Mathematical Foundations of Markov Chains

Thursday, September 17, 2015
2:04 PM

Reading:

- Lawler Ch. 1

Homework 1 due Friday, October 2 at 5 PM.
Office hours today are moved to 6-7 PM.

Let's revisit the stream of iid random variables denoting the occurrence/nonoccurrence of incidents during time intervals $[(n-1)\Delta t, n\Delta t)$. This discrete time stochastic process is often called a a Bernoulli process.

Let's contemplate a continuous-time limit of the Bernoulli process:

- $\Delta t \downarrow 0$
- $p = r\,\Delta t + o(\Delta t)$ where $\frac{o(\Delta t)}{\Delta t} \downarrow 0$ as $\Delta t \downarrow 0$
  - That is, the incidents happen with average rate r, or equivalently, with mean time $\frac{1}{r}$ between incidents
- simulated by "Poisson Experiment" on "Random" site

Let's revisit our key questions, now in this continuum limit:

1. How many incidents occur up to a time t?
   - Poisson distributed with mean $\mu = rt$
     - applying Poisson limit theorem to the binomial distribution with small success probability $p = r\Delta t$ with large number of trials $n = \frac{t}{\Delta t}$.
2. Starting from time t, how long do I need to wait until the next incident?
   - exponentially distributed random variable with mean $\frac{1}{r}$
     - just continuous-time limit of geometric distribution, with $p = r\Delta t, n = t/\Delta t, \Delta t \downarrow 0$

We'll discuss this more later in the class but the continuous-time limit of the Bernoulli process is called the Poisson point process

- If one uses nonhomogenous success probability p_n, then one can take limits that would give rise to nonhomogenous Poisson point process which has a rate $r = r(t)$ which varies in time.

Stochastic processes that are simply a stream of iid random variables (in discrete time) are useful for setting up models, but most systems of interest accumulate memory in some form, so these stochastic processes are usually not useful for modeling the output of a stochastic model. Iid models are often good for input.

Markov processes are one of the most widely used frameworks for stochastic models which allow for memory in your system variables, but in a sufficiently limited way that much theory and calculation can be done efficiently. The idea is that independence of random variables makes calculations easier, and Markov processes still refer to independent random variables, even though the Markov process itself carries memory. Intuitively, the simplicity of a Markov process is that once you know its current state, then the randomness in the system acts as an independent variable to update the state of the system to the next time step.

- Memory is encoded only by the current state of a selected set of variables; the other variables act like independent noise (not necessarily identically distributed!)

We will develop this idea concretely first for the simplest context of a discrete parameter domain and finite state space.

Finite State, Discrete-Time (FSDT) Markov chain

For concreteness take parameter domain:
$T = \{0,1,2,3, \dots \}$ (though once in a while, we'll want
$T = \{\dots, -2, -1, 0, 1, 2, \dots \}$

State space:
$S = \{1, \ldots, M\}$

There are two equivalent ways to define a FSDT Markov chain.

Stochastic update rule (Resnick Sec. 2.1)

You define updating function f_n:

$X_n \in S, \quad Z_n \in S_Z$

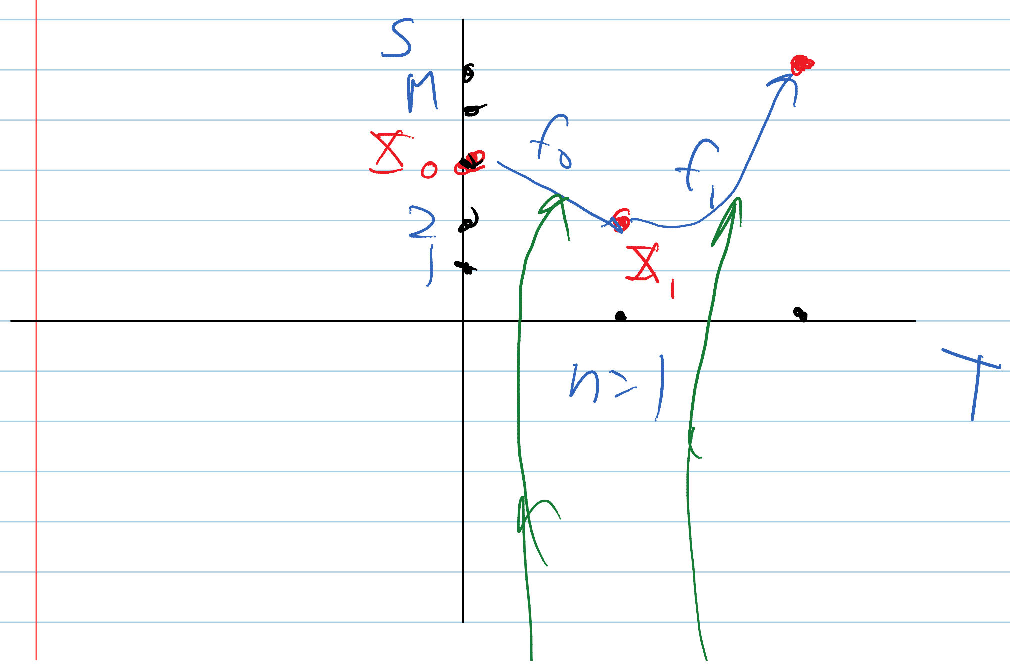
$f_n : S \times S_Z \to S$

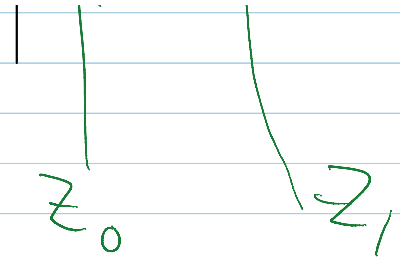
where S_Z is the state space for some ``noise'' in the system.

In particular, we define a sequence of noise $\{Z_n\}_{n=0}^{\infty}$ which are iid random variables with some probability distribution on noise state space S_Z.

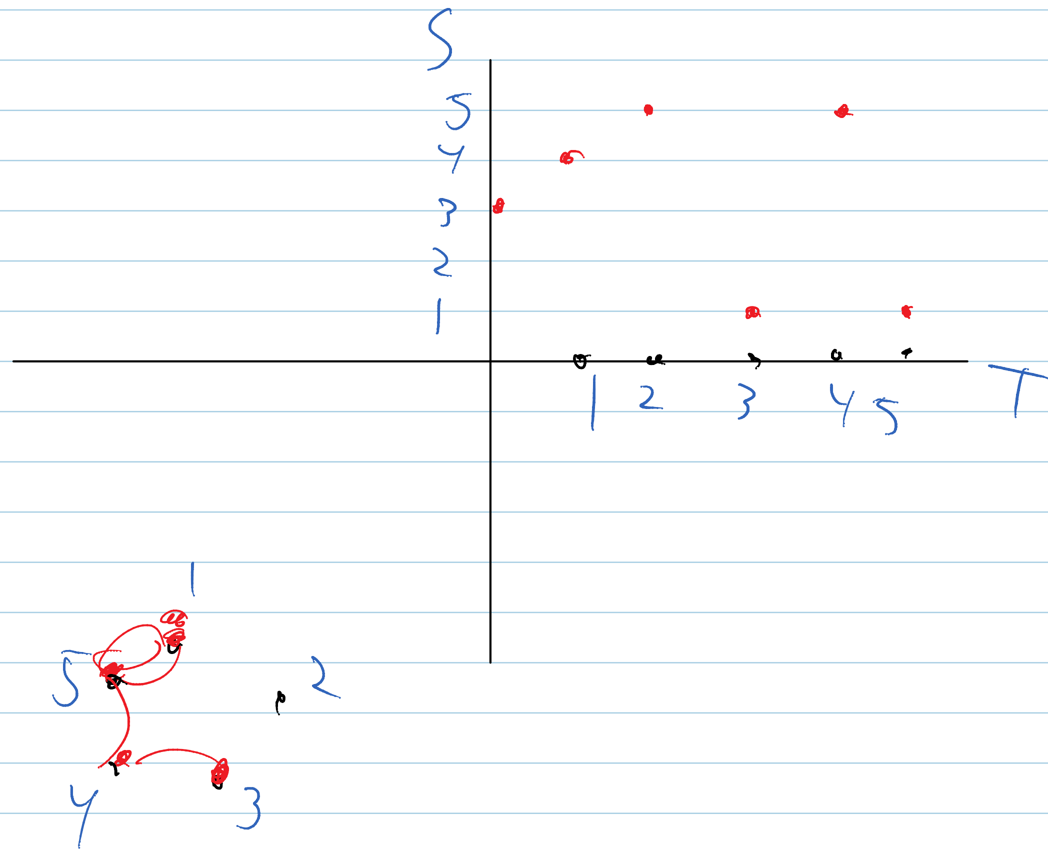
Then the dynamics of the FSDT Markov chain is given by:

$X_{n+1} = f_n(X_n, Z_n)$ for $n = 0,1,2, \ldots$





Notice that in this formulation, the memory is carried only by the value of the variable X_n.
For example, consider $X_{n+1} = X_n + Z_n$ (mod M) which is a periodic random walk.



Here we see that the step size Z_n are independent (say $S_Z = \{\pm 1\}$, with each value equally likely), while the current location X_n does carry some memory of the past noise. But notice that in determining the future evolution of X_n, I do not need to refer to past values of X_{n-1}, X_{n-2} etc.

An alternative way to formulate a FSDT MC is through probability transition matrices.

For each $n \in T$, and each $i, j \in S$, we define a matrix:

$$P_{ij}^{(n)} \equiv P(X_{n+1} = j | X_n = i)$$

This is what we call a probability transition matrix. For the periodic random walk example given above, this matrix might look like:

destination

$$P^{(n)} = \begin{array}{cc} & \begin{array}{ccccccc} 1 & 2 & 3 & \cdots & & & M \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ \vdots \\ M \end{array} & \begin{pmatrix} 0 & 1/2 & 0 & \cdots & & 0 & 1/2 \\ 1/2 & 0 & 1/2 & \cdots & & 0 & 0 \\ & 1/2 & 0 & 1/2 & & & \\ & & 1/2 & 0 & 1/2 & & \\ 1/2 & 0 & & & \ddots & \ddots & \end{pmatrix} \end{array}$$

current

Row i gives the probability distribution for X_{n+1}, when $X_n = i$.

In general, the probability transition matrix can depend on the epoch n, just as the stochastic update rule $f^{(n)}$ can also depend on epoch, but much more powerful theory is available when the Markov chain is time-homogenous which means that the dynamical rules are independent of epoch. So then for time-homogenous Markov chain,

- $f^{(n)} \equiv f$
- $P^{(n)} \equiv P$

Most of the class will be concerned with time-homogenous Markov chains (analogous to autonomous dynamical systems and differential equations)

Actually besides the probability transition matrix, we must assert the Markov property, which in discrete time, reads:

$$P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i_n)$$
$$= P(X_{n+1} = j | X_n = i_n)$$

In other words, to determine the probability rule to update my current state, the past states don't matter separately. All the past state influence on the future is assumed to be contained in the current state $X_n = i_n$ of the Markov chain. In other words, the variable X carries all the relevant memory in the system needed for the future probabilistic updates.

Markov property has very powerful mathematical consequences while being reasonable enough to be practical.

- Skorokhod embedding says that essentially any stochastic process can be formulated as a Markov process (i.e. a stochastic process with the Markov property) provided we use a rich enough state space (i.e., use enough variables to keep track of the memory).
  - In practice what this means, is that often a stochastic process X_n that does not satisfy the Markov property can be reformulated in terms of a stochastic process on a richer state space that keeps track of enough information about the memory in X_n so that this augmented stochastic process satisfies the Markov property.
    - If $X_{n+1} = f_n(X_n, X_{n-1}, X_{n-2}, Z_n)$, this is not Markov but the stochastic process $Y_n = (X_n, X_{n-1}, X_{n-2})$ would satisfy the Markov property. $Y_{n+1} = (f_n(X_n, X_{n-1}, X_{n-2}, Z_n), X_n, X_{n-1})$, and note that this is a function of Y_n and the independent noise variable Z_n
    - Or if a random walk had a notion of momentum, i.e., the next step could be correlated with the direction of the previous step, then one could augment the random walk description by defining a stochastic process $Y_n = (X_n, Z_n)$ or more generally some notion $Y_n = (X_n, V_n)$ where the stochastic process V_n encodes the momentum of the random walk.

So in practice, one can define a Markov process if one can choose a suitably rich state space such that the value of the state variable (which can be a vector of information) contains all the relevant memory.

The Markov property looks like it has a directionality to it ("future, given the present and the past, only depends on the present") but in fact it doesn't because one can show that if one wants to run the Markov chain backwards in time, one can prove an equivalent statement ("past, given the present and the future, depends only on the present")

More broadly, here is a version of the Markov property that is equivalent to what is conventionally stated, but relates future and past in more symmetric way:

"Given the present, the future and the past are conditionally independent."