CS 2800: Discrete Math Oct. 19, 2011

Lecture

Lecturer: John Hopcroft *Scribe:Doo San & June*

Review

Note: the next homework will have a thought problem at the bottom of each homework. No credit given for answer.

Basic Probability Terms

These are the terms you need to know to understand the basic principles of probability.
Probability Space : All possible outcomes
Probability Mass Function : The probability that the outcome i happens
Event : A subset of Probability Space

Example

Let's think about the situation that we are rolling a die.
Probability Space : $S = \{1, 2, 3, 4, 5, 6\}$
Probability Mass Function : $f(i) = \frac{1}{6}$
Event : $\{1, 2, 3\}$

Note

1. Probability of a particular event E : $P(E) = \sum_{i \in E} f(i)$

2. Birthday problem can be understood with the similar concept with collisions in hashing. It can be roughly worked out to $\sqrt{n}$, where n is the number of your buckets. In the birthday problem, it would be 365 buckets, or roughly 20 people until someone has the same birthday.

Continuous Probability

In the situation described above, there are a finite number of outcomes with weight, $f(i) > 0$. We call these probabilities discrete. Now, we need to think about infinite sample spaces - i.e. spaces where an infinite number of events could happen. In this case, each point source must have weight 0. Before we move on, let's think about why. We can assume that the outcomes in this case are all equally likely. The probability of each of these events can be formulated as $\frac{1}{N(S)} = \frac{1}{\infty} = 0$. Therefore, we need to introduce the new concept of probability density function of which the area under the curve is the probability of the outcomes in the certain range.

The difference in probability mass function $f(x)$ and the probability density function $P(x)$ can be summarized as below.

1. $f(i)$: The probability that the particular outcome i happens

2. $\int_a^b P(x)dx$: The probability that the outcomes between a and b happen.

3. $F(x) = \int_{-\infty}^{x} P(x)dx$: Cumulative distribution function; the probability that all the outcomes less than x happen. In this sense, $\int_{a}^{b} P(x)dx$ can be reformulated as $F(b) - F(a)$.

With all the prior knowledge mentioned above, now we can move onto the event in a certain range. First of all, we can march through the example of the rationals in the range 0 to 1.

Example 1 : Positive Integers (Discretely Infinite Case)

Pick an integer i with probability proportional to $\frac{1}{i^2}$. we can do this because $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\Pi^2}{6}$, so we can normalize and have a probability space by dividing the certain area under probability density function in continuous probability by $\frac{\Pi^2}{6}$. However, consider the case that the probability is proportional to $\frac{1}{i}$. Do you think it can be normalized?

The answer is *No*. It is important to revisit why it's not. As you learned from calculus, the sum of the sequence $s_i = \frac{1}{i}$ is diverging, which means the sum itself goes to infinity. Therefore, when you try to normalize by dividing the area under probability density function by the infinite sum of each probability, it always turns out to be 0.

Now, how about picking a real number between 0 and 1?

Example 2 : Picking a real number between 0 and 1 (Continuously Infinite Case)

We want to pick a real uniformly at random, $P(x) = 1$. The probability we pick any given real number is $0 = \frac{1}{\infty}$, as there are infinite number of them. However, here is where the range comes in. We can ask the probability what we picked was less than $\frac{1}{2}$. Recall this idea of cumulative distribution function in continuous probability :

$$F(b) - F(a) = P(a \leq x \leq b) = \int_{a}^{b} P(x)dx$$

In this sense, the desired range of outcomes can be found by solving the following inequality.

$$\int_{0}^{\frac{1}{2}} 1dx = \frac{1}{2}$$

Similarly, you can ask the question of, what is the probability you pick a random real in the range $\left[\frac{1}{2}, 1\right]$. You'll find that it is $\frac{1}{2}$, so the probability of picking a real less than $\frac{1}{2}$ is the same as picking a real greater than $\frac{1}{2}$. This is because we are picking the real number uniformly at random. You could just as well compare dividing the range 0 to 1 into quarters and find that a real from each quarter is just as likely. It's uniform!

So, that is how we deal with infinite spaces, over ranges, with integrals.

Disjoint Events

Definition 1. *If two events A and B are disjoint, when an outcome i occurs, it can never be the case that* $i \in A$ *and* $i \in B$.

Suppose A and B are disjoint events. Hence, intuitively:
[use wikipedia for a diagram of disjoint events]

$$P(A \cup B) = P(A) + P(B)$$

That was straightforward. However, what if A and B are not disjoint? Remember inclusion/exclusion? We can draw the venn-diagram to reach the same intuitive conclusion:

[use wikipedia for a diagram of non-disjoint events]

$$P(A \cup B) = P(A) + P(B) = P(A \cap B)$$

Let us check that this aligns with our intuition for when A and B were disjoint, since when they are disjoint $A \cap B = \phi$.

Now if we were to continue in the same way that inclusion/exclusion wanted us to. It would be the long train of:

$$P(A_1 \cup A_2 \cup A_3 \dots) = P(A_1) + P(A_2) + P(A_3) \dots - P(A_1 \cap A_2) - P(A_2 \cap A_3) - \dots + P(A_1 \cap A_2 \cap A_3) + \dots$$

Thought Experiment

This material was covered in class for your own knowledge, but we do *NOT* epxect you to deal with this particular topic in the exam or in the homework assignment.

Let's take a look at the concept of Union bound :

$$P(A_1 \cup A_2 \dots) \leq \sum_{i=1}^{n} P(A_i)$$

In most cases, this does not give a useful bound. To reason this out, you can think about two-event cases. When does the equality hold and when does the inequality hold? All these cases are relatively trivial. In the last few years, a field has developed in Learning Theory that helps us with this problem.

Conditional Probability

Definition 2. *Let A and B be events with $P(B) \neq 0$. The conditional probability of event A given that the event B happens is :*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

You can get to this equation with the intuitive venn-diagram. If we know that B has occured, we can normalize our probability space down to event B occuring. We can consider the probability of the event A *under this consition* as the proportion of the event B that happens in A as well.

In the real world, more than one event happens together like the weather, the number of car accidents, etc. If we know one of the events, for example the weather is terrible, how does it affect the likelihood of car accidents? We can apply conditional probability to these situations.

Example : Consider the set $\{1, 2, 3, 4\}$

We are drawing two numbers from the given set.

(1) What is the probability that a person draws $1, 2$?

Solution. There are 6 possibilites to draw two numbers from the given set as below.

$$\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}$$

Therefore, the probability of drawing $1, 2$ is $\frac{1}{6}$.

(2) What is the probability of drawing $1, 2$ *given that* one of the numbers is 1?

Solution. Define the corresponding events.
E is the event that drawing $1, 2$, while F is the event that drawing 1.

$$E = \{\{1, 2\}\}$$

$$F = \{\{1, 2\}, \{1, 3\}, \{1, 4\}\}$$

The desired probability we want to have is $P(E|F)$. $P(E|F)$ can be calculated by $\frac{P(E \cap F)}{P(F)}$ from the definition of conditional probability. From the definition of the probability of a particular event, $P(E \cap F) = \frac{1}{6}$ and $P(F) = \frac{1}{2}$. Therefore, $P(E|F) = \frac{1}{3}$.

Do not get confused by conditional probability and the intersection of two events! Conditional probability is considering the situation that one event occurs given that the other event already occurred. However, the intersection of two events is considering the situation that two events occur simulataneously. Usually, the problems to deal with conditional probability is marked with certain expressions such as *if* or *given that*.

Independent Events

Conditional Probability tells us how the probability can change when we are given more information. However, if more knowledge of B does not change the probability of A, then that is a special type of change - none! We call this as A is independent of B.

Definition 3. *The event A is independent of the event B when it satisfies either of these equivalent conditions below :*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

$$P(A \cap B) = P(A)P(B)$$

Note : Think about why these two conditions above are equivalent.

An example is flipping a coin. Even if you have seen 10 flips turn up heads, the next flip has the same probability, $\frac{1}{2}$ of heads and $\frac{1}{2}$ tails. Now, realize that a certain path resulting in all heads is far less likely than a path resulting in a half heads and a half tails, but this is only because there are far more paths that result in half heads and half tails. The probabilty of getting any given path is exactly the same (i.e. HHHHH, is just as likely as HTHTT to have a probability of $\frac{1}{32}$).

Before we move on, we need to address the difference between *disjoint events* and *independent events*. Independence here does *NOT* necessarily mean that two events do not have the same outcomes. Rather, independence means independence from the given precondition. For two disjoint events, there must not be any outcomes that reside in both of them. When you get confused dealing independence and disjoint, the rule of thumb is always to go back to the definition of these two concepts.

Indpendence on n Events

Definition 4. *A sequence of events $A_1, A_2, \ldots A_n$ are mutually independent if and only if for every subset of events $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$:*

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\ldots$$

Note : The definition is over all subsequences, not just on pairwise ones.

Preview : Hashing

We never looked at the hash function. We just said modular arithmetic would be good. The next class will cover more details about hashing.