Lecture-00: Introduction

1 Deterministic and stochastic models

Evolution of a **deterministic** system is characterized by a set of equations, with each run leading to the same outcome given the same initial conditions. Evolution of a **stochastic** system is at least partially random, and each run of the process leads to potentially a different outcome. Each of these different runs are called a **realization** or a **sample path** of the stochastic process.

We are interested in modeling, analysis, and design of stochastic systems. Following are some of the stochastic systems from different disciplines of science and engineering.

- Evolution of number of molecules due to chemical reaction, where the time to form new molecules is uncertain and it depends on density of other molecules.
- Financial commodities like stock prices, currency exchange rates fluctuate with time. These can be modeled by random walks. One can provide probabilistic predictions and optimal buying and selling strategies using these models.
- Machines that detect photons, have a dead time post a successful detection. This adds uncertainty in estimating photon density. These processes can be modeled by an *on-off* process.
- A contagious disease can spread very quickly across a region. This is similar to a content getting viral on internet. One can model spread of epidemics on network by Urn models.
- Counting number of earthquakes that occur everyday at a certain location. These can be modeled by a counting process, and inter-arrival time of the quakes can be estimated to make probabilistic predictions.
- A mother cell takes a random amount of time to subdivide and create a daughter cell. A daughter cell takes certain random time to mature, and become a mother cell. A mother cell dies after certain number of sub-divisions. One is interested in finding out the asymptotic behavior of population density.
- Popularity of a page depends on how quickly one can reach it from other pages on the Internet. Equilibrium distribution of certain random walks on graphs can be used to estimate page ranks on the web.

2 Stochastic modeling

2.1 Page rank

The set of webpages in the internet can be denoted by a finite set V. If a page $j \in V$ can be reached by another page $i \in V$ through hyperlinks, we denote it by $i \sim j$. The set of directed links is denoted by $\overrightarrow{E} = \{(i,j) \in V \times V : i \sim j\}$. We can assume $(i,i) \in \overrightarrow{E}$ for each $i \in V$, and the probability of going to page j from page i is $P_{ij} = P(e)$ where $(i,j) = e \in \overrightarrow{E}$ and $\sum_{j:(i,j)\in\overrightarrow{E}} P_{ij} = 1$. Then, we can model a random walk on web-pages on the internet by a Markov chain $(X_n \in V : n \in \mathbb{N})$ with transition matrix $P \in [0,1]^{V\times V}$ on the directed graph $G = (V, \overrightarrow{E})$. Let π be the stationary distribution of this Markov chain, then the page-ranks are the indices corresponding to sorted values of the stationary distribution in decreasing order.

Therefore, to return a search query, one should be able to find the stationary distribution of the pages and sort them. For large N, as in the case of internet, one doesn't know the transition matrix P a priori. And even if the whole transition matrix was known, finding the stationary distribution would take $O(N^3)$ computations, a very large number. This follows from the following approximation $\pi \approx \pi_0 P^n$ for large n. Another way to approximate stationary distribution is by observing the following almost sure equality due to strong law of large numbers,

$$\pi_i = \lim_{N \in \mathbb{N}} \frac{1}{N} \sum_{n=1}^{N} 1_{\{X_n = i\}} \approx \frac{1}{N} \sum_{n=1}^{N} 1_{\{X_n = i\}} \quad \text{for large } N. \tag{1}$$

This method sometimes fails for approximately null recurrent (large number of vertices with small associated stationary probabilities), and reducible (disconnected) graphs. One way to fix this is to perturb the transition probability matrix P in the following fashion,

$$T = (1-\beta)P + \beta Q, \tag{2}$$

where $Q_{ij} = 1/|V|$ for all $i, j \in V$. This implies that with probability β, one can jump to any other page in V uniformly at random, and with probability $1-\beta$ it jumps according the graph structure. Typically, the transition matrix P is unknown, and hence $P_{ij} = \frac{1}{|\{j \in V : (i,j) \in \overrightarrow{E}\}|}$.

2.2 Gambling

One can model the gambler's fortune by a random walk $(S_n \in \mathbb{Z} : n \in \mathbb{N})$, where $(X_n \in \mathbb{Z} : n \in \mathbb{N})$ denote the size of her winnings, and S_0 her initial fortune. One is interested in designing optimal gambling strategies. For example, gambler decides she would quit gambling when she has at least b units, or she is broke. Let H be this stopping time,

$$H \triangleq \inf\{n \in \mathbb{N} : S_n \geqslant b \text{ or } S_n \leqslant 0\}. \tag{3}$$

We are interested in finding mean of this stopping time, or stopped process S_H as a function of S_0 and b. Martingales are popularly used to find mean of stopping times, and stopped processes. Some non-trivial examples of stopping times for a Bernoulli process $(X_n \in \{0,1\} : n \in \mathbb{N})$ are

$$T_1 = \inf\{n \in \mathbb{N} : X_n = 1\}, \quad T_{10} = \inf\{n \in \mathbb{N} : X_n = 1, X_{n-1} = 0\}, \quad T_{101} = \inf\{n \in \mathbb{N} : X_n = 1, X_{n-1} = 0, X_{n-2} = 1\}.$$

For small pattern sizes, one can find the mean hitting times by forming Markov chains from the Bernoulli process. For example, $((X_n, X_{n-1}) : n \in \mathbb{N})$ forms a Markov chain.

2.3 Population modeling

Suppose a population where each organism lives for an *iid* random time period of X units with common distribution function F. Just before dying, each organism produces a number of offsprings N, an *iid* discrete random variable with common distribution P. Let $X(t)$ denote the number of organisms alive at time t. The stochastic process $(X(t),\ t \geqslant 0)$ is called an age-dependent branching process. We are interested in computing $M(t) = \mathbb{E}X(t)$ when $m = \mathbb{E}[N] = \sum_{j \in \mathbb{N}} jP_j$. This is a popular model in biology for population growth of various organisms. We will show that $M(t) \approx Ce^{\alpha t}$ for large t, where the constant

$$C = \frac{m-1}{m^2 \alpha \int_{\mathbb{R}_+} xe^{-\alpha x} dF(x)},$$

and α is the unique positive solution to the equation $\int_{x \in \mathbb{R}_+} e^{-\alpha x} dF(x) = \frac{1}{m}$.

2.4 Polya's urn scheme

Let (W_n, B_n) denotes the number of white and black balls in urn at the end of nth draw. In each draw a ball is picked from the urn at random, and returned to the urn along with another ball of the same color. Given the initial condition (W_0, B_0), can we say anything about the limiting value of random ratio $\lim_{n \in \mathbb{N}} \frac{W_n}{W_n + B_n}$?