Lecture-02: Probability Review

1 Probability Review

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

1. Event space \(\mathcal{F}\) is a \(\sigma\)-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 is additive for countably disjoint events.

Example 1.1 (Discrete \(\sigma\)-algebra). For a finite sample space \(\Omega\), the event space \(\mathcal{F} = \{A : A \subseteq \Omega\}\) consists of all subsets of sample space \(\Omega\).

Example 1.2 (Borel \(\sigma\)-algebra). If the sample space \(\Omega = \mathbb{R}\), then a Borel \(\sigma\)-algebra is generated by half-open intervals by complements and countable unions. That is, \(\mathcal{B} = \sigma(\{(-\infty, x] : x \in \mathbb{R}\})\). We can see that \((x, \infty)\) belong to \(\mathcal{B}\) for each \(x \in \mathbb{R}\) by closure under complements. From closure under countable unions, we have open intervals \((-\infty, x) = \bigcup_{n \in \mathbb{N}} (-\infty, x - \frac{1}{n})\) also belonging to \(\mathcal{B}\). We can also show that singletons belong to \(\mathcal{B}\), since \(\{x\} = \bigcap_{n \in \mathbb{N}} [x - \frac{1}{n}, \infty) \cap (-\infty, x + \frac{1}{n}]\).

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 set sequence respectively as

\[
\limsup_{n} A_n = \bigcup_{n \geq k} A_k, \quad \liminf_{n} A_n = \bigcap_{n \geq k} A_k.
\]

It is easy to check that \(\liminf A_n \subseteq \limsup A_n\). We say that limit of set sequence exists if \(\limsup A_n \subseteq \liminf A_n\), and the limit of the set sequence in this case is \(\limsup A_n\).

Theorem 1.3. 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 \(A_n \uparrow A = \bigcup_{n \in \mathbb{N}} A_n\). We can define disjoint sets \(\{E_n : n \in \mathbb{N}\}\), where

\[
E_1 = A_1, \quad E_n = A_n \setminus A_{n-1}, \quad n \geq 2.
\]

The disjoint sets \(E_n\)’s satisfy \(\bigcup_{i=1}^n E_i = A_n\) for all \(n \in \mathbb{N}\) and \(\bigcup_{n \in \mathbb{N}} E_n = \bigcup_{n \in \mathbb{N}} A_n\). 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.
2 Random variables

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 $\{ \omega \in \Omega : X(\omega) \leq x \} = X^{-1}(-\infty,x] \subseteq \mathcal{F}$. That is, $X^{-1}(\mathbb{B}) \subseteq \mathcal{F}$. The distribution function $F : \mathbb{R} \to [0,1]$ for this random variable $X$ is defined as $F(x) = (P \circ X^{-1})(-\infty,x], \forall x \in \mathbb{R}$.

**Theorem 2.1.** 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

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

**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

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

□

**Example 2.2.** One of the simplest random variables are indicator functions $1 : \mathcal{F} \times \Omega \to \{0,1\}$. For each event $A \in \mathcal{F}$, we can define an indicator function as

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A. \end{cases}$$

2.1 Expectation

Let $g : \mathbb{R} \to \mathbb{R}$ be a function. Then, the expectation of $g(X)$ for a random variable $X$ with distribution function $F$ is defined as

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

**Remark 1.** Recall that probabilities are defined only for events. For a random variable $X$, the probabilities are defined for generating events $X^{-1}(-\infty,x] \subseteq \mathcal{F}$ by $F(x) = P \circ X^{-1}(-\infty,x].$

**Remark 2.** The expectation is only defined for random variables. For an event $A$, the probability $P(A)$ equals expectation of the indicator random variable $1_A$.

3 Stochastic Processes

Let $(\Omega, \mathcal{F}, P)$ be a probability space. For an arbitrary index set $T$ and state space $\mathcal{X} \subseteq \mathbb{R}$, a random process is a measurable map $X : (\Omega, T) \to \mathcal{X}$. For each $t \in T$, we have a random variable $X_t \triangleq \{X(t, \omega) : \omega \in \Omega\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, and random process $X$ is a collection of random variables $(X_t \in \mathcal{X} : t \in T)$. For each $\omega \in \Omega$, the map $X_{t_\omega} \triangleq \{X_t(\omega) : t \in T\}$ denotes a sample path of the process $X$.

3.1 Classification

State space $\mathcal{X}$ can be countable or uncountable, corresponding to discrete or continuous valued process. If the index set $T$ is countable, the stochastic process is called discrete-time stochastic process or random sequence. When the index set $T$ is uncountable, it is called continuous-time stochastic process. The index set $T$ 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.
Example 3.1. We list some examples of each such stochastic process.

i. Discrete random sequence: brand switching, discrete time queues, number of people at bank each day.

ii. Continuous random process: stock prices, currency exchange rates, waiting time in queue of nth arrival, workload at arrivals in time sharing computer systems.

iii. Discrete random process: counting processes, population sampled at birth-death instants, number of people in queues.

iv. 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 a random process, we can either put a measure on sample paths, or equip the collection of random variables with a joint measure. We are interested in identifying the joint distribution of a stochastic process \( X_t \) for any finite set \( S \subseteq T \) we focus on the product sets of the form

\[
\bigotimes_{s \in S} (-\infty, x_s] \times \mathbb{R}.
\]

We can define a finite dimensional distribution for any finite set \( S \subseteq T \) and \( x_S = \{x_s \in \mathbb{R} : s \in S\} \),

\[
F_S(x_S) = P\left(\bigcap_{s \in S} \{\omega \in \Omega : X_s(\omega) \leq x_s\}\right) = P\left(\bigcap_{t \in T} X_t^{-1}(-\infty, x_t]\right).
\]

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.

\[
m_X(t) \triangleq \mathbb{E}X_t, \quad R_X(t, s) \triangleq \mathbb{E}X_tX_s, \quad C_X(t, s) \triangleq \mathbb{E}(X_t - m_X(t))(X_s - m_X(s)).
\]

Example 3.2. Some examples of simple stochastic processes.

i. \( X_t = A \cos 2\pi t \), where \( A \) is random.

ii. \( X_t = \cos(2\pi t + \Theta) \), where \( \Theta \) is random and uniformly distributed between \((-\pi, \pi]\).

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 \( x, y \in \mathbb{R} \), the events \( \{X \leq x\} \) and \( \{Y \leq y\} \) are independent. That is,

\[
P\{X(\omega) \leq x, Y(\omega) \leq y\} = P\{X(\omega) \leq x\}P\{Y(\omega) \leq y\}.
\]
A stochastic process \( X \) is said to be \textbf{independent} if for all finite subsets \( S \subseteq T \), we have
\[
P(\{X_s \leq x_s, s \in S\}) = \prod_{s \in S} P(X_s \leq x_s).
\]

Two stochastic processes \( X, Y \) for the common index set \( T \) are \textbf{independent random processes} if for all finite subsets \( I, J \subseteq T \)
\[
P(\{X_i \leq x_i, i \in I\} \cap \{Y_j \leq y_j, j \in J\}) = P(\{X_i \leq x_i, i \in I\})P(\{Y_j \leq y_j, j \in J\}).
\]

### 3.4 Filtration

A net of \( \sigma \)-algebras \( \mathcal{F}_t = \{\mathcal{F}_t \subseteq \mathcal{F} : t \in T\} \) is called a \textbf{filtration} when the index set \( T \) is totally ordered and the net is non-decreasing, that is for all \( s \leq t \) we have \( \mathcal{F}_s \subseteq \mathcal{F}_t \). Consider a real-valued random process \( X \) indexed by the ordered set \( T \) on the probability space \( (\Omega, \mathcal{F}, P) \). The process \( X \) is called \textbf{adapted} to the filtration \( \mathcal{F}_\bullet \), if for each \( t \in T \), we have the random variable \( X_t \in \mathcal{F}_t \) or \( X_t^{-1}(\{x\}) \in \mathcal{F}_t \) for each \( x \in \mathbb{R} \).

We can define a natural filtration \( \mathcal{F}_\bullet = \{\mathcal{F}_t \subseteq \mathcal{F} : t \in T\} \) indexed by totally ordered \( T \) for the random process \( X = (X_s : s \in T) \), where \( \mathcal{F}_t = \sigma(X_s, s \leq t) \) is the information about the process till index \( t \) and the process \( X \) is adapted to its natural filtration by definition.

If \( X = (X_t : t \in T) \) is an independent process with the associated natural filtration \( \mathcal{F}_\bullet \), then for any \( t > s \) and events \( A \in \mathcal{F}_s \), the random variable \( X_t \) is independent of the event \( A \). This is just a fancy way of saying \( X_t \) is independent of \((X_u, u \leq s)\). Hence, for any random variable \( Y \in \mathcal{F}_t \), we have
\[
\mathbb{E}[\mathbb{E}[X_t | \mathcal{F}_s]] = \mathbb{E}[\mathbb{E}[X_t | Y]] = \mathbb{E}X_t \mathbb{E}Y.
\]

### 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.

#### 4.1 Independent and identically distributed processes

Let \( \{X_t : t \in T\} \) be an independent and identically distributed (iid) random process, with a common distribution \( F(x) \). Then, the finite dimensional distribution for this process for any finite \( S \subseteq T \) can be written as
\[
F_S(x) = P(\{X_\omega(t) \leq x, s \in S\}) = \prod_{s \in S} F(x).
\]

It’s easy to verify that the first and the second moments are independent of time indices. Since \( X_t = X_0 \) in distribution,
\[
m_X = \mathbb{E}X_0, \quad R_X = \mathbb{E}X_0^2, \quad C_X = \text{Var}(X_0).
\]

#### 4.2 Stationary processes

A stochastic process \( X_t \) is \textbf{stationary} if all finite dimensional distributions are shift invariant, that is for finite \( S \subseteq T \) and \( t > 0 \), we have
\[
F_S(x) = P(\{X_\omega(t) \leq x, s \in S\}) = P(\{X_{\omega+t} \leq x, s \in S\}) = F_{t+S}(x).
\]

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

#### 4.3 Markov processes

A stochastic process \( X_t \) is \textbf{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
\[
P(\{X_u \leq x_u \} | \mathcal{F}_t) = P(\{X_u \leq x_u \} | \sigma(X_t)).
\]

We will study this process in detail in coming lectures.
4.4 Lévy processes

A right continuous with left limits stochastic process $X = (X_t : t \in T \subseteq \mathbb{R}_+)$ with $X_0 = 0$ almost surely, is a Lévy process if the following conditions hold.

(L1) The increments are independent. For any $0 \leq t_1 < t_2 < \cdots < t_n < \infty$, $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

(L2) The increments are stationary. For any $s < t$, $X_t - X_s$, is equal in distribution to $X_{t-s}$.

(L3) Continuous in probability. For any $\varepsilon > 0$ and $t \geq 0$ it holds that $\lim_{h \to 0} P(|X_{t+h} - X_t| > \varepsilon) = 0$.

Example 4.1. Two examples of Lévy processes are Poisson process and Wiener process. The distribution of Poisson process at time $t$ is Poisson with rate $\lambda t$ and the distribution of Wiener process at time $t$ is zero mean Gaussian with variance $t$.

Theorem 4.2. A Lévy process has infinite divisibility. That is, for all $n \in \mathbb{N}$

$$Ee^{\theta X_t} = \left(Ee^{\theta X_t/n}\right)^n.$$  

Further, if the process has finite moments $\mu_n(t) = EX^n_t$ 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

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

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

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

\[\square\]