

3 Discrete Stochastic Processes (Markov Process)

3.1 Probability Review

To work with a stochastic driven situation we first need to understand a few things about probability. We'll start with some terminology.

The source of the randomness is called a **Trial**. We usually associate it with some action, like rolling a die, flipping a coin, spinning a spinner, generating a random number, etc. For our situations, we'll associate the result of a trial with a number and then let $S \subset \mathbf{R}$ be the set of all possible outcomes. For example, with a regular six-sided die, $S = \{1, 2, 3, 4, 5, 6\}$, for a fair coin $S = \{0, 1\}$ with 0 for tails, 1 for heads, and for a spinner $S = [0, 2\pi)$ representing the measure of the angle (in radians). The set S is called the **Sample Space**. Any subset of S is an **Event**, while a single element is called an **Outcome**. In terms of notation, we use capital letters, like X to denote an outcome of a trial before it happens and lowercase, e.g. x , for the result of a trial. So X is a **Random Variable** as initially all we know is that $X \in S$.

To work with such random processes we need more information about the random variable X ; so we introduce a **Probability Function** P . P is a function defined on subsets of the sample space S that satisfies certain conditions. The basic way to construct P is to divide S up into disjoint parts, i.e. write $S = \cup_k U_k$ where $U_i \cap U_j = \emptyset$ when $i \neq j$, and then set $P(U_i) = p_i$ such that $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$. For example, with a fair six-sided die, we set $U_k = \{k\}$, $k = 1, 2, \dots, 6$, and set $P(U_k) = P(\{k\}) = \frac{1}{6}$. This function P is additive over disjoint unions, so, for example $P(U_1 \cup U_3) = P(U_1) + P(U_3)$. This means we assume the events represented by the U_i are independent. In terms of notation, for an event $U = \{k\}$, i.e. a single outcome, we may write $P(U)$ or $P(k)$ or $\text{Prob}(X = k)$ or $\text{Prob}(k)$. In some cases we may use P and Prob interchangeably.

One of the most important results in probability is the idea of **Conditional Probability**. In this case you want the probability of one event under the circumstance that you know another event has occurred. This is not a cause and effect situation, but one where you are restricting the sample space to some subset. For example, the probability of rolling a 2 on a die is $\frac{1}{6}$, but the probability of rolling a 2 on a die, if you know that the result is even is $\frac{1}{3}$. Let $U, V \subset S$, i.e. let U and V be events, then for the probability of U restricted to V (or under the condition of V), is written as $P(U | V)$ or $\text{Prob}(X \in U | X \in V)$. It is calculated as

$$P(U | V) = \frac{P(U \cap V)}{P(V)}.$$

So in our dice example

$$\text{Prob}(X = 2 | X \in \{2, 4, 6\}) = \frac{\text{Prob}(X = 2)}{\text{Prob}(X \in \{2, 4, 6\})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

This helps us figure out more complex probabilities by simply considering various cases. If $S = \cup U_k$ where the U_k are disjoint events, then for any event $U \subset S$, we have

$$P(U) = \sum_i P(U | U_k)P(U_k).$$

In the simplest cases $P(U | U_k)$ is either 0 or 1.

We distinguish in probability two sub-cases based on the sample space S : Either S is discrete, i.e. $S = \{x_1, \dots, x_n\}$ (finite) or $S = \{x_1, x_2, \dots\}$ (infinite), or it is continuous (or mixed), i.e. S contains a non-empty interval of values. The main difference between the two types is how we make certain calculations. In the discrete case we use sums, and in the continuous case, integrals. There are also a few notational differences. In the discrete case we are typically interested in $\text{Prob}(X = x_k)$ and we denote this by π_k with $\pi = (\pi_1, \dots)$ denoting the probability distribution for our sample space. The conditions on P , let us know that $0 \leq \pi_k \leq 1$ and $\sum_k \pi_k = 1$. We can also talk about a **Cumulative Distribution** $F(j) = \text{Prob}(X \leq x_j)$, assuming the x_k are in increasing order. Then we have $F(j) = \sum_{k=1}^j \pi_k$. If we have a value v_k for each event, like a cost or a reward, then we can talk about the **Expected Value** for a random variable X , by calculating $E[X] = \sum_k v_k \pi_k$.

In the continuous case, the probability of individual outcomes is 0, so we have to work with intervals, and thus instead of working with a distribution π , we work with a **Probability Density Function** f . f is defined by the relationship that $\text{Prob}(a < x \leq b) = \int_a^b f(x) dx$. From the properties of probabilities, we know that $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x) dx = 1$. The cumulative distribution is calculated as $F(b) = \int_{-\infty}^b f(x) dx$ and if $v(x)$ is the value associate with the outcome x , we have the expected value $E[X] = \int_{-\infty}^{\infty} v(x) f(x) dx$.

3.2 Discrete Stochastic Process

Now instead of a single discrete random variable X , consider a sequence of random variables $\{X(k)\}$. Just like with a DDS, we assume that $X(k+1)$ depends on prior values, the main difference being that the dependence is stochastic. As an example, let $X(n)$ be the total of rolling n fair six-sided dice. Then, if we had $X(k)$ for some k , then we'd know that $X(k+1)$ would be one of the values $X(k) + 1, \dots, X(k) + 6$ with equal probability. We could even write this like a DDS as $X(k+1) = X(k) + r$ where r is a random value from $\{1, 2, 3, 4, 5, 6\}$.

In a Stochastic Process, we are not interested in the value of $X(k)$, but in the distribution of the possible values of $X(k)$. Thus we study $\pi(k) = (\pi_j(k))$ where $\pi_j(k) = \text{Prob}(X(k) = x_j)$. If we write out the possible changes for $\pi(k)$ using our result from conditional probability, we get

$$\pi_j(k+1) = \sum_i \text{Prob}(X(k+1) = x_j \mid X(k) = x_i) \text{Prob}(X(k) = x_i).$$

To simplify the notation, let $T(k)$ be a matrix with entries $T_{j,i} = \text{Prob}(X(k+1) = x_j \mid X(k) = x_i)$ and then our model for the changes in $\pi(k)$ is simply

$$\pi_j(k+1) = \sum_i T_{j,i}(k) \pi_i(k) \text{ or } \pi(k+1) = T(k) \pi(k).$$

The matrix $T(k)$ is called the **Transition Matrix** for this process.

One of the first simplifications that we can make is to assume that the transition matrix doesn't change over time, i.e. $T(k) = T(0)$ for all k . Another way to think about this assumption is that it says that the possible transitions only depend on the current status and does not depend on when or how we got there. Although this assumption seems excessive, it is actually quite appropriate in many cases. For example, in almost all games, your options when it is your turn do not depend on how many turns you've had or how you got to your current spot on the board. You just roll the dice, spin a spinner or draw a card and make the appropriate move from where you are. This

assumption is also reasonable when modeling most sports games (at least for short term models) and for many biological and physical models. This assumption is called the **Markov** assumption and the resulting process is called a **Markov Process**.

3.3 Building a Markov Process

As a result of this simplification, our model building effort only involves construction the transition matrix T . One tool for doing this is to use a **State Diagram**. A state diagram is like a box model except that we use circles instead of rectangles which now represent states (or outcomes) that can occur, we can draw arrows from a state back to itself, and it is assumed that each state influences only the flow out of itself. The fundamental difference between the two is that in a state model we are marking out the possible paths of a single entity rather than looking at changes in a large population.

We start with the sample space S which represents all the possible outcomes (states) we can be in over all time (or at least over the time frame we are considering). This can be somewhat challenging as we need to make sure we have all the possibilities and that we've organized them in some logical way to make constructing T easy. For example if we are looking at the total of n dice rolls, then we would take $S = \{0, 1, 2, 6N\}$ where N is the maximum number of rolls we will consider. We include 0 as we have a total of 0 before we roll any dice. For a board games, we would set S to all the possible places on the board. For a social or biological situation, S would be all the possible states, values or locations. Large and complex sets of states are possible, for example, if we had a 4×4 grid of locations and 3 different states for each location, we'd have 48 different possible states. If we are modeling a sporting event where each team could score 0 – 20 points, we could have $21^2 = 441$ possible states. In some physical models, we could easily have millions of states.

Once we have the sample space S , we want to order or number the states from 1 to N so that we can construct the appropriate transition matrix T . Sometimes it is easy to number the states as they have a natural order. If they don't have a natural order, then you have to think of some scheme that easily maps the different states to the natural numbers. There are two techniques that can help with this. First, sometimes it helps to include extra states that will never be reached, but help you number. For example, suppose you are looking at a sports playoff which consists of 5 games and thus the winner is the first to win 3 games. Thus final results like 5-0 would never occur as the series would stop at 3-0. However, it might make it easier to say the states are the 16 possibilities where each has 0 – 3 wins or the 36 possibilities where each has 0 – 5 wins. Another case is if the states represent different physical locations, then you may want to extend the locations to a rectangular grid as it is easier to number. The second technique is for multidimensional states, i.e. you can most easily represent a state with an ordered pair (p, q) (like our sports or location examples). If the states are of the form (p, q) with $0 \leq p, q \leq M$ then you can order the $(M + 1)^2$ states by assigning to the state (p, q) the value $p + q * (M + 1) + 1$. This can be extended to higher dimensions.

Now we have our sample space and it is numbered from 1 to N . For our state diagram, we draw N circles, numbering them from 1 to N and, if possible, labeling them with their state. You may have to rearrange you circles as you figure out how they are related. Then from the mechanisms in the system of interest, draw appropriate arrows from state to state. If the probability is known for that transition, label the arrow with that value. Since some action always takes place, if the total of all the probabilities moving out of a state is less than one, then we include an arrow back into