Exclusive OR (XOR) and hardware random number generators

Robert B Davies

February 28, 2002

1 Introduction

The *exclusive or* (XOR) operation is commonly used to reduce the bias from the bits generated by a *hardware random number generator*. Typically, the uncorrected bits generated by a hardware random number generator will have *expectation* (long run average value) different from the ideal value of $\frac{1}{2}$. Also, adjacent bits may be *correlated*. In this note, I find the expectations and correlations of various combinations of random bits using the XOR operator under a variety of assumptions about the means and correlations of the original variables. Specifically, I am interested in the effectiveness of the XOR operator for reducing *bias* (the deviation of the expectation from $\frac{1}{2}$), and what happens when the successive bits are correlated.

The symbols X, Y, Z etc. will denote random bits (taking the values 0 or 1 – in probability theory they would be known as the outcomes of *Bernoulli* trials).

The results are in section 3 and the derivations of these results are in section 4.

2 Notation and basics

2.1 Exclusive OR

I use the symbol $\otimes$ to denote the exclusive-or operation. So $X \otimes Y = 1$ if just one of X and Y is equal to 1; otherwise $X \otimes Y = 0$.

	$Y = 0$	$Y = 1$
$X = 0$	$X \otimes Y = 0$	$X \otimes Y = 1$
$X = 1$	$X \otimes Y = 1$	$X \otimes Y = 0$

The XOR operation is commutative

$$X \otimes Y = Y \otimes X$$

and associative

$$X \otimes (Y \otimes Z) = (X \otimes Y) \otimes Z.$$

2.2 Expectation, variance and covariance

$E(X)$ denotes the *expected value* or *mean value* of X, that is, the average value of a large number of repeated trials. In the case of random bits $E(X) = \Pr(X = 1)$ where Pr denotes probability.

$\mathrm{var}(X) = E[\{X - E(X)\}^2]$ denotes the *variance* of X. Variance is a measure of the variability of a random variable. In the case of random bits $\mathrm{var}(X) = E(X)\{1 - E(X)\}$. This reaches a maximum value of $\frac{1}{4}$ when $E(X) = \frac{1}{2}$.

$\mathrm{cov}(X, Y) = E[\{X - E(X)\}\{Y - E(Y)\}]$ denotes the *covariance* of X and Y and

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}} \tag{1}$$

denotes the *correlation* of X and Y. Correlations are always between –1 and 1. When two variables are identical their correlation is equal to 1. When the two variables are *statistically independent* the correlation is zero. In the case of random bits $\mathrm{cov}(X, Y) = \Pr(X = 1, Y = 1) - \Pr(X = 1)\Pr(Y = 1)$.

When the expected value, $E(X)$, of a bit, X, is close to $\frac{1}{2}$, the variance, $\mathrm{var}(X)$, is very close to $\frac{1}{4}$ and so $\mathrm{corr}(X, Y)$ is close to $4 \times \mathrm{cov}(X, Y)$. This sometimes provides convenient approximate simplification.

2.3 Correlation and independence

A sequence of random variables $X_1, X_2, \ldots$ are statistically independent if the following is satisfied for each of variables X_i: the variables apart from X_i do not provide any useful information for predicting the value of X_i.

In the case of pairs of random bits *correlation = zero* and independence are equivalent. However when more than two random bits are involved, independence implies zero correlation but not vice versa. Here is a simple example. Consider X, Y and $X \otimes Y$ where X and Y are random bits with expected value $\frac{1}{2}$. Each pair of X, Y and $X \otimes Y$ are uncorrelated with each other, but X, Y and $X \otimes Y$ are not independent since given any two of these variables one can calculate the third.

3 Main results

3.1 Independent biased pairs

If X and Y are independent random bits with $E(X) = \mu$ and $E(Y) = \nu$ then

$$E(X \otimes Y) = \mu + \nu - 2\mu\nu = \tfrac{1}{2} - 2(\mu - \tfrac{1}{2})(\nu - \tfrac{1}{2}). \tag{2}$$

Thus if μ and ν are *close* to $\frac{1}{2}$ then $E(X \otimes Y)$ is *very close* to $\frac{1}{2}$. For example, if $\mu = \nu = 0.6$ then $E(X \otimes Y) = 0.48$.

In all cases $|E(X \otimes Y) - \frac{1}{2}| \leq \min(|\mu - \frac{1}{2}|, |\nu - \frac{1}{2}|)$ and so the XOR operation always reduces bias (except when one of the variables is fixed) *when the component bits are independent.*

Presenting the same result when $\mu = \nu$: if X and Y are independent and $E(X) = E(Y) = \mu$ then

$$E(X \otimes Y) = 2\mu(1 - \mu) = \tfrac{1}{2} - 2(\mu - \tfrac{1}{2})^2. \tag{3}$$

3.2 Independent sequences

Suppose we have n independent random bits each with expected value μ. Then the expected value of the result of XORing all of these variables is

$$\tfrac{1}{2} + (-2)^{n-1}(\mu - \tfrac{1}{2})^n. \tag{4}$$

The following table shows the bias (expected value – 0.5) for various values of the mean, μ, and number bits, n.

μ	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
0.51	−0.0002	0.000004	−0.00000008	0.0000000016	−0.00000000003
0.52	−0.0008	0.000032	−0.00000128	0.0000000512	−0.00000000205
0.53	−0.0018	0.000108	−0.00000648	0.0000003888	−0.00000002333
0.54	−0.0032	0.000256	−0.00002048	0.0000016384	−0.00000013107
0.55	−0.0050	0.000500	−0.00005000	0.0000050000	−0.00000050000
0.56	−0.0072	0.000864	−0.00010368	0.0000124416	−0.00000149299
0.57	−0.0098	0.001372	−0.00019208	0.0000268912	−0.00000376477
0.58	−0.0128	0.002048	−0.00032768	0.0000524288	−0.00000838861
0.59	−0.0162	0.002916	−0.00052488	0.0000944784	−0.00001700611
0.60	−0.0200	0.004000	−0.00080000	0.0001600000	−0.00003200000

3.3 Correlated pairs

If $E(X) = \mu$, $E(Y) = \nu$ and X and Y have correlation ρ then

$$\begin{aligned} E(X \otimes Y) &= \tfrac{1}{2} - 2(\mu - \tfrac{1}{2})(\nu - \tfrac{1}{2}) - 2\rho\sqrt{\mu(1-\mu)\nu(1-\nu)} \qquad (5) \\ &\approx \tfrac{1}{2} - 2(\mu - \tfrac{1}{2})(\nu - \tfrac{1}{2}) - \tfrac{1}{2}\rho \end{aligned}$$

(assuming μ, ν are near $\frac{1}{2}$). Thus a small amount of correlation can add significant bias to the result.

3.4 Independent correlated pairs

Suppose $E(X_1) = E(X_2) = E(Y_1) = E(Y_2) = \mu$, the pair (X_1, X_2) is independent of the pair (Y_1, Y_2) and $\text{cov}(X_1, X_2) = \text{cov}(Y_1, Y_2) = c$.

For example, X_1 and X_2 might be successive observations from a random number generator and so are slightly correlated with each other. Y_1

and Y_2 are from a second identical generator that operates completely independently of the first one.

Then

$$\operatorname{cov}(X_1 \otimes Y_1, X_2 \otimes Y_2) = 4\{c^2 + 2c(\mu - \tfrac{1}{2})^2\}. \tag{6}$$

In terms of correlation (assuming μ near $\frac{1}{2}$)

$$\operatorname{corr}(X_1 \otimes Y_1, X_2 \otimes Y_2) \approx \rho^2 + 8\rho(\mu - \tfrac{1}{2})^2 \tag{7}$$

where ρ is the correlation. Thus, if ρ is small and μ is close to $\frac{1}{2}$ then $\operatorname{corr}(X_1 \otimes Y_1, X_2 \otimes Y_2)$ is very small. If μ is not close to $\frac{1}{2}$ then some of the correlation will persist.

3.5 Words

Suppose we have a sequence of bits

$$W_1, W_2, \ldots, W_n, X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n, Z_1, Z_2, \ldots, Z_n$$

from a random number generator. We are going to amalgamate then into words **W**, **X**, **Y**, **Z**, each with $n > 2$ bits, and then bit-wise XOR words **W** and **X** and bit-wise XOR words **Y** and **Z**. Here, I am letting **W** etc. denote the sequences $W_1, W_2, \ldots, W_n$, etc.

W_1	W_2	$\ldots$	W_n	Y_1	Y_2	$\ldots$	Y_n
X_1	X_2	$\ldots$	X_n	Z_1	Z_2	$\ldots$	Z_n
			$\Downarrow$				
$W_1 \otimes X_1$	$W_2 \otimes X_2$	$\ldots$	$W_n \otimes X_n$	$Y_1 \otimes Z_1$	$Y_2 \otimes Z_2$	$\ldots$	$Y_n \otimes Z_n$

Suppose the correlation structure is such that only adjacent bits are significantly correlated. Then formulae (6) and (7) give us the expected values and correlation structure of the combined word except for the correlation between $W_1 \otimes X_1$ and $W_n \otimes X_n$; $Y_1 \otimes Z_1$ and $Y_n \otimes Z_n$; $W_n \otimes X_n$ and $Y_1 \otimes Z_1$.

Consider $W_1 \otimes X_1$ and $W_n \otimes X_n$ (or equivalently $Y_1 \otimes Z_1$ and $Y_n \otimes Z_n$). Suppose $E(W_1) = E(W_n) = E(X_1) = E(X_n) = \mu$ and $W_1, (W_n, X_1), X_n$ are independent but W_n and X_1 have covariance c and correlation ρ. Then

$$\operatorname{cov}(W_1 \otimes X_1, W_n \otimes X_n) = 4c(\mu - \tfrac{1}{2})^2. \tag{8}$$

In terms of correlation (assuming μ near $\frac{1}{2}$)

$$\operatorname{corr}(W_1 \otimes X_1, W_n \otimes X_n) \approx 4\rho(\mu - \tfrac{1}{2})^2. \tag{9}$$

This will be small under the same kind of conditions that make (7) small.

Now consider $W_n \otimes X_n$ and $Y_1 \otimes Z_1$. The correlation structure is the same as above so that

$$\operatorname{corr}(W_n \otimes X_n, Y_1 \otimes Z_1) \approx 4\rho(\mu - \tfrac{1}{2})^2. \tag{10}$$

So arranging the bits into bytes, for example, and XORing the bytes gives a better performance than XORing adjacent bits as in section 3.3. I have not proved that there is no hidden dependence of the type found in section 3.6.2. While I think it is unlikely that there is hidden dependence, this does need to be proved.

3.6 Flawed correctors

Suppose we have a sequence of independent random bits $X_1, \ldots, X_n$ with expected value $\mu \neq \frac{1}{2}$. Following section 3.1 we can reduce the bias by forming the sequence $X_1 \otimes X_2$, $X_3 \otimes X_4$, $X_5 \otimes X_6, \ldots$. Of course, we have only half as many corrected bits as raw bits. This section considers two flawed methods of attempting to produce about as many corrected bits as raw bits.

3.6.1 Accumulated randomness

Let $Y_1, \ldots, Y_n$ be defined by $Y_1 = X_1$ and

$$Y_i = X_i \otimes Y_{i-1} \text{ for } i > 1.$$

We are attempting to produce a corrected sequence of random bits by taking the exclusive-or of all the bits so far observed. Y_{n-1} and Y_n, are the last two bits in our *corrected* series. They will have expected values very close to $\frac{1}{2}$ if n is large. However, they will have correlation

$$\operatorname{corr}(Y_{n-1}, Y_n) = -2(\mu - \tfrac{1}{2}) \tag{11}$$

so these *corrected* numbers have just replaced bias by correlation unless we use only every second Y_i.

3.6.2 Pair-wise exclusive-or

Let $Y_2, \ldots, Y_n$ be defined by

$$Y_i = X_{i-1} \otimes X_i.$$

Then

$$E(Y_i) = \tfrac{1}{2} - 2(\mu - \tfrac{1}{2})^2 \tag{12}$$

and

$$\operatorname{corr}(Y_i, Y_{i+1}) = \frac{2(\mu - \frac{1}{2})^2}{1 - 2\mu + 2\mu^2} \approx 4(\mu - \tfrac{1}{2})^2 \tag{13}$$

(assuming near μ near $\frac{1}{2}$). If $|i - j| > 1$ then $\operatorname{corr}(Y_i, Y_j) = 0$.

Apparently this method of correction has taken n raw bits and provided $n - 1$ corrected bits with a substantial improvement in the bias with just a minor increase in pair-wise correlation (assuming the bias of the raw bits is not too large). This is illusory. The correlation is not picking up the full extent of the dependence. In section 4.2.7 I define a random bit, S, which can be calculated from $Y_2, \ldots, Y_{n-1}$ such that, for large n

$$\mathrm{corr}(Y_n, S) \approx 2|\mu - \tfrac{1}{2}| \tag{14}$$

so that the deviation from unbiasedness and independence is really similar to that in the original series, which, of course, is what one would expect. Of course, there is no problem if we take only every second Y_i.

On the other hand, suppose we used the series $X_1 \otimes X_2$, $X_2 \otimes X_3$, $X_4 \otimes X_5$, $X_5 \otimes X_6$, $X_7 \otimes X_8, \ldots$. That is, we are omitting every *third* Y_i. Then the bias and correlations are of order $(\mu - \frac{1}{2})^2$ and there can be no hidden dependence. So this might be a reasonable way of increasing the yield of corrected bits.

3.7 Correlated triple

Suppose $E(X) = E(Y) = E(Z) = \mu$, $\mathrm{corr}(X, Y) = \mathrm{corr}(Y, Z) = \rho$ and the process X, Y, Z is Markov so that, for example, the *conditional expectation* $E(Z|X, Y) = E(Z|Y)$. Then

$$\begin{aligned} E(X \otimes Y \otimes Z) &= \tfrac{1}{2} + 4(\mu - \tfrac{1}{2})^3 + 4\rho(\mu - \tfrac{1}{2})\mu(1-\mu)(2-\rho) \\ &\approx \tfrac{1}{2} + 4(\mu - \tfrac{1}{2})^3 + 2\rho(\mu - \tfrac{1}{2}) \end{aligned} \tag{15}$$

if μ is near $\frac{1}{2}$ and ρ is near 0.

This formula is important if we are going to correct by XORing three bits and auto-correlation is present. If ρ is very small it is not going to present a problem but if it is of a similar size to $\mu - \frac{1}{2}$ then it may have a significant effect on the bias.

In general, there is no reason to suppose the process of random bits is exactly Markov. However, if ρ is small, and the dependence between non-adjacent bits very small, the Markov assumption is probably close enough.

4 Derivation of formulae

4.1 Calculation trick

Suppose X can take values 0 or 1.

Let $a(X) = 1 - 2X$. So $X = \{1 - a(X)\}/2$ and $a(X)$ takes the values 1 and -1 corresponding to X's 0 and 1. Then

$$a(X \otimes Y) = a(X)a(Y)$$

where $\otimes$ denotes the XOR operation. The usefulness of this is that we know how to do manipulate multiplication in probability calculations but doing

XOR calculations directly is awkward and unfamiliar. Also

$$\begin{aligned}
E\{a(X)\} &= 1 - 2E(X) \\
\operatorname{var}\{a(X)\} &= 4\operatorname{var}(X) \\
\operatorname{cov}\{a(X), a(Y)\} &= 4\operatorname{cov}(X, Y) \\
\operatorname{corr}\{a(X), a(Y)\} &= \operatorname{corr}(X, Y)
\end{aligned}$$

so we can transform expectations and variances between X and $a(X)$.

Also note that $a(X)^2 = 1$.

4.2 The derivations

4.2.1 Formula (2)

Suppose X and Y are independent, $E(X) = \mu$ and $E(Y) = \nu$ then

$$\begin{aligned}
E(X \otimes Y) &= \tfrac{1}{2} - \tfrac{1}{2}E\{a(X \otimes Y)\} \\
&= \tfrac{1}{2} - \tfrac{1}{2}E\{a(X)a(Y)\} \\
&= \tfrac{1}{2} - \tfrac{1}{2}E\{a(X)\}E\{a(Y)\} \\
&= \tfrac{1}{2} - \tfrac{1}{2}(1 - 2\mu)(1 - 2\nu) \\
&= \tfrac{1}{2} - 2(\mu - \tfrac{1}{2})(\nu - \tfrac{1}{2}).
\end{aligned}$$

4.2.2 Formula (5)

Suppose X and Y have covariance c and $E(X) = \mu$ and $E(Y) = \nu$.

$$\begin{aligned}
E(X \otimes Y) &= \tfrac{1}{2} - \tfrac{1}{2}E\{a(X \otimes Y)\} \\
&= \tfrac{1}{2} - \tfrac{1}{2}E\{a(X)a(Y)\} \\
&= \tfrac{1}{2} - \tfrac{1}{2}[E\{a(X)\}E\{a(Y)\} + 4c] \\
&= \tfrac{1}{2} - \tfrac{1}{2}(1 - 2\mu)(1 - 2\nu) - 2c \\
&= \tfrac{1}{2} - 2(\mu - \tfrac{1}{2})(\nu - \tfrac{1}{2}) - 2c.
\end{aligned}$$

4.2.3 Formula (6)

Suppose $E(X_1) = E(X_2) = E(Y_1) = E(Y_2) = \mu$, the pair (X_1, X_2) is independent of the pair (Y_1, Y_2) and $\mathrm{cov}(X_1, X_2) = \mathrm{cov}(Y_1, Y_2) = c$ then

$$
\begin{aligned}
&\mathrm{cov}(X_1 \otimes Y_1, X_2 \otimes Y_2)\\
&= \tfrac{1}{4}\,\mathrm{cov}\{a(X_1 \otimes Y_1), a(X_2 \otimes Y_2)\}\\
&= \tfrac{1}{4}\,\mathrm{cov}\{a(X_1)a(Y_1), a(X_2)a(Y_2)\}\\
&= \tfrac{1}{4}[E\{a(X_1)a(Y_1)a(X_2)a(Y_2)\} - E\{a(X_1)a(Y_1)\}E\{a(X_2)a(Y_2)\}]\\
&= \tfrac{1}{4}[E\{a(X_1)a(X_2)\}E\{a(Y_1)a(Y_2)\}\\
&\qquad - E\{a(X_1)\}E\{a(Y_1)\}E\{a(X_2)\}E\{a(Y_2)\}]\\
&= \tfrac{1}{4}[\{4c + (1-2\mu)^2\}^2 - (1-2\mu)^4]\\
&= 4\{c^2 + 2c(\mu - \tfrac{1}{2})^2\}.
\end{aligned}
$$

4.2.4 Formula (8)

Suppose $E(W_1) = E(W_n) = E(X_1) = E(X_n) = \mu$ and $W_1, (W_n, X_1), X_n$ are independent but W_n and X_1 have covariance c. Then

$$
\begin{aligned}
&\mathrm{cov}(W_1 \otimes X_1, W_n \otimes X_n)\\
&= \tfrac{1}{4}\,\mathrm{cov}\{a(W_1 \otimes X_1), a(W_n \otimes X_n)\}\\
&= \tfrac{1}{4}\,\mathrm{cov}\{a(W_1)a(X_1), a(W_n)a(X_n)\}\\
&= \tfrac{1}{4}[E\{a(W_1)a(X_1)a(W_n)a(X_n)\} - E\{a(W_1)a(X_1)\}E\{a(W_n)a(X_n)\}]\\
&= \tfrac{1}{4}[E\{a(W_1)\}E\{a(W_n)a(X_1)\}E\{a(X_n)\}\\
&\qquad - E\{a(W_1)\}E\{a(X_1)\}E\{a(W_n)\}E\{a(X_n)\}]\\
&= \tfrac{1}{4}(1-2\mu)^2[\{4c + (1-2\mu)^2\} - (1-2\mu)^2]\\
&= 4c(\mu - \tfrac{1}{2})^2.
\end{aligned}
$$

4.2.5 Formula (11)

Suppose X_n and Y_{n-1} are independent, $E(X_n) = \mu$ and $E(Y_{n-1}) = \frac{1}{2}$. Then

$$
\begin{aligned}
&\mathrm{corr}(Y_{n-1}, X_n \otimes Y_{n-1})\\
&= \mathrm{corr}\{a(Y_{n-1}), a(X_n \otimes Y_{n-1})\}\\
&= \mathrm{cov}\{a(Y_{n-1}), a(X_n \otimes Y_{n-1})\}\\
&= \mathrm{cov}\{a(Y_{n-1}), a(X_n)a(Y_{n-1})\}\\
&= E\{a(Y_{n-1})a(X_n)a(Y_{n-1})\} - E\{a(Y_{n-1})\}E\{a(X_n)\}E\{a(Y_{n-1})\}\\
&= E\{a(X_n)\}\\
&= 1 - 2\mu.
\end{aligned}
$$

4.2.6 Formula (13)

Suppose X_1, X_2, X_3 are independent with expected value μ. Then

$$\begin{aligned}
&\operatorname{cov}(X_1 \otimes X_2, X_2 \otimes X_3) \\
&= \tfrac{1}{4} \operatorname{cov}\{a(X_1)a(X_2), a(X_2)a(X_3)\} \\
&= \tfrac{1}{4}[E\{a(X_1)a(X_2)a(X_2)a(X_3)\} - E\{a(X_1)a(X_2)\}E\{a(X_2)a(X_3)\}] \\
&= \tfrac{1}{4}\{(1-2\mu)^2 - (1-2\mu)^4\} \\
&= 4(\mu - \tfrac{1}{2})^2\mu(1-\mu).
\end{aligned}$$

Divide by the variance of $X_1 \otimes X_2$ to get formula (13).

4.2.7 Formula (14)

Suppose $X_1, \ldots, X_n$ are a sequence of independent random bits with expected value μ. Let $Y_2, \ldots, Y_n$ be defined by

$$Y_i = X_{i-1} \otimes X_i.$$

Let

$$\begin{aligned}
Z_{n-1} &= Y_{n-1} = X_{n-1} \otimes X_{n-2}, \qquad (16) \\
Z_{n-2} &= Z_{n-1} \otimes Y_{n-2} = X_{n-1} \otimes X_{n-3}, \\
Z_{n-3} &= Z_{n-2} \otimes Y_{n-3} = X_{n-1} \otimes X_{n-4}, \\
&\ldots, \\
Z_2 &= Z_3 \otimes Y_2 = X_{n-1} \otimes X_1
\end{aligned}$$

and define

$$S = 1 \text{ if } \frac{1}{n-2}\sum_{i=2}^{n-1} Z_i > \tfrac{1}{2} \text{ and } S = 0 \text{ otherwise.}$$

S can be calculated from $Y_2, \ldots, Y_n$. From formula (16)

$$\frac{1}{n-2}\sum_{i=2}^{n-1} Z_i = \frac{1}{n-2}\sum_{i=1}^{n-2} X_i \text{ if } X_{n-1} = 0$$

and

$$\frac{1}{n-2}\sum_{i=2}^{n-1} Z_i = 1 - \frac{1}{n-2}\sum_{i=1}^{n-2} X_i \text{ if } X_{n-1} = 1.$$

If n is large

$$\frac{1}{n-2}\sum_{i=1}^{n-2} X_i > \tfrac{1}{2} \text{ with a high probability if } \mu > \tfrac{1}{2}$$

and

$$\frac{1}{n-2}\sum_{i=1}^{n-2} X_i < \tfrac{1}{2} \text{ with a high probability if } \mu < \tfrac{1}{2}.$$

Putting all this together, if n is large, we have (with a high probability)

$$S = 1 \text{ if } X_{n-1} = 0 \text{ and } \mu > \tfrac{1}{2} \text{ or if } X_{n-1} = 1 \text{ and } \mu < \tfrac{1}{2}$$

and

$$S = 0 \text{ if } X_{n-1} = 1 \text{ and } \mu > \tfrac{1}{2} \text{ or if } X_{n-1} = 0 \text{ and } \mu < \tfrac{1}{2}.$$

Hence (assuming μ is close to but not exactly $\frac{1}{2}$)

$$\begin{aligned}\operatorname{corr}(Y_n, S) &\approx -\operatorname{corr}(X_n \otimes X_{n-1}, X_{n-1}) \times \operatorname{sign}(\mu - \tfrac{1}{2})\\ &\approx 2|\mu - \tfrac{1}{2}|\end{aligned}$$

since

$$\begin{aligned}&\operatorname{corr}(X_n \otimes X_{n-1}, X_{n-1})\\ &\approx \operatorname{cov}\{a(X_n)a(X_{n-1}), a(X_{n-1})\}\\ &= E\{a(X_n)a(X_{n-1})a(X_{n-1})\} - E\{a(X_n)a(X_{n-1})\}E\{a(X_{n-1})\}\\ &= (1-2\mu) - (1-2\mu)^3 \approx -2(\mu - \tfrac{1}{2}).\end{aligned}$$

4.2.8 Formula (15)

Suppose $E(X) = E(Y) = E(Z) = \mu$, $\operatorname{corr}(X, Y) = \operatorname{corr}(Y, Z) = \rho$ and the process is Markov so that the *conditional expectation* $E(Z|X, Y) = E(Z|Y)$. Now

$$\begin{aligned}E(Z|Y=1) &= \Pr(Z=1, Y=1)/\Pr(Y=1)\\ &= E(YZ)/\mu\\ &= \{\rho\mu(1-\mu) + \mu^2\}/\mu\\ &= \rho - \rho\mu + \mu\end{aligned}$$

and

$$\begin{aligned}E(Z|Y=0) &= \Pr(Z=1, Y=0)/\Pr(Y=0)\\ &= \{\Pr(Z=1) - \Pr(Z=1, Y=1)\}/\Pr(Y=0)\\ &= \{\mu - E(YZ)\}/(1-\mu)\\ &= \{\mu - \rho\mu(1-\mu) - \mu^2\}/(1-\mu) = -\rho\mu + \mu\end{aligned}$$

so that $E(Z|Y) = \rho Y - \rho\mu + \mu$. Hence

$$E\{a(Z)|Y\} = 1 - 2(\rho Y - \rho\mu + \mu) = \rho a(Y) + (1-2\mu)(1-\rho). \qquad (17)$$

Then

$$\begin{aligned}
&E(X \otimes Y \otimes Z) \\
&= \tfrac{1}{2} - \tfrac{1}{2}E\{a(X \otimes Y \otimes Z)\} \\
&= \tfrac{1}{2} - \tfrac{1}{2}E\{a(X)a(Y)a(Z)\} \\
&= \tfrac{1}{2} - \tfrac{1}{2}E[a(X)a(Y)E\{a(Z)|X,Y\}] \\
&= \tfrac{1}{2} - \tfrac{1}{2}E[a(X)a(Y)E\{a(Z)|Y\}] \\
&= \tfrac{1}{2} - \tfrac{1}{2}E[a(X)a(Y)\{\rho a(Y) + (1-2\mu)(1-\rho)\}] \\
&= \tfrac{1}{2} - \tfrac{1}{2}E\{\rho a(X) + a(X)a(Y)(1-2\mu)(1-\rho)\} \\
&= \tfrac{1}{2} - \tfrac{1}{2}(1-2\mu)[\rho + \{(1-2\mu)^2 + 4\rho\mu(1-\mu)\}(1-\rho)] \\
&= \tfrac{1}{2} - \tfrac{1}{2}(1-\rho)(1-2\mu)^3 - \tfrac{1}{2}\rho(1-2\mu)\{1 + 4\mu(1-\mu)(1-\rho)\} \\
&= \tfrac{1}{2} + 4(1-\rho)(\mu - \tfrac{1}{2})^3 + \rho(\mu - \tfrac{1}{2})\{1 + 4\mu(1-\mu)(1-\rho)\} \\
&= \tfrac{1}{2} + 4(\mu - \tfrac{1}{2})^3 + \rho(\mu - \tfrac{1}{2})\{1 + 4\mu(1-\mu)(1-\rho) - 4\mu^2 + 4\mu - 1\} \\
&= \tfrac{1}{2} + 4(\mu - \tfrac{1}{2})^3 + 4\rho(\mu - \tfrac{1}{2})\mu(1-\mu)(2-\rho).
\end{aligned}$$