Lecture 01: Introduction to Stochastic Processes

1 Probability Review

A **probability space** $(\Omega, \mathcal{F}, P)$ consists of set of all possible outcomes denoted by Ω and called a **sample space**, a collection of subsets $\mathcal{F}$ of sample space, and a non-negative set function probability $P : \mathcal{F} \to [0, 1]$, with the following properties.

1. The collection of subsets of $\mathcal{F}$ is a σ-algebra, that is it contains an empty set and is closed under complements and countable unions.
2. Set function P satisfies $P(\Omega) = 1$, and for every countable pair-wise disjoint collection $\{A_n \in \mathcal{F} : n \in \mathbb{N}\}$, we have
$$P(\bigcup_n A_n) = \sum_n P(A_n).$$

There is a natural order of inclusion on sets through which we can define monotonicity of probability set function P. To define continuity of this set function, we define limits of sets. For a sequence of sets $\{A_n : n \in \mathbb{N}\}$, we define **limit superior** and **limit inferior** of this sequence respectively as

$$\limsup_n A_n = \bigcap_n \bigcup_{k \geq n} A_k, \qquad \liminf_n A_n = \bigcup_n \bigcap_{k \geq n} A_k.$$

We say that limit exists if the limit superior and limit inferior are equal, and is equal to the limit of the sequence of sets.

Theorem 1.1. *Probability set function is monotone and continuous.*

Proof. Let $A \subseteq B$ both subsets be elements of $\mathcal{F}$, then from the additivity of probability over disjoint sets A and $B \setminus A$, we have

$$P(B) = P(A \cup B \setminus A) = P(A) + P(B \setminus A) \geq P(A).$$

Monotonicity follows from non-negativity of probability set function, that is since $P(B \setminus A) > 0$. For continuity from below, we take an increasing sequence of sets $\{A_n : n \in \mathbb{N}\}$, such that $A_n \subseteq A_{n+1}$ for all n. Then, it is clear that $A_n \uparrow A = \cup_n A_n$. We can define disjoint sets $\{E_n : n \in \mathbb{N}\}$, where

$$E_1 = A_1, \qquad E_n = A_n \setminus A_{n-1},\ n \geq 2.$$

The disjoint sets E_n's satisfy $\cup_{i=1}^n E_i = A_n$ for all $n \in \mathbb{N}$ and $\cup_n E_n = \cup_n A_n$. From the above property and the additivity of probability set function over disjoint sets, it follows that

$$P(A) = P \cup_n E_n = \sum_{n \in \mathbb{N}} P(E_n) = \lim_{n \in \mathbb{N}} \sum_{i=1}^n P(E_i) = \lim_{n \in \mathbb{N}} P \cup_{i=1}^n E_i = \lim_{n \in \mathbb{N}} P(A_n).$$

For continuity from below, we take decreasing sequence of sets $\{A_n : n \in \mathbb{N}\}$, such that $A_{n+1} \subseteq A_n$ for all n. We can form increasing sequence of sets $\{B_n : n \in \mathbb{N}\}$ where $B_n = A_n^c$. Then, the continuity from above follows from continuity from above. $\square$

A real valued **random variable** X on a probability space $(\Omega, \mathcal{F}, P)$ is a function $X : \Omega \to \mathbb{R}$ such that for every $x \in \mathbb{R}$, we have $\{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}(-\infty, x] \in \mathcal{F}$. The **distribution function** $F : \mathbb{R} \to [0, 1]$ for this random variable X is defined as

$$F(x) = (P \circ X^{-1})(-\infty, x],\ \forall x \in \mathbb{R}.$$

Let $g : \mathbb{R} \to \mathbb{R}$ be a function. Then, the **expectation** of $g(X)$ is defined as

$$\mathbb{E}g(X) = \int_{x \in \mathbb{R}} g(x) dF(x).$$

Theorem 1.2. *Distribution function F of a random variable X is non-negative, monotone increasing, continuous from the right, and has countable points of discontinuities. Further, if $P \circ X^{-1}(\mathbb{R}) = 1$, then*

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1.$$

Proof. Non-negativity and monotonicity of distribution function follows from non-negativity and monotonicity of probability set function, and the fact that for $x_1 < x_2$

$$X^{-1}(-\infty, x_1] \subseteq X^{-1}(-\infty, x_2].$$

Let $x_n \downarrow x$ be a decreasing sequence of real numbers. Then, the right continuity of distribution function follows from the continuity from above of probability set functions. We take decreasing sets $\{A_n : n \in \mathbb{N}\}$, where

$$A_n = \{\omega \in \Omega : X(\omega) \leq x_n\}.$$

$\square$

2 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.

3 Stochastic Processes

A collection of random variables $\{X_t \in \mathcal{X} : t \in T\}$ each defined on the same probability space $(\Omega, \mathcal{F}, P)$ is called a **random process** for an arbitrary index set T and arbitrary state space $\mathcal{X}$.

3.1 Classification

If the state space T is countable, the stochastic process is called **discrete**-time stochastic process or random sequence. When the state space T is uncountable, it is called **continuous**-time stochastic process. However, T doesn't have to be time, if the index set is space, and then the stochastic process is spatial process. When $T = \mathbb{R}^n \times [0, \infty)$, stochastic process $X(t)$ is a spatio-temporal process. State space $\mathcal{X}$ can also be countable or uncountable.

We list some examples of each such stochastic process.

- Discrete random sequence: brand switching, discrete time queues, number of people at bank each day.
- Continuous random sequence: stock prices, currency exchange rates, waiting time in queue of nth arrival, workload at arrivals in time sharing computer systems.
- Discrete random process: counting processes, population sampled at birth-death instants, number of people in queues.
- Continuous random process: water level in a dam, waiting time till service in a queue, location of a mobile node in a network.

3.2 Specification

To define a measure on collection of random variables, we need to know it's joint distribution $F : \mathbb{R}^T \to [0, 1]$. To this end, for any $x \in \mathbb{R}^T$ we need to know

$$F(x) = P\left(\bigcap_{t \in T} \{\omega \in \Omega : X_t(\omega) \leq x_t\}\right).$$

When the index set T is infinite, any function of the above form would be zero if x_t is finite for all $t \in T$. Therefore, we only look at the values of $F(x)$ when $x_t \in \mathbb{R}$ for indices t in a finite set S and $x_t = \infty$ for all $t \notin S$. We can define a **finite dimensional distribution** for any finite set $S \subseteq T$ and $x_S = \{x_s \in \mathbb{R} : s \in S\}$,

$$F_S(x) = P\left(\bigcap_{s \in S} \{\omega \in \Omega : X_s(\omega) \leq x_s\}\right).$$

Set of all finite dimensional distributions of the stochastic process $\{X_t : t \in T\}$ characterizes its distribution completely. Simpler characterizations of a stochastic process $X(t)$ are in terms of its moments. That is, the first moment such as mean, and the second moment such as correlations and covariance functions.

$$m_X(t) \triangleq \mathbb{E}X_t, \qquad R_X(t, s) \triangleq \mathbb{E}X_t X_s, \qquad C_X(t, s) \triangleq \mathbb{E}(X_t - m_X(t))(X_s - m_X(s)).$$

Some examples of simple stochastic processes.

i. $X_t = A \cos 2\pi t$, where A is random. The finite dimensional distribution is given by

$$F_S(x) = P\left(\{A \cos 2\pi s \leq x_s, s \in S\}\right).$$

The moments are given by

$$m_X(t) = (\mathbb{E}A) \cos 2\pi t, \quad R_X(t, s) = (\mathbb{E}A^2) \cos 2\pi t \cos 2\pi s, \quad C_X(t, s) = \mathrm{Var}(A) \cos 2\pi t \cos 2\pi s.$$

ii. $X_t = \cos(2\pi t + \Theta)$, where Θ is random and uniformly distributed between $(-\pi, \pi]$. The finite dimensional distribution is given by

$$F_S(x) = P\left(\{\cos(2\pi s + \Theta) \leq x_s, s \in S\}\right).$$

The moments are given by

$$m_X = 0, \qquad R_X(t, s) = \frac{1}{2} \cos 2\pi (t - s), \qquad C_X(t, s) = R_X(t, s).$$

iii. $X_n = U^n$ for $n \in \mathbb{N}$, where U is uniformly distributed in the open interval $(0, 1)$.

iv. $Z_t = At + B$ where A and B are independent random variables.

3.3 Independence

Recall, given the probability space $(\Omega, \mathcal{F}, P)$, two events $A, B \in \mathcal{F}$ are **independent events** if

$$P(A \cap B) = P(A)P(B).$$

Random variables X, Y defined on the above probability space, are **independent random variables** if for all $x, y \in \mathbb{R}$

$$P\{X(\omega) \le x, Y(\omega) \le y\} = P\{X(\omega) \le x\}P\{Y(\omega) \le y\}.$$

Two stochastic process X_t, Y_t for common index set T are **independent stochastic processes** if for all finite subsets $I, J \subseteq T$

$$P\left(\{X_i \le x_i, i \in I\} \cap \{Y_j \le y_j, j \in J\}\right) = P\left(\{X_i \le x_i, i \in I\}\right) P\left(\{Y_j \le y_j, j \in J\}\right).$$

3.4 Examples of Tractable Stochastic Processes

In general, it is very difficult to characterize a stochastic process completely in terms of its finite dimensional distribution. However, we have listed few analytically tractable examples below, where we can completely characterize the stochastic process.

3.4.1 Independent and identically distributed processes

Let $\{X_t : t \in T\}$ be an independent and identically distributed (iid) random process, with common distribution $F(x)$. Then, the finite dimensional distribution for this process for any finite $S \subseteq T$ can be written as

$$F_S(x) = P\left(\{X_s(\omega) \le x_s, s \in S\}\right) = \prod_{s \in S} F(x_s).$$

It's easy to verify that the first and the second moments are independent of time indices. Since $X_t = X_0$ in distribution,

$$m_X = \mathbb{E}X_0, \qquad R_X = \mathbb{E}X_0^2, \qquad C_X = \text{Var}(X_0).$$

3.4.2 Stationary Processes

A stochastic process X_t is **stationary** if all finite dimensional distributions are shift invariant, that is for finite $S \subseteq T$ and $t > 0$, we have

$$F_S(x) = P(\{X_s(\omega) \le x_s, s \in S\}) = P(\{X_{s+t}(\omega) \le x_s, s \in S\}) = F_{t+S}(x).$$

In particular, all the moments are shift invariant. Since $X_t = X_0$ and $(X_t, X_s) = (X_{t-s}, X_0)$ in distribution, we have

$$m_X = \mathbb{E}X_0, \qquad R_X(t-s, 0) = \mathbb{E}X_{t-s}X_0, \qquad C_X(t-s, 0) = R_X(t-s, 0) - m_X^2.$$

3.4.3 Markov Processes

A stochastic process X_t is **Markov** if conditioned on the present state, future is independent of the past. That is, for any ordered index set T containing any two indices $u > t$, we have

$$P(\{X_u(\omega) \le x_u | X_s, s \le t\}) = P(\{X_u(\omega) \le x_u | X_t\}).$$

We will study this process in detail in coming lectures.

3.4.4 Lévy Processes

A stochastic process X_t indexed by positive reals is **Lévy** if the following conditions hold.

i. $X_0 = 0$, almost surely.

ii. The increments are independent: For any $0 \le t_1 < t_2 < \cdots < t_n < \infty$, $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

iii. The increments are stationary: For any $s < t$, $X_t - X_s$, is equal in distribution to X_{t-s}.

iv. Continuous in probability: For any $\epsilon > 0$ and $t \ge 0$ it holds that $\lim_{h\to 0} P(|X_{t+h} - X_t| > \epsilon) = 0$.

Two examples of Lévy processes are Poisson process and Wiener process. The distribution of Poisson process at time t is Poisson with rate λt and the distribution of Wiener process at time t is zero mean Gaussian with variance t.

Theorem 3.1. *A Lévy process has infinite divisibility. That is, for all $n \in \mathbb{N}$*

$$\mathbb{E}e^{\theta X_t} = \left(\mathbb{E}e^{\theta X_{t/n}}\right)^n .$$

Further, if the process has finite moments $\mu_n(t) = \mathbb{E}X_t^n$ then the following Binomial identity holds

$$\mu_n(t+s) = \sum_{k=0}^{n} \binom{n}{k} \mu_k(t)\mu_{n-k}(s).$$

Proof. The first equality follows from the independent and stationary increment property of the process, and the fact that we can write

$$X_t = \sum_{k=1}^{n} X_{\frac{kt}{n}} - X_{\frac{(k-1)t}{n}} .$$

Second property also follows from the the independent and stationary increment property of the process, and the fact that we can write

$$X_{t+s}^n = (X_t + X_{t+s} - X_t)^n = \sum_{k=0}^{n} \binom{n}{k} X_t^n (X_{t+s} - X_t)^{n-k}.$$

□