

DE 13: Stochastic Processes

Madras School of Economics

Srikanth Pai

Last updated: October 8, 2024

Contents

Fundamentals of Probability	3
1.1 Random variables and their distribution	8
1.2 Joint p.m.fs	9
1.3 Expectation	9
Introduction to Stochastic Processes	9
Discrete time Markov Chains	15
Poisson point process	17
4.1 Cramer-Lundberg Ruin theory	23
Continuous time Markov Chains	24
Martingales	24

Preface

I am a curious person with little knowledge of finance and economics, so I urge the reader to not take my finance/economics ideas seriously. It is essentially a layman's view of the subject. I love to learn new ideas, so I will be delighted to be corrected. So, you can always tell me where I am wrong by writing an email or telling me directly if you spot me somewhere at MSE.

The goal of mathematical training is to make ideas precise. So when mathematics is written, the definitions are stated clearly, precise claims are proposed, and then these claims are proved rigorously. A collection of examples usually motivates the questions. The examples are also used to test the claims. It is always worthwhile to study examples that are important to humanity. After all, we all live in this world briefly, so why squabble our precious time and energy on inconsequential examples? I.I.D processes, Markov processes, Martingales and Brownian motions are essential examples of stochastic processes that are widely studied because they arise very often when we look at the world around us. Natural questions about these processes are proposed in this text, and the answers are stated precisely using a theorem-proof format. We will also apply the theorems to many problems and learn to compute efficiently.

※ Fundamentals of Probability

The primary objective of probability theory is to provide a framework for studying the chances of interesting events that happen in the real world. We view all such interesting phenomena as experiments. Imagine we are studying an experiment that involves chance and we can contemplate the collection of all possible outcomes of the experiment. The plan is to model real life events using **set theory**. A textbook example of an experiment is tossing a coin indefinitely and here is a quick dictionary:

English	Math Translation
Collection of all outcomes	Sample space Ω
An outcome is realized/The experiment is performed	Pick $\omega \in \Omega$
An event E	E is a subset of Ω
An experiment is performed and the event E has occurred	Pick $\omega \in \Omega$ and $\omega \in E$
An event E implies event F	$E \subseteq F$
The chance of E happening is p	$\Pr(E) = p$

In other words, the main actors of the story are **events** and **probability** assigned to every event. In school we learn to add, subtract, and multiply numbers. The rules governing these numbers is called arithmetic, but when the rules are generalized to variables, we call the subject 'Algebra'. In the same way there is an associated algebra of manipulating events by forming composite events using simple events. We implement this algebra using algebra of sets. The analog of addition is union, multiplication is intersection and negation is complement (*Don't take the analogy too literally!*). Imagine an experiment has been performed and an outcome has been realized. This means we have picked an element $\omega \in \Omega$. Here is a quick dictionary between composite events and algebra of sets: For a countable collection, $A_1, A_2, \dots, A_n \dots$, of subsets,

Atleast one of the events A_i happens	$\omega \in A_1 \cup A_2 \cup \dots \cup A_n \cup \dots$
All the events A_i simultaneously happen	$\omega \in A_1 \cap A_2 \cap \dots \cap A_n \dots$
Event A_1 has not occurred	$\omega \in A_1^c$

In your first year math course, you were introduced to a new operation: limits! Calculus is founded on the notion of limits of functions and the interactions of limits with algebra of variables. In a similar way, we want to be able to work with limits in probability. So above we have considered the dictionary for countable collection of events.

It is natural to wish that given basic events, the countable unions, countable intersections and complements of these basic events are also events. This wish leads to the following definition.

Definition 1.1. Let Ω be a given set. If \mathcal{F} is a collection of subsets of Ω that are closed under countable unions, countable intersections and complements, then we say \mathcal{F} is a σ -algebra.

The central focus of probability theory is to explore questions of the type: What is the chance associated to an event? Given a sample space Ω and any event E in a σ -algebra \mathcal{F} , we want to assign a number $\Pr(E)$ that measures the chance that the event E occurs. The assignment of chance to events transforms algebra of manipulating events to the algebra of real numbers. The rules governing the transformations is decreed in the axioms of probability theory:

Definition 1.2. A triple $(\Omega, \mathcal{F}, \Pr)$, where Ω is a set, \mathcal{F} is a σ -algebra of subsets and $\Pr : \mathcal{F} \rightarrow \mathbb{R}$ is a function, that satisfy

1. (Axiom 1) For any E in \mathcal{F} , $\Pr(E) \geq 0$.
2. (Axiom 2) $\Pr(\Omega) = 1$.
3. (Axiom 3) If $A_1, A_2, \dots, A_n, \dots$ in \mathcal{F} are a countable or a finite collection of disjoint subsets, then

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i).$$

is called a *probability space*.

The sample space Ω represents the collection of all outcomes of an experiment and we can consider it synonymous with the experiment itself. The collection of subsets \mathcal{F} represents a collection of interesting events. Finally the function \Pr is called a *probability measure* that measures chances of events. The first axiom states that chances of events is a non-negative number. The second axiom says that any outcome is surely in the sample space. The third axiom states: The chance that atleast one of the events in a countable (or finite) collection of mutually exclusive events occurs is the sum of the chances of individual events. It mimics the principle in geometry “The total area of the union of disjoint regions is the sum of the areas of its parts.” Or we can simply remember it as “the whole is the sum of its parts”.

A few immediate consequences of the axioms:

1. Let ϕ denote the empty set, then

$$\Pr(\phi) = 0.$$

Proof: Apply Axiom 3 to two disjoint sets $A_1 = A_2 = \phi$ to get

$$\Pr(\phi) + \Pr(\phi) = \Pr(\phi \cup \phi)$$

$$\Pr(\phi) + \Pr(\phi) = \Pr(\phi)$$

$$\Pr(\phi) = 0.$$

2. (Monotonicity) Let A, B denote the two elements of \mathcal{F} with $A \subset B$ then

$$\Pr(A) \leq \Pr(B).$$

Proof: Apply Axiom 3 to two disjoint sets $A_1 = A, A_2 = A^c \cap B$ to get

$$\Pr(A_1) + \Pr(A_2) = \Pr(A_1 \cup A_2)$$

$$\Pr(A) + \Pr(A^c \cap B) = \Pr(B)$$

Since by Axiom 1, $\Pr(A^c \cap B) \geq 0$, we have

$$\Pr(A) \leq \Pr(B).$$

3. (Complement rule) Let A be an event then

$$\Pr(A) + \Pr(A^c) = 1.$$

Proof: Apply Axiom 3 to two disjoint sets $A_1 = A, A_2 = A^c$ to get

$$\Pr(A_1) + \Pr(A_2) = \Pr(A_1 \cup A_2)$$

$$\Pr(A) + \Pr(A^c) = \Pr(\Omega)$$

Since by Axiom 2, $\Pr(\Omega) = 1$, so we are done.

4. (Union rule) Let A, B be events then

$$\Pr(A \cup B) + \Pr(A \cap B) = \Pr(A) + \Pr(B).$$

Proof: Exercise!

Lemma. Let A_1, A_2, \dots be a sequence of events and let $A = \bigcup_{n=1}^{\infty} A_n$, then

$$\Pr(A) = \lim_{n \rightarrow \infty} \Pr(A_n).$$

Proof. Let $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \dots$ which is a disjoint union (draw Venn diagrams!).

From proof of Monotonicity rule above, note that

$$\Pr(A_{i+1} \setminus A_i) = \Pr(A_{i+1}) - \Pr(A_i).$$

Apply Axiom 3 to the disjoint union presentation of A to get

$$\begin{aligned}
 \Pr(A) &= \Pr(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \dots) \\
 &= \Pr(A_1) + \Pr(A_2 \setminus A_1) + \Pr(A_3 \setminus A_2) \dots \\
 &= \Pr(A_1) + (\Pr(A_2) - \Pr(A_1)) + (\Pr(A_3) - \Pr(A_2)) \dots \\
 &= \lim_{n \rightarrow \infty} \Pr(A_n)
 \end{aligned}$$

□

Let us see three important classes of examples.

Example 1.3. Let $\Omega = \{\omega_1, \omega_2, \dots\}$ be a countable/finite set and let $f : \Omega \rightarrow [0, 1]$ be a function with the property

$$\sum_{i=1}^{\infty} f(\omega_i) = 1.$$

Let \mathcal{F} be the set of all subsets of Ω . Check that \mathcal{F} is a σ -algebra. Let E be any element of \mathcal{F} , define

$$\Pr(E) := \sum_{\omega \in E} f(\omega).$$

Now check the the three axioms. The first two are straightforward. The last axiom requires some real analysis.

Here are some examples of f :

1. (Bernoulli distribution) Set $\Omega = \{0, 1\}$ and let p be a real number between 0 and 1. Let $f(1) = p$ and $f(0) = 1 - p$.
2. (Binomial distribution) For a positive integer n , set $\Omega = \{0, 1, 2, 3, \dots, n\}$. Let a real number p between 0 and 1 be given. Then let

$$f(k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

3. (Geometric distribution) Set $\Omega = \mathbb{Z}_{\geq 0}$. For a real number p between 0 and 1, let

$$f(k) = p(1 - p)^k.$$

Example 1.3 is the setup of discrete probability theory. Coin tosses, and rolling dice experiments fit into the above framework since the sample space of outcomes is a finite set in these cases.

Example 1.4. Let $\Omega = \mathbb{R}$ and let \mathcal{F} be the smallest σ -algebra that contains all intervals of the form $(a, b]$ for all reals a, b . For any non-negative continuous function $f : \mathbb{R} \rightarrow [0, 1]$ such that

$$\int_{-\infty}^{\infty} f(x) dx = 1,$$

we define

$$\Pr\{(a, b]\} = \int_a^b f(x) dx$$

Note that the proof of axiom 3 is highly technical.

Here are some examples of f :

1. (Uniform distribution) Let $f(x) = \frac{1}{b-a}$ for $a < x < b$ and $f(x) = 0$ otherwise.
2. (Exponential distribution) For a positive real number λ , let $f(x) = \lambda e^{-\lambda x}$ for $x > 0$ and $f(x) = 0$ otherwise.
3. (Normal distribution) For a real numbers μ, σ^2 , let

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Example 1.4 is the setup of the probability theory of continuous random variables. Picking a random real number in an interval, and the random time it takes for a device to break down are examples of continuous random variables.

Example 1.5. Let $(\Omega, \mathcal{F}, \Pr)$ be a probability space and let C be an element of \mathcal{F} , then we can define a new triple $(C, \mathcal{F}_C, \Pr_C)$ as follows:

1. C is the sample space of the new triple.
2. \mathcal{F}_C is the set of all subsets of Ω of the form $A \cap C$ where A is in \mathcal{F} . Clearly any element of \mathcal{F}_C is a subset of C . Prove that \mathcal{F}_C is a σ -algebra.
3. Let A be an element of \mathcal{F}_C . Define

$$\Pr_C(A) := \frac{\Pr(A \cap C)}{\Pr(C)}.$$

Check that the axioms are satisfied (straightforward exercise). We quickly remark that usually we know this measure as

$$\Pr_C(A) = \Pr(A \mid C).$$

Example 1.5 shows that given an event C associated with a probability space, we can define a new probability space that represents “conditional probabilities”. The probability measure \Pr_C is called *the conditional probability measure*.

1.1 Random variables and their distribution

First we define a random variable.

Definition 1.6. Given a probability space $(\Omega, \mathcal{F}, \Pr)$, a random variable $X : \Omega \rightarrow \mathbb{R}$ is a function so that the subsets $\{a < X \leq b\}$ belong to \mathcal{F} for any reals a, b .

Given a random variable X defined on a probability space $(\Omega, \mathcal{F}, \Pr)$, define the *cumulative distribution function (c.d.f)* of X as

$$F_X(b) := \Pr(X \leq b).$$

Note that this means $\Pr(a < X \leq b) = F_X(b) - F_X(a)$.

Few quick properties of c.d.fs:

1.

$$\lim_{b \rightarrow \infty} F_X(b) = 1, \quad \lim_{b \rightarrow -\infty} F_X(b) = 0.$$

2. F_X is a non-decreasing function.

3. If $F_X(x)$ is differentiable, then

$$f(x) := \frac{dF_X}{dx}$$

is called the *probability density function* (or p.d.f) of X . Examples of p.d.f.s f are given in Example 1.4.

4. In case X is a discrete random variable, then F_X is piecewise constant and if the jumps occur at x_1, x_2, \dots , then the magnitude of the jump at x_i is called the p.m.f $f(x_i)$, i.e.

$$f(x_i) := F(x_i) - \lim_{x \rightarrow x_i^-} F(x).$$

Note that the sum of the p.m.f values should be 1.

If F_X is a differentiable function and $F'_X = f$, then we have

$$\Pr\{a < X \leq b\} = \int_a^b f(x) dx.$$

Let us informally discuss the meaning of density function.

What is the meaning of probability density function? Intuitively, density is a measure of how densely something is packed in a given region. We have seen that probability assigned to a point in a continuous sample space can be zero for some random variables. So in Example 1.4 we decided to assign probability to intervals. This means instead of

asking “What is the chance that $X = x$?”, we have to settle for the question “What is the chance that $x - \frac{\epsilon}{2} < X \leq x + \frac{\epsilon}{2}$?” for a small error of ϵ . If a random variable X has p.d.f $f(x)$, then

$$\Pr(x - \frac{\epsilon}{2} < X \leq x + \frac{\epsilon}{2}) \approx f(x)\epsilon$$

assuming $f(x)$ is constant in the small interval. Since ϵ is the length of the interval $I = (x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2})$, we see that

$$f(x) = \frac{\Pr(X \in I)}{\text{length}(I)}.$$

In words, we can say that at a point x , $f(x)$ measures the chance at x per unit length.

1.2 Joint p.m.fs

1.3 Expectation

In this chapter, we will collect facts that will be used in the upcoming chapters.

Show sum of finite number of i.i.d Bernoullis is Binomial.

Proposition 1.7. If we pick disjoint subsets of T and sum the values of i.i.d process on each subset of T , the resulting set of random variables is still independent.

※ Introduction to Stochastic Processes

When we look at the world around us, we see interesting phenomenon that changes over time. Often the variations in these processes have unknown or random causes. The theory of probability is very useful as a tool to analyze such processes and study of these processes using the tools of probability theory is called the theory of *stochastic processes*.

Stochastic processes are crucial in economics and finance, capturing the inherent randomness in markets and economic activities. From modeling stock prices using geometric Brownian motion to predicting economic growth through random walks, these processes provide powerful tools to understand and forecast complex systems. Landmark results, such as the Black-Scholes-Merton model for option pricing and the Efficient Market Hypothesis, deeply rely on stochastic processes. These models not only offer profound insights but also drive practical applications, underscoring the natural and indispensable role of randomness in economic and financial analysis.¹

Let us start with the definition of a stochastic process.

¹Thanks ChatGPT.

Definition 2.1. A stochastic process $\{X_t\}_{t \in T}$ is a collection of random variables indexed by an *indexing set* T . The parameter t is generally called *time*.

Please note that to specify a stochastic process, you have to specify a set T (indexing set) and then define a collection of random variables X_t for each t in T . In order to understand this definition, we will now see many natural examples.

Example 2.2. Suppose a coin is tossed forever. Let $T = \mathbb{N}$ denote the set of all natural numbers. Let N_k denote the number of tails in the k th toss for each natural number k . Note that N_k is either 0 or 1. Now $\{N_k\}_{k \in T}$ is a stochastic process. All the random variables that make up this process are independent and identically distributed. So this is an example of an *i.i.d process*.

Example 2.3. Suppose a coin is tossed forever. Let $T = \mathbb{N}$ denote the set of all natural numbers. Let S_k denote the number of tails in the first k tosses for each natural number k . Note that $S_k \in \{0, 1, 2, 3, \dots, k\}$. Now $\{S_k\}_{k \in T}$ is a stochastic process. This stochastic process is an example of a *discrete time Markov chain* which is the focus of Section 3. Note that this stochastic process takes discrete values (hence ‘chain’) and the indexing set is also discrete.

This example relates to the previous example since $S_k = N_1 + N_2 + \dots + N_k$.

Example 2.4. Suppose we stand at the MSE entrance and keep track of the ACs that are taken out of MSE for repair. Let $T = [0, \infty)$ denote the set of all non-negative real numbers. Let A_t denote the number of ACs that have gone for repair in the interval $[0, t]$ for each $t \in T$ with $A_0 = 0$. Note that A_t is a non-negative integer. Now $\{A_t\}_{t \in T}$ is a stochastic process.

This stochastic process is an example of a *continuous time Markov chain*. This stochastic process takes discrete values (hence ‘chain’) but the indexing set is not discrete. We will study this in detail in Section 5.

Interestingly, we can obtain another stochastic process from $\{A_t\}_{t \in T}$. If we assume that atmost 1 AC is carried out at any instant of time, then define T_k as the time elapsed between the k th AC and $k+1$ th AC being carried out. This produces a stochastic process of ‘waiting times’ $\{T_k\}_{k \in \mathbb{N}}$. If the stochastic process of waiting times is an i.i.d process with an exponential distribution, then the continuous time Markov chain is called a *Poisson process* and it is the focus of the Section 4.

Example 2.5. Let T be the set of natural numbers. A gambler’s fortune X_n after playing the n th betting game gives us a stochastic process $\{X_n\}_{n \in T}$. If we assume all the betting games are fair games, then we get a *Martingale* process which will be studied in 6.

Example 2.6. Suppose you are very interested in a particular stock. Let $T = [0, \infty)$ denote the set of all non-negative real numbers. Let P_t denote the stock price at time t for each $t \in T$. Note that P_t will be considered to be a real number and we will plot the stock price chart as a continuous function (when considered over very long periods of time). Now $\{P_t\}_{t \in T}$ is a stochastic process.

Recall that while studying the theory of probability, we start with an experiment and then collect all possible outcomes of the experiment into a set called the *sample space*, denoted by Ω . An outcome is usually denoted by the greek letter ω . Practically speaking, these outcomes are randomly realized when we perform the experiment. We measure how random the outcomes are by assigning probabilities to these outcomes (either intuitively or guided by statistical inferences from real world data). Any function of these random outcomes will be random and are usually termed as *random variables*. In terms of mathematical symbols, a function $X : \Omega \rightarrow \mathbb{R}$ is a random variable.

The examples above contained a real-life experiment we wished to model. However, it might only sometimes be possible to list the experiment's outcomes in the above examples. For instance, in the case of stock prices, the experiment is the *the processes of the entire world!!*. The outcomes of this experiment could range anywhere from individual emotions to global policies. It is not clear what the sample space would be in this case. At any rate, that should not stop us from studying the behaviours of stocks. In this course, we often ignore the sample space point of view for computation purposes and work directly with random variables.

However from the point of view of conceptual understanding, the sample space point of view clarifies the nature of a stochastic process. So let us consider Example 2.3. Since, in this case, it is not hard to describe the sample space, it is a useful example to understand the nature of a stochastic process.

Example 2.7. Suppose a coin is tossed forever. In this case, the a random outcome of this experiment is an infinite list of heads and tails. Let us write collect all these outcomes and write the set formally as

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots) | \omega_i \in \{H, T\} \text{ for all } i\}.$$

An example outcome is an infinite string of tosses

$$\omega = (H, T, T, H, T, H, T, T, T, \dots) = HTTHTHTTTT\dots$$

where all the tosses are tails from the seventh toss onwards. Note that we also suppress the commas and the parantheses and simply write the infinite list as an infinite string.

Let $T = \mathbb{N}$ denote the set of all natural numbers. Let S_k denote the number of tails in the first k tosses for each natural number k . Clearly each S_k is a random variable, i.e. $S_k : \Omega \rightarrow \mathbb{R}$. For instance, consider the outcome ω considered earlier and fix k as 10. The value of

$$S_{10}(\omega) = S_{10}(HTTHTHTTTT \dots) = 10.$$

In fact, can you see that

$$S_k(\omega) = k - 3 \quad \text{for } k \geq 7?$$

So we see that the stochastic process S is a function of time (k) and the random outcome (ω) of the experiment. All the sources of randomness is captured in ω . Once you fix ω , the sequence $\{S_k(\omega)\}_{k \in \mathbb{N}}$ is a deterministic process. In our example if we fix $\omega = (H, T, T, H, T, H, T, T, T, T, \dots) = HTTHTHTTTT \dots$, the sequence is $(0, 1, 2, 2, 3, 3, 4, 5, 6, 7 \dots)$.

The summary of the discussion is that if you fix omega, the stochastic process is a deterministic process and if you fix time index k the stochastic process at that instant of time returns a random variable S_k .

Stochastic processes studied in this text have nice properties described below.

Definition 2.8.

1. Given a stochastic process $\{X_t\}_{t \in T}$, a difference random variable $X_b - X_a$ for $b > a$ (in T) is called an *increment*. The increment is the net change in the value of the process in the interval $[a, b]$ contained in T . We say that $X_b - X_a$ is *the increment associated to the interval $[a, b]$* .
2. A stochastic process $\{X_t\}_{t \in T}$ has the *stationary increment property* if the distribution of the increment random variable associated to any interval $[t + h, t]$ in T is purely a function of h and does not depend on t . In other words, the distribution $X_{t+h} - X_t$ depends only on h and not on t .
3. A collection of intervals of the form $[t_0, t_1], [t_1, t_2], [t_2, t_3], \dots, [t_{k-1}, t_k]$ for some integer k will be called *successive*.
4. A stochastic process $\{X_t\}_{t \in T}$ has the *independent increment property* if the increments associated to for any collection of successive intervals are independent random variables.

Let us see an immediate example and a non-example.

Example 2.9. The stochastic process $\{S_k\}_{k \in \mathbb{N}}$ in Example 2.3 has both stationary and independent increments. Let us prove these claims.

First, we show that the process has *stationary increments*. Since time is discrete we will use the standard symbol k instead of t . For $k, h \in \mathbb{N}$, consider the distribution of

$$S_{k+h} - S_k = N_{k+1} + N_{k+2} + \dots + N_{k+h}.$$

Since N_k form a i.i.d Bernoulli(p) process, the sum of h random variables is distributed as Binomial(h, p). Thus the r.v. $S_{k+h} - S_k$ is distributed as Binomial(h, p). Clearly this is independent of k . Thus we have stationary increments.

Next, we show that the process has *independent increments*. Since time is discrete we will use the standard symbol k instead of t . For any collection of successive intervals $[k_0, k_1], [k_1, k_2], [k_2, k_3], \dots, [k_{l-1}, k_l]$, we have to show that increments associated to the intervals are independent random variables. But note that the increment associated to the interval $[k_0, k_1]$ is

$$S_{k_1} - S_{k_0} = N_{k_0+1} + N_{k_0+2} + \dots + N_{k_1}.$$

Note that this increment is the sum of the independent random variables N_i in the interval $(k_0, k_1]$. Since the successive intervals $(k_i, k_{i+1}]$ are disjoint, the increments in the successive intervals are sum of i.i.d processes on disjoint subsets of T . So by Proposition 1.7, the increments are independent.

Example 2.10. The stochastic process $\{N_k\}_{k \in \mathbb{N}}$ in Example 2.2 has stationary increments and but not independent increments. Let us prove these claims.

First, we show that the process has *stationary increments*. Since time is discrete we will use the standard symbol k instead of t . For $k, h \in \mathbb{N}$, consider the distribution of

$$D = N_{k+h} - N_k.$$

Since N is an i.i.d Bernoulli(p) process, the difference random variable has a pmf supported on $\{1, 0, -1\}$ and the pmf is

$$p_D(1) = \Pr\{N_{k+h} = 1, N_k = 0\} = p(1-p) = p_D(-1), p_D(0) = p^2 + (1-p)^2.$$

Thus the r.v. $S_{k+h} - S_k$ is distributed as Binomial(h, p). Clearly this is independent of k . Thus we have stationary increments.

Next, we show that the process has *independent increments*. Since time is discrete we will use the standard symbol k instead of t . For any collection of successive intervals $[k_0, k_1], [k_1, k_2], [k_2, k_3], \dots, [k_{l-1}, k_l]$, we have to show that increments associated to the intervals are independent random variables. But note that the increment associated to the interval $[k_0, k_1]$ is

$$S_{k_1} - S_{k_0} = N_{k_0+1} + N_{k_0+2} + \dots + N_{k_1}.$$

Note that this increment is the sum of the independent random variables N_i in the interval $(k_0, k_1]$. Since the successive intervals $(k_i, k_{i+1}]$ are disjoint, the increments in the successive

intervals are sum of i.i.d processes on disjoint subsets of T . So by Proposition 1.7, the increments are independent.

Suppose $N(t)$ is a process valued in whole numbers and $t \in \mathbb{R}$. Further if $N(t)$ has independent increment property and the distribution of the increment $N(t + s) - N(t) \sim \text{Poisson}(\lambda s)$, then $N(t)$ is called a *Poisson point process*. Note that this means $N(\omega)$ is an increasing piecewise constant function.

Suppose $W(t)$ is a process valued in real numbers, $t \in \mathbb{R}$ and $W(\omega)$ is a continuous function. Further if $W(t)$ has independent increment property and the distribution of the increment $W(t + s) - W(t) \sim \text{Normal}(0, s)$, then $W(t)$ is called a *Weiner process*.

※ Discrete time Markov Chains

For the purposes of this chapter, a discrete set is a set whose elements can be written as a finite or infinite list. For instance, $\mathbb{Z}, \mathbb{Z}_{\geq 0}, \mathbb{N}$ are discrete sets. ²

Definition 3.1. A stochastic process $\{X_n\}_{n \in \mathbb{W}}$ with X_n valued in a discrete set D is called a *discrete time Markov chain* (DTMC) if

$$\Pr\{X_n = a_n \mid X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \dots, X_1 = a_{n-1}\} = \Pr\{X_n = a_n \mid X_{n-1} = a_{n-1}\}$$

for all whole numbers n and for all $a_1, a_2, \dots, a_{n-1}, a_n \in D$.

A probability of the form $\Pr\{X_n = a_n \mid X_{n-1} = a_{n-1}\}$ is called a *transition probability*. If $\Pr\{X_n = i \mid X_{n-1} = j\}$ does not depend on n , then we say that the DTMC is homogenous.

We will only consider homogenous DTMCs. We start with a nice example of a DTMC.

Example 3.2. You walk into a casino with Rs.2000 to play a fair betting game. You can win or lose Rs. 1000 per game. You plan to play the game until you hit Rs.3000 for the first time and then take your winnings. Let X_n denote the amount of money you have at the end of the n th game. Note that the range of the random variable X_n is $\{0, 1000, 2000, 3000\}$. Note that outcome every game is independent of outcomes of the rest of the games.

$$\begin{aligned} \Pr\{X_n = a_n \mid X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \dots\} &= \frac{\Pr\{X_n = a_n, X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \dots\}}{\Pr\{X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \dots\}} \\ &= \frac{\Pr\{X_n = a_n, X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \dots\}}{\Pr\{X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \dots\}} \end{aligned}$$

We can represent this data by a 'chain'. /*insert figure*/

Example 3.3. Suppose a coin is tossed forever. Let S_k denote the number of tails in the first k tosses for each natural number k . Note that S_k is a whole number. Now we prove the stochastic process $\{S_k\}_{k \in \mathbb{T}}$ is a Markov chain. Let N_k denote the number of tails in the k th toss and let $S_k = N_1 + N_2 + \dots + N_k$.

Definition 3.4. Let $\{X_n\}$ be a homogenous DTMC with countable, ordered set S of states. Let n be any natural number, define the n -step transition probability matrix $P^{(n)}$, whose rows and columns are labelled by S in the same order, as

$$P_{ij}^{(n)} := \Pr\{X_{k+n} = i \mid X_k = j\}.$$

We will denote the one-step transition probability matrix (t.p.m) by $P = P^{(1)}$.

²The mathematically accurate terminology is 'countable sets'. The precise definition is: T is a countable set if T is in bijection with natural numbers.

Theorem 3.5. Let $\{X_n\}$ be a DTMC with set of ordered set of states S and t.p.m P . Then

$$P^{(n)} = P^n.$$

where P^n is the matrix P raised to the power of n .

Proof. By definition $P^{(1)} = P$, suppose we have shown $P^{(n-1)} = P^{n-1}$ We will show the ij th entry match on both sides.

$$P_{ij}^{(n)} = \Pr\{X_n = i \mid X_0 = j\}$$

□

※ Poisson point process

Definition 4.1. Let $\{N(t) \mid t \in [0, \infty)\}$ be a stochastic process with $N(t) \in \{0, 1, 2, 3, \dots\}$. The stochastic process $N(t)$ is called a *Poisson point process* if

1. $N(0) = 0$.
2. $N(t + s) - N(s) \sim \text{Poisson}(\lambda t)$ for some positive real number λ and all reals s, t .
3. $N(t)$ has the independent increment property.

Note that the Poisson point process is a continuous time process but $N(t)$ takes values from a discrete set.

Definition 4.2. Let T_i denote the first time t when $N(t) = i$ for $i = 0, 1, 2, 3, \dots$ and the random variables are called *time of arrivals*. The random variables $W_{i+1} := T_{i+1} - T_i$, for $i = 0, 1, 2, 3, \dots$ are called *interarrival times* (or *waiting times*).

The following theorem establishes few basic properties of Poisson point processes.

Theorem 4.3. The following facts are immediate:

1. $N(t)$ has the stationary increment property.
2. Let n be a natural number. The interarrival times $\{W_1, W_2, W_3, \dots, W_n\}$ are independent random variable and each random variable has $\text{Exponential}(\lambda)$ as its distribution.
3. The distribution of 'time of arrival' T_n is $\text{Erlang}(n, \lambda)$.

Proof.

1. Since by second property in the definition of Poisson process $N(t + s) - N(s) \sim \text{Poisson}(\lambda t)$, we see that the distribution of the increment is not dependent on s .
2. We will do it for the case of two variables W_1, W_2 . The general proof is similar. Clearly the interarrival times are non-negative valued. So we will consider w_1, w_2 as non-negative real numbers.

$$\begin{aligned}
 \Pr\{W_2 > w_2 \mid W_1 = w_1\} &= \Pr\{T_2 > T_1 + w_2 \mid T_1 = w_1\} \\
 &= \Pr\{T_2 > w_1 + w_2 \mid T_1 = w_1\} \\
 &= \Pr\{N(w_1 + w_2) - N(w_1) = 0 \mid N(w_1) - N(0) = 1, N(t) - N(0) = 0 \text{ for } t < w_1\} \\
 &= \Pr\{N(w_1 + w_2) - N(w_1) = 0\}
 \end{aligned}$$

$$= e^{-\lambda w_2}$$

The first equality follows from definition of interarrival times and the third equality follows from the definition of arrival times. The fourth equality follows from independent increment property since interval $(w_1, w_1 + w_2]$ is disjoint from $(0, t]$ for all $t \leq w_1$. The fifth equality follows from the second property of the Poisson process, i.e. the distribution of $N(w_1 + w_2) - N(w_1)$ is $\text{Poisson}(\lambda w_2)$.

Now we note that

$$\Pr\{W_2 > w_2 \mid W_1 = w_1\} = e^{-\lambda w_2}.$$

Thus W_1, W_2 are independent random variables and further

$$\Pr\{W_2 > w_2\} = e^{-\lambda w_2}$$

for $w_2 > 0$. Thus $W_2 \sim \text{Exponential}(\lambda)$.

A similar calculation shows

$$\begin{aligned} \Pr\{W_1 > w_1\} &= \Pr\{T_1 > w_1\} \\ &= \Pr\{N(w_1) - N(0) = 0\} \\ &= e^{-\lambda w_1} \end{aligned}$$

Thus $W_1 \sim \text{Exponential}(\lambda)$.

3. Since $T_n = W_n + W_{n-1} + \dots + W_1$ and the sum of n i.i.d $\text{Exponential}(\lambda)$ variables is $\text{Erlang}(n, \lambda)$. Note that the pdf of $\text{Erlang}(n, \lambda)$ is

$$f(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{n-1}}{(n-1)!} \quad \text{for } x > 0.$$

□

The Erlang distribution is the special case of the Gamma distribution. You are supposed to memorise the p.d.f of Gamma distribution. You should know the mean, median, variance and mode of the Gamma distribution.

It turns out that any calculation in Poisson process theory is executed by rewriting events in terms of increments of disjoint intervals. So let us setup some notations and a dictionary to aid our calculations:

Definition 4.4. Let $N(t)$ be a Poisson point process with rate λ , then for $0 < a < b < c < d < \dots$ let us write

$$N[a, b, c, d, e \dots] = [p, q, r, s \dots]$$

Events in terms of (inter-)arrival times and process	Events in terms of increments of disjoint intervals
$\{T_1 > t_1\}$	$\{N[0, t_1] = [0]\}$
$\{T_1 = t_1, N(t) = 2\}$	$\{N[0, s, t_1, t] = [0, 1, 1] \text{ for all } s < t_1\}$
$\{t_1 \leq T_1 < t_1 + dt_1, t_2 \leq T_2 < t_2 + dt_2, N(t) = 3\}$	$\{N[0, t_1, t_1 + dt_1, t_2, t_2 + dt_2, t] = [0, 1, 0, 1, 1]\}$

Table 1: Few examples of moving between equivalent events.

to mean

$$N(b) - N(a) = p$$

$$N(c) - N(b) = q$$

$$N(d) - N(c) = r$$

$$N(e) - N(d) = s$$

$$. = .$$

$$. = .$$

A dictionary of equivalent events that translates events of arrival times to events of independent increments of the Poisson process is shown in Table 1.

In order to cut down on rigorous mathematical arguments, we will use the following (intuitive) formulae involving infinitesimals (dx, dt, \dots):

1. Let X be a random variable with p.d.f f , then

$$f(x) dx = \Pr\{x < X \leq x + dx\}.$$

2. Let X_1, X_2, \dots, X_n be a collection of n random variables with joint p.d.f f , then

$$f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \Pr\{x_1 < X_1 \leq x_1 + dx_1, x_2 < X_2 \leq x_2 + dx_2, \dots, x_n < X_n \leq x_n + dx_n\}.$$

3. We will use $f(t) + dt = f(t)$ only if $f(t) \neq 0$ and $dt + O(dt^2) = dt$. For example $e^{-\lambda dt} = 1 - \lambda dt + O(dt^2) = 1 - \lambda dt = 1$

Remark 4.5. Strictly speaking the above formulae are ridiculous and can easily confuse a careless novice. So one needs to take care while using infinitesimals. Learn from discussions in class and chat with TAs if you are confused. If at any step you feel worried, then go back to computing cdfs and differentiate cdfs to get pdfs.

Let us see an example of how to use the above intuitive rules.

Example 4.6. Let $N(t)$ be a Poisson process with rate λ , then the joint distribution of arrival times T_1, T_2 is computed below. Let $0 < t_1 < t_2$, then

$$\begin{aligned}
 f_{T_1, T_2}(t_1, t_2) dt_1 dt_2 &= \Pr\{t_1 < T_1 \leq t_1 + dt_1, t_2 < T_2 \leq t_2 + dt_2\} \\
 &= \Pr\{N[0, t_1, t_1 + dt_1, t_2, t_2 + dt_2] = [0, 1, 0, 1]\} \\
 &= (e^{-\lambda t_1})(e^{-\lambda dt_1} \lambda dt_1)(e^{-\lambda(t_2 - t_1 - dt_1)})(e^{-\lambda dt_2} \lambda dt_2) \\
 &= (e^{-\lambda t_1})(e^{-\lambda dt_1} \lambda dt_1)(e^{-\lambda(t_2 - t_1 - dt_1)})(e^{-\lambda dt_2} \lambda dt_2) \\
 &= e^{-\lambda t_2} \lambda^2 dt_1 dt_2.
 \end{aligned}$$

So comparing we see that for $0 < t_1 < t_2$, we have

$$f_{T_1, T_2}(t_1, t_2) = e^{-\lambda t_2} \lambda^2.$$

The joint pdf vanishes for any other pair t_1, t_2 .

Here is a list of exercises for you to practice.

Lemma 4.7. Let $N(t)$ be a Poisson process with rate λ , then

$$\begin{aligned}
 \Pr\{N[0, t_1, t_1 + dt_1] = [0, 1]\} &= e^{-\lambda t_1} \lambda dt_1 \\
 \Pr\{N[0, t_1, t_1 + dt_1, t] = [0, 1, 0]\} &= e^{-\lambda t} \lambda dt_1 \\
 \Pr\{N[0, t_1, t_1 + dt_1, t] = [0, 1, 1]\} &= e^{-\lambda t} \lambda^2 (t - t_1) dt_1 \\
 \Pr\{N[0, t_1, t_1 + dt_1, t_2, t_2 + dt_2] = [0, 1, 0, 1]\} &= e^{-\lambda t_2} \lambda^2 dt_1, dt_2 \\
 \Pr\{N[0, t_1, t_1 + dt_1, t_2, t_2 + dt_2, t] = [0, 1, 0, 1, 0]\} &= e^{-\lambda t} \lambda^2 dt_1, dt_2 \\
 \Pr\{N[0, t_1, t_1 + dt_1, t_2, t_2 + dt_2, t] = [0, 1, 0, 1, 1]\} &= e^{-\lambda t} \lambda^3 (t - t_2) dt_1, dt_2
 \end{aligned}$$

For $k > 1$ $\Pr\{N[dt] = [k]\} = 0$.

Example 4.8. Let us calculate a couple of related distributions.

1. Let $n \geq 1$. The conditional p.d.f

$$f_{T_1|N(t)}(s | N(t) = n) = \frac{n}{t} \left(1 - \frac{s}{t}\right)^{n-1}$$

for $0 < s < t$ and 0 otherwise. This tells us that if it is known a point has arrived in the interval $[0, t]$, then the arrival time is uniformly distributed in the interval $[0, t]$.

2. The conditional p.d.f

$$f_{T_2|T_1, N(t)}(t_2 | T_1 = t_1, N(t) = n) = \frac{n-1}{t-t_1} \left(1 - \frac{t_2 - t_1}{t - t_1}\right)^{n-1}, \quad t_1 < t < t_2$$

for $t_1 < t_2 < t$ and 0 otherwise.

Proof. 1. Assume $0 < s < t$,

$$\begin{aligned}
 f_{T_1|N(t)}(s \mid N(t) = n) ds &= \Pr\{s < T_1 < s + ds \mid N(t) = n\} \\
 &= \frac{\Pr\{s < T_1 < s + ds, N(t) = n\}}{\Pr\{N(t) = n\}} \\
 &= \frac{\Pr\{N[0, s, s + ds, t] = [0, 1, n - 1]\}}{\Pr\{N(t) = n\}} \\
 &= \frac{e^{-\lambda t} (\lambda ds)^{\frac{(\lambda(t-s))^{n-1}}{(n-1)!}}}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} \\
 &= \frac{n}{t} \left(1 - \frac{s}{t}\right)^{n-1} ds
 \end{aligned}$$

Clearly if $s > t$,

$$f_{T_1|N(t)}(s \mid N(t) = 1) ds = \Pr\{s < T_1 < s + ds \mid N(t) = 1\} = 0.$$

2. Assume $t_1 < t_2 < t$,

$$\begin{aligned}
 f_{T_2|T_1, N(t)}(t_2 \mid T_1 = t_1, N(t) = n) dt_2 &= \Pr\{t_2 < T_2 < t_2 + dt_2 \mid T_1 = t_1, N(t) = n\} \\
 &= \frac{\Pr\{t_2 < T_2 < t_2 + dt_2, N(t) = n \mid T_1 = t_1\}}{\Pr\{N(t) = n \mid T_1 = t_1\}} \\
 &= \frac{\Pr\{t_2 < T_2 < t_2 + dt_2, N(t) - N(t_1) = n - 1 \mid T_1 = t_1\}}{\Pr\{N(t) - N(t_1) = n - 1 \mid T_1 = t_1\}} \\
 &= \frac{\Pr\{t_2 < T_2 < t_2 + dt_2, N(t) - N(t_1) = n - 1\}}{\Pr\{N(t) - N(t_1) = n - 1\}} \\
 &= \frac{\Pr\{N[t_1, t_2, t_2 + dt_2, t] = [0, 1, n - 2]\}}{\Pr\{N(t) - N(t_1) = n - 1\}} \\
 &= \frac{e^{-\lambda(t-t_1)} (\lambda dt_2)^{\frac{(\lambda(t-t_2))^{n-2}}{(n-2)!}}}{e^{-\lambda(t-t_1)} \frac{(\lambda(t-t_1))^{n-1}}{(n-1)!}} \\
 &= \frac{n-1}{t-t_1} \left(1 - \frac{t_2-t_1}{t-t_1}\right)^{n-2} dt_2
 \end{aligned}$$

□

From the above examples, we see that for T_1 is uniformly distributed in $[0, t]$ given $N(t) = 1$. If it is given that $N(t) \geq 2$, then the first arrival time T_1 is not uniformly distributed in $[0, t]$.

Exercise 1. A couple of exercises for the reader:

1. Show that

$$f_{T_1|N(t)}(s \mid N(t) = 0) = \lambda e^{-\lambda(s-t)} \text{ for } s \geq t$$

and it is zero elsewhere.

2. Show that for $n > 0$,

$$\mathbb{E}(T_1 \mid N(t) = n) = \frac{t}{n+1}.$$

3. Compute

$$\mathbb{E}(T_1 \mid N(t) = 0).$$

4. Compute $\mathbb{E}(T_2 \mid T_1, N(t))$.

Supposing we knew that $N(t) = n$, then what is the distribution of the arrival times T_1, T_2, \dots, T_n ? The answer is interesting.

Theorem 4.9. The conditional pdf of interarrival times T_1, T_2, \dots, T_n given $N(t) = n$ is given by: For $0 < t_1 < t_2 < \dots < t_n < t$

$$f(t_1, t_2, \dots, t_n \mid N(t) = n) = \frac{n!}{t^n}$$

Elsewhere the pdf is zero. Note the pdf is constant (or uniform) in the region $0 < t_1 < t_2 < \dots < t_n < t$.

Using the techniques shown in the previous examples, this above theorem is easily proved.

A remarkable property of the Poisson process is that if split the arrivals of a Poisson process by flipping a biased coin, the two resulting processes are independent Poisson processes!! I don't know any other stochastic processes with this property. So if you want extra credit, then think about classifying all stochastic processes with this property.

Theorem 4.10. Let $N(t)$ be a Poisson process with rate λ , suppose every arrival is labelled 1 with probability p and 0 with probability $1 - p$. Suppose this each arrival is independently labelled. Let $N_1(t)$ denote the number of points labelled 1 in $[0, t]$, and $N_0(t)$ denote the number of points labelled 0 in $[0, t]$. Then $N_0(t), N_1(t)$ are independent Poisson processes with rates $p\lambda$ and $(1 - p)\lambda$.

Proof. It is clear that $N_0(t) + N_1(t) = N(t)$. We will use this fact freely.

Suppose there are totally n arrivals. The probability that k of them are labelled 1 and $n - k$ are labelled 0 is given by the Binomial distribution:

$$\Pr\{N_1(t) = k, N_0(t) = n - k \mid N(t) = n\} = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\begin{aligned}
\frac{\Pr\{N_1(t) = k, N_0(t) = n - k, N(t) = n\}}{\Pr\{N(t) = n\}} &= \binom{n}{k} p^k (1-p)^{n-k} \\
\frac{\Pr\{N_1(t) = k, N_0(t) = n - k\}}{\Pr\{N(t) = n\}} &= \binom{n}{k} p^k (1-p)^{n-k} \\
\Pr\{N_1(t) = k, N_0(t) = n - k\} &= \binom{n}{k} p^k (1-p)^{n-k} \Pr\{N(t) = n\} \\
\Pr\{N_1(t) = k, N_0(t) = n - k\} &= \left(\frac{n!}{k!(n-k)!} \right) p^k (1-p)^{n-k} \left(\frac{e^{-\lambda t} (\lambda t)^n}{n!} \right) \\
\Pr\{N_1(t) = k, N_0(t) = n - k\} &= \left(\frac{e^{-p\lambda t} (p\lambda t)^k}{k!} \right) \left(\frac{e^{-(1-p)\lambda t} ((1-p)\lambda t)^{n-k}}{(n-k)!} \right)
\end{aligned}$$

Thus for every real $t > 0$, $N_0(t), N_1(t)$ are independent random variables and they are distributed as $\text{Poisson}(p\lambda)$ and $\text{Poisson}((1-p)\lambda)$ respectively. \square

Another cool property of Poisson processes is that sum of two independent Poisson process is a Poisson process with sum of the rates. This will be an exercise problem.

4.1 Cramer-Lundberg Ruin theory

The following section is an selective adaptation of [?] and we advise the reader to refer to those notes. What follows is a poor imitation of those notes and at times I have shamelessly lifted a few sentences from those notes.

In a *collective risk model* there are a number of contracts for risks, like insurance against thefts, accidents, natural calamities and other adverse events. A company might be selling a collection of such contracts and we are interested in avoiding the ruin of the insurance company.

Assume that the number of claims that appear in the interval $[0, t]$ is denoted by $N(t)$. The i th claim that appears at time T_i causes a payment of X_i (called the i th claim size). We assume that the collection of claim sizes $\{X_i | i \in \mathbb{Z}_{\geq 0}\}$ are independent random variables and further, we also assume that the random variables of stochastic process $N(t)$ are independent of the claim sizes $\{X_i | i \in \mathbb{Z}_{\geq 0}\}$.

Let $t \geq 0$. The aggregate claim amount paid by the company in the interval $[0, t]$ is given by

$$S(t) := \sum_{i=1}^{N(t)} X_i.$$

Expand on this exercise:

1. Let r be the fixed interest rate in the market. Define the discounted sum:

$$S_0(t) := \sum_{i=1}^{N(t)} e^{-rT_i} X_i.$$

Explain the meaning of $S_0(t)$ and the significance of this quantity.

2. Assume $N(t)$ is a Poisson point process with rate λ . If the average claim size is \bar{X} . Show that the expectation of $S_0(t)$ is

$$\mathbb{E}(S_0(t)) = \frac{\lambda}{r}(1 - e^{-rt})\bar{X}.$$

3. Let $p(t)$ denote the premium income in the time interval $[0, t]$ and suppose the company had an initial capital of u units. Then the company's capital balance function is defined as

$$U(t) = u + p(t) - S(t).$$

What is the average value of $U(t)$?

The first time t such that $U(t) < 0$ is called the ruin time. **Discuss the main theorem of ruin after introducing martingales?**

※ Continuous time Markov Chains

※ Martingales

References

- [1] S. Ramasubramanian, *Poisson Process and Insurance: An Introduction*, 2005. Available at: <https://www.isibang.ac.in/~statmath/ram/ASPDelhi05.pdf>.