Lecture 1: Finite Markov Chains. Branching process.

October 9, 2007

Antonina Mitrofanova

A *Stochastic process* is a counterpart of the deterministic process. Even if the initial condition is known, there are many possibilities how the process might go, described by probability distributions. More formally, a Stochastic process is a collection of random variables $\{X(t), t \in T\}$ defined on a common probability space indexed by the index set T which describes the evolution of some system.

One of the basic types of Stochastic process is a *Markov process*. The Markov process has the property that conditional on the history up to the present, the probabilistic structure of the future does not depend on the whole history but only on the present. The future is, thus, conditionally independent of the past.

Markov chains are Markov processes with *discrete* index set and countable or finite state space.

Let $\{X_n, n \geq 0\}$ be a Markov chain , with a discrete index set described by n. Let this Markov process have a finite state space $S = \{0, 1, \dots, m\}$. At the beginning of the process, the initial state should be chosen. For this we need an initial distribution $\{p_k\}$, where

$$P[X_0 = k] = p_k$$

$$p_k \geq 0$$

and $\sum_{k=0}^{m} p_k = 1$. After the initial state is chosen, the choice of the next state is defined by transition probabilities p_{ij}, which is the probability of moving to a state j given that we are currently in state i. Observe that the choice of the next state only depends on a current state and not on any states prior to that (in other words, it depends purely on the recent history).

The transition from state to state is usually governed by the transition matrix P.

$$P = \begin{pmatrix} p_{00} & p_{01} & \cdots & \\ p_{10} & p_{11} & \cdots & \\ \vdots & & \ddots & \end{pmatrix}$$

where $p_{ij} \geq 0$ is a transition probability from state i to state j. Precisely, it is a probability going to state j given that you are currently in state i. Observe that each row corresponds to transition probabilities from the state i to all other states and should satisfy the following

$$\sum_{j=0}^{m} p_{ij} = 1 \tag{1.1}$$

Therefore, if the state space is $S = \{0, 1, \dots, m\}$, then the transition matrix P is $(m + 1) \times (m + 1)$ dimensional.

Let us describe some of the formal properties of the above construction. In particular, for any $n \geq 0$

$$P[X_{n+1} = j | X_n = i] = p_{ij} \tag{1.2}$$

As a generalization of the above fact, we show that we may condition on the history with no change in the conditional probability, provided the history ends in state i. More specifically,

$$P[X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i] = p_{ij} \tag{1.3}$$

Now we can define a Markov chain formally:

Definition 1 *Any process $\{X_n, n \geq 0\}$ satisfying the (Markov) properties of equations 1.2 and 1.3 is called a Markov chain with initial distribution $\{p_k\}$ and transition probability matrix P.*

The Markov property can be recognized by the following finite dimensional distributions:

$$P[X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k] = p_{i_0} p_{i_0 i_1} p_{i_1 i_2 \ldots} p_{i_{k-1} i_k}$$

1 Some examples of Markov chains

1.1 The Branching process

A Branching process is a Markov process that models a population in which each individual in generation n produces a random number of offsprings for generation $n + 1$.

The basic ingredient is a density $\{p_k\}$ on the non-negative integers. Suppose an organism at the end of his lifetime produces a random number ξ of offsprings so that the probability to produce k offsprings is

$$P(\xi = k) = p_k$$

for $k = 0, 1, 2, \ldots$. We assume that $p_k \geq 0$ and $\sum_{k=0}^{\infty} p_k = 1$. All offsprings are to act independently of each other.

In a Simple branching process, a population starts with a progenitor (who forms population number 0). Then he is split into k offsprings with probability p_k ;these k offsprings constitute the first generation. Each of these offsprings then independently split into a random number of offsprings, determined by the density $\{p_k\}$ and so on. The question of ultimate extinction (where no individual exists after some finite number of generations) is central in the theory of branching processes.

The formal definition of the model follows: let us define a branching process as a Markov chain $\{Z_n\} = Z_0, Z_1, Z_2, \ldots$, where Z_n is a random variable describing the population size at the n'th generation. The Markov property can be reasoned as: in the n'th generation the Z_n individuals independently give rise to some number of offsprings $\xi_{n+1,1}, \xi_{n+1,2}, \ldots, \xi_{n+1,Z_n}$ for the $n + 1$st generation. $\xi_{n,j}$ can be though as the number of members of n'th generation which are offsprings of the j'th member of the $(n - 1)$ generation. Observe that $\{\xi_{n,j}, n \geq 1, j \geq 1\}$ are identically distributed (having a common distribution $\{p_k\}$) non-negative integer-valued random variables.

Thus, the cumulative number produced for the $n + 1$st generation is

$$Z_{n+1} = \xi_{n+1,1} + \xi_{n+1,2} + \cdots + \xi_{n+1,Z_n}$$

Thus the probability of any future behavior of the process, when its current state is known exactly, is not altered by any additional knowledge concerning its past behavior.

Generally speaking, we define a branching process $\{Z_n, n \geq 0\}$ by

$$Z_0 = 1$$

$$Z_1 = \xi_{1,1}$$

$$Z_2 = \xi_{2,1} + \cdots + \xi_{2,Z_1}$$

$$\vdots$$

$$Z_n = \xi_{n,1} + \cdots + \xi_{n,Z_{n-1}}$$

Observe that once the process hits zero, it stays at zero. In other words, if $Z_k = 0$, then $Z_{k+1} = 0$.

It is of interest to show what is the expected size of generation n, given that we started with one individual in generation zero. Let μ be the expected number of children for each individual, $E[\xi] = \mu$ and σ^2 be the variance of the offspring distribution, $Var[\xi] = \sigma^2$. let us denote the expected size of the generation Z_n by $E[Z_n]$ and its variance by $Var[Z_n]$.

1.1.1 On the random sums

Let us recall some facts about a random sum of the form $X = \xi_1 + \xi_2 + \cdots + \xi_N$, where N is a random variable with probability mass function $p_N(n) = P\{N = n\}$ and all ξ_k are independent and identically distributed random variables. Let N be independent of $\xi_1, \xi_2, \ldots$ We assume that ξ_k and N have finite moments, which determine mean and variance for the random sum:

$$E[\xi_k] = \mu$$

$$Var[\xi_k] = \sigma^2$$

$$E[N] = \nu$$

$$Var[N] = \tau^2$$

Then, $E[X] = \mu\nu$ and $Var[X] = \nu\sigma^2 + \mu^2\tau^2$.

We will show the derivation for the expectation of X, $E[X]$, which uses the manipulations with conditional expectations:

$$E[X] = \sum_{n=0}^{\infty} E[X|N = n]p_N(n) \quad (1)$$

$$= \sum_{n=1}^{\infty} E[\xi_1 + \xi_2 + \cdots + \xi_N|N = n]p_N(n) \quad (2)$$

$$= \sum_{n=1}^{\infty} E[\xi_1 + \xi_2 + \cdots + \xi_n|N = n]p_N(n) \quad (3)$$

$$= \sum_{n=1}^{\infty} E[\xi_1 + \xi_2 + \cdots + \xi_n]p_N(n) \quad (4)$$

$$= \sum_{n=1}^{\infty} (E[\xi_1] + E[\xi_2] + \cdots + E[\xi_n])p_N(n) \quad (5)$$

$$= \sum_{n=1}^{\infty} (\mu + \mu + \cdots + \mu)p_N(n) \quad (6)$$

$$= \sum_{n=1}^{\infty} \mu n p_N(n) \quad (7)$$

$$= \mu \sum_{n=1}^{\infty} n p_N(n) = \mu\nu \quad (8)$$

$$\quad (9)$$

The direct application of the above fact to the branching process gives:

$$E[Z_{n+1}] = \mu\nu = \mu E[Z_n]$$

and

$$Var[Z_{n+1}] = \nu\sigma^2 + \mu^2\tau^2 = E[Z_n]\sigma^2 + \mu^2 Var[Z_n]$$

Now, if we apply recursion to the expectation (with the initial condition $Z_0 = 1$ at $E[Z_0] = 1$), we obtain:

$$E[Z_n] = \mu^n$$

Therefore, the expected number of members in generation n is equal to μ^n, where μ is the expected number of offspring for each individual.

Observe that the extinction time directly depends on the value of μ. When $\mu > 1$, then the mean population size increases geometrically; however, if $\mu < 1$, then it is decreases geometrically, and it remains constant if $\mu = 1$.

The population extinction occurs if the population size is reduced to zero. The random time of extinction N is the first time n for which $Z_n = 0$ and $Z_k = 0$ for all $k \geq N$. In Markov chain, this state is called the *absorbtion state*.

We define

$$u_n = P\{N \leq n\} = P\{Z_n = 0\}$$

as the probability of extinction at or prior to the n'th generation.

Let us observe the process from the beginning: we begin with the single individual in generation 0 , $Z_0 = 1$. This individual then gives rise to k offsprings. Each of these offsprings will then produce its own descendants. Observe that if the original population is to die out in generation n, then each of these k lines of descent must die out in at most $n - 1$ generations.

Since all k subpopulations are independent and have the same statistical properties as the original population, then the probability that any of them dies out in $n - 1$ generation is u_{n-1}. And the probability that all of them will die out in the $n - 1$'th population (with the independence assumption) is

$$(u_{n-1})^k$$

Then, after weighting this by the probability of having of k offsprings and summing according to the law of total probability, we obtain:

$$u_n = \sum_{k=0}^{\infty} p_k (u_{n-1})^k$$

for $n = 1, 2, \dots$. It is not possible to die out in generation 0 since we must have the first individual (parent). For generation one, however, the probability of being extinct is $u_1 = p_0$, which is the probability of having no (zero) offsprings.

Usually, the Branching process is used to model reproduction (say, of the bacteria) and other system with similar to reproduction dynamics (the likelihood of survival of family names (since it is inherited by the son's only; in this case p_k is the probability of having k male offsprings), electron multipliers, neutron chain reaction). Basically, it can be thought as a model of population growth in the absence of environmental pressures.

1.1.2 Branching Processes and Generating Functions

Generating functions are extremely helpful in solving sums of independent random variables and thus provide a major tool in the analysis of branching processes.

Let us again consider an integer-valued random variable ξ whose probability distribution is given by

$$P\{\xi = k\} = p_k$$

for $k = 0, 1, \dots$.

The generating function is basically a power series whose coefficients give (define) the sequence $\{p_0, p_1, p_2, \dots\}$, such as

$$\phi(s) = p_0 + sp_1 + s^2 p_2 + \cdots = \sum_{k=0}^{\infty} p_k s^k$$

To specify that a specific generation function corresponds to a sequence (in our case, probability distribution), we write

$$\langle p_0, p_1, p_2, \dots \rangle \longleftrightarrow p_0 + sp_1 + s^2 p_2 + \dots$$

We call $\phi(s)$ a generating function of the distribution $\{p_k\}$; sometimes it is also called a generating function of ξ. We denote

$$\phi(s) = \sum_{k=0}^{\infty} p_k s^k = E[s^\xi]$$

for $0 \leq s \leq 1$.

Example: Say for a sequence $\langle 1, 1, 1, 1 \ldots \rangle$, the generating function is $1 + s + s^2 + s^3 + \cdots = \frac{1}{1-s}$.

Differentiation of the above function gives

$$\frac{d}{dx}(1 + s + s^2 + s^3 + \ldots) = \frac{d}{dx}\left(\frac{1}{1-s}\right)$$

$$1 + 2s + 3s^2 + \cdots = (\frac{1}{1-x})^2$$

$$\langle 1, 2, 3, 4 \ldots \rangle \longleftrightarrow (\frac{1}{1-x})^2$$

In fact, now it is a generating function of the sequence $\langle 1, 2, 3, 4 \ldots \rangle$

Let us observe the following relationship between the probability mass function $\{p_k\}$ and generating function $\phi(s)$:

$$p_k = \left.\frac{1}{k!}\frac{d^k\phi(s)}{ds^k}\right|_{s=0}$$

For example, let us derive p_0:

$$\phi(s) = p_0 + sp_1 + s^2 p_2 + \ldots$$

If we let $s = 0$, then it is clear that $p_0 = \phi(0)$. Now, let us derive p_1. For this purpose we take a derivative of $\phi(s)$:

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \ldots$$

And again letting s be 0, we get

$$p_1 = \left.\frac{d\phi(s)}{ds}\right|_{s=0}$$

Let us say that if $\xi_1, \xi_1, \ldots, \xi_n$ are independent random variables, which have generating functions $\phi_1(s), \phi_2(s), \ldots, \phi_n(s)$, then the generating function of their sum $Z = \xi_1 + \xi_1 + \cdots + \xi_n$ is the product

$$\phi_Z(s) = \phi_1(s)\phi_2(s)\ldots\phi_n(s)$$

Additionally, the moment of nonnegative random variable may be found by differentiating the generating function and evaluating it at $s = 1$, which actually equivalent to the expectation of the random variable. Taking the derivative of $p_0 + p_1 s + p_2 s^2 + \ldots$ gives:

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \ldots$$

And evaluating it at 1 produces:

$$\left.\frac{d\phi(s)}{ds}\right|_{s=1} = p_1 + 2p_2 + 3p_3 + \cdots = E[\xi]$$

Consecutively, it is possible to find the variance $Var[\xi]$of the random variable ξ: The second derivative of the generating function $\phi(s)$ is

$$\frac{d^2\phi(s)}{ds^2} = 2p_2 + 3(2)p_3 s + 4(3)p_4 s^2 + \ldots$$

and evaluating it at 1 gives

$$\left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} = 2p_2 + 3(2)p_3 + 4(3)p_4 + \dots \tag{10}$$

$$= \sum_{k=2}^{\infty} k(k-1)p_k = E[\xi(\xi-1)] \tag{11}$$

$$= E[\xi^2 - \xi] = E[\xi^2] - E[\xi] \tag{12}$$

$$\tag{13}$$

Rearranging the above equality gives:

$$E[\xi^2] = \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + E[\xi] \tag{14}$$

$$= \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + \left.\frac{d\phi(s)}{ds}\right|_{s=1} \tag{15}$$

$$\tag{16}$$

and the variance $Var[\xi]$ is

$$Var[\xi] = E[\xi^2] - \{E[\xi]\}^2 \tag{17}$$

$$= \left[\left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + \left.\frac{d\phi(s)}{ds}\right|_{s=1}\right] - \left[\left.\frac{d\phi(s)}{ds}\right|_{s=1}\right]^2. \tag{18}$$

$$\tag{19}$$

Generating Functions and Extinction Probabilities: Let us define $u_n = Pr\{X_n = 0\}$ as a probability of extinction by state n. Using the recurrence established before, we can define it in terms of generating functions:

$$u_n = \sum_{k=0}^{\infty} p_k (u_{n-1})^k = \phi(u_{n-1})$$

Thus, knowing the generating function $\phi(s)$, we can compute the extinction probabilities (starting at $u_0 = 0$; then $u_1 = \phi(u_0)$, $u_2 = \phi(u_1)$ etc.).

It is interesting that the extinction probabilities converge upwards to the smallest solution to the equation $\phi(s) = s$. We denote n_∞ as the smallest solution to $\phi(s) = s$. In fact, it gives the probability that the population becomes distinct in some finite time.

Speaking differently, the probability of extinction depends on weather or not the generating function crosses the 45 degree angle line ($\phi(s) = s$), which can be determined from the slope of the generating function at $s = 1$:

$$\phi^{'}(1) = \left.\frac{d\phi(s)}{ds}\right|_{s=1}$$

If the slope is less then or equal to one, then no crossing takes place and the probability of eventual extinction is $u_\infty = 1$. However, if the slope exceeds one then the equation $\phi(s) = s$ has a smaller solution (less than one) and thus extinction is not a certain event. In this case, this smaller solution corresponds to a probability of extinction.

Observe that the slope of the generating function $\phi(s)$ is $E[\xi]$. Therefore, the above rules (of the probability of extinction) can be applied with respect to $E[\xi]$ in a similar way.

1.1.3 Generating Functions and Sums of Independent Random Variables

Let $\xi_1, \xi_2, \dots$ be independent and identically-distributed nonnegative integer-values random variables with generating function $\phi(s) = E[s^\xi]$. Then the sum of these variables $\xi_1 + \xi_2, \dots, \xi_m$ has a generating function:

$$E[s^{\xi_1+\xi_2+\dots+\xi_m}] = E[s^{\xi_1} s^{\xi_2} \dots s^{\xi_m}] = E[s^{\xi_1}]E[s^{\xi_2}] \dots E[s^{\xi_m}] = [\phi(s)]^m$$

Now, let N be non-negative integer-valued random variable , independent of $\xi_1, \xi_2, \ldots$, with generating function $g_N(s) = E[s^N]$. We consider a random sum

$$X = \xi_1 + \xi_2 + \cdots + \xi_N$$

and let $h_X(s) = E[s^X]$ be a generating function of X. Then,

$$h_X(s) = g_N[\phi(s)]$$

And applying it to a general branching process gives

$$\phi_{n+1}(s) = \phi_n(\phi(s)) = \phi(\phi_n(s))$$

That is we obtain a generating function for a population size Z_n at generation n by repeated substitution in the generating function of the offspring distribution.

For general initial population size $Z_0 = k$, the generating function is

$$\sum_{j=0}^{\infty} P[X_n = j | Z_0 = k] s^j = [\phi_n(s)]^k$$

It basically corresponds to the sum of k independent lines of descents. In other words, the branching process evolves as the sum of k independent branching processes (one for each initial parent).

References:

Howard M. Taylor, Samuel Karlin, An Introduction to Stochastic modelling, pp 70-73 (on random sums), III (ch 8) 177-198 (on branching process)

Sidney I. Resnick, Adventures in Stochastic Processes, I (pp 1-59) - preliminaries, II (pp 60-67)- the Markov chains and Branching process.