

# The Vacillating Mathematician

## 2. A Stochastic Version

### ***K B Athreya***



**K B Athreya is Professor at both the Department of Mathematics and the Department of Statistics, Iowa State University, Ames, Iowa, USA. His research interests are probability theory and stochastic processes. The present article was written when he was on sabbatical leave at JNCASR, Indian Institute of Science and Indian Statistical Institute, Bangalore.**

In the first part of this article, the author described the deterministic version of the *Vacillating Mathematician*. Stochastic generalizations of this idea lead to interesting Markov chain problems.

Recall the first stochastic generalization of the *Vacillating Mathematician* given at the end of the first part of this article.

### Model I

Our mathematician starts at 0, goes half way through and then flips a fair coin. If the coin comes out *heads* she continues towards one and if the coin comes out *tails* she turns back towards 0. Again half way through whatever direction she is headed she flips a fair coin and either continues in that direction or goes in the opposite direction.

We let  $X_n$  denote the position at the  $n^{\text{th}}$  change point. Then, given  $X_n$ ,

$$X_{n+1} = \begin{cases} X_n + \frac{(1 - X_n)}{2} & \text{with probability } 1/2 \\ \frac{X_n}{2} & \text{with probability } 1/2 \end{cases} \quad (1)$$

independent of  $X_0, X_1 \dots X_{n-1}$ .

Thus the distribution of  $X_{n+1}$  given  $X_n$  depends only on  $X_n$  and does not depend on  $X_0, X_1 \dots X_{n-1}$  or  $n$ . Therefore, the sequence  $\{X_n\}_0^\infty$  is a *Markov Chain* with *stationary transition probabilities*.

We may rewrite (1) as

$$X_{n+1} = \frac{X_n}{2} + \frac{1}{2}\delta_{n+1} \quad (2)$$

where  $\{\delta_n\}_1^\infty$  is a sequence of independent and identically distributed random variables (i.i.d.r.v.) (see Karandikar in

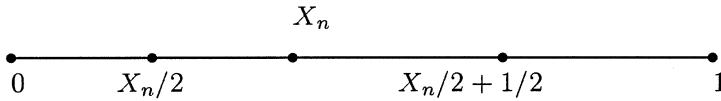
The previous article of this series was:

1. Where does she end up? January 1997.



Suggested Reading) such that

$$P(\delta_n = 1) = P(\delta_n = 0) = 1/2.$$



The sequence of positions of the Vacillating Mathematician is an autoregressive sequence of order one.

Figure 1

This is a random difference equation of the type known as *autoregressive sequence of order one*. The general form is given by

$$X_{n+1} = \rho X_n + \epsilon_{n+1}, \quad n = 0, 1, 2, \dots \quad (3)$$

where  $\rho$  is a constant and  $\{\epsilon_n\}_1^\infty$  is a sequence of i.i.d. r.v.

Iterating (3) yields

$$\begin{aligned} X_{n+1} &= \rho(\rho X_{n-1} + \epsilon_n) + \epsilon_{n+1} \\ &= \rho^2 X_{n-1} + \rho \epsilon_n + \epsilon_{n+1} \\ \text{ie.,} \\ X_{n+1} &= \rho^{n+1} X_0 + \rho^n \epsilon_1 + \rho^{n-1} \epsilon_2 + \dots + \rho \epsilon_n + \epsilon_{n+1} \end{aligned} \quad (4)$$

Since the  $\{\epsilon_n\}_1^\infty$  are i.i.d.r.v. the random variable

$$Y_n \equiv \sum_1^{n+1} \rho^{j-1} \epsilon_j$$

has the same distribution as  $X_{n+1} - \rho^{n+1} X_0 = \sum_1^{n+1} \rho^{n+1-j} \epsilon_j$  (see Karandikar in Suggested Reading). Now if  $|\rho| < 1$  (which is true in our case:  $\rho = 1/2$ )  $\rho^{n+1} X_0 \rightarrow 0$  and so the random sequence  $\{X_n\}$  has the same limit behaviour as  $\{Y_n\}$ . Suppose  $|\epsilon_j| \leq K < \infty$  for all  $j$  (in our case  $|\epsilon_j| \leq 1/2$ ) then since  $|\rho| < 1$ , the infinite series  $\sum_1^\infty \rho^{j-1} \epsilon_j$



converges absolutely to a random variable  $Y$ . Thus going back to (4) we can conclude that  $X_n$  converges in distribution to  $Y$ . That is,  $\lim_n P(a \leq X_n \leq b) = P(a \leq Y \leq b)$  for all  $a, b$  such that  $P(Y = a) = 0 = P(Y = b)$ . But in our case, ie., for equation (2),

$$Y = \sum_{j=1}^{\infty} \frac{\delta_j}{2^j}.$$

Since  $\{\delta_j\}_1^{\infty}$  are 0 or 1 valued random variables,  $Y$  represents a number in  $[0, 1]$  whose expansion to the base 2 is the sequence  $\{\delta_j\}_1^{\infty}$ . In general, given an integer  $k > 1$  and  $x$  in  $[0,1]$  define  $a_1(x) \equiv [kx]$  where for any  $y \geq 0$ ,  $[y]$  is that integer  $n$  such that  $n \leq y < n + 1$ . Define  $x_1 = x - \frac{a_1(x)}{k}$ . Then  $0 \leq x_1 < 1/k$  and let  $a_2(x) = [k^2x_1]$ . Recursively, let  $x_r = x - \sum_{j=1}^{r-1} \frac{a_j(x)}{k^j}$  for  $r \geq 2$  and  $a_r(x) = [k^r x]$ . Then it can be shown that  $0 \leq x_r < \frac{1}{k^r}$  and  $a_j(x)$  assumes values in  $\{0, 1, 2, \dots, k - 1\}$ . Thus as  $r \rightarrow \infty$  we get  $x = \sum_1^{\infty} \frac{a_j(x)}{k^j}$ . This expansion of  $x$  is known as *expansion to the base k* and  $a_j(x)$  is called the  $j$ th digit. This expansion is called *binary* for  $k = 2$ , *ternary* for  $k = 3$  and *decimal* for  $k = 10$ . To get familiar with this try to expand the numbers  $1/2, 2/3, 5/8$  etc given in decimal system with respect to other bases such as 2, 3 etc. The additional fact that  $\{\delta_j\}_1^{\infty}$  are i.i.d.r.v. with  $P(\delta_j = 1) = 1/2 = P(\delta_j = 0)$  implies that  $Y$  has *uniform distribution* on  $[0,1]$  (see Billingsley or Feller in Suggested Reading). That is,  $P(a \leq Y \leq b) = (b - a)$  for all  $0 \leq a \leq b \leq 1$ . Thus  $P(a < X_n \leq b) \rightarrow (b - a)$  for all  $0 \leq a \leq b \leq 1$ . Since  $\{X_n\}_0^{\infty}$  is a Markov Chain on  $[0,1]$  and has the uniform distribution as its limit distribution it can be shown (see part d of Theorem 1 in the Markov Chains section, part I of this article) that the empirical frequency of visits to  $[a, b]$  by the sequence  $\{X_n\}_0^{\infty}$ , ie.

$$\frac{1}{n} \sum_1^n I_{[a,b]}(X_j)$$

How often does  
the vacillating  
mathematician visit  
any  
nondegenerate  
interval in  $[0,1]$ ?

where  $I_A(x) = 1$  if  $x \in A$  and 0 if  $x \notin A$  converges to  $(b - a)$  for almost all trajectories  $\{X_n\}_0^{\infty}$ . Thus the sequence  $\{X_n\}_0^{\infty}$  visits every nondegenerate interval  $[a, b]$  in  $[0,1]$  infinitely many times. This entails that the sequence  $X_n$  will have the entire interval  $[0,1]$  as its set of *limit points*. This



is in sharp contrast to the deterministic case where there were only two limit points, namely,  $1/3$  and  $2/3$ .<sup>1</sup>

**Model - II**

Suppose we change Model - I as follows. Each time the mathematician has to decide on the direction she uses a biased coin with probability  $p(\neq 1/2)$  for heads and then goes a fraction  $\alpha$  of the distance in that direction. Thus given  $X_n, X_{n-1} \dots X_0$ ,

$$X_{n+1} = \begin{cases} X_n + \alpha(1 - X_n) & \text{with probability } p \\ X_n(1 - \alpha) & \text{with probability } (1 - p). \end{cases} \quad (5)$$

This is equivalent to writing

$$\begin{aligned} X_{n+1} &= (1 - \alpha)X_n + \alpha \delta_n \\ \text{where } \delta_n &= 1 \quad \text{with probability } p \\ &= 0 \quad \text{with probability } (1 - p) \end{aligned}$$

and  $\{\delta_n\}_1^\infty$  are i.i.d.r.v.

This is again a special case of (3) and so from the discussion following (3) we see that  $X_n$  will converge in distribution to

$$Y = \sum_1^\infty (1 - \alpha)^{j-1} \alpha \delta_j.$$

If  $\alpha = 1/2$  then  $Y$  belongs to the set  $A$  of numbers in  $[0,1]$  whose binary expansion has digits 0 and 1 in such a way that the proportion  $\frac{1}{n} \sum_1^n \delta_j$  converges to  $p$ . There is a theorem due to E Borel (see Billingsley in Suggested Reading) which says that the set  $B$  of numbers in  $[0,1]$  whose binary expansion has digits 0 and 1 such that the proportion  $\frac{1}{n} \sum_1^n \delta_j$  converges to  $1/2$  has length (ie. measure) one. In fact, Borel's theorem says something stronger. For any given  $k$  and any pattern  $a_1 a_2 \dots a_k$  of 0's and 1's and real number  $x$  in  $[0, 1]$  with base 2 expansion  $x = \sum_1^\infty \frac{\delta_j}{2^j}$  let  $\eta_j = 1$  if  $(\delta_{j+1}\delta_{j+2} \dots \delta_{j+k}) = (a_1 a_2 \dots a_k)$  and 0 otherwise. Then  $\frac{1}{n} \sum_{j=1}^n \eta_j$  converges to  $\frac{1}{2^k}$  as  $n \rightarrow \infty$  for almost all  $x$  in  $[0,1]$ . That is, the set of  $x$ 's for which this does not hold has length or measure zero. Now this has an amusing interpretation.<sup>2</sup>

<sup>1</sup> The destination of a stochastic vacillating mathematician can be very strange, indeed!

<sup>2</sup> Suppose that there is some catastrophe and all the books in the world were destroyed. Then all is not lost. Say we want to recover the Bhagavad Gita. Have a monkey type away at a Devanagiri typewriter. The entire Bhagavad Gita text will show up at some finite time (almost surely). So if we have atleast one person left who can recognize this great work the monkey will recreate it for us.



If  $S$  is the smallest closed set with  $P(Y \in S) = 1$  we say that the *support of the random variable  $Y$  is  $S$* . Thus if  $p \neq 1/2$  then our set  $A$  has *length* or *measure* zero and since  $S \subset A$ , the length of  $S$  is zero as well.

If  $\alpha = 2/3$  then  $Y$  belongs to the *Cantor set* ie. all those numbers in  $[0,1]$  whose expansions to base 3 omits digit one altogether. That is, all  $x = \sum_1^{\infty} \frac{a_j}{3^j}$  where  $a_j$  is either 0 or 2. The Cantor set  $C$  was introduced by the nineteenth century mathematician Greger Cantor (see Rudin in Suggested Reading). It is constructed as follows. Start with the unit interval  $[0,1]$ . Delete the middle third, i.e. the open interval  $(1/3, 2/3)$ . What is left is the union of the closed intervals  $[0,1/3]$  and  $[2/3,1]$ . Next delete the middle thirds of  $[0,1/3]$  and  $[2/3,1]$  and so on indefinitely. What is left of  $[0,1]$  after this process of deletion is the Cantor set. Since at the  $n^{\text{th}}$  stage we delete  $2^{n-1}$  intervals of length  $(1/3)^n$  each, the total length of all the deleted intervals is  $\sum_{n=1}^{\infty} 2^{n-1} \left(\frac{1}{3}\right)^n = \frac{1/3}{1 - 2/3} = 1$ . Thus the ‘length’ of what is left over of  $[0,1]$ , i.e., of the Cantor set, has to be zero. However, in another sense the Cantor set is as big as the interval  $[0,1]$ . Namely, it can be put into one to one correspondence with  $[0,1]$ . For  $\alpha = \frac{2}{3}$  and  $p = \frac{1}{2}$ ,  $S$  is all of the Cantor set. For  $\alpha = \frac{2}{3}$  and  $p \neq \frac{1}{2}$ ,  $S$  is a proper subset of the Cantor set  $C$ .

It can be shown that the set of limit points of  $\{X_n\}$  coincide with the support of  $Y$ . Thus we have examples in the stochastic case where the limit point set of  $\{X_n\}$  can be the whole interval  $[0,1]$  or a subset of ‘length’ zero like the Cantor set etc. The Cantor set  $C$  is an example of a fractal set (see Barnsley in Suggested Reading of Part 1).

Note also that if  $X_0 < 1$  then in none of the above models, deterministic or stochastic, does the mathematician reach her office, ie., the point 1. But in the stochastic case she gets arbitrarily close to it.

### Model - III

Let us change the source of randomness. Suppose the mathematician, instead of going half way through to her destination, goes a random fraction of the distance and



then changes her mind, turns around and goes in the opposite direction and goes a random fraction. Let us assume that each time this random fraction is chosen from  $[0,1]$  independently of the history so far, and that the random fraction while going towards one has distribution function  $F$  and the one while going towards zero has distribution  $G$ . To formalise this we introduce two sequences of random variables, i.e.,  $\{Y_n\}_1^\infty$  that are i.i.d.r.v. with distribution  $F$  and  $\{Z_n\}_1^\infty$  that are i.i.d.r.v. with distribution  $G$ .

The positions of the vacillating mathematician have now changed to form a random autoregressive series.

Let  $X_0$  be the initial position of our mathematician (it could be zero). Then she proceeds towards one going a random fraction  $Y_1$  of the distance from  $X_0$  to 1. That is,

$$X_1 = X_0 + (1 - X_0)Y_1.$$

Now she changes her mind and goes towards zero a random fraction  $Z_1$  of the distance from  $X_1$  to zero. That is,

$$X_2 = X_1(1 - Z_1).$$

More generally,

$$\begin{aligned} X_{2n+1} &= X_{2n} + (1 - X_{2n})Y_n \text{ for } n \geq 0 \\ X_{2n} &= X_{2n-1}(1 - Z_n) \text{ for } n \geq 1. \end{aligned} \tag{6}$$

Letting  $U_n = X_{2n-1}$  and  $V_n = X_{2n}$  then

$$\begin{aligned} U_{n+1} &= X_{2n}(1 - Y_n) + Y_n \\ &= U_n(1 - Z_n)(1 - Y_n) + Y_n \equiv U_n R_n + Y_n, \text{ say.} \end{aligned} \tag{7}$$

This is a random difference equation known as *random autoregressive series of order one*. The random autoregression parameter  $R_n$  and the error  $Y_n$  are not independent but the pair sequence  $(R_n, Y_n)$ ,  $n = 1, 2, \dots$  are i.i.d. random vectors. Now iterating (6)

$$\begin{aligned} U_{n+1} &= R_n(U_{n-1}R_{n-1} + Y_{n-1}) + Y_n \\ &= R_n R_{n-1}(U_{n-2}R_{n-2} + Y_{n-2}) + R_n Y_{n-1} + Y_n. \end{aligned}$$

Thus

$$\begin{aligned} U_{n+1} &= R_n R_{n-1} \dots R_1 U_1 + R_n R_{n-1} \dots R_2 Y_1 \\ &\quad + R_n R_{n-1} \dots R_3 Y_2 + \dots + R_n R_{n-1} Y_{n-2} \\ &\quad + R_n Y_{n-1} + Y_n. \end{aligned} \tag{8}$$



**Suggested Reading**

- ◆ W Feller. *An Introduction to Probability Theory and its Applications*. Wiley Eastern. Vol.I. 1968, Vol.II. 1970.
- ◆ W Rudin. *Principles of Mathematical Analysis*. McGraw Hill. N.Y., 1972.
- ◆ P Billingsley. *Probability and Measure*. Second Edition. John Wiley and Sons. N.Y., 1990.
- ◆ R Karandikar. *Resonance*. Vol.1, No.2, Feb. 1996.

Since both  $Y_n$  and  $Z_n$  are random variables with values in  $[0,1]$ ,  $R_n = (1 - Y_n)(1 - Z_n)$  satisfies

$$P(0 \leq R_n \leq 1) = 1$$

and  $P(R_n = 1) < 1$  unless  $P(Y_n = 1) = 1 = P(Z_n = 1)$  (a case we exclude). Since  $\{R_n\}_1^\infty$  is an i.i.d.r.v. sequence with values in  $[0,1]$ , by the law of large numbers,  $\frac{1}{n} \sum_{i=1}^n \log R_j \rightarrow E \log R_1 < 0$  with probability one and hence  $R_1 R_2 \dots R_n \rightarrow 0$ . Also since  $\{(R_n, Y_n)\}_1^\infty$  is an i.i.d. sequence

$$\sum_{j=2}^n Y_{j-1} R_j R_{j+1} \dots R_n + Y_n \tag{9}$$

has the same probability distribution as

$$Y_1 + \sum_{j=2}^n R_1 R_2 \dots R_{j-1} Y_j. \tag{10}$$

Let  $\eta_j = R_1 R_2 \dots R_j Y_{j+1}$ ,  $j \geq 1$ . Then  $E \eta_j = (E R_1)^j (E Y_1)$  and so  $E \left( \sum_2^\infty \eta_j \right) = \sum_2^\infty E(\eta_j) < \infty$  since  $E R_1 < 1$ . So  $\sum_1^\infty \eta_j$  converges (with probability one) and hence

$$Y_1 + \sum_{j=2}^n \eta_j \tag{11}$$

converges (with probability one) to, say,  $Y$ . This implies by (9) and (10) that

$$Y_n + \sum_{j=2}^n Y_{j-1} R_j \dots R_n \tag{12}$$

converges in distribution to  $Y$ . Since  $U_1 R_1 \dots R_n \rightarrow 0$  it follows that  $U_n$  converges in distribution to  $Y$  as in (11). Again as before the support set  $S$  of the random variable  $Y$  will be the set of limit points of the sequence  $\{X_n\}$ . An interesting open question is to determine all possible sets  $S$  that can arise this way.

Address for Correspondence  
 KB Athreya  
 Department of Statistics  
 Iowa State University  
 Ames, Iowa, IA 50011  
 USA

