Applied Stochastic Processes

2018A7PS0193P

April 22, 2021

Chapter 1

Fundamentals

1.1 Stochastic Processes

Definition 1. A stochastic process is a probability model that describes the evolution of a system evolving randomly in time.

Definition 2. A random variable is a mapping $X : \Omega \to \mathbb{R}$ that assigns a real number $X(\omega)$ to each outcome $\omega \in \Omega$, where Ω is the sample space.

A stochastic process can be given by a collection of random variables $\{X(t), t \in T\}$, where T is called the **index set**. If T is countable (observed at discrete times), we get a **discrete time stochastic process**. On the other hand, if T is uncountable (observed continuously), then we get a **continuous time stochastic process**.

Definition 3. The state space of a stochastic process is defined as the set of all possible values that the random variables $X(t)$ can assume.

1.2 Elementary Probability

For a recap of elementary probability, refer to the notes from Applied Statistical Methods.

1.3 Transformation of Random Variables

> **Lemma 1.1.** Let X have a continuous, strictly increasing CDF F. Let $U \sim$ Uniform(0,1). If $Y = F^{-1}(U)$, then Y also has the CDF F.

The lemma above allows us to transform U to any other random variable, as long as it has a continuous and strictly increasing CDF. Say we had an algorithm to define a uniform random variable in the range (0,1), now we have a way to generate random variables from a different distribution.

1.4 Moment Generating Functions

Definition 4. The moment generating function $\phi(t)$ of the random variable X is defined for all values t as $\phi(t) = E[e^{tx}]$.

The moment generating functions of some oft-used distributions are as follows:

- Moment generating function of Binomial(n, p) is:

$$\phi(t) = (pe^t + 1 - p)^n$$

- Moment generating function of Poisson(λ) is:

$$\phi(t) = e^{\lambda(e^t - 1)}$$

- Moment generating function of Exponential(λ) is:

$$\phi(t) = \frac{\lambda}{\lambda - t}$$

- Moment generating function for Normal(μ, σ^2) is:

$$e^{t\mu + \frac{1}{2}\sigma^2 t^2}$$

- Moment generating function for Uniform(a, b) is :

$$\frac{e^{tb} - e^{ta}}{t(b - a)}$$

> **Theorem 1.2.** The moment generating function of the sum of independent random variables is the product of the individual moment generating functions.

So, this means that $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$, as long as $X \perp Y$ (this notation means that they are independent).

1.5 Conditional distributions

The conditional probability distribution of Y given the occurrence of the value x of X is given by:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

where $f_{X,Y}(x,y)$ is the joint distribution and $f_X(x)$ is the marginal density of X.

The conditional expectation of X given Y is:

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x,y)dx$$

1.6 Markov's and Chebyshev's Inequality

> **Theorem 1.3** (Markov's Inequality). Let X be a non-negative random variable and suppose that $E(X)$ exists. For any $t > 0$,
>
> $$P(X > t) \leq \frac{E(X)}{t}$$

Proof.
$$E(X) = P(X < a) \cdot E(X|X < a) + P(X \geq a) \cdot E(X|X \geq a)$$

Here, $E(X|X < a)$ is larger than 0 since X is a non negative random variable. $E(X|X \geq a)$ is larger than a, since it only considers values larger than a. Hence,

$$E(X) \geq P(X < a) \cdot 0 + P(X \geq a) \cdot E(X|X \geq a) \geq a \cdot P(X \geq a)$$

This directly leads to Markov's Inequality. □

Theorem 1.4 (Chebyshev's Inequality). Let $\mu = E(X)$ and $\sigma^2 = V(X)$. Then

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

$$P(|Z| \geq k) \leq \frac{1}{k^2}$$

where $Z = (X - \mu)/\sigma$ and $t > 0$.

Proof.

$$
\begin{aligned}
\sigma^2 &= E((X - \mu)^2) \\
&= E((X - \mu)^2 | k\sigma \leq |X - \mu|)P(k\sigma \leq |X - \mu|) + E((X_\mu)^2 | k\sigma > |X - \mu|)P(k\sigma > |X - \mu|) \\
&\geq (k\sigma)^2 P(k\sigma \leq |X - \mu|) + 0 \cdot P(k\sigma > |X - \mu|) \\
&= k^2 \sigma^2 P(k\sigma \leq |X - \mu|)
\end{aligned}
$$

Now the inequality follows from dividing by $k^2 \sigma^2$.   □

Chebyshev's Inequality is a more general version of Markov's Inequality, applicable for any random variable X.

1.7   Convergence of Random Variables

Let $X_1, X_2, ...$ be a sequence of random variables and let X be another random variable. Let F_n be the CDF of X_n and F be the CDF of X.

We say that X_n converges to X in probability, denoted by $X_n \xrightarrow{P} X$, if for every $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \to 0$$

as $n \to \infty$.

We say that X_n converges to X in distribution, denoted by $X_n \rightsquigarrow X$, if

$$\lim_{n \to \infty} F_n(t) = F(t)$$

for all t for which F is continuous.

Lemma 1.5. If $X_n \xrightarrow{P} X$, then $X_n \rightsquigarrow X$.

The above lemma is provided without proof, as it is beyond the scope of the course.

Exercise. Let $X_n \sim N(0, \frac{1}{n})$. Prove that this series converges to

$$F_X(n) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

in distribution.

Solution. Let $t > 0$, and define the standard normal variable $Z_n = \sqrt{n}X_n$, so $Z_n \sim N(0, 1)$. So,

$$\begin{aligned} F_{X_n}(t) &= P(X_n \leq t) \\ &= P(Z_n \leq \sqrt{n}t) \\ &= \int_{-\infty}^{\sqrt{n}t} f(x)dx \end{aligned}$$

where $f(x)$ is the PDF of Z_n.

It is clear that as $n \to \infty$, we get the following distribution:

$$F_{X_n}(t) = \begin{cases} 0 & t < 0 \\ 0.5 & t = 0 \\ 1 & t > 0 \end{cases}$$

So, $F_{X_n} \rightsquigarrow F_X \forall t \in \mathbb{R} - \{0\}$. ∎

Chapter 2

Markov Chains

2.1 Introduction

Definition 5. A stochastic process with a finite number of state spaces $S = \{0, 1, ..., N\}$ and a countable index state $T = \{t_0, t_1, t_2, ...\}$ is a Markov chain if

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0)$$

$$= P(X_{n+1} = j | X_n = i)$$

$$= \mathbf{P}_{ij}$$

This means that the probability of a transition from state i to a state j is completely dependent on i and j, and not on any history. The resulting transition matrix is called the **one step transition matrix**, denoted by $\mathbf{P}$.

> **Lemma 2.1.**
> $$\sum_{j=0}^{\infty} \mathbf{P}_{ij} = 1 \forall i \in S$$

Of course, this matrix means that we could represent a Markov Chain by a digraph, or even Petri Nets (out of syllabus).

2.2   State Probabilities

Let $\mathbf{\Pi}^{(n)}$ be a row vector such that $\mathbf{\Pi}_i^{(n)}$ is the probability that after n transitions, we are at state i. This is known as the **marginal probability distribution**. This vector can be recursively computed by the formula:

$$\mathbf{\Pi}^{(n)} = \mathbf{\Pi}^{(n-1)}\mathbf{P}$$

So, we get

$$\mathbf{\Pi}^{(n)} = \mathbf{\Pi}^{(0)}\mathbf{P}^n$$

$\mathbf{P}^n$ is known as the nth step transition probability matrix (denoted by $p^{(n)}$), where:

$$\mathbf{P}_{ij}^n = p_{ij}^{(n)} = P(X_n = j | X_0 = i)$$

This means that $p^{(n)}$ solves the question - after n transitions starting from i, what is the probability that I will be in state j?

A problem we are facing with this approach is that we have to calculate $\mathbf{P}^n$ fast enough. It is not enough to use binary exponentiation and find it in $O(d^3 \log n)$, since we are using matrices with large size, so while it would be fast in exponent, it would be slow doing matrix multiplication. Instead, we use **Eigendecomposition** (read FDS notes).

Using eigendecomposition, we can decompose $\mathbf{P}$ into:

$$\mathbf{P} = Q\Lambda Q^{-1}$$

Raising it to power of n, we get the decomposition to be:

$$\mathbf{P}^n = Q\Lambda^n Q^{-1}$$

Since Λ is a diagonal matrix, we can find the exponent even faster in $O(d \log n)$, hence speeding up the process.

Theorem 2.2 (Chapman-Kolmogorov Equation)**.**

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)}$$

Proof.

$$p_{ij}^{(n+m)} = P(X_{n+m} = j | X_0 = i)$$

$$= \sum_{k=0}^{\infty} P(X_{n+m} = j, X_n = k | X_0 = i)$$

$$= \sum_{k=0}^{\infty} P(X_{n+m} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i)$$

$$= \sum_{k=0}^{\infty} P_{kj}^m P_{ik}^n$$

$\square$

Theorem 2.3. If $\mathbf{P}$ is a transition matrix for a finite state Markov chain, it has at least one eigenvalue as 1. All the other eigenvalues have an absolute value $|\lambda_i| \leq 1$.

Definition 6. The stationary distribution of a Markov Chain is a row vector $\mathbf{\Pi}$ such that

$$\mathbf{\Pi} \cdot \mathbf{P} = \mathbf{\Pi}$$

So, the stationary distribution $\mathbf{\Pi}$ is the left eigenvector of $\mathbf{P}$ such that it's eigenvalue is 1.

Definition 7. The probability distribution $\mathbf{\Pi}$ is called the limiting distribution of a Markov chain if:

$$\mathbf{\Pi}_j = \lim_{n \to \infty} P(X_n = j | X_0 = i)$$

for all $i, j \in S$ and

$$\sum_{j \in S} \mathbf{\Pi}_j = 1$$

The limiting distribution may not always exist, but if it does, it is equivalent to the stationary distribution. In fact, the limiting distribution does not depend on the start state.

The limiting distribution only exists if $\lim_{n \to \infty} P^n$ has all equal rows, and that row will be equal to the limiting distribution.

Exercise. Consider a Markov chain with two possible states $S = \{0, 1\}$ and the transition matrix:

$$\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$

where $0 < a < 1, 0 < b < 1$. Find the limiting distribution π of this Markov Chain.

Solution. Let $\pi = [\pi_0, \pi_1]$. Then, we can write:

$$\pi = \pi P$$

From this, we get the equation:

$$\pi_0 a = \pi_1 b$$

Since π must be a valid probability distribution,

$$\pi_0 + \pi_1 = 1$$

Thus, we can obtain a unique solution:

$$\pi = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$$

∎

2.3   Occupancy Time and First Entrance Time

Definition 8. Let $N_{ij}^{(n)}$ be the number of times a discrete time Markov Chain visits a state j starting from state i over a given time span of n. The occupancy time for state j starting from i is:

$$T_{ij}^{(n)} = E(N_{ij}^{(n)})$$

Theorem 2.4. The occupancy times matrix $\mathbf{T}^{(n)} = \sum_{k=0}^{n} \mathbf{P}^k$.

Definition 9. Let $f_{ij}^{(n)}$ be the probability the Markov Chain visits a state j for the first time starting from a state i after n transitions. Then f is the First Entrance Time Matrix.

Of course, the probability that a system starting at i will ever reach j is:

$$F_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$$

F_{ij} is called the probability of eventual return.

Theorem 2.5 (First Entrance Theorem)**.**

$$p_{jk}^{(n)} = \sum_{r=0}^{n} f_{jk}^{(r)} p_{kk}^{(n-r)}, n \geq 1$$

where

$$p_{kk}^{(0)} = 1, f_{jk}^{(0)} = 0, f_{jk}^{(1)} = p_{jk}$$

The proof of this is rigorous and uses the strong Markov Property, so it is not included here.

Definition 10. The mean time of eventual return is the expected number of transitions to go from i to j. This is denoted by μ_{ij} and is given by:

$$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}$$

μ_{ii} is called the mean recurrence time. This value is not guaranteed to converge.

2.4 Classification of States

Definition 11 (accessibility)**.** We say that i reaches j (or j is accessible from i) if $P_{ij}^n > 0$ for some n, and we denote it by $i \to j$.

Definition 12 (communicability)**.** If $i \to j$ and $j \to i$, then we write $i \leftrightarrow j$ and we say that i and j communicate.

Theorem 2.6. The communication relation satisfies the following properties:

1. $i \leftrightarrow i$ (reflexive)

2. $i \leftrightarrow j \Rightarrow j \leftrightarrow i$ (symmetric)

3. If $i \leftrightarrow j$ and $\jmath \leftrightarrow k$ then $i \leftrightarrow k$ (transitive)

4. The set of states χ can be written as a disjoint union of classes $\chi = \chi_1 \cup \chi_2 \cup ...$ where two states i and j communicate with each other if and only if they are in the same class. (equivalence class)

It is hence an equivalence relation.

Remark. This has it's own mathematical proof from the definition, but it is far more intuitive to think of the strongly connected components in a directed graph.

If all states communicate with each other, then the chain is said to be **irreducible**. A set of states is **closed** if, once you enter that set of states you never leave. In a more mathematical sense, a set B is closed if for all $j \in B^C$, there is no $i \in B$ such that $i \to j$. A closed set consisting of a single state is called an **absorbing state**.

Definition 13. State i is recurrent or persistent if

$$P(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1$$

Otherwise, the state is transient.

This definition means that if a state is recurrent, then if we start at a state i, we will definitely return to that state after some $n > 0$ steps, i.e, $F_{ii} = 1$. Since this is essentially a recursive process (start from state i, return to state i, start again), the process will re-enter i again and again and again.

However, if a state i is transient, there is some probability $p < 1$ (p is in fact F_{ii}) that it will re-enter the state. Hence, the probability that the process will be in state i for exactly n time periods is:

$$p^{n-1}(1 - p)$$

A recurrent state is **null** if $\mu_{ii} = \infty$. Otherwise it is called **positive** or non null. In a finite state Markov chain, all recurrent states are positive recurrent.

> **Theorem 2.7.** A state i is recurrent if and only if $\sum_n P_{ii}^n = \infty$. A state i is transient if and only if $\sum_n P_{ii}^n < \infty$

Proof. From our discussion, it is clear that a state i is recurrent if and only if the expected number of time periods that the process is in state i is infinite. Let I_n be such that:

$$I_n = \begin{cases} 1, & \text{if } X_n = i \\ 0, & \text{if } X_n \neq i \end{cases}$$

Hence, $\sum_{n=0}^{\infty} I_n$ is the number of periods that a process in a state i.

$$E\left(\sum_{n=0}^{\infty} I_n | X_0 = i\right) = \sum_{n=0}^{\infty} E(I_n | X_0 = i)$$
$$= \sum_{n=0}^{\infty} P(X_n = i | X_0 = i)$$
$$= \sum_{n=0}^{\infty} P_{ii}^n$$

Hence, if the state is recurrent, this value must be infinite. Otherwise, it would be recurrent.

$\square$

Note that this theorem also implies that if i is transient, $P_{ii}^n \to 0$ as $n \to \infty$.

Exercise. Prove that in a finite state Markov Chain, all states cannot be transient

Solution. Let us assume we have a Markov Chain where all states are transient. If this were the case, there would be some finite amount of time after which state i, will never be visited, for every state i. Hence, after some finite time, no states will be visited. But, the process must be in some state, so we have arrived at a contradiction. Hence all states cannot be transient. ∎

Theorem 2.8. If state i is recurrent and $i \leftrightarrow j$, then j is recurrent.

Proof. To prove this, notice that since $i \leftrightarrow j$, there exists integers k and m such that $P_{ij}^k > 0, P_{ji}^m$. For any integer n,

$$P_{jj}^{m+n+k} \geq P_{ji}^m P_{ii}^n P_{ij}^k$$

This is because the LHS is the probability of going from state j to state j in $m+n+k$ steps, but the RHS requires this to occur along a particular path. Summing over all n,

$$\sum_{n=0}^{\infty} P_{jj}^{m+n+k} \geq P_{ji}^m P_{ij}^k \sum_{n=0}^{\infty} P_{ii}^n = \infty$$

This is straightforward from theorem 2.7. From that same theorem, j is recurrent. $\square$

This means that if i is a recurrent state, then if every state in it's equivalence class is also recurrent, i.e. it is a class property.

Theorem 2.9. A set B is closed if and only if $P_{ij} = 0 \forall j \in B^C, i \in B$.

Remark. This comes directly from thinking of the Markov Chain as a graph.

Theorem 2.10. If B is a closed set, then B is the union of equivalence classes in the Markov Chain.

Proof. Assume this is not true. Then, there is some states $i \in B$ and $j \in B^C$ such that $i \leftrightarrow j$. By definition of a closed state, there must be no $j \in B^C$ such that $i \to j$. This generates a contradiction. Hence, $j \in B$. $\qquad \square$

Definition 14. The period of a state i is $d_i = gcd(\{n : P_{ii}^n > 0\})$. If $d_i > 1$, the state is periodic, and is otherwise aperiodic if $d_i = 1$.

Periodicity is a property of a class. Every state in the same class has the same period.

Definition 15. A state is ergodic if it is persistent, non-null and aperiodic. A chain is ergodic if all it's states are ergodic.

Ergodism is a class property, since periodicity and persistency are both class properties.

Theorem 2.11. An irreducible Markov chain always has a limiting distribution.

Exercise. Consider a gambler who at each play of the game has probability p of winning one unit and probability $q = 1 - p$ of losing one unit. Assuming that successive plays of the game are independent , what is the probability that starting with i units, the gambler's fortune will reach N before reaching 0?

Solution. Let $\{X_n, n = \{0, 1, 2, ...\}\}$ be a Markov chain, where X_n is the player's fortune at time n. Then, the transition probability P is given by:

$$P_{00} = P_{NN} = 1$$

$$P_{i,i+1} = p = 1 - P_{i,i-1}, i = 1, 2, ..., N - 1$$

As such, our Markov chain has 3 classes - $\{0\}, \{1, 2, 3, ..., N - 1\}, \{N\}$. The second class is transient, the others are recurrent. Hence, at some point the gambler will go broke or hit the jackpot of N. If P_i is the probability of reaching N from i eventually, then:

$$P_i = pP_{i+1} + qP_{i-1}, i = 1, 2, ..., N - 1$$

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}$$

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1})$$

Of course, $P_0 = 0, P_N = 1$. From our recurrence,

$$P_i - P_{i-1} = \left(\frac{q}{p}\right)^{i-1} P_1$$

Adding these equations over all i,

$$P_i - P_1 = P_1\left[\left(\frac{q}{p}\right) + \ldots + \left(\frac{q}{p}\right)^{i-1}\right]$$

Hence,

$$P_i = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)}P_1 & \frac{q}{p} \neq 1 \\ iP_1 & \text{if } \frac{q}{p} = 1 \end{cases}$$

Using the fact that $P_N = 1$,

$$P_1 = \begin{cases} \frac{1-(q/p)}{1-(q/p)^N} & \frac{q}{p} \neq 1 \\ \frac{1}{N} & \text{if } \frac{q}{p} = 1 \end{cases}$$

Hence,

$$P_i = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)^N} & p \neq \frac{1}{2} \\ \frac{1}{N} & p = \frac{1}{2} \end{cases}$$

We can see that if $p \leq \frac{1}{2}$, the gambler will definitely go broke against an infinitely rich adversary. Otherwise, there is a positive probability of increasing indefinitely. ∎

Chapter 3

Poisson Processes

3.1 Author's Note

Proofs will be omitted in this chapter. For most theorems they are long and rigorous, and would be better understood directly from the textbook than being written here.

3.2 Memorylessness

Definition 16. Let X be a continuous random variable whose values lie in the non negative real numbers. The probability distribution of X is **memoryless** if for any non negative real numbers t and s, we have:

$$P(X > t + s | X > t) = P(X > s)$$

Hence, the distribution of a "waiting time" until a certain event does not depend on how much time has elapsed already.

> **Theorem 3.1.** A positive continuous random variable X is memoryless if and only if $X \sim Exp(x; \lambda)$ for some λ.

3.3 Counting Process

Definition 17. A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ represents the total number of events that occur by time t.

From it's definition,

- $N(t) \geq 0$

- $N(t)$ is integer valued

- If $s < t$ then $N(s) \leq N(t)$

- For $s < t$, $N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$

A counting process has **independent increments** if the numbers of events that occur in disjoint time intervals are independent. A counting process has **stationary increments** if the distribution of the number of events that occur in any time interval only depends on the length of that time interval.

3.4   Poisson Process

Definition 18. The counting process $\{N(t), t\}$ is said to be a Poisson Process having rate $\lambda, \lambda > 0$ if

1. $N(0) = 0$

2. The process has independent increments

3. The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all $s, t \geq 0$

$$P(N(t + s) - N(s) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

   This means it has stationary increments.

The definition of a Poisson process can also be encoded in the form of three postulates, which must be present for a counting process to be a Poisson process:

1. **Independence** : $N(t + h) - N(t)$ is independent of the number of occurrences prior to that interval

2. **Homogeneity in time** : $p_n(t)$ depends only on the length t of the interval and is independent of where the interval is situated, where:

$$p_n(t) = P(N(t + s) - N(s) = n)$$

3. **Regularity** : In an interval of infinitesimal length h, the probability of exactly one occurrence $\lambda h + o(h)$ and that of more than one occurrence is $o(h)$

We can prove that these definitions are equivalent.

Theorem 3.2. The probability distribution of a Poisson Process is:

$$p_n(t) = e^{\lambda t} \frac{(\lambda t)^n}{n!}$$

3.5   Interarrival times

Let Z_i be the time between two occurrences t_{i-1} and t_i. Then Z_i is called the **inter-arrival time**.

Theorem 3.3. Z_i in a Poisson process (with mean λt) are identically independently distributed random variables which follow the negative exponential distribution with mean $1/\lambda$.

The proof of this is beyond the scope of the course. In fact, the converse also holds - if the inter-arrival times follow the negative exponential distribution, then the process must be a Poisson process.

Since inter-arrival times follow an exponential distribution, it also follows that they are memoryless.

Lemma 3.4. If $Z_1, \cdots, Z_n$ are i.i.d with $Z_i \sim Exp(z; \lambda)$ then $S_n = \sum_{i=1}^{n} Z_i$ is distributed according to a Gamma distribution with n and λ, i.e.,

$$f_X(t; n, \lambda) = \frac{\lambda^n}{\Gamma(\alpha)} t^{n-1} e^{-\lambda t}$$

The condition for the above theorem to apply is always going to be present when considering Poisson processes

3.6   Properties of a Poisson Process

A Poisson process has the following properties:

1. For a Poisson process with rate λ,

$$E[N(t)] = Var(N(t)) = \lambda t$$

2. The sum of two independent Poisson processes is a Poisson process

3. A random selection from a Poisson process yields a Poisson Process. Suppose that $N(t)$, the number of occurrences of an event E in an interval length t is a Poisson process with parameter λ. Suppose also that each occurrence of E has a constant probability p of being recorded and that that the recording of an occurrence is independent of that of other occurrences and also of $N(t)$. If $M(t)$ is the number of occurrences recorded in an interval of length t, then $M(t)$ is also a Poisson process with parameter λp

3.7   Poisson Cluster Process

Let us suppose that several events can happen simultaneously at an instance, i.e. we have a cluster of occurrences at a point. We assume that:

1. The number $N(t)$ of clusters in time t constitute a Poisson process with mean rate λ

2. Each cluster has a random number of occurrences, i.e. the number X_i of occurrences in the ith cluster is a random variable. The number of occurrences in the different clusters are independent and follow the same probability distribution $Pr(X_i = k) = p_k$, having probability generating function:

$$P(s) = \sum_{k=1}^{\infty} p_k s^k$$

Theorem 3.5. If $M(t)$ denotes the total number of occurrences in an interval of length t under the given conditions, then the generating function of $M(t)$ is given by:

$$G(P(s)) = \exp[\lambda t\{P(s) - 1\}]$$

3.8   Non-homogeneous Poisson Processes

In a non-homogeneous Poisson process, the rate λ varies with time, i.e. $\lambda(t)$ is a function of t. $P_n(t)$ is the probability that n events occur by time t.

> **Theorem 3.6.**
> $$P_n(t) = \frac{1}{n!}(m(t))^n e^{-m(t)}$$
> where
> $$m(t) = \int_0^t \lambda(x)dx$$

From this, we can find that:
$$E[N(t)] = m(t)$$

3.9 Yule-Furry Process

Here we consider that λ is a function of n, the population size at an instant. We assume that:

$$p_k(h) = Pr(N(t+h) - N(t) = k|N(t) = n) = \lambda_n h + o(h), k = 1$$
$$= o(h), k \geq 2$$
$$= 1 - \lambda_n h + o(h), k = 0$$

where $\lambda_n = n\lambda$ and the initial condition is $p_1(0), p_i(0) = 0$ for $i \neq 1$. The general process where λ is a function of n is called a pure birth process, but if $\lambda_n = n\lambda$, then it is a Yule-Furry Process.

An example of such a process is a population of bacteria that can give birth to new members but cannot die.

> **Theorem 3.7.** A Yule-Furry Process is given by the following characteristics
>
> $$p_n(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1}$$
>
> $$E[N(t)] = e^{\lambda t}$$
>
> $$Var(N(t)) = e^{\lambda t}(e^{\lambda t} - 1)$$

This means that starting with a single individual, the population size at time t follows a geometric distribution with mean $e^{\lambda t}$.

Exercise. Let the waiting time T_n be the time at which the population first reaches the

value n. Then, if there is only one individual at the start, prove that the density function of T_n is:

$$f(t) = \lambda(n-1)e^{-\lambda t}(1 - e^{-\lambda})^{n-2}$$

Solution.

$$1 - F_{T_n}(t) = P[T_n > t] = P[N(t) \leq n - 1]$$

$$= \sum_{k=1}^{n} -1 p_k(t)$$

$$= e^{-\lambda t} \sum_{k=1}^{n} -1(1 - e^{-\lambda t})^{k-1}$$

$$= e^{-\lambda t} \sum_{k=0}^{n-2} (1 - e^{-\lambda t})^{k}$$

$$= 1 - (1 - e^{-\lambda t})^{n-1}$$

Therefore,

$$F_{T_n}(t) = (1 - e^{-\lambda t})^{n-1}$$

or

$$f_{T_n}(t) = F'_{T_n}(t) = \lambda e^{-\lambda t}(n-1)(1 - e^{-\lambda t})^{n-2}$$

∎

Let us suppose that there is also immigration in addition to birth, such that in an interval of infinitesimal length h, the probability of a new member being added (by immigration) is $vh + o(h)$.

Then the generating function of the probability is:

$$P(s,t) = \frac{s^i e^{-\lambda t} e^{-v(1-s)t}}{(1 - s(1 - e^{-\lambda t}))^i}$$

3.10 Birth And Death Process

The probability that the number of births between t and $t + h$ is k, given that the number of individuals at epoch t is n is given by:

$$p(k, h|n, t) = \begin{cases} \lambda_n h + o(h) & k = 1 \\ o(h) & k \geq 2 \\ 1 - \lambda_n h + o(h) & k = 0 \end{cases}$$

The probability that the number of deaths between t and $t + h$ is k, given that the number of individuals at epoch t is n, is given by:

$$q(k, h|n, t) = \begin{cases} \mu_n h + o(h) & k = 1 \\ o(h) & k \geq 2 \\ 1 - \mu_n h + o(h) & k = 0 \end{cases}$$

This is called a **Birth and Death Process**. A birth increases population by one, while a death decreases it by one.

If $\lambda_n = n\lambda$ and $\mu_n = n\mu$, then the process is called a **linear growth process**. Then, the PGF is given by:

$$P(t, s) = \sum_{n=0}^{\infty} p_n(t) s^n$$
$$= \left[\frac{\mu(1 - s) - (\mu - \lambda s)e^{-(\lambda - \mu)t}}{\lambda(1 - s) - (\mu - \lambda s)e^{-(\lambda - \mu)t}} \right]^i$$

where i is the initial population size.

The probability of population extinction is obviously $p_0(t)$. Then,

$$p_0(t) = P(t, 0) = \left[\frac{\mu}{\lambda} \right]^i$$

as $t \to \infty$.

The mean population size will be:

$$E[N(t)] = i e^{t(\lambda - \mu)}$$

Chapter 4

Brownian Motion

4.1 Wiener Process

Consider a particle performs a random walk such that in a small interval of time of duration Δt, the displacement of the particle to the right or to the left by a small amount Δx. The probability of the particle moving right by Δx is p and to the left $-\Delta x$ is q, such that $p + q = 1$. Both p and q are independent of time t. This displacement at the ith step is denoted by Z_i. Then, the total displacement is:

$$X(t) = \sum_{i=1}^{n(t)} Z_i$$

Each "step" is obviously a time interval of Δt, so $n(t) = t/\Delta t$. First, let us compute the expected value:

$$
\begin{aligned}
E[X(t)] &= \sum_{i=1}^{n} E[Z_i] \\
&= \sum_{i=1}^{n} (p - q)\Delta x \\
&= n(p - q)\Delta x \\
&= \frac{t(p - q)\Delta x}{\Delta t}
\end{aligned}
$$

Now, let us compute the variance:

$$V[X(t)] = \sum_{i=1}^{n} V[Z_i]$$

$$= \sum_{i=1}^{n} E[Z_i^2] - E[Z_i]^2$$

$$= \sum_{i=1}^{n} (\Delta x)^2 (p+q) - \Delta x^2 (p-q)^2$$

$$= \sum_{i=1}^{n} (\Delta x)^2 [1 - (p-q)^2]$$

$$= \sum_{i=1}^{n} 4pq(\Delta x)^2$$

$$= 4npq(\Delta x)^2$$

$$= \frac{4pqt(\Delta x)^2}{\Delta t}$$

Let $X(t)$ is such that as $\Delta t \to 0$, $V[X(t)] \to \sigma^2 t$. This means that:

$$\lim_{\Delta t \to 0} \frac{4pqt(\Delta x)^2}{\Delta t} = \sigma^2 t$$

$$\lim_{\Delta t \to 0} \frac{4pq(\Delta x)^2}{\Delta t} = \sigma^2$$

For this limit to exist,

$$\Delta x = \sigma (\Delta t)^{1/2}$$

Also Let $X(t)$ is such that

$$\lim_{\Delta t \to 0} E[X(t)] = \mu t$$

$$\lim_{\Delta t \to 0} \frac{(p-q)\Delta x}{\Delta t} = \mu$$

$$\lim_{\Delta t \to 0} \frac{p-q}{\Delta t} \frac{\sigma}{2\sqrt{pq}} (\Delta t)^{1/2} = \mu$$

This equation is satisfied when:

$$p = \frac{1}{2}(1 + \mu(\Delta t)^{1/2}/\sigma)$$

$$q = \frac{1}{2}(1 - \mu(\Delta t)^{1/2}/\sigma)$$

Since Z_i are i.i.d, the sum of them $X(t)$ for large n is normal with mean μt and variance $\sigma^2 t$ from the central limit theorem. Note that here t refers to the interval $[0, t]$. If we consider the interval $[s, t]$, then the mean would be $\mu(t - s)$ and variance $\sigma^2(t - s)$.

Definition 19 (Wiener Process). The stochastic process $\{X(t), t > 0\}$ is called a Wiener process with drift μ and variance σ^2 if

1. $X(t)$ has independent increments. This means that for every pair of disjoint intervals (t, s) and (u, v), the variables $X(t) - X(s)$ and $X(u) - X(v)$ are independent.

2. Every increment is normally distributed with mean $\mu(t - s)$ and variance $\sigma^2(t - s)$.

A Wiener process is also called a **Brownian Motion**. If $X(0) = 0, \mu = 0$ and $\sigma = 1$, then it is called a **standard Brownian motion**.

Since $X(t) - X(0)$ is normally distributed with mean μt and variance $\sigma^2 t$, the **transition probability density function** f is given by:

$$f(x_0, x; t) = P[x \leq X(t) < x + \delta x | X(0) = x_0]$$
$$= \frac{1}{\sigma \sqrt{2\pi t}} \exp\left\{-\frac{(x - x_0 - \mu t)^2}{2\sigma^2 t}\right\} \delta x$$

Exercise. Find $Cov[X(s), X(t)]$ where X is a standard Brownian motion and $s \leq t$.

Solution.

$$X(t) = X(s) + (X(t) - X(s))$$

$$X(s)X(t) = X(s)^2 + X(s)(X(t) - X(s))$$

$$Cov[X(s), X(t)] = E[X(s)X(t)] - E[X(s)]E[X(t)]$$

We already know that X is a standard Brownian Motion, so $E[X(t)] = 0 \forall t$. Therefore

$$Cov[X(s), X(t)] = E[X(s)X(t)]$$

$$Cov[X(s), X(t)] = E[X(s)^2] + E[X(s)(X(t) - X(s))]$$

All increments of a Brownian motion are independent. Hence, the right term becomes 0.

$$Cov[X(s), X(t)] = E[X(s)^2] = Var[X(s)] = s$$

■

Exercise. Let $X(t)$ be a standard Brownian motion.

$$Y(t) = tX\left(\frac{1}{t}\right)$$

$$Y(0) = 0$$

Is Y a Brownian motion?

Solution. We want to show that $Y(t)$ is a normal distribution, since the difference between two Gaussian distributions is also Gaussian.

$$
\begin{aligned}
F_{Y(t)}(x) &= P[Y(t) \le x] \\
&= P[tX(1/t) \le x] \\
&= P[X(1/t) \le x/t] \\
&= \frac{\sqrt{t}}{\sqrt{2\pi}} \int_{-\infty}^{x/t} e^{-\frac{y^2 t}{2}} dy \\
&= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{x} e^{-\frac{\eta^2}{2t}} d\eta
\end{aligned}
$$

where $\eta = yt$. From this,

$$f_Y(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

which is a normal distribution with 0 mean and variance t.

The difference between two intervals is independent since the difference between X is independent. ∎

4.2   Diffusion Process

Let $f(x,t)\Delta x$ be the probability of finding the particle in a Wiener Process at $(x, x + \Delta x)$ at time t.

$$f(x,t)\Delta x = pf(x - \Delta x, t - \Delta t)\Delta x + qf(x + \Delta x, t - \Delta t)\Delta x$$

$$f(x,t) = p\left[f(x,t) - \frac{\partial f}{\partial x} - \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} + 0(\Delta t)\right] + q\left[f(x,t) + \frac{\partial f}{\partial x} - \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} + 0(\Delta t)\right]$$

The above result comes from Taylor Expansion.

$$f(x, t) = f(x, t)(p + q) - \frac{\partial f}{\partial x} \Delta x(p - q) - \frac{\partial f}{\partial t} \Delta t(p + q) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (\Delta x)^2 (p + q)$$

Cancelling some terms and using the values of p and q, we get the following partial differential equation:

$$\frac{\partial f}{\partial t} + \mu \frac{\partial f}{\partial x} = \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial x^2}$$

This is called the **Advection Diffusion Equation**. $\sigma^2/2$ is called the diffusion coefficient.

Solving the diffusion equation will yield the transition probability density function already discussed.

If $\mu = 0$, it is called the heat equation.

4.3 Hitting Times and the Maximum Variable

Let T_a denote the first time the Brownian motion process hits a. T_a obviously follows some distribution. Rigorously finding this distribution is a very involved task, and beyond the scope of the course. Instead, we can use a simpler argument.

Let $a > 0$. We seek $F_{T_a}(t) = P(T_a \leq t)$.

$$P(X(t) \geq a) = P(X(t) \geq a | T_a \leq t) P(T_a \leq t) + P(X(t) \geq a | T_a > t) P(T_a > t)$$

If $T_a \leq t$, then the process hits a at some point in $[0, t]$ and, by symmetry, it is just as likely to be above a or below a at time t.

$$P(X(t) \geq a | T_a \leq t) = \frac{1}{2}$$

Note that the idea of symmetry is called the **principle of reflection**. We have just assumed it to make sense (because it does, intuitively), but it can be proven to be true as well.

The second term of the first equation is obviously 0. Hence,

$$P(T_a \leq t) = 2P(X(t) \geq a)$$

$X(t) = N[0, \sigma\sqrt{t}]$. Hence.

$$F_{T_a}(t) = 2P(X(t) \geq a)$$
$$= \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/2t} dx$$
$$= \frac{2}{\sqrt{2\pi}} \int_{|a|/\sqrt{t}}^\infty e^{-y^2/2} dy$$

This is also equivalent to:

$$F_{T_a}(t) = 2\left[1 - \phi\left[\frac{a}{\sigma\sqrt{t}}\right]\right]$$

where ϕ is the cumulative distribution function of a standard normal.

Another random variable of interest is the maximum value a process attains in $[0, t]$. It's distribution is given by

$$P\{\max_{0 \leq s \leq t} X(s) \geq a\} = \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^\infty e^{-y^2/2} dy$$

4.4   Brownian Bridge

If $X(t)$ is a Brownian motion, then a Brownian Bridge $Z(t)$ is given by:

$$Z(t) = X(t) - tX(1), 0 \leq t < 1$$

This can be generalized and better expressed as:

$$Z(t) = (X(t)|X(T) = 0), t \in [0, T]$$

This means that the Brownian bridge is a process pinned to the origin at both ends. For now, we will only consider the case where $T = 1$, since this is what was discussed in class.

$Z(t)$ is obviously normal since it is a linear combination of normal variables. It's expectation will be given by:

$$E[Z(t)] = E[X(t) - tX(1)]$$
$$= \mu t - t\mu$$
$$= 0$$

Let $0 \leq s < t \leq 1$. Then

$$
\begin{aligned}
Cov[Z(s), Z(t)] &= Cov[X(s) - sX(1), X(t) - tX(t)] \\
&= Cov[X(s), X(t)] - sCov[X(1), X(t)] - tCov[X(s), X(1)] + stCov[X(1), X(1)] \\
&= V[X(s)] - sV[X(t)] - tV[X(s)] + stV[X(1)] \\
&= \sigma^2 s - s\sigma^2 t - t\sigma^2 s + st\sigma^2 \\
&= \sigma^2 s(1 - t)
\end{aligned}
$$

4.5 Ornstein-Uhlenbeck Process

A Weiner process is called a **stationary covariance process** if $Cov[X(t), X(t + \tau)]$ if it is independent of t. This means that the covariance is a function of τ. The process we will now define is an example of this.

Let $X(t)$ be a standard Brownian motion. Then

$$
Y(t) = \frac{1}{\sqrt{g(t)}} X(\sigma^2 g(t))
$$

where $g(0) = 1$ and is a strictly increasing function.

First let us find the expectation of Y.

$$
\begin{aligned}
E[Y(t)] &= \frac{1}{\sqrt{g(t)}} E[X(\sigma^2 g(t))] \\
&= 0
\end{aligned}
$$

Next, let us calculate the variance of Y.

$$
\begin{aligned}
V[Y(t)] &= \frac{1}{g(t)} V[X(\sigma^2 g(t))] \\
&= \frac{1}{g(t)} \sigma^2 g(t) \\
&= \sigma^2
\end{aligned}
$$

Finally, we can find the covariance of Y.

$$Cov[Y(t), Y(t+\tau)] = \frac{1}{\sqrt{g(t)g(t+\tau)}} Cov[X(\sigma^2 g(t)), X(\sigma^2 g(t+\tau))]$$

$$= \frac{1}{\sqrt{g(t)g(t+\tau)}} \sigma^2 g(t)$$

$$= \sigma^2 \left(\frac{g(t)}{g(t+\tau)}\right)^{1/2}$$

We want this process to be stationary covariant, i.e,

$$g(t+\tau) = g(t)g(\tau)$$

Luckily, there is a simple function for this:

$$g(t) = e^{\mu t}$$

Hence,

$$Cov[Y(t), Y(t+\tau)] = \sigma^2 e^{-\mu\tau/2}$$

Putting it all together,

$$Y(t) = \frac{1}{e^{\mu t/2}} X[\sigma^2 e^{\mu t}]$$

is a stationary covariance process called the **Ornstein-Uhlenbeck Process**.

4.6   Quadratic Variations

Let $W(t)$ be a standard Brownian motion. Let Π_n be a partition on interval $[0, t]$, such that it divides the segment into equally spaced intervals of size $\Delta t = t/n$. Let us define

$$Q_{\Pi_n} = \sum_{j=1}^{n} [W(t_j) - W(t_{j-1})]^2$$

Taking it as n tends to infinity,

$$[W, W](t) = \lim_{n\to\infty} Q_{\Pi_n}$$

$$= \lim_{n\to\infty} \sum_{j=1}^{n} [W(t_j) - W(t_{j-1})]^2$$

This is called the **quadratic variation**.

Obviously, $Q_{\Pi_n}(t)$ is a stochastic process. Is $[W,W](t)$ a stochastic variable? It is, but at the same time, it does not have much in the way of randomness. This is given in the following theorem

Theorem 4.1. $[W,W](t) = t$ "almost surely".

Proof. We know that $W(t_i) - W(T_{i-1})$ is a normal variable with mean 0 and variance $t_i - t_{i-1}$.

$$
\begin{aligned}
E[Q_{\Pi_n}] &= \sum_{i=1}^{n} E[(W(t_i) - W(t_{i-1}))^2] \\
&= \sum_{i=1}^{n} V[W(t_i) - W(t_{i-1})] \\
&= \sum_{i=1}^{n} (t_i - t_{i-1}) \\
&= t
\end{aligned}
$$

Now let us handle the variance.

$$
V[Q_{\Pi_n}] = \sum_{i=1}^{n} \left[E[(W(t_i) - W(t_{i-1}))^4] - E[(W(t_i) - W(t_{i-1}))^2]^2 \right]
$$

Let X be a normal distribution $N(0, \sigma^2)$. Then we can prove that $E(X^4) = 3\sigma^4$. Hence,

$$
\begin{aligned}
V[Q_{\Pi_n}] &= \sum_{i=1}^{n} (3(t_i - t_{i-1})^2 - (t_i - t_{i-1})^2) \\
&= 2 \sum_{i=1}^{n} (t_i - t_{i-1})^2 \\
&= 2n(\delta t)^2 \\
&= \frac{2t^2}{n}
\end{aligned}
$$

Taking limit as $n \to \infty$,

$$
E[[W,W](t)] = t
$$

$$
V[[W,W](t)] \to 0
$$

This is what we mean by "almost surely". This process is almost deterministic. $\qquad \square$

This gives us a nice result - the value of the quadratic variation of a Brownian motion is finite.

Theorem 4.2. The quadratic variation of a function f that is differentiable and whose derivative f' is also continuous, is 0.

Proof. TODO   □

4.7 Ito's Calculus and Stochastic Differential Equations

Lemma 4.3 (Ito's Lemma)**.** Let X_t be a standard Weiner process and $f(t, x)$ be a twice differentiable function in x. Then,

$$df(t, X_t) = \left[\partial_t f(t, X_t) + \frac{1}{2}\partial_{xx}^2 f(t, X_t)\right] dt + \partial_x f(t, X_t) dX_t$$

The proof for this is too complicated to include in this course.

Lemma 4.4 (Ito's Second Lemma)**.** Let X be a stochastic process such that:

$$dX = \theta(t)dt + \Delta(t)dW$$

and $f(t, x)$ is another stochastic process,

$$df = \left[f_t(t, X_t) + \frac{1}{2}f_{xx}(t, X(t))\Delta^2(t)\right] dt + f_x(t, X(t))dX$$

This is a generalization of Ito's Lemma.

Exercise. Solve the following stochastic differential equation:

$$dX = \theta(\mu - X)dt + \sigma dW$$

where W is a standard Weiner Process.

Solution. Comparing with Ito's Second Lemma,

$$\Delta(t) = \sigma, \theta(t) = \theta(\mu - X)$$

■