Lecture-01: 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.

Example 1.1 (Differential equation). We consider a system that evolves deterministically in continuous time as the solution the following differential equation for $t > 0$

$$\frac{dx_t}{dt} = \alpha - \beta x_t,$$

with initial condition $x_0 = a$. It follows that $x_t = \frac{\alpha}{\beta}(1 - e^{-\beta t}) + ae^{-\beta t}$ for all $t \geqslant 0$.

Example 1.2 (Difference equation). We consider a system that evolves deterministically in discrete time as the solution the following difference equation for $n \in \mathbb{N}$

$$X_n = AX_{n-1} + B,$$

with initial condition $X_0 = C$. It follows that $X_n = A^n C + \sum_{k=0}^{n-1} A^k B$.

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.

Example 1.3 (Stochastic differential equation). We consider a system that evolves stochastically in continuous time as the solution the following differential equation for $t > 0$

$$\frac{dX_t}{dt} = \alpha X_t + dB_t,$$

with initial condition $X_0 = a$ and B_t being a Wiener process. It follows that the evolution of X_t is not deterministic.

Example 1.4 (Stochastic difference equation). We consider a system that evolves stochastically in discrete time as the solution the following difference equation for $n \in \mathbb{N}$

$$S_n = S_{n-1} + X_n,$$

with initial condition $S_0 = 0$ and $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ being an *i.i.d.* random step size sequence. The process $S : \Omega \to \mathbb{R}^{\mathbb{N}}$ is a random walk.

2 Motivating examples

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.

3 Stochastic modeling

Definition 3.1 (Set of discrete probability measures). For any finite set $\mathcal{X}$, the set of probability measures over $\mathcal{X}$ is defined by

$$\mathcal{P}(\mathcal{X}) \triangleq \left\{ \nu \in [0,1]^{\mathcal{X}} : \sum_{x \in \mathcal{X}} \nu_x = 1 \right\}.$$

3.1 Gambler's ruin

One can model many gambling games with random walks, where wins or losses on each bet can be thought of as a random step. That is, one can model the gambler's fortune by a random walk $S : \Omega \to \mathbb{Z}^{\mathbb{N}}$ such that $S_n \triangleq S_0 + \sum_{i=1}^{n} X_i$ for all $n \in \mathbb{N}$. The random sequence $X : \Omega \to \mathbb{Z}^{\mathbb{N}}$ denote the size of her winnings, and S_0 is 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,

$$H \triangleq \inf \{ n \in \mathbb{N} : S_n \geqslant b \text{ or } S_n \leqslant 0 \}. \tag{1}$$

One question of interest is finding the probability of the gambler getting bankrupt before she can quit gambling. That is, finding the $P\{S_H \leqslant 0\}$. Other questions of interest are finding the mean of this stopping time H and the mean of the stopped process S_H as a function of initial capital S_0 and quitting threshold b. These questions are related to hitting times of a random walk. Random walks have deep relations to Brownian motion.

3.2 Patterns

Martingales are popularly used to find mean of stopping times and stopped processes. Some non-trivial examples of stopping times for a Bernoulli process $X : \Omega \to \{0,1\}^{\mathbb{N}}$ are

$$\begin{aligned} T_1 &\triangleq \inf\{n \in \mathbb{N} : X_n = 1\}, \\ T_{10} &\triangleq \inf\{n \in \mathbb{N} : (X_n, X_{n-1}) = (1,0)\}, \\ T_{101} &\triangleq \inf\{n \in \mathbb{N} : (X_n, X_{n-1}, X_{n-2}) = (1,0,1)\}. \end{aligned}$$

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.

3.3 Page rank

The set of webpages in the internet can be denoted by a finite set V. We are interested in finding the rank of a page denoted by r_i for each page $i \in V$. 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 $E = \{(i,j) \in V \times V : i \sim j\}$. One can assume that the importance of a page is measured by the number of referrals to this page. For example,

$$r_i \propto d_{\text{in}}(i) = \sum_{j \in V} \mathbb{1}_{\{(j,i) \in E\}}.$$

However, this measure ignores the importance of the referrals. An easy way to fix that is to assume that the importance of a page is proportional to the aggregate sum of weighted number of referrals, where each referral is weighted by the importance of referring page distributed equally among all the referrals it makes. That is,

$$r_i \propto \sum_{j \in V} \frac{r_j}{d_{\text{out}}(j)} \mathbb{1}_{\{(j,i) \in E\}}.$$

If we define a matrix $H \in [0,1]^{V \times V}$ where $H_{j,i} \triangleq \frac{1}{d_{\text{out}}(j)} \mathbb{1}_{\{(j,i) \in E\}}$ for all $j, i \in V \times V$, then it is easy to see that $\sum_{i \in V} H_{ji} = 1$ for all $j \in V$ and $r = rH$. That is, H is a transition probability matrix with $H_{j,i}$ being the one-step transition probability from page j to page i. Thus, we can model a directed random walk $X : \Omega \to V^{\mathbb{N}}$ over the web-pages on the internet denoted by the directed graph $G = (V, E)$. We note that X is a Markov chain with transition matrix $H \in [0,1]^{V \times V}$ and the ranking vector $r \in \mathcal{P}(V)$ is its invariant distribution. 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 V, as in the case of internet, one doesn't know the transition matrix H a priori. And even if the whole transition matrix was known, finding the invariant distribution would take $O(|V|^3)$ computations, a very large number. This follows from the following approximation $r \approx \pi_n = \pi_0 H^n$ for large n. Another way to approximate stationary distribution is by observing the following almost sure equality due to strong law of large numbers,

$$r_i = \lim_{N \in \mathbb{N}} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\{X_n = i\}} \approx \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\{X_n = i\}} \quad \text{for large } N. \tag{2}$$

This method sometimes fails for approximately null recurrent chain where there are a large number of vertices with small associated stationary probabilities, and reducible or disconnected graphs. One way to fix this is to perturb the transition probability matrix H in the following fashion,

$$G \triangleq (1 - \beta)H + \beta \mathbf{1} q^T, \tag{3}$$

where $q \in \mathcal{P}(V)$ is a personalized search probability distribution and β is the teleportation probability. This implies that with teleportation probability β, one can jump to any other page $j \in V$ with probability q_j, and with probability $1 - \beta$ it jumps according the graph structure.

3.4 Population modeling

Suppose a population where each organism lives for an independent and identical distributed (*i.i.d.*) 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 $X : \Omega \to \mathbb{Z}_+^{\mathbb{R}_+}$ 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. These type of models can answer questions related to survival of species. 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 $\int_{x \in \mathbb{R}_+} e^{-\alpha x} dF(x) = \frac{1}{m}$.

3.5 Queues

Queues are complex stochastic processes and consist of two stochastic processes arrival and service, coupled non-linearly through a buffer. Number of arrivals and arrival instants could be discrete or continuous random variable. For a discrete arrival case, arrival process can be characterized by the time epochs of discrete arrivals, denoted $A : \Omega \to \mathbb{R}_+^{\mathbb{N}}$. Similarly, service requirement of each incoming arrival can also be a discrete or continuous random variable, denote by $S_n : \Omega \to \mathbb{R}_+$ for nth arrival. Queue can have a finite or infinite waiting area, and can be served by single or multiple servers. Important performance metrics for queues are mean waiting time of arrivals and mean queue length. These metrics are affected by the service policy that determines how to serve incoming arrivals. Few important service policies are first come first out, last in first out, processor sharing etc. Queues have applications in operations research, industrial engineering, telecommunications networks, among others.

3.6 Urn models

There are balls and urns in these models, and one is interested in distribution of balls in urns when one can randomly throw balls into urns. Balls can be of multiple colors and may or may not be replaced after putting into urns. These models have applications in influence maximization and epidemic control, where urns can denote influence or infection and balls can denote the individuals of a population.

For example in Polya's urn scheme, there are balls of two colors white and black in a single urn. Let the pair (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 $\lim_{n \in \mathbb{N}} \frac{W_n}{W_n + B_n}$?