Notes on the theory of stochastic processes

Samuel Baltz

2019

Contents

1 Markov chains and transition probabilities

Definition 1. [2: p. 2] X_n is a **discrete time Markov chain** over the state space S with **transition matrix** $p(i,j)$ if, for any states $j, i, i_{n-1}, \ldots, i_0 \in S$,

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

Remark 1. This represents a strong idea of history independence.

Definition 2. [2: p. 2] In state space S, the probability of going from state $i \in S$ to state $j \in S$ in one step is given by the **transition probability**

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

Definition 3. [2: p. 6] In state space S, an **absorbing state** is any state $x \in S$ with transition probability

$$p(x,x) = 1$$

Definition 4. [2: p. 8] In state space S, the probability of going from state $i \in S$ to state $j \in S$ in m steps, where $m \in \mathbb{N}$ and $m > 1$, is called the **multistep transition probability** and is given by

$$p^m(i,j) = P(X_{n+m} = j | X_n = i)$$

Theorem 1. *[2: p.10] The m-step transition probability $P(X_{n+m} = j | X_n = i)$ is the m^{th} power of the transition matrix p.*

Example 1. [2: p. 11-12] Consider the following matrix of transition probabilities:

$$p = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

By Theorem 1, the probabilities of reaching each state from each other state in exactly two steps are given by

$$p^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0.24 & 0 & 0.16 & 0 \\ 0.36 & 0 & 0.48 & 0 & 0.16 \\ 0 & 0.36 & 0 & 0.24 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Using Theorem 1, we can offer some interpretations of p^2. Consider the following examples:

$\hookrightarrow$the probability of going from state one to state 1 in 2 steps is 1; this is because, if we begin at state 1, we will certainly visit state 1 in the next step, and again in the following step.

$\hookrightarrow$the probability of going from state 2 to state 1 in 2 steps is 0.6; there is no alternative path to start at state 2 and reach state 1 in 2 steps other than to immediately transition to state 1, and once state 1 has been reached it cannot be left. So this is identical to the probability of moving to state 1 from state 2 in 1 step.

$\hookrightarrow$the probability of going from state 3 to state 3 in 2 steps is 0.36. One path is by going to step 2 with probability 0.6, and then returning to step 3 with independent probability 0.4, so this path has probability 0.24 of occurring. The only other 2-step path back to state 3 is to go to step 4 with probability 0.4, and then return with independent probability 0.6, giving this path also probability 0.24. Because these two paths are mutually exclusive, we can sum them to obtain the probability of either one occurring, which gives a net probability of 0.48 that a Markov chain at state 3 will return to state 3 in 2 steps.

Remark 2. [2: p. 12] When we are interested in the behaviour of a Markov chain for a fixed intitial state, we will use the notation $P_x(A)$ to represent $P(A|X_0 = x)$. Later E_x will represent $E[P(A|X_0 = x)]$.

2 Stopping and return times

Definition 5. [2: p. 13] Let $T_y = \min\{n \geq 1 : X_n = y\}$ be the **time of first return to** state y (noting that this definition does not count time 0), and let $\rho_{yy} = P_y(T_y < \infty)$ be the probability that X_n returns to y given that it starts at y.

Remark 3. [2: p. 13] By the Markov property, we can immediately notice that the probability that X_n returns to y at least ρ_{yy}^2, since the chain can always move to y from y, and then do so again.

Definition 6. [2: p. 13] We say that T is a **stopping time** if we can formalize "stop at time n", when $\{T = n\}$, by looking at the values of the Markov chain up to n, $X_0, X_1, \ldots, X_n$.

Remark 4. [2: p. 13] To see that T_y in Remark 3 is a stopping time, note that

$$\{T_y = n\} = \{X_1 \neq y, \ldots, X_{n-1} \neq y, X_n \neq y\}$$

which can be determined using only $X_0, X_1, \ldots, X_n$.

Theorem 2. *[2: p. 13] Suppose that T is a stopping time. Given that $T = n$ and $X_T = y$, any other information about $X_0, \ldots, X_T$ is irrelevant for predicting the future, and $X_{T+k}, k \geq 0$ behaves like the markov chain with initial state y.*

Definition 7. [2: p. 14] Let $T_y^1 = T_y$, and for $k \geq 2$ call $T_y^k = \min\{n > T_y^{k-1} : X_n = y\}$ the **time of the $\mathbf{k^t}$ return to** y.

Remark 5. With this we can establish that $P_y(T_y^k < \infty) = \rho_{yy}^k$.

3 Classification of states

Definition 8. [2: p. 14] A state y is called **transient** if $\rho_{yy} < 1$, so that the probability of returning to y k times has $\rho_{yy}^k \to 0$ as $k \to \infty$. So, eventually the Markov chain will not return to y again.

Definition 9. [2: p. 14] A state y is called **recurrent** if the probability of returning to y k times has $\rho_{yy}^k = 1$. So, the Markov chain returns to y infintely many times.

Definition 10. [2: p. 15] State x **communicates with** state y, denoted $x \to y$, if there is a positive probability of reaching y when starting from x, so

$$\rho_{xy} = P_x(T_y < \infty)$$

satisfies $\rho_{xy} > 0$.

Theorem 3. *[2: p. 16] If $\rho_{xy} > 0$, but $\rho_{yx} < 1$, then x is transient.*

Definition 11. [2: p. 17] A set A of states is **closed** if, for any $i \in A$ and $j \notin A$, $p(i,j) = 0$.

Definition 12. [2: p. 17] A set A of states is **irreducible** if $i \to j$ $\forall i, j \in A$.

Definition 13. A set A of states is **finite** if $|A| < \infty$.

Theorem 4. *[2: p. 17] If A is a finite, closed, and irreducible set, then all states in A are reccurent.*

Theorem 5. *[2: p. 17] If a state space S isfinite, then S can be written as a disjoint union $T \cup R_1 \cup \cdots \cup R_k$, where T is a set of transient states and the R_i, $1 \leq i \leq k$, are closed irreducible sets of recurrent states.*

Example 2. [2: p. 16-17] Consider the transition probability matrix

$$\begin{bmatrix} 0.7 & 0 & 0 & 0 & 0.3 & 0 & 0 \\ 0.1 & 0.2 & 0.3 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.3 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0.6 & 0 & 0 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Let us classify states using just Theorem 3 and Theorem 4. We have

$\{1, 5\}$ is a finite, closed, and irreducible set, so by Theorem 4, 1 and 5 are recurrent.

$\rho_{21} > 0$ but $\rho_{12} < 1$, so by Theorem 3, 2 is transient.

$\rho_{35} > 0$ but $\rho_{53} < 1$, so by Theorem 3, 3 is transient.

$\{4, 6, 7\}$ is a finite, closed, and irreducible set, so by Theorem 4, 4, 6, and 7 are recurrent.

Theorem 6. *[2: p. 19] y is recurrent iff*

$$\sum_{n=1}^{\infty} p^n(y, y) = E_y N_y$$

satisfies

$$\sum_{n=1}^{\infty} p^n(y, y) = \infty$$

[5: p. 92] Consequently, y is transient iff

$$\sum_{n=1}^{\infty} p^n(y, y) < \infty$$

Corollary 7. *[5: p. 92] If state i is recurrent, and $i \to j$, then j is recurrent also.*

Example 3. [5: p. 92] Consider the transition matrix

$$\begin{bmatrix} 0 & 0 & 0.5 & 0.5 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Because $0 \to 3$, $0 \to 2$, $3 \to 1$, $2 \to 1$, and $1 \to 0$, the full state space is finite, closed, and irreducible, so by Theorem 4 all states are recurrent.

Example 4. [5: p. 92] Consider the transition matrix

$$\begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0.25 & 0.25 & 0 & 0 & 0.5 \end{bmatrix}$$

$\{0, 1\}$ is finite, closed, and irreducible, so by Theorem 4, 0, 1 are recurrent.

$\{2, 3\}$ is finite, closed, and irreducible, so by Theorem 4, 2, 3 are recurrent.

$\rho_{40} > 0$ but $\rho_{04} < 1$, so by Theorem 3, 4 is transient.

4 Limit behaviours

Remark 6. [2: p. 20] What happens when a Markov chain begins in a random state? It turns out that we can obtain the distribution over states in the following step simply by multiplying the transition probabilities by that initial distribution. Consider the initial distribution over states i given by the row vector

$$q(i) \equiv P(x_0 = i)$$

And we seek the distribution over states in the following step conditional on the values in the first step, so

$$P(X_n = j) = \sum_i P(X_0 = i, X_n = j)$$

$$P(X_n = j) = \sum_i P(X_0 = i) P(x_n = j | X_0 = i)$$

$$P(X_n = j) = \sum_i q(i) p^n(i, j)$$

Example 5. [2: p. 21] Suppose we have the transition matrix

$$p = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$

together with the initial distribution over states

$$q = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}$$

Then the distribution in the next step is given by

$$\begin{bmatrix} 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.32 & 0.68 \end{bmatrix}$$

Definition 14. [2: p. 21] If $qp = q$, then q is called a **stationary distribution** and is denoted π.

Remark 7. [2: p. 21] Stationary distributions are a very special case, because by the Markov property in Definition 1, if the distribution at time n is the same as the distribution at the time $n + 1$, then it will continue to be the distribution for all time.

Example 6. [2: p. 21-22] Consider the transition matrix in Example 5. We seek the stationary distribution, if one exists, so we want to solve

$$\begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix}$$

$$\begin{bmatrix} 0.6\pi_1 + 0.2\pi_2 & 0.4\pi_1 + 0.8\pi_2 \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix}$$

With the additional property that $\pi_1 + \pi_2 = 1$, we have

$$\begin{bmatrix} \pi_1 & \pi_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Theorem 8. *[2: p. 22] In general, the transition probabilities in the stationary distribution of any two-state Markov chain, with transition probabilities denoted*

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

are always given by

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

Theorem 9. *[2: p. 25] Suppose that a $k \times k$ transition matrix p is irreducible. Then there is a unique solution to $\pi p = \pi$ with $\sum_x \pi_x = 1$, and we have $\pi_x > 0 \; \forall x$.*

Remark 8. [2: p. 26] We can set aside transient states because $y$transient $\implies p^n(x,y) \to 0$ for any initial state x, so we can restrict our attention to Markov chains that contain one closed set of recurrent states.

Definition 15. [2: p. 27] The **period** of a state is the largest number that will divide all the n, $n \geq 1$, for which $p^n(x,x) > 0$. So, the period is the greatest common divisor of $I_x \equiv \{n \geq 1 : p^n(x,x) > 0\}$.

Lemma 10. *[2: p. 29] $p(x,x) > 0 \implies x$ has period 1.*

Lemma 11. *[2: p. 29] If $\rho_{xy} > 0$ and $\rho_{yx} > 0$ then x and y have the same period.*

Notation 1. [2: p. 30] We use the following shorthand for properties of Markov chains:
↪I: p is irreducible
↪A: aperiodic, all states have period 1
↪R: all states are recurrent
↪S: there is a stationary distribution π

Theorem 12. *[2: p. 31] Suppose I, A, S. Then as $n \to \infty$, $p^n(x,y) \to \pi(y)$.*

Theorem 13. *[2: p. 31] Suppose I and R. Then there is a stationary measure with $\mu(x) > 0\ \forall x$.*

Theorem 14. *[2: p. 31] Suppose I, R. Let $N_n(y)$ be the number of visits to y up to time n. Then,*

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y}$$

Theorem 15. *[2: p. 31] Suppose I, S. Then*

$$\pi(y) = \frac{1}{E_y T_y}$$

which implies that the stationary distribution is unique.

5 Special Markov chains

Definition 16. [2: p. 34] A transition matrix p is called **doubly stochastic** if its columns sum to 1, so $\sum_x p(x, y) = 1$

Theorem 16. *[2: p. 35] If p is a doubly stochastic transition probability for a Markov chain with N states, then the uniform distribution $\pi(x) = \frac{1}{N}$*

Definition 17. [2: p. 37] π satisfies the **detailed balance condition** if

$$\pi(x)p(x, y) = \pi(y)p(y, x)$$

Remark 9. The detailed balance condition says that whatever is going from one state to another is exactly counterbalanced by whatever is going to that state from the other. It is a special case of a stationary distribution.

Example 7. [2: p. 37] Consider

$$\begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.3 & 0.1 & 0.6 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$$

There is no stationary distribution with detailed balance because $\pi(1)p(1, 3) = 0$ but $p(1, 3) > 0$, which would force all the $\pi(i)$ to be 0. Since the chain is doubly stochastic, by Theorem 16 the stationary distribution is $\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$

6 Exit distributions and exit times

Remark 10. An exit distribution answers the question: what is the probability that a chain currently in one state will eventually exit by means of another state?

Example 8. [2: p. 52] Consider the Markov chain with states {Freshman, Sophomore, Graduate, Drop Out} and the associated transition probabilities

$$\begin{bmatrix} 0.25 & 0.6 & 0 & 0.15 \\ 0 & 0.2 & 0.7 & 0.1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Let $h(x)$ be the probability that a student currently in state x eventually graduates. First consider what happens in one step from states 1 and 2:

$$h(1) = 0.25h(1) + 0.6h(2)$$

$$h(2) = 0.2h(2) + 0.7$$

Which together imply

$$h(2) = \frac{7}{8}$$

so that

$$h(1) = 0.7$$

Considering just these one-step transition probabilities works because it doesn't matter how long a student remains a Freshman; by the Markov property in Definition 1, no matter how many times the student has transitioned from being a Freshman to being a Freshman, they will always have the same probability of then becoming a Sophomore and graduating. Similarly, the probability of a sophomore graduating is independent of the chain's history up to that point. So the probability of graduating if you are a sophomore is always the probability of graduating immediately, plus the independent probability of becoming a sophomore again but eventually graduating, which is $0.2 \cdot h(2)$.

Example 9. [2: p. 53] Consider the tennis game with a possible difference in scores of $\{2, 1, 0, -2, -2\}$, where the first player to be 2 ahead of their opponent will win, and with associated transition probabilities

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let $h(x)$ be the probability that the server will win when the score is x. Then,

$$h(x) = \sum_y p(x,y)h(y)$$

with $h(2) = 1$ and $h(-2) = 0$, we then have

$$h(1) = 0.6 + 0.4h(0)$$

$$h(0) = 0.6h(1) + 0.4h(-1)$$

$$h(-1) = 0.6h(0)$$

Solving we get $h(0) = 0.6923$. Now we could proceed by substitution, but this is a good moment to introduce the more general solution method. First let's rearrange the above equations to get all of the variables on the left-hand side in a regular order, as:

$$h(1) - 0.4h(0) + 0h(-1) = 0.6$$

$$-0.6h(1) + h(0) - 0.4h(-1) = 0$$

$$0h(1) - 0.6h(0) + h(-1) = 0$$

With the equations in this form, we can now read them off into a matrix:

$$\begin{bmatrix} 1 & -0.4 & 0 \\ -0.6 & 1 & -0.4 \\ 0 & -0.6 & 1 \end{bmatrix} \begin{bmatrix} h(1) \\ h(0) \\ h(-1) \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0 \\ 0 \end{bmatrix}$$

To write this more generally, we want to note that with the k nonabsorbing states $C \in \{1, 0, -1\}$ and $r(x,y)$ the restriction of p to the set of $x, y \in C$, the matrix written above is simply $I_{k \times k} - r_{k \times k}$. This immediately suggests how to solve for the exit distribution from every state:

$$\begin{bmatrix} h(1) \\ h(0) \\ h(-1) \end{bmatrix} = (I - r)^{-1} \begin{bmatrix} 0.6 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} h(1) \\ h(0) \\ h(-1) \end{bmatrix} = \begin{bmatrix} 0.8769 \\ 0.6923 \\ 0.4154 \end{bmatrix}$$

Remark 11. An exit time answers the question: how long should we expect it to take for a chain to reach an absorbing state?

Example 10. [2: p. 58] We return to Example 8 which had states $\{\text{Freshman}, \text{Sophomore}, \text{Graduate}, \text{Drop Out}\}$ and transition probabilities

$$\begin{bmatrix} 0.25 & 0.6 & 0 & 0.15 \\ 0 & 0.2 & 0.7 & 0.1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now we ask: how long do we expect it to take for a student to graduate or drop out? Let $g(x)$ be the expected time for a student starting in state x. Then $g(G) = g(D) = 0$. Consider again what happens in one step:

$$g(1) = 1 + 0.25g(1) + 0.6g(2)$$

$$g(2) = 1 + 0.2g(2)$$

so

$$g(2) = 1.25$$

and

$$g(1) \approx 2.33$$

Example 11. [2: p. 59] We return to Example 9 which had states representing the possible difference in scores $\{2, 1, 0, -2, -2\}$, where the first player to be 2 ahead of their opponent will win, and with associated transition probabilities

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Now we ask: what's the expected number of points for the game to end? Let's try solving it using the general method that goes through the restriction matrix. Since -2 and 2 are the absorbing states, the restriction method should omit these two states. This leaves the restricted matrix

$$\begin{bmatrix} 0 & 0.4 & 0 \\ 0.6 & 0 & 0.4 \\ 0 & 0.6 & 0 \end{bmatrix}$$

Next we seek $I - r$, which is given by

$$\begin{bmatrix} 1 & -0.4 & 0 \\ -0.6 & 1 & -0.4 \\ 0 & -0.6 & 1 \end{bmatrix}$$

Then $(I - r)^{-1}$ can be found using one method or another to be:

$$\begin{bmatrix} \frac{19}{13} & \frac{10}{13} & \frac{4}{13} \\ \\ \frac{15}{13} & \frac{25}{13} & \frac{10}{13} \\ \\ \frac{9}{13} & \frac{15}{13} & \frac{19}{13} \end{bmatrix}$$

A game begins at 0 points, so the expected number of points before the game ends one way or another is

$$g(0) = \frac{15}{13} + \frac{25}{13} + \frac{10}{13}$$

$$g(0) = 3.846$$

Example 12. Call T_{HT} the number of times needed to flip a coin to get a Heads and then a Tails consecutively. We have states $\{HH, HT, TH, TT\}$ and transition probabilities

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Then $I - r$ is given by

$$\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

And then

$$\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

7 Poisson processes

Recall the exponential distribution:

Definition 18. The PDF of a random variable τ that has an exponential distribution with parameter $\lambda > 0$ is given by

$$f_\tau(t) = \begin{cases} \lambda e^{-\lambda t} & \forall t \geq 0 \\ 0 & \forall t < 0 \end{cases}$$

Because the exponential property is memoryless, we use it to model events that are time-homogeneous. It is uniquely used for this application because it is essentially the only memoryless distribution. [3: Lecture 9]

Example 13. [3: Lecture 9] Consider a bus line without a schedule. From historical observations we note that the distribuition of the wait times until a bus arrives does not depend on the time when you arrive at the bus stop. Then the wait time should be modeled using an exponential distribution $Exp(\lambda)$, where λ measures how often on average buses arrive.

Example 14. [3: Lecture 9] In a hardware store, customers first have to go to server 1 to get goods, and then have to go to server 2 to pay for those goods. Suppose that the times for the two activities are exponentially distributed with rates λ and μ. First, compute the expected amount of time it takes Bob to get his goods and pay if, when he comes in, there is a customer named Al with server 1 and no one at server 2.

Denote the time that Bob spends at servers 1 and 2 by S_1 and S_2. Additionally denote the time it takes Al to get the goods from server 1, after Bob enters the store, by T_1, and the time it takes for Al to pay for the goods by T_2. Then, $S_1 \sim Exp(\lambda)$, and $S_2, T_2 \sim Exp(\mu)$. By the memoryless property, $T_1 \sim Exp(\lambda)$. Notice that T_1, T_2, S_1, S_2 are independent.

The amount of time that it takes Bob to get his goods and pay is

$$T = T_1 + S_1 + \max(T_2 - S_1, 0) + S_2$$

Then,

$$E[T] = E[T_1] + E[S_1] + E[\max(T_2 - S_1, 0)] + E[S_2]$$

$$E[T] = \frac{1}{\lambda} + \frac{1}{\lambda} + E[\max(T_2 - S_1, 0)] + \frac{1}{\mu}$$

By properties of conditionals and an extended memorylessness property,

$$E[\max(T_2 - S_1, 0)] = P(T_2 \leq S_1)E[0|T_2 \leq S_1] + P(T_2 > S_1)E[T_2 - S_1|T_2 > S_1]$$

$$E[\max(T_2 - S_1, 0)] = P(T_2 > S_1)E[T_2 - S_1 | T_2 > S_1]$$

$$E[\max(T_2 - S_1, 0)] = \frac{\lambda}{\lambda + \mu} \cdot \frac{1}{\mu}$$

So

$$E[T] = \frac{2}{\lambda} + \frac{\lambda}{\lambda + \mu} \cdot \frac{1}{\mu} + \frac{1}{\mu}$$

Definition 19. [3: Lecture 9] A Poisson process $(N_t)_{t \geq 0}$ with intensity (rate) $\lambda > 0$ is a stochastic process defined $\forall\ t \geq 0$ by

$$N_t = \max\{n \geq 0 : T_n \leq t\} \left(= \sum_{n=1}^{\infty} \mathbb{1}_{\{T_n \leq t\}} \right)$$

where $T_n = \tau_1 + \tau_2 + \cdots + \tau_n$, with independent random variables $\tau_i \sim Exp(\lambda)$.

Remark 12. A Poisson process counts the number of independent exponentially distributed events that have happened by a certain time.

Theorem 17. *[3: Lecture 9] A Poisson process (N_t) has independent increments: the random variables $(N_{t_1} - N_{t_0}), \ldots, (N_{t_n} - N_{t_{n-1}})$ are independent, for all $n \geq 1$ and all $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n < \infty$.*

Example 15. [3: Lecture 9] The independence of increments allows us to do something very powerful: compute the joint distribution of the values of a Poisson process at multiple times. So for example, consider (N_t) a Poisson process with intensity $\lambda = 2$. Then

$$P(N_2 = 1, N_{2.5} = 3, N_4 = 4) = P(N_4 - N_{2.5} = 1 | N_{2.5} = 3, N_1 = 1)P(N_{2.5} - N_2 = 2 | N_2 = 1)P(N_2 = 1)$$

$$P(N_2 = 1, N_{2.5} = 3, N_4 = 4) = P(N_4 - N_{2.5} = 1)P(N_{2.5} - N_2 = 2)P(N_2 = 1)$$

$$P(N_2 = 1, N_{2.5} = 3, N_4 = 4) = P(N_{1.5} = 1)P(N_{0.5} = 2)P(N_2 = 1)$$

$$P(N_2 = 1, N_{2.5} = 3, N_4 = 4) = e^{-2 \cdot 1.5} \frac{(2 \cdot 1.5)^1}{1!} e^{-2 \cdot 0.5} \frac{(2 \cdot 0.5)^2}{2!} e^{-2 \cdot 2} \frac{(2 \cdot 2)^1}{2!}$$

Example 16. [1: p. 76] Let N be a Poisson process with rate $\lambda = 8$, and we want to compute $P(N_{2.5} = 17, N_{3.7=22}, N_{4.3=36})$. This we can decompose into

$$P(N_{2.5} = 17, N_{3.7=22}, N_{4.3=36}) = P(N_{4.3} - N_{3.7} = 14)P(N_{3.7} - N_{2.5} = 5)P(N_{2.5} = 17)$$

$$P(N_{2.5} = 17, N_{3.7=22}, N_{4.3=36}) = P(N_{0.6} = 14)P(N_{1.2} = 5)P(N_{2.5} = 17)$$

$$P(N_{2.5} = 17, N_{3.7=22}, N_{4.3=36}) = e^{8\cdot 0.6}\frac{(8\cdot 0.6)^{14}}{14!}e^{8\cdot 1.2}\frac{(8\cdot 1.2)^{5}}{5!}e^{8\cdot 2.5}\frac{(8\cdot 2.5)^{17}}{17!}$$

Theorem 18. *[3: Lecture 9] A stochastic process $(N_t)_{t\geq 0}$ is a Poisson process with parameter λ iff*

$\hookrightarrow N_0 = 0$

$\hookrightarrow N_{t+s} - N_s \sim Pois(\lambda t)\ \forall t, s \geq 0$

$\hookrightarrow (N_t)$ *has independent increments*

Now we will consider compound and non-homogeneous Poisson processes.

Example 17. [3: Lecture 10] Consider customers arriving at a restaurant. Assume that the Poisson process governing the arrival of cars has intensity λ, and each variable Y_i has a known distribution: $P(Y_i = k) = p_k$, for $k = 1, 2, 3, 4$, with some nonnegative numbers $\{p_k\}$ such that $\sum_{k=1}^{4} p_k = 1$. Then, the average number of customers who have arrived in the restaurant by time t is given by

$$E[S_t] = E[N_t]E[Y_i] = \lambda t \sum_{k=1}^{4} k p_k$$

And the variance of the number of customers is

$$Var(S_t) = \lambda t E[Y_i^2]$$

$$Var(S_t) = \lambda t \sum_{k=1}^{4} k^2 p_k$$

Remark 13. [3: Lecture 10] We can do something surprising and very powerful, called thinning: we can split the Poisson process into several independent Poisson processes which each measure the subprocess which has value exactly k. Each of these processes has the simple intensity λp_k.

Theorem 19. *[2: p. 109] If $s < t$ and $0 \le m \le n$, then*

$$P(N(s) = m | N(t) = n) = \binom{n}{m} \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m}$$

So the conditional distribution of $N(s)$ given $N(t) = n$ is binomial with parameters $n, \frac{s}{t}$.

Remark 14. [3: Lecture 10] The opposite of thinning, in which we take several independent Poisson processes and sum them to obtain a new Poisson process, is called superposition.

Example 18. [3: Lecture 10] Concert tickets are being sold at a kiosk. Female and male customers arrive according to two independent Poisson processes (N_t^f) and (N_t^m), with rates 30 and 20 per hour respectively. Suppose that each customer regardless of gender can, independently, buy either 1 ticket with probability $\frac{1}{2}$, 2 tickets with probability $\frac{2}{5}$, or 3 tickets with probability $\frac{1}{10}$. Let Z_k be the number of customers each of whom bought k tickets within the first hour.

First, what is the joint distribution of (Z^1, Z^2, Z^3)? We apply thinning and superposition to obtain

$$P(Z^1 = i, Z^2 = j, Z^3 = k) = e^{-50 \cdot \frac{1}{2}} \frac{(50 \cdot \frac{1}{2})^i}{i!} e^{-50 \cdot \frac{2}{5}} \frac{(50 \cdot \frac{2}{5})^j}{j!} e^{-50 \cdot \frac{1}{10}} \frac{(50 \cdot \frac{1}{10})^k}{k!}$$

Next, what is the probability that the first three customers are female? Simply measure these with binary rvs that identify the first 3 customers, and note that these rvs are independent and each have probability $\frac{3}{5}$. So the joint probability is $\frac{27}{125}$.

Finally, suppose the kiosk reports to the main office that 1000 tickets are sold between 3 and 9 pm, each customer buys one ticket, kiosk closes at 9. Customers enter the concert hall right after buying a ticket, and the concert starts at a time that is uniformly distributed in the last hour, between 8 and 9 pm. What is the probability that half the people were on time? Call $t = 0$ 3 pm, and call T the random time when the concert starts. Conditioning on the value of T and applying the conditioning fact for a Poisson process that it is binomially distributed,

$$P(N_T^1 + 2N_T^2 + 3N_T^3 = 500 | N_6^1 + 2N_6^2 + 3N_6^3 = 1000, N_6^2 = N_6^3 = 0) = P(N_t^1 = 500 | N_6^1 = 1000)$$

$$P(N_T^1 + 2N_T^2 + 3N_T^3 = 500 | N_6^1 + 2N_6^2 + 3N_6^3 = 1000, N_6^2 = N_6^3 = 0) = \int_5^6 P(N_t^1 = 500 | N_6^1 = 1000) P(T \in dt)$$

$$P(N_T^1 + 2N_T^2 + 3N_T^3 = 500 | N_6^1 + 2N_6^2 + 3N_6^3 = 1000, N_6^2 = N_6^3 = 0) = \int_5^6 \binom{1000}{500} \left(\frac{t}{6}\right)^{500} \left(1 - \frac{t}{6}\right)^{500} dt$$

Example 19. [5: p. 122] Suppose that people immigrate into a territory at a Poisson rate of $\lambda = 1$ per day.

What is the expected time until the 10th immigrant arrives? This is simply $E[S_{10}] = 10 \cdot \frac{1}{\lambda} = 10$.

What is the probability that the elapsed time between the 10th and 11th arrivals is greater than 2 days? This is $P(T_{11} > 2) = e^{-2\lambda} = e^{-2}$.

Example 20. [5: p. 124] If immigrants to area A arrive at a Poisson rate of ten per week, and if each immigrant is English descent with probability $\frac{1}{12}$, then what is the probability that no people of English descent will emigrate to area A during the month of February?

Note first that February consists of 4 weeks, giving $t = 4$, and we also have $\lambda = 10$, $p = \frac{1}{12}$, and seek the probability that exactly $k = 0$. This is given by the PMF of the Poisson sum:

$$P(S_n = 0) = e^{-10\cdot\frac{1}{12}\cdot 4}\frac{(10 \cdot \frac{1}{12} \cdot 4)^0}{0!}$$

$$P(S_n = 0) = e^{-10\cdot\frac{1}{12}\cdot 4}$$

Example 21. [5: p. 132] Suppose that families immigrate to some area at a Poisson rate $\lambda = 2$ per week. If the number of people in each family are independent with values $\{1, 2, 3, 4\}$ with respective probabilities $\{\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\}$, then what is the expected value and variance of the number of individuals migrating to this area during a fixed five week period?

The expected value of the compound Poisson process with $t = 5$ and $\lambda = 2$ is

$$E[S_n] = 5 \cdot 2\left(\frac{1}{6} \cdot 1 + \frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3 + \frac{1}{6} \cdot 4\right)$$

While the variance is given by

$$Var(S_n) = 5 \cdot 2\left(\frac{1}{6} \cdot 1^2 + \frac{1}{3} \cdot 2^2 + \frac{1}{3} \cdot 3^2 + \frac{1}{6} \cdot 4^2\right)$$

8 Continuous-time Markov chains

Now we extend our study of Markov chains from chains with a discrete time index $n = 0, 1, 2, \cdots$ to chains with a continuous time parameter $t \geq 0$, but still with discrete state spaces.

Remark 15. We will also call these **Markov processes**.

Updating the discrete Markov property given in Definition 1, we define continuous Markov chains as follows:

Definition 20. [2: p. 139] X_t is a Markov chain, with $t \geq 0$, if for any $0 \leq s_0 < s_1 < \cdots < s_n < s$ and possible states $i_0, \cdots, i_n, i, j$ it holds that

$$P(X_{t+s} = j | X_s = i, X_{s_n} = i_n, \cdots, X_{s_0} = i_0) = P(X_t = j | X_0 = i)$$

Remark 16. [4] Say that at time t, we know that $X(t) = i$. If X is a Markov chain, then our capacity to predict what it will do in the future should not be improved by information about how long the process has been in state i; it needs to be independent of history. The only continuous distribution that has this property is the exponential distribution, so the amount of time that a continuous-time Markov chain spends in some state i will need to have distribution $exponential(\lambda_i)$, where λ_i is some nonegative real number. We also assume that $\exists M < \infty$ so that $\lambda_i < M \ \forall i \in S$.

Remark 17. [4] We can consider a continuous-time Markov chain as being composed of two pieces: a discrete-time Markov chain (called a **jump chain** or an **embedded chain**), which provides the probability of transitions between each state, together with a holding time λ_i which governs how much time is spent in every state.

Definition 21. [2: p. 140] In discrete time Markov chains, Definition 2 of discrete transition probabilities $p(i, j)$ captured the probability of jumping from state i to state j in any one step. In continuous time there is no first time $t > 0$ to base our definition on, so instead introduce for each $t > 0$ a **transition probability** which represents

$$p_t(i, j) = P(X_t = j | X_0 = i)$$

Theorem 20. *[2: p. 140] In continuous time, Markov transition probabilities satisfy the* ***Chapman-Kolmogorov equations****:*

$$\sum_k p_s(i, k) p_t(k, j) = p_{s+t}(i, j)$$

Remark 18. Theorem 20 is the continuous-time analog of Definition 4.

Definition 22. [2: p. 149] A continuous Markov chain X_t is called **irreducible** if for any two states i and j there exists a sequence of states $k_0 = i$, $k_1, \ldots, k_n = j$ so that $q(k_{m-1}, k_m) > 0$ for $1 \leq m < n$

Theorem 21. *[3: Lecture 13] A continuous markov process is irreducible iff its embedded Markov chain is reducible.*

Definition 23. [2: p. 149] π is called a **stationary distribution** of a continuous Markov chain if $\pi p_t = \pi \ \forall t > 0$

Definition 24. [3: Lecture 12] For a regular Markov process with jump rates $(q(i,j))$, its **generator** is a matrix $Q = (Q(i,j))$, defined as

$$Q(i,j) = \begin{cases} q(i,j) & j \neq i \\ -\lambda_i = -\sum_{j \neq i} q(i,j) & j = i \end{cases}$$

So that

1. The off-diagonal entries of the generator are non-negative
2. The diagonal elements of the generator are non-positive
3. The entries in each row sum to zero

Remark 19. [3: Lecture 12] Any square matrix satisfying Definition 24 is the generator of some Markov process.

Theorem 22. *[3: Lecture 13] π is a stationary distribution of a Markov process with generator Q iff*

$$\pi Q = 0$$

Theorem 23. *[3: Lecture 13] An irreducible Markov process, with a finite state space, has a unique stationary distribution, and all of the entires in this distribution are strictly positive.*

Theorem 24. *[3: Lecture 13] All continuous-time Markov processes are aperiodic.*

Theorem 25. *Assume that an irreducible Markov process, with transition probabilities* $(p_t(i,j))$*, has a stationary distribution* π*. Then, for all states* i *and* j*, we have:*

$$\lim_{t\to\infty} p_t(i,j) = \pi(j)$$

So

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X_s=j\}} ds = \pi(j)$$

Example 22. [2: p. 150] Say that the weather in Los Angeles has three states: sunny, rainy, and smoggy, and that it changes as follows:

- It is sunny for an exponential number of days, with mean 3, then it becomes smoggy
- It is smoggy for an exponential number of days with mean 4, then it becomes rainy
- It is rainy for an exponential number of days with mean 1, then it becomes sunny

First, we write the generator matrix. To do so, we will use the fact that the rate of jump *away* from each state is given by the rate of the corresponding exponential distribution, and the rate is the inverse of the distribution's mean [3: Lecture 13]. That provides the entries in the principal diagonal. Then, rule 2 of Definition 24, together with the fact that there is only one non-zero off-diagonal in each row, to write the off-diagonals.

$$\begin{bmatrix} -\frac{1}{3} & \frac{1}{3} & 0 \\ 0 & -\frac{1}{4} & \frac{1}{4} \\ 1 & 0 & -1 \end{bmatrix}$$

To find the stationary distribution, work in [2: p. 150] and [3: Lecture 13] show that it is sufficient to solve the system

$$\pi = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} A^{-1}$$

Where A is the matrix obtained by replacing the last column in Q by a column of ones, so

$$A = \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} & 1 \\ 0 & -\frac{1}{4} & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Solving gives $\pi_1 = \frac{3}{8}, \quad \pi_2 = \frac{4}{8}, \quad \pi_3 = \frac{1}{8}$.

To also find the long-run fraction of sunny days, we use the procedure in Theorem 25, as:

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X_s=j\}} ds = \pi(1)$$

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X_s=j\}} ds = \frac{3}{8}$$

Example 23. [2: p. 150] Take the transition probability matrix

$$\begin{bmatrix} -3 & 2 & 1 & 0 \\ 0 & -5 & 5 & 0 \\ 1 & 0 & -2.5 & 1.5 \\ 6 & 0 & 0 & -6 \end{bmatrix}$$

Then the stationary distribution is

$$\pi = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -3 & 2 & 1 & 1 \\ 0 & -5 & 5 & 1 \\ 1 & 0 & -2.5 & 1 \\ 6 & 0 & 0 & 1 \end{bmatrix}^{-1}$$

The result is $\pi_1 = \frac{10}{29}, \quad \pi_2 = \frac{4}{29}, \quad \pi_3 = \frac{12}{29}, \quad \pi_4 = \frac{3}{29}$.

Definition 25. [2: p. 152] The continuous-time analog of the detailed balance condition in Definition 17 is

$$\pi(k)q(k,j) = \pi(j)q(j,k) \quad \forall j \neq k$$

Theorem 26. *[2: p. 152] If Definition 25 holds, then π is a stationary distribution.*

Example 24. [2: p. 153] Consider a birth and death chain on $S = \{0, 1, \ldots, N\}$, $N \leq \infty$ so that

$$\begin{cases} q(n, n+1) = \lambda_n & n < N \\ q(n, n-1) = \mu_n & n > 0 \end{cases}$$

Then we can establish recursively that

$$\pi(n) = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1}\pi(0)$$

Example 25. [3: Lecture 13] and similarly [2: p. 154] A barber cuts hair with a rate of 3 people per hour. Equivalently, the time of each haircut is an independent random variable with rate 3, or mean $\frac{1}{3}$. The shop has 2 chairs. Customers arrive according to a Poisson process with rate 2 per hour. If there are no chairs currently, the customer leaves. Let X_t be the number of customers in the shop (either waiting or having their hair cut) at time t, so $S = \{0, 1, 2, 3\}$. What is the stationary distribution of X_t?

We want to begin by constructing a generator for X_t. Such a generator is

$$\begin{cases} q(n,n+1) = 2 & X_t = 0,1,2 \\ q(n,n-1) = 3 & X_t = 1,2,3 \end{cases}$$

So the corresponding matrix is

$$\begin{bmatrix} -2 & 2 & 0 & 0 \\ 3 & -5 & 2 & 0 \\ 0 & 3 & -5 & 2 \\ 0 & 0 & 3 & -3 \end{bmatrix}$$

Then the matrix A is given by

$$A = \begin{bmatrix} -2 & 2 & 0 & 1 \\ 3 & -5 & 2 & 1 \\ 0 & 3 & -5 & 1 \\ 0 & 0 & 3 & 1 \end{bmatrix}$$

But inverting A looks like a pain (although it would certainly work), and we know that this is a birth and death process as in Example 24. So, we can use the detailed balance condition from Definition 25, as:

$$2\pi(0) = 3\pi(1), \quad 2\pi(1) = 3\pi(2), \quad 2\pi(2) = 3\pi(3)$$

Which after some effort produces the solutions

$$\pi(0) = \frac{27}{65}, \quad \pi(1) = \frac{18}{65}, \quad \pi(2) = \frac{12}{65}, \quad \pi(3) = \frac{8}{65}$$

Theorem 27. *[3: Lecture 13] Let (X_t) be an irreducible Markov process, with stationary distribution π, and let (N_t) be a Poisson process with rate λ. Assume also that state i has the following property: whenever (X_t) starts from i, the time it spends there, until the first jump away, is independent of (N_t). Denote by $\{T_k\}$ the arrival (jumping) times of (N_t) and by $X_{T_k^-}$ the value of the Markov process just before T_k. Then, $\mathbb{1}_{\{X_{T_k^-}=i\}}$ takes one value iff the jump time T_k occurs at the time when the Markov process is at state i, and we have:*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{T_k \le t} \mathbb{1}_{\{X_{T_k^-}=i\}} = \pi(i)\lambda$$

$$\lim_{t\to\infty} \frac{1}{t} N_t = \lambda$$

$$\lim_{t\to\infty} \frac{\sum_{T_k \le t} \mathbb{1}_{\{X_{T_k^-}=i\}}}{N_t} = \pi(i)$$

Example 26. [2: p. 154] A factory has 3 machines in use and one repairman. Suppose each machine works for an exponential amount of time with mean 60 days between breakdowns, but each breakdown requires an exponential repair time with mean 4 days. What is the long-run fraction of time all three machines are working?

Let X_t be the number of working machines. Since there is one repairman we have the following generator:

$$\begin{cases} q(n,n+1) = \frac{1}{4} & 0,1,2 \\ q(n,n-1) = \frac{1}{60} & 1,2,3 \end{cases}$$

Setting $\pi(0) = c$ where c is some constant and applying the recursion formula from Example 24 we have

$$\pi(1) = \frac{\lambda_0}{\mu_1}\pi(0)$$

$$\pi(1) = \frac{\frac{1}{4}}{\frac{1}{60}}c$$

$$\pi(1) = 15c$$

Then

$$\pi(2) = 225c$$

$$\pi(3) = 1125c$$

Then solve back for c with $\pi(0) + \pi(1) + \pi(2) + \pi(3) = 1$ yielding $\pi(3) = \frac{1125}{1382}$ $\pi(2) = \frac{225}{1382}$ $\pi(1) = \frac{30}{1382}$ $\pi(0) = \frac{2}{1382}$.

Definition 26. [3: Lecture 14] The **time of the first visit** to state i (also called the **hitting time** of state i) is

$$V_i = \min\{t \geq 0 : X_t = i\}$$

Definition 27. [3: Lecture 14] The **time of the first visit** to a set of states $A \subset S$ (also called the **hitting time** of set A) is

$$V_A = \min\{t \geq 0 : X_t \in A\}$$

9 Martingales

Definition 28. [3: Lecture 16] A discrete time stochastic process $(M_n)_{n=0}^{\infty}$ is a **martingale** with respect to a sequence of random variables $(X_n)_{n=0}^{\infty}$ if

$\hookrightarrow (M_n)_{n=0}^{\infty}$ is adapted to $(X_n)_{n=0}^{\infty}$, so $M_n = f_n(X_1, \ldots, X_n)\ \ \forall n$, where f_n is some deterministic function

$\hookrightarrow E|M_n| < \infty\ \ \forall n \geq 0$

$\hookrightarrow E[M_{n+1}|\mathcal{F}_n^X] = M_n\ \ \forall n \geq 0$

Definition 29. [3: Lecture 17] Consider a stopping time τ wrt rvs $(X_n)_{n=0}^{\infty}$, which is a random variable with values in $\mathbb{N}$ such that $\forall n \in \mathbb{N}$, $\mathbb{1}_{\tau \leq n}$ is a function of $X_1, \ldots, X_n$. Then, we can define the stopped martingale obtained by calculating the value of a martingale stopped at τ, as:

$$M_{n\wedge\tau} = M_n \mathbb{1}_{n<\tau} + M_\tau \mathbb{1} \geq \tau$$

Theorem 28. *[3: Lecture 17] Optional Sampling Theorem: Let $(M_n)_{n=0}^{\infty}$ be a Martingale, and let τ be a stopping time with respect to the same information. Then the stopped process $(M_{n\wedge\tau})_{n=0}^{\infty}$ is a martingale as well. In particular, $E[M_{n\wedge\tau}] = E[M_0]$. Analogous results hold for submartingales and supermartingales.*

Theorem 29. *[3: Lecture 17] Dominated Convergence Theorem: Let the random variables X, Z, and $\{X_n\}_{n=1}^{\infty}$ satisfy $|X_n| \leq Z$ for all $n \geq 1$ and with $E[Z] < \infty$, and $\lim_{n\to\infty} = X$ with probability 1. Then, $\lim_{n\to\infty} E[X_n] = E[X]$.*

Theorem 30. *[3: Lecture 18] Monotone Convergence Theorem: Let X and $\{X_n\}$ be random variables, such that, with probability one: $X_n \leq X_{N+1}$ for each n, and $\lim_{n\to\infty} X_n = X$. Then, $\lim_{n\to\infty} E[X_n] = E[X]$.*

Theorem 31. *[3: Lecture 18] Wald's identity: Let (X_n) be iid rvs, with mean μ. Let τ be a stopping time with respect to (X_n), such that $E[\tau] < \infty$. Consider*

$$S_n = \sum_{i=1}^{n} X_i$$

for all $n \geq 1$. Then,

$$E[S_\tau] = \mu E[\tau]$$

Theorem 32. *[2: p. 200] Maximal Inequality for Supermartingales: Let* (X_n) *be a supermartingale, such that* $X_N \geq 0\ \forall n$*, and consider some arbitrary constant* $\lambda > 0$*. Then*

$$P(\max_{n \geq 0} X_n > \lambda) \leq \frac{E[X_0]}{\lambda}$$

Theorem 33. *[3: Lecture 18] Doob's maximal inequality: Let* (X_n) *be a submartingale, such that* $X_n \geq 0 \ \ \forall n$*, and consider arbitrary constants* $\lambda > 0$ *and* $p \geq 1$*. Then,*

$$P(\max_{0 \leq k \leq n} X_k > \lambda) \leq \frac{E[X_n^p]}{\lambda^p}$$

References

[1] Erhan Çinlar. *Introduction to stochastic processes*. Dover publications, 1975.

[2] Richard Durrett. *Essentials of stochastic processes*. Springer, 2012.

[3] Sergey Nadtochiy. Lecture notes for math 526: Stochastic processes. 2017.

[4] Hossein Pishro-Nik. *Introduction to probability, statistics, and random processes*. 2014.

[5] Sheldon M. Ross. *Introduction to probability models*. Academic Press, 1972.