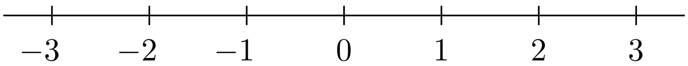
8 — STOCHASTIC PROCESSES

The word *stochastic* is derived from the Greek $\sigma\tau o\chi\alpha\sigma\tau\iota\kappa o\varsigma$, meaning 'to aim at a target'. Stochastic processes involve state which changes in a random way. A *Markov process* is a particular kind of stochastic process. Using discrete time the state of the process at time $n+1$ depends only on its state at time n. The classic example of a stochastic process is the *random walk*...

Random Walk

The simplest form of the random walk problem imagines a line marked out in unit steps or paces from some origin:



A person or other object starts at the origin and then makes a sequence of steps, some to the right and some to the left, at random.

It is reasonable to think of a sequence of turns. At each turn a weighted coin is tossed and if it lands heads one step is taken to the right and if it lands tails one step is taken to the left.

In the analysis below assume:

- Probability of a left-step (tails) is q
- Probability of a right-step (heads) is p

where $p + q = 1$

Consider a walk which consists of a total of n steps or turns. Let X be a random variable whose value, r, is the number of those n steps which are to the right.

Given a total of n steps, each of which has a probability p of being a right-step, the probability of there being r right-steps is given by the Binomial distribution:

$$\mathrm{P}(X = r) = \binom{n}{r} p^r q^{n-r} \tag{8.1}$$

Usually one is interested in the net displacement. Call this k measured in net steps to the right of the origin. Clearly:

$$\text{Net displacement to the right} = \text{Total right-steps} - \text{Total left-steps}$$

Since the total number of left-steps is $n - r$ the net displacement k can be expressed algebraically:

$$k = r - (n - r) = 2r - n \qquad \text{hence } r = \tfrac{1}{2}(n + k)$$

Rewriting (8.1):

$$\mathrm{P}\big(X = \tfrac{1}{2}(n+k)\big) = \binom{n}{\frac{1}{2}(n+k)} p^{\frac{1}{2}(n+k)} q^{\frac{1}{2}(n-k)} \tag{8.2}$$

An incomplete interpretation is the following: for fixed n, this is the probability that the number of right-steps is such as to give a net displacement of k steps to the right.

Interpretation of the Probability

Since X is the number of right-steps, its value must be an integer. Therefore the probability $\mathrm{P}\big(X = \frac{1}{2}(n+k)\big)$ requires the term $\frac{1}{2}(n+k)$ to be an integer. Accordingly $n+k$ (and hence $n-k$) are *even* in the expression for the probability:

$$\mathrm{P}\big(X = \tfrac{1}{2}(n+k)\big) = \binom{n}{\frac{1}{2}(n+k)} p^{\frac{1}{2}(n+k)} q^{\frac{1}{2}(n-k)}$$

Further, $-n \leqslant k \leqslant n$.

This accords with common sense. If the total number of steps is 2 the net displacement must be one of the three possibilities: two steps to the left, back to the start, or two steps to the right. These correspond to values of $k = -2, 0, +2$. Clearly it is impossible to get more than two units away from the origin if you take only two steps and it is equally impossible to end up exactly one unit from the origin if you take two steps.

The following table shows the probabilities associated with the different possible values of k for $n = 1, 2, 3, 4$:

n	k	$\mathrm{P}(\text{net} = k)$
1	-1	q
	1	p
2	-2	q^2
	0	$2pq$
	2	p^2
3	-3	q^3
	-1	$3pq^2$
	1	$3p^2q$
	3	p^3
4	-4	q^4
	-2	$4pq^3$
	0	$6p^2q^2$
	2	$4p^3q$
	4	p^4

For given n, $\mathrm{P}(\text{net} = k)$ is the probability that the net displacement is k units to the right of the origin. For each n any missing value of k (such as $k = 2$ when $n = 3$) is impossible and $\mathrm{P}(\text{net} = k) = 0$.

Notice that for each n the tabulated probabilities total 1. Thus for $n = 3$ the sum of the probabilities is $q^3 + 3pq^{+}3p^2q + p^q = (q+p)^3 = 1$.

Expected Displacement and Drift

Given that X is distributed Binomial(n,p), the expectation $\mathrm{E}(X) = np$. This is also the expectation $\mathrm{E}\big(\frac{1}{2}(n+k)\big)$, so:

$$np = \mathrm{E}(X) = \mathrm{E}\big(\tfrac{1}{2}(n+k)\big) = \tfrac{1}{2}\big(n + \mathrm{E}(k)\big)$$

Hence:

$$\mathrm{E}(k) = 2np - n = n(2p-1) = n\big(2p - (p+q)\big) = n(p-q)$$

If $p = q$ the expected displacement is zero but if $p \neq q$ the expected displacement is non-zero and the walk is not expected to end at the starting point. This phenomenon is known as *drift*. The expected net displacement is proportional to the number of steps so the longer the walk the greater the drift.

The term *recurrent random walk* is used to describe a random walk which is certain to return to the starting point in a finite number of steps. In the present case, the random walk is recurrent if and only if $p = q = \frac{1}{2}$.

The term *transient random walk* is used to describe a random walk which has a non-zero probability of never returning to the starting point. In the present case, the random walk is transient if $p \neq q$.

Corollary

A footnote to the random walk analysis is to consider the probability of landing on the origin at step n. Clearly n must be even and $k = 0$ so, from (8.2):

$$\mathrm{P}\big(X = \tfrac{1}{2}n\big) = \begin{cases} \dbinom{n}{\frac{1}{2}n} p^{\frac{1}{2}n} q^{\frac{1}{2}n}, & \text{if } n \text{ even} \\ 0, & \text{otherwise} \end{cases}$$

Remember that X is the number of right-steps. When this is $\frac{1}{2}n$ the number of right-steps is obviously the same as the number of left-steps; thus $\mathrm{P}(X = \frac{1}{2}n)$ is exactly equivalent to P(return to origin at step n).

The Gambler's Ruin Problem

Many stochastic processes are disguised variants of the random walk problem. One of the best-known variants is the Gambler's Ruin problem. You suppose there are two gamblers, A and B, and they each have a pile of pound coins:

A has initial capital of £n

B has initial capital of £$(a-n)$

Play then proceeds by a sequence of turns. At each turn:

Probability(A wins the turn) $= p$

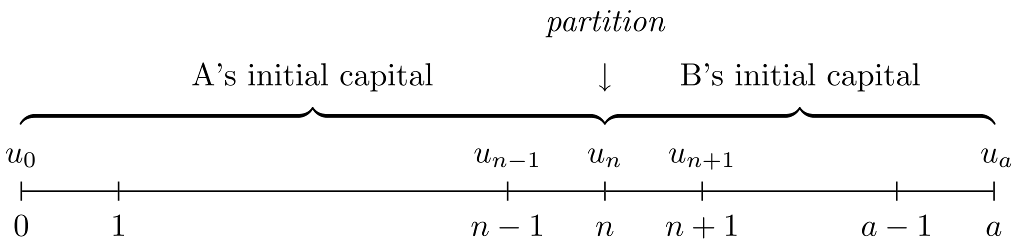
Probability(B wins the turn) $= q$

where $p + q = 1$

At the end of each turn one pound is transferred from the loser's pile of pound coins to that of the winner. Note that the total capital is £a and that stays constant.

The game ends when one player is ruined and has no money left.

A diagrammatic representation of the game at the start is:



The horizontal line is a units long and is marked off at unit intervals numbered $0, 1, 2, \ldots, a$. Point n is marked as the *partition*. The n units of line to the left of the partition represent A's initial capital and the $a-n$ units of line to the right of the partition represent B's initial capital.

At each turn the partition moves one place to the right if A wins (this outcome has probability p) and one place to the left if B wins (this outcome has probability q).

Let u_n = probability that A ultimately wins starting from n

Let v_n = probability that B ultimately wins starting from n

It is important to note that $u_n + v_n$ *may* be less than unity. In this kind of problem there is often the possibility of there being *no* winner. There could be a non-zero probability of the game going on for ever with the partition moving backwards and forwards but never reaching 0 or a.

The only respectable way of tackling this problem is to determine u_n and v_n separately and check whether or not they sum to 1.

Probability that A wins

First, extend the notation u_n so that, for example:

Let u_{n+1} = probability that A ultimately wins starting from $n+1$

Let u_{n-1} = probability that A ultimately wins starting from $n-1$

Consider the position of the partition after the first turn:

- The partition is at $n+1$ with probability p and from $n+1$ the probability of ultimately winning is u_{n+1}.
- The partition is at $n-1$ with probability q and from $n-1$ the probability of ultimately winning is u_{n-1}.

Now $p+q$ necessarily sum to 1 (unlike $u_n + v_n$ about which the sum is still in doubt) so after one turn the partition *must* be at $n+1$ or $n-1$ with probabilities u_{n+1} and u_{n-1} respectively so:

$$u_n = p\,u_{n+1} + q\,u_{n-1} \tag{8.3}$$

This is a homogeneous difference equation. In words:

The probability of A ultimately winning from $n =$

(probability of first turn landing on $n+1$) $\times$ (probability of winning from $n+1$) $+$
(probability of first turn landing on $n-1$) $\times$ (probability of winning from $n-1$)

The difference equation (8.3) holds for $n = 1, 2, \ldots, (a-1)$ but since the game ends when $n = 0$ or $n = a$ the equation does not hold for u_0 or u_a. Either leads to an invalid right-hand side.

Points 0 and a on a random walk are known as *absorbing barriers*. No walk can pass these barriers.

The absorbing barriers lead to the *boundary conditions*:

$u_0 = 0$ probability that A wins when B has all the capital

$u_a = 1$ probability that A wins when B is out of capital

From (7.4), the general solution to the difference equation is:

$$u_n = A_1(1)^n + A_2\left(\frac{q}{p}\right)^n \qquad \text{provided} \quad p \neq q \tag{8.4}$$

Given the boundary conditions $u_0 = 0$ and $u_a = 1$:

$$A_1 + A_2 = 0 \qquad \text{and} \qquad A_1 + A_2\left(\frac{q}{p}\right)^a = 1$$

Solve for A_1 and A_2 and back substitute in (8.4) to give:

$$u_n = \frac{\left(\frac{q}{p}\right)^n - 1}{\left(\frac{q}{p}\right)^a - 1} \tag{8.5}$$

Further discussion of this will be postponed until after a solution has been found for $v_n \ldots$

Probability that B wins

As with u_n, first extend the notation v_n so that, for example:

Let $v_{n+1} =$ probability that B ultimately wins starting from $n+1$

Let $v_{n-1} =$ probability that B ultimately wins starting from $n-1$

The difference equation is set up in exactly the same way as for u_n and is now:

$$v_n = p\,v_{n+1} + q\,v_{n-1}$$

The absorbing barriers now lead to different boundary conditions:

$v_0 = 1$ probability that B wins when A is out of capital

$v_a = 0$ probability that B wins when A has all the capital

The general solution has exactly the same form as (8.4):

$$v_n = A_1(1)^n + A_2\left(\frac{q}{p}\right)^n \qquad \text{provided} \quad p \neq q \tag{8.6}$$

Given the boundary conditions $v_0 = 1$ and $v_a = 0$:

$$A_1 + A_2 = 1 \qquad \text{and} \qquad A_1 + A_2\left(\frac{q}{p}\right)^a = 0$$

Solve for A_1 and A_2 and back substitute in (8.6) to give:

$$v_n = \frac{\left(\frac{q}{p}\right)^n - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^a} \tag{8.7}$$

Now that both u_n and v_n are determined, their sum can be computed:

$$u_n + v_n = \frac{\left(\frac{q}{p}\right)^n - 1 + \left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^n}{\left(\frac{q}{p}\right)^a - 1} = 1$$

Happily the sum is unity so there really is a winner.

Observation about the Solutions

If both players have a reasonable amount of capital to start with the outcome is very sensitive to the sign of the difference between p and q . . .

The assumptions are that the total capital a is fairly large and that $0 \ll n \ll a$ (so that the partition is initially not close to 0 or a).

If $p > q$ then $\frac{q}{p} < 1$ and when $\frac{q}{p}$ is raised to the powers n and a small values result. By inspection of (8.5) and (8.7) $u_n \to 1$ and $v_n \to 0$. A small rightward bias makes it very likely that the partition will end up at the right-hand end.

If $p < q$ then $\frac{p}{q} < 1$ and when $\frac{p}{q}$ is raised to the powers n and a small values result. By inspection of (8.5) and (8.7) $u_n \to 0$ and $v_n \to 1$. A small leftward bias makes it very likely that the partition will end up at the left-hand end.

The fair Case

If $p = q$ the solutions found for u_n and v_n are invalid because the constants A_1 and A_2 in the general solutions (8.4) and (8.6) are not independent.

From (7.5), the appropriate general solution for u_n now is:

$$u_n = (A_1 + A_2 n)\,(1)^n$$

Using the same boundary conditions $u_0 = 0$ and $u_a = 1$:

$$A_1 = 0 \qquad \text{and} \qquad A_2\,a = 1$$

Solving and substituting gives:

$$u_n = \frac{n}{a}$$

The appropriate general solution for v_n is likewise:

$$v_n = (A_1 + A_2 n)\,(1)^n$$

Using the boundary conditions $v_0 = 1$ and $v_a = 0$:

$$A_1 = 1 \qquad \text{and} \qquad 1 + A_2 a = 0$$

Solving and substituting gives:

$$v_n = \frac{a-n}{a}$$

Again, happily, $u_n + v_n = 1$ and there really is a winner.
Notice that both solutions show that the probability of each player winning is equal to that player's share of the capital. Gamblers say that the probability of winning is proportional to the initial share of the *stake*.

The Expected Length of a Game — I

Assume that the length of a game is finite and:

Let d_n turns be the expected duration of play when starting from n

Extend this notation so that:

d_{n+1} turns is the expected duration of play when starting from $n+1$

d_{n-1} turns is the expected duration of play when starting from $n-1$

Consider the situation after one turn. The partition is either at $n+1$ (and the probability of this outcome is p) where there are expected to be d_{n+1} further turns or at $n-1$ (and the probability of this outcome is q) where there are expected to be d_{n-1} further turns.
Thus the expected number of turns from n is the very first turn plus either d_{n+1} more or d_{n-1} more. This gives the inhomogeneous difference equation:

$$d_n = 1 + p\,d_{n+1} + q\,d_{n-1} \tag{8.8}$$

Notice the special case $p = 1$ and $q = 0$ when the expected duration from n is simply the first turn plus d_{n+1} more.

The boundary conditions now are $d_0 = 0$ and $d_a = 0$. No further turns are to be expected if the partition has reached an absorbing barrier.

From (7.6), the general solution to the new difference equation is:

$$d_n = A_1 + A_2\left(\frac{q}{p}\right)^n + \frac{n}{q-p} \qquad \text{provided} \quad p \neq q \tag{8.9}$$

Given the boundary conditions $d_0 = 0$ and $d_a = 0$:

$$A_1 + A_2 = 0 \qquad \text{and} \qquad A_1 + A_2\left(\frac{q}{p}\right)^a + \frac{a}{q-p} = 0$$

Solve for A_1 and A_2 and back substitute in (8.9) to give:

$$d_n = \frac{n}{q-p} - \frac{a}{q-p} \cdot \frac{1-\left(\frac{q}{p}\right)^n}{1-\left(\frac{q}{p}\right)^a}$$

This is the expected duration of play when $p \neq q$.

The Expected Length of a Game — II

When $p = q$ the general solution (8.9) is invalid. From (7.7), the appropriate solution is:

$$d_n = A_1 + A_2 n - n^2 \tag{8.10}$$

Given the boundary conditions $d_0 = 0$ and $d_a = 0$:

$$A_1 = 0 \qquad \text{and} \qquad A_2 a - a^2 = 0$$

Solve for A_1 and A_2 and back substitute in (8.10) to give:

$$d_n = n(a-n)$$

This is a remarkable result. Suppose the total capital a is £1000 but player A starts with just £1 (so $n = 1$) whereas player B starts with £999. Since the partition starts off just one unit from the left-hand end, there is a probability of $\frac{1}{2}$ that the game will be over at the first turn. Nevertheless the expected duration of play is $1.(1000-1)$ or 999 turns.

Although the probability of player A winning is only $\frac{1}{1000}$ a mere £1 investment provides entertainment that is expected to last 999 turns!

Glossary

The following technical terms have been introduced:

stochastic processes	recurrent random walk	absorbing barrier
random walk	transient random walk	boundary conditions
drift	partition	stake

Exercises — VIII

1. Consider a variant of the Gambler's Ruin problem. To decide the outcome of each turn, the players are using a fat coin which sometimes lands on its edge. Such an outcome is deemed a draw for the turn and no money changes hands. There are now three probabilities relating to the outcome of each turn:

   Probability(A wins the turn) $= p$

   Probability(B wins the turn) $= q$

   Probability(Turn is a draw) $= r$

   Necessarily $p + q + r = 1$ and, in this question, assume that $p \neq q$. It is possible that $p = r$ or that $q = r$ but such coincidences turn out not to matter. It is not immediately clear whether the introduction of turns that have no effect alters the probabilities that A ultimately wins or that B ultimately wins but it seems likely that the duration of play (measured in turns) will be increased.

   Complete the following tasks:

   (*a*) Modify equation (8.3) and determine the probability u_n that A ultimately wins starting from n.

   (*b*) In a similar manner determine the probability v_n that B ultimately wins starting from n.

   (*c*) Could the two results have been predicted at the outset? In what circumstances will the game never finish?

   (*d*) Modify equation (8.8) and determine d_n the duration of play starting from n. If the ratio $p : q$ is kept constant while the value of r is increased steadily, does the duration of play lengthen in a way that might have been predicted at the outset?

2. Suppose that at each turn the two players play one round of the paper-scissors-stone game. The outcome may be a win for A, a win for B or a draw and money changes hands only when there is clear win.

   Complete the following tasks:

   (*a*) Determine the probabilities p, q and r for a round of the paper-scissors-stone game and note that $p = q$.

   (*b*) Rework question 1 for the case $p = q$.

   (*c*) Use the values of p, q and r appropriate for the paper-scissors-stone game in the expressions derived for u_n, v_n and d_n. Could the results have been predicted at the outset?