Lecture 1: An Introduction

A stochastic process describes the evolution of a random variable over time. If we take snapshots at discrete time intervals, we actually consider a sequence $Z_1, \ldots, Z_n$ of dependent variables. We use real indices for stochastic processes in a continuous time model.

The course outline yields a glimpse of the topics and approaches in this area covered in this course. These notes provide a more detailed introduction. To begin, we present a partial analysis of a simple stochastic process.

A Classic Random Walk

Imagine playing a game which consists of flipping a fair coin repeatedly, where each time heads appears you win a dollar, and each time tails appears you lose a dollar. Then, if X_i is your winning on the i^{th} flip, the X_i are independent identically distributed (i.i.d.) variables each of which is 1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$. We set $S_n = \sum_{i=1}^{n} X_i$. I.e, S_n is your net winnings after n wagers. Clearly, $\mathbf{E}(S_n) = \sum_{i=1}^{n} \mathbf{E}(X_i) = 0$ and S_n is distributed like $n - 2Bin(n, \frac{1}{2})$ where $Bin(m, a)$ is the sum of m i.i.d variables each of which is 1 with probability a and 0 with probability $1 - a$.

So, the behavior of S_n for a particular n is well understood. We are interested in how S_n evolves over time. For this particular walk, many such questions can easily be answered by exploiting symmetry and the following simple property, which Grimmett and Stirzaker refer to as homogeneity: $Pr(S_{i+j} = t + a | S_j = a) = Pr(S_i = t)$.

As an example we study the questions: *What is the probability that you are ever ahead?* and *How long do you expect to wait before you are ahead*?

To this end, we let $p_{>0}$ be the probability of the event $\{\exists t | S_t > 0\}$, and let $e_{>0}$ be $\sum_{t=1}^{\infty} t Pr(\{S_t > 0\} \cap \{\forall i < t,\ S_i \leq 0\})$. We are interested in computing these values.

Since X_1 is equally likely to be 1 or -1, we see $p_{>0} = \frac{1}{2} + \frac{1}{2} Pr(\{\exists t | S_t > 1\})$. But by applying homogeneity at the first time the walk hits 1, we see that $Pr(\{\exists t | S_t > 1\}) = p_{>0}^2$. So $p_{>0} = \frac{1+p_{>0}^2}{2}$ and we obtain: $p_{>0} = 1$.

In the same vein, $e_{>0} = \frac{1}{2}(1) + \frac{1}{2}(1 + 2e_{>0})$. I.e. $e_{>0} = e_{>0} + 1$ and so $e_{>0}$ is infinite.

Thus, you will certainly be ahead at some point but you expect to wait forever before this happens.

Random Walks and Concentration Inequalities

More generally, a random walk is obtained from a sequence of independent identically distributed random steps which are vectors $X_0, \ldots, X_n$ in $\mathcal{Z}^d$. The location after i steps denoted S_i is simply $S_{i-1} + X_i$ with S_0 usually but not always zero.

We are interested in the distribution of each S_i, and also in how the walk behaves over time. E.g: How often does it revisit its starting point?, How often in one dimension is a new maximum (or new point on the convex hull in 2 or more dimensions) obtained?

We note that if X is a vector chosen from the same distribution as the X_i then $E(S_n) = nE(X)$. The Central Limit Theorem tells us that for distributions X which have a finite mean, the probability that S_n is very far away from its expected value is small. We can use this to show that for walks where $E(X) \neq 0$ we have that the probability we return to 0 is strictly less than 1 and that the expected number of returns to 0 is finite.

If we impose some further restrictions on the step distribution then we can apply *Concentration Inequalities* with more bite than the Law of Large Numbers to tell us even more about the distribution of the S_n.

Markov Chains

Sometimes the step taken as we randomly move through Z^d depends on the current location S_n. For example, we may want to perform a random walk on a finite subset of the lattice which we are not allowed to leave, so although the step distribution may be i.i.d. for points in the interior, for points near the boundary, certain directions will be forbidden. In analyzing such walks, we will define a step distribution for each point in Z_d.

Equivalently, we consider any countable state space Ω and relabel the states so that Ω is either the positive integers or $\{1, 2, 3, \ldots, n\}$ for some n. We assume that the distribution of the i^{th} step taken is completely determined by the value of S_i. Thus, we have a transition matrix P where p_{ij} gives the probability that we move to state j in the next step given that we are in state i. Now, we choose some initial state S_0 and take steps according to P. So, for all t, p_{ij} is the probability that $S_{t+1} = j$ given that $S_t = i$.

Of course, a random walk is a Markov chain, but these more general walks can be used to model a wide variety of random processes which don't look like walks at all. We give two examples here and will see many more in the course.

Example: Forcasting the Weather

We have two states rainy (0) and sunny (1). We ASSUME that the probability of rain tomorrow ONLY depends on whether or not it rained today NOT on the weather of previous days.

The transition matrix for this Markov Chain might be: $\begin{pmatrix} .3 & .7 \\ .4 & .6 \end{pmatrix}$

Find the probability $P(X_2 = 0 | X_0 = 1)$.

Example: Transforming a Process into a Markov Chain

Next we ASSUME that the probability of rain tomorrow depends on whether or not it rained today AND whether or not it rained yesterday.

For instance assume we have the following conditional probabilities.

$$\begin{aligned} P(X_{n+1} = rain|X_n = rain, X_{n-1} = rain) &= 0.7 \\ P(X_{n+1} = rain|X_n = rain, X_{n-1} = sunny) &= 0.5 \\ P(X_{n+1} = rain|X_n = sunny, X_{n-1} = rain) &= 0.4 \\ P(X_{n+1} = rain|X_n = sunny, X_{n-1} = sunny) &= 0.2 \end{aligned}$$

Why doesn't the stochastic process X_n, $n = 0, 1, \ldots$ describe a Markov Chain?

Let (,) denote (state of weather yesterday, state of weather today) Then define the following states,
0 = (rain, rain)
1 = (sunny, rain)
2 = (rain, sunny)
3 = (sunny, sunny)

Convince yourself that we now have a Markov Chain with transition matrix,

$$\begin{pmatrix} .7 & 0 & .3 & 0 \\ .5 & 0 & .5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & .2 & 0 & .8 \end{pmatrix}$$

Rather than starting a Markov chain at a fixed point, we may start it at a random point chosen from some initial distribution S_0, and then the distribution of the point after 1 step is S_0P and after t steps is S_0P^t.(Note, we are abusing notation here bu using S_0 for a state above and a distribution on states here).

Thus, for a given t and S_0, we have some idea of how S_t behaves. Once again, we will be interested in the evolution of this variable over time. As with random walks, we will be particularly interested in the probability that we return to our starting point, and how long this takes. A distribution π on Ω is called *stationary* if we return to it in one step, i.e. if $\pi P = \pi$.

We will also be interested in determining if S_t approaches a limit as t goes to infinity. If this is to occur then the limit distribution must be stationary.

The key to analyzing Markov chains is to exploit the fact that they satisfy temporal homogeneity and the Markov Property: $\forall i, j, k, x, y \approx Pr(S_{i+j} = y|S_i = x) = Pr(S_{k+j} = y|S_k = x)$ even though they do not satisfy the stronger homogeneity condition mentioned earlier.

Renewal Theory

To introduce renewal theory, we consider a random walk which only moves in one direction. X_0 is a non-negative integer which indicates the starting point. $X_1, X_2, \ldots$ are i.i.d positive integers.

Renewal theory studies random walks of this type, asking two questions:

1) What is the probability that we visit a state i, and

2) How many steps do we take before we first exceed i.

It also studies similar questions when the X_i and S_i are permitted to be real.

We briefly mention a classical example:

Example: Light Bulb Example

A fresh light bulb is observed until burnout at which time it is instantaneously replaced by another fresh light bulb. This is a pure renewal process where X_i represents the lifetime of the i^{th} light bulb.

Memoryless Walks in R^d

All of the processes we have considered so far are memoryless walks in R^d. I.e. the distribution of our choice of step is determined only by our current location.

We will study one other such process, Markov Chains with delays, but we do not have time to discuss these here.

In studying all of these processes, we concentrate on a common theme: how does the process evolve over time?

To answer this question, we will need to use the following tools:

Temporal Homogeneity, Conditioning, Symmetry, Concentration Inequalities, Generating Functions, Differential Equations, and something called Coupling.

Galton-Watson Branching Processes

Start with a single parent. Create a family tree by iteratively choosing for each descendant, an i.i.d (non-negative integer) number of children with common distribution Z. Z_i is the number of descendents in the i^{th} generation. We are interested in the behavior of Z_i and, in particular, the extinction probability, i.e. the probability that some Z_i is 0 and hence that the family tree is finite.

As shown in class, the probability that the number of offspring in a family tree built in this way is exactly t is equal to the probability that a random walks in Z^1 with step distribution $Z - 1$ first goes negative at time t.

Thus, the same techniques can be used to study random walks and branching processes. Along with symmetry, homogeneity, and concentration inequalities, these include more sophisticated tools such as generating functions and differential equations.