Contents

The book is optimized for looking up facts. However, it contains pointers to the end of the books that give proof. I am debating whether or not to move the examples into a separate section as well. For now they are kept with the facts.

1 Probability

1.0.0.1 Sample Space and Outcome We perform random experiments and the sample space is the set of possible outcomes.
For example, consider rolling a die. The set of possible outcomes are:

$$S = \{1, 2, 3, 4, 5, 6\}$$

1.0.0.2 Event An event is a subset of the sample space. An example event is rolling a die and getting an even odd outcome:

$$E = \{1, 3, 5\}$$

1.0.0.3 Disjunction of Events The event E occurs if E_1 or E_2 occur. Another way to imagine this is the union of events: $E = E_1 \cup E_2$.

1.0.0.4 Conjunction of Events The event E occurs if E_1 and E_2 occur. Another way to imagine this is the intersection of events:
$E = E_1 \cap E_2$. Some alternative ways of writing this are:

$$P(E) = P(E_1 \cap E_2)$$
$$P(E) = P(E_1 \wedge E_2)$$
$$P(E) = P(E_1 \, , E_2)$$
$$P(E) = P(E_1 \, E_2)$$

1.0.0.5 Mutually Exclusive Events Events E_1 and E_2 are mutually exclusive if only one of them can occur in a single experiment. For example, the event rolling an even number and the event rolling an odd number on a die are mutually exclusive events:

$$E_{even} \cap E_{odd} = \{2, 4, 6\} \cap \{1, 3, 5\} = \varnothing$$

1.0.0.6 Axioms of Probability The are the rules we accept as truth without proof. We build probability untop of these axioms.

1. $0 \leq P(E) \leq 1$, for any event E. In the smallest case, the event cannot occur which is inidicated by a probability of 0. In the largest case, the event always occurs, which is indicated by the probability of 1.

2. $P(S) = 1$, where S is the sample space. The sample space contains all possible outcomes for each experiment. It's reasonable to accept that an event from the sample space always occurs.

3. For a potentially infinite set of mutually exclusive events E_1, E_2, ...

   $$P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

   It makes senses that events that do not share outcomes for a single event, can have their probabilities added to arrive at the probability of combining the outcomes from the events.

1.0.0.7 Properties From the above axioms, we get the following useful properties (TODO proof):

1. For any event E, let $\overline{E}$ be the complement of E. More concretely, $\overline{E} = S - E$, where S is the sample space. Then E and $\overline{E}$ are mutually exclusive.

2. $P(\varnothing) = 0$ You can never get none of the outcomes of the sample space.

3. If E_1 and E_2 are mutually exclusive events then

   $$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1\,, E_2)$$

1.0.0.8 Conditional Probability

We use conditional probability to model the probability given knowing some circumstance has happened. Given two event E and F, the conditional probability, the probability of F given E has occured, is defined as $(P(E) \neq 0)$:

$$P(F|E) = \frac{P(E, F}{P(E)}$$

An example is what is the probability of rolling a 3, given that we rolled an odd number. Let $F = \{3\}$ and $E = \{1, 3, 5\}$:

$$P(F|E) = \frac{P(E, F)}{P(E)} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2}$$

1.0.0.9 Joint Probability

In queing theory, we often have to use multiple sample sapce. The theory in this book so far has covered only a single probability space.
Suppose we have two sample spaces S_1 and S_2. The outcomes of the joint probability space are the tuples that result from the cross product of the two sample spaces:

$$S_{joint} = S_1 \times S_2$$

For example, consider rolling a die and a coin

$$\begin{aligned}
S_{die} =& \{1, 2, 3, 4, 5, 6\} \\
S_{coin} =& \{H, T\} \\
S =& S_{die} \times S_{coin} \\
S =& \{(1, H), (2, H), (3, H), (4, H), (5, H), (6, H), \\
& (1, T), (2, T), (3, T), (4, T), (5, T), (6, T)\}
\end{aligned}$$

1.0.0.10 Marginal Probability

Given the joint probabilties, we might want to compute the probabilties of only one of the sample spaces.
For example, suppose that we know the joint probability of the number of jobs at server 1 and server two and we want to compute the probability of the number of jobs at server 1 only. We can apply Marginal probability to determine that.

2 Stochastic Processes

A family of random variables, indexed by time

2.0.1 Classifications

2.0.1.1 State Space The set of possible values (states).

2.0.1.2 Discrete State Space Example: the number of jobs in the system. ($S = 0, 1, 2, 3...$). We will only deal with the discrete case in this class. To make notation easier the state is usually identified by the number.

2.0.1.3 Continuous State Space Example: motion of a particle. We will not be studying this in this class.

2.0.1.4 Time Parameter There are two ways to observe times in Stochastic processes.

2.0.1.5 Discrete Time Parameter For example, we consider the states at X_0, X_1, ... X_n. For example, looking at the state of the system i^{th} hour.

2.0.1.6 Continuous Time Parameter The states are function of time t ($X(t)$).

2.1 Discrete State Space and Time

There might be a dependency between the previous time interval X_i's and the states those time interval can be in that need to be model in the current X_n.

$$P(X_{n+1} = j | X_n = i_n \, , X_{n-1} = i_{n-1} \, , ... \, , X_1 = i_1 \, , X_0 = i_0)$$

The number of dependency combinations is exponential because
X_{n+1} depends on X_n to X_0
and X_n depends on X_{n-1} to X_o
and so on.

2.1.0.7 Markov Chain As a result of the exponential size, we make a simplifying assumption. We only use the latest information. X_{n+1} only depends on X_n. Now we are left with transition probabilities:

$$P(X_{n+1}|X_n) = P(X_{n+1} = j|X_n = i_n\,, X_{n-1} = i_{n-1}\,, ...\,, X_1 = i_1\,, X_0 = i_0)$$

Given $P(X_0 = i)$ for all i's, we can compute any state. However, notice that the formula depends on n, the discrete time that has pasted so far, which make analysis difficult still. For example, in the 9am one hour interval the number of jobs in a login system tends to be higher than at 2am.

2.1.0.8 Homogeneous Markov Chains We now make the assumption that that the transistion probabilities do not depend on time. For example, the transition probabilies for the number of webcrawling robots requesting a webpage remain the same despite the time. So we can write

$$P(X_{n+1} = j|X_n = i) = P(X_n = j|X_{n-1} = i), \forall n\, i\, j \geq 0$$

Which is abbreviated to P_{ij}

2.1.0.9 State Transistion Diagram TODO draw diagram and place it here

2.1.0.10 Chapman-Kolmogorov Equation Let $P_{ij}^{(m)}$ be the m step transistion probability from state i to j defined as:

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

To take m steps to go from (possibly visiting m in an intermediate state multiple itmes), you can take m-1 to steps. At step m-1, you can arrive at any state k. To go from each state k at the m-1 step to the m step, you just apply the transisition probability.
We can sum the probabilities because the probabililites of going from state k in the m-1 step to state j are mutually exclusive (they are mutually exclusive because they are different states).
Lastly, we can mulitple the probability of going to state k in m-1 steps by the probability of going to state j because the probabilities are independent.

$$P_{ij}^{(m)} = \sum_k p_{ik}^{(m-1)} p_{kj}$$

Which is exactly the same as taking 1 step, and then m-1 steps.

$$P_{ij}^{(m)} = \sum_k p_{ik} p_{kj}^{(m-1)}$$

For a derivation of this, see (3.2.0.19).

2.1.0.11 Irreducible Markov Chain allows every state to be reached from every other state for all pairs of states i and j. More concretely:

$$\forall i \forall j \neq i : \exists m_{ij} : P_{ij}^{(m_{ij})} > 0$$

2.1.0.12 Recurrent State : State j is recurrent if after leaving state j then you are guarenteed to eventually comeback.
Let $f_j^{(n)}$ be the probability that you first return to state j in n steps.
Notice that $f_j^{(0)} = 0$ because it's impossible to comeback without taking any steps. Also notice that $f_j^{(1)} > 0$ is only possible is the transistion pointing to itself $P_{jj} > 0$.
j is recurrent if and only if

$$f_j = \sum_{n=1}^{\infty} f_j^{(n)} = 1$$

2.1.0.13 Recurrent Non-null A state j is recurrent non-null if and only if we get back to state j, but it does not take forever. More concretely:

j is recurrent non-null if and only if j is recurrent ($f_j = 1$) and

$$M_j = \sum_{n=1}^{\infty} n f_j^{(n)}$$

Where M_j is the expected number of steps to come back.

2.1.0.14 Recurrent Null j is recurrent null if and only if j is recurrent and j is not recurrent non-null.
TODO EXAMPLE WITH DIAGRAM
TODO EXAMPLE WITH DIAGRAM

2.1.0.15 Periodicity A state j is periodic if and only if the only way to come back to state j is to take r, 2r, 3r, ... , cr, steps.

If a state j is not periodic, it's called aperiodic

If state j as a self loop ($p_{jj} > 0$), then state j is aperiodic

If the system is a irreducible Markov Chain, and contains a self loop, then all states j are aperiodic.

2.1.0.16 State Probability Let X_n be the random variable for the state at interval n, then in a homogeneous Markov Chain we have that:

$$\pi_j^{(n}) = P(X_n = j) \text{ - at step j}$$

Let p_{ij} be the transition probability for going from state i to state j (independant of time, so it's the same for all intervals (X_n), then

$$\pi_j^{(n+1)} = \sum_{i=0}^{\infty} p_{ij}\pi_i^{(n)} \text{ (by applying total probability)}$$

Given initial conditions $\pi_j^{(0)}\forall j$, we can compute all $\pi_j^{(i)}$ apply the above formula recursively.

2.1.0.17 Equilibrium Probability Therom (I made up this name. It seems better than 'Fundamental Theorm.)
If a homogeneous Markov Chain is irreducible and a periodic, then there exists a limiting probability (equilibrium):

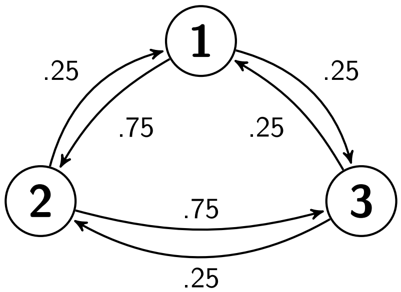
$$\pi_j = \lim_{n\to\infty} \pi_j^{(n)}$$

which is independant of the initial conditions $\pi_j^{(0)}$.

Moreover, if all states j are recurrent non-null, then π_j is non zero and can be uniquely etermined from the equations:

$$\pi_j = \sum_i p_{ij}\pi_i, \forall j$$

$$\sum_j \pi_j = 1$$



Let the initial state be (1). This is the 'initial condition'

n	0	1	2	3	4	...	∞
$\pi_1^{(n)})$	1	0	.25	.187	.203		.2
$\pi_2^{(n)})$	0	.25	.062	.359	.254		.28
$\pi_3^{(n)})$	0	.75	.688	.454	.543		.52

To get the steady state column, you have to solve the system of equations that results from the above theorm.

2.1.0.17.1 Example

2.1.0.18 Memoryless Property

, the residence time in a state.

3 Derivations

3.1 Probability

3.2 Stochastic Processes

3.2.0.19 Chapman-Kolmogorov Equation For context see (2.1.0.10). TODO