it significantly. In terms of equation (15), the values shown in the plot would reflect a value of $\beta\epsilon$ for the Mindsong devices of roughly -10^{-3} .

Although the Mindsong mask was evidently intended to cancel the effects of autocorrelation by combining all the possible balanced 8-bit sequences, it does not succeed in cancelling even its leading-order contribution, because the resulting value of $\langle I_n I_{n+1} \rangle$ is non-zero. The same is true of the mask used for the Orion devices (and in fact this simple alternating mask maximises the absolute value of $\langle I_n I_{n+1} \rangle$). But despite this, the Orion devices produce a much closer approximation to ideal behaviour because two independent bitstreams are combined, which in effect multiplies the small errors together. If an even closer approach to ideality were required, it could be achieved by combining three or more independent bitstreams rather than two.

Although the equations presented here are based on a particular simple model of autocorrelation and bias, the main conclusion - that a closer approach to ideality can be obtained by combining two or more independent bitstreams - is expected to be valid more generally.

References

Peter Amalric Bancel (2016). Searching for Global Consciousness: Supplementary Materials - Experimental and analytical details. Available at ResearchGate.

Roger Nelson (1999-2020). The Global Consciousness Project: Meaningful Correlations in Random Data. (Project website.)