CS 330 Discussion - Probability

March 24 2017

1 Fundamentals of Probability

1.1 Random Variables and Events

A random variable X is one whose value is non-deterministic. For example, suppose we flip a coin and set $X = 1$ if we get heads, and $X = 0$ if we get tails. Then, X is a random variable since its value is not known prior to the flip.

The set of values X can take and their associated probabilities is referred to as a probabilistic distribution. Variables can be discrete (e.g. the result of a die) or continuous (e.g. the distance between two objects). For this course, we focus on discrete distributions.

We can define an event E as something that may or may not happen, and then discuss the probability it happens as $Pr[E]$. When we discuss random variables, we may describe the probability they take on certain values using this notation. For example, $Pr[X = x]$ signifies the probability of the event "$X = x$".

We can define events which are functions of other events as well. For example, given two events A, B the event where both A and B happen can be written as $A \cap B$ (the intersection, or logical AND). The event where at least one of A, B happens is $A \cup B$ (the union, or logical OR). We can also define the complement of an event, A^C. A^C is the event where A does not happen.

Note that:

- $Pr[A^C] = 1 - Pr[A]$
- $(A \cap B)^C = A^C \cup B^C$
- $(A \cup B)^C = A^C \cap B^C$
- $Pr[A \cup B] \leq Pr[A] + Pr[B]$ (this is the union bound, and generalized to multiple events)
- $Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$ (can be proven pictorially via Venn diagram)

We can also define events conditionally on other events already having occurred. For example, given that we rolled an even number on a die, what's the probability we rolled a 6? The conditional probability of B given A is written as $Pr[B|A]$. More formally, $Pr[B|A] = \frac{Pr[B \cap A]}{Pr[A]}$.

Some conditional probabilities can be very cleanly defined in terms of each other. Note that $Pr[B|A]Pr[A] == Pr[A \cap B] = Pr[A|B]Pr[B]$. Then, $Pr[B|A] = \frac{Pr[A|B]Pr[B]}{Pr[A]}$. This is Bayes' rule. For example, if we know the probability a drug-using individual tests positively for drugs, if a test has false positives determining the probability someone who tests positive has actually taken drugs isn't easy. However, if we also know the probability an individual tests positive overall and the probability an individual has taken drugs, Bayes' rule lets us relate the two quantities easily.

1.2 Mutual Exclusion and Independence

Two events are said to be mutually exclusive if one happening means the other cannot happen. For example, rolling a 1 and rolling an even number on a die are mutually exclusive events. Formally, A and B are mutually exclusive if $Pr[A \cap B] = 0$.

Two events are independent if knowing whether one happened or not does not give us any additional knowledge as to whether the other happened or not. For example, while knowing that I rolled an even number on a die gives us more info about whether I rolled a 6 or not, knowing that I flipped a coin and got heads tells me nothing about whether I rolled a 6 or not. Formally, A and B are independent if $Pr[A|B] = Pr[A]$ or equivalently, $Pr[A \cap B] = Pr[A]Pr[B]$. The latter equation extends to multiple independent events. That is, $Pr[E_1 \cap E_2 ... E_n] = \prod_i Pr[E_i]$ if all E_i are independent. In general, two events which have no effect on each other are independent. So, if we have many events we know occur independently, we can calculate the probability of the joint set of events occurring by just multiplying each of the individual events' probabilities of occurring.

1.3 Expected Values

For a random variable X, the expected value can be denoted as $E[X]$. The expected value can be thought of as the average outcome. Formally, $E[X] = \sum_x Pr[X = x] * x$. For example, if I flip a fair coin and set $X = 1$ if the coin gets heads and $X = 0$ otherwise, $E[X] = .5$. We can also write expected values conditionally - for example, the expected value of a die roll given I got an even number. The expected value of X conditioned on A happening is $E[X|A]$. For some set of outcomes of some event $E_1 ... E_n$ and some associated random variable, $E[X] = \sum_i Pr[E_i]E[X|E_i]$.

Some random variables can be written as the sum of other random variables. For example, suppose I flip n coins, and denote $X_1, X_2 ... X_n$ to be 1 if the ith coin lands heads and 0 if it lands tails. Suppose X is a random variable which counts the number of heads I flip. I can write $X = X_1 + X_2 ... X_n$. When a random variable is the sum of other random variables, its expected value is the sum of their expected values. i.e. $X = \sum_i X_i \rightarrow E[X] = \sum_i E[X_i]$. This is known as linearity of expectation. It holds even when the X_i are dependent on each other, making it a very strong property.

2 The Geometric Distribution

Suppose we flip a coin that lands heads with probability p and tails otherwise, until it lands heads. How many flips will this take? If X is the number of flips it takes, then in order for X to equal i, we need to flip $i-1$ tails (if we flip heads before the ith flip, we stop) and then 1 heads. So, $Pr[X = i] = (1-p)^{i-1}p$ for all positive integers i. A random variable with this distribution is said to be geometrically distributed - the set of probabilities is referred to as a geometric distribution. Formally, $X \sim Geo(p)$.

The geometric distribution is useful in that it lets us predict how many times we may have to repeat some trial until it succeeds. For example, in randomized algorithms we might run a while loop until some condition is. If the chance of that condition being met on each iteration is constant, the number of iterations the while loop takes is geometrically distributed.

Note that $Pr[X > b] = (1-p)^b$ (we make more than b flips if the first b flips produce tails).

The expected value of the geometric distribution is $\frac{1}{p}$. Intuitively this makes sense, but we can also prove it quite simply. With probability p, on the first flip, we get heads and so $X = 1$. Otherwise, if we flip tails, the number of remaining flips is still $E[X]$, so now X's conditional expected value is $(1 + E[X]))$. In other words, $E[X] = Pr[X = 1] * E[X|X = 1] + Pr[X > 1] * E[X|X > 1] = p * 1 + (1-p)(1 + E[X])$. This is solved by $E[X] = \frac{1}{p}$.

3 Probability Problems

3.1 Coupon Collector

We're drawing coupons from a machine, and there are n types of coupons. In one draw, a coupon is picked at random and each type of coupon has the same probability to be picked. What is the expected number of draws to get all n types of coupons?

Let X_i be the random variable representing the number of draws to get the i the distinct coupon after we have got $(i-1)$ distinct coupons. We have X_i follows a geometric distribution and the probability of success for each trial is:

$$p_i = \frac{n-i+1}{n}$$

The expectation of X_i is

$$E[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$$

.

Let X be the number of draws to get all coupons. Then we have

$$X = X_1 + X_2 + \dots X_n$$

$$E[X] = \sum_{i=1}^{n} E[X_i]$$

$$E[X] = n[\frac{1}{1} + \frac{1}{2} \dots \frac{1}{n}] \in \Theta(n\ log\ n)$$

3.2 Birthday Paradox

There are k people and each person has a random birthday from n days. How large should k be such that the chance that two people have the same birthday is at least α

Let X be the event that every person has a distinct birthday. The ith person has $(n-i+1)$ options to have a distinct birthday. Thus, the probability to have a distinct birthday is $\frac{n-i+1}{n}$. Then:

$$Pr[X] = \frac{n}{n}\frac{n-1}{n}\frac{n-2}{n} \dots \frac{n-k+1}{n}$$

$$Pr[X] = 1(1-\frac{1}{n})(1-\frac{2}{n}) \dots (1-\frac{k-1}{n})$$

Calculus shows us $e^{-x} \approx 1-x$ for small x. Then:

$$Pr[X] \approx e^{-(0+\frac{1}{n}+\frac{2}{n} \dots + \frac{k-1}{n})}$$

$$Pr[X] \approx e^{-\frac{k(k-1)}{n}}$$

Then the minimum k needed is roughly that which solves

$$e^{-\frac{k(k-1)}{n}} = 1-\alpha$$

.

3.3 Balls in Bins

Suppose we have n balls and m bins, and we toss the balls uniformly at random into the bins. This is called a balls and bins process, and is a popular model in practice. For example, hashing algorithms can use a balls and bins process to analyze collisions. For this section, let X_i be the number of balls in ith bin after all balls are thrown.

3.3.1 Balls to fill all bins

How large should n be before all bins have a ball in expectation? This is a restatement of the coupon collector problem, so we get $n = \Theta(m\ log\ m)$.

3.3.2 $E[X_i]$

What is $E[X_i]$? Note that since all bins are equivalent from the perspective of throwing balls, $E[X_i]$ is the same for all i. Then,

$$\sum_{i=1}^{m} E[X_i] = n$$

and thus

$$E[X_i] = \frac{n}{m}$$

.

3.3.3 Most filled bin

How large is the most filled bin in expectation? That is, what is $E[max_i X_i]$? This problem is particularly interesting for hashing in that it asks what's the expected worst case runtime of checking a bucket in the hashtable.

For simplicity, assume there are also n bins (that is, $m = n$). In addition, we will only try to upper bound the expected value.

The probability a fixed bin has more than k balls is the probability that some set of k balls all landed in this bin. There are $\binom{n}{k}$ such sets, so by union bound, this is

$$Pr(X_i \geq k) \leq \binom{n}{k} (\frac{1}{n})^k$$

The operation $\binom{n}{k}$ is the choose operation. It computes the number of subsets of size k we can choose from a set of size n.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Where $i! = 1 * 2 * \ldots i$.

A useful inequality known as Stirling's approximation bounds $\binom{n}{k}$

$$(\frac{n}{k})^k \leq \binom{n}{k} \leq (\frac{en}{k})^k$$

Then:

$$Pr(X_i \geq k) \leq (\frac{en}{k})^k (\frac{1}{n})^k = (\frac{e}{k})^k$$

.

Then, $max_i X_i$ is at least k if any X_i is at least k:

$$Pr(max_i X_i \geq k) = Pr(\cup_{i=1}^{n} X_i \geq k)$$

By union bound:

$$Pr(max_i X_i \geq k) \leq n(\frac{e}{k})^k$$

.

We know that $max_i X_i \leq n$. Then, we can upper bound $E[max_i X_i]$ as follows:

$$E[max_i X_i] \leq k * Pr(max_i X_i < k) + n * Pr(max_i X_i \geq k) \leq n^2(\frac{e}{k})^k + k$$

The first inequality is arrived at by taking the expected value formula and increasing all values less than k to k, and all other values to n. The second is arrived at by bounding the first probability by 1, and the second as shown above.

We will then choose k in order to remove the first half of the sum. Asymptotically, this will be the same as if we had optimized for k

$$n^2(\frac{e}{k})^k \leq 1$$

$$(\frac{e}{k})^k \leq \frac{1}{n^2}$$

$$(\frac{k^k}{e^k}) \geq n^2$$

$$k^k \approx poly(n)$$

$$k \; log \; k = O(log \; n)$$

.

This happens to be solved by $k = O(\frac{log \; n}{log \; log \; n})$.

$$k \; log \; k = \frac{log \; n}{log \; log \; n} * log(\frac{log \; n}{log \; log \; n})$$

$$= \frac{log \; n}{log \; log \; n} * (log \; log \; n - log \; log \; logn)$$

.

$log \; log \; logn$ is vanishingly small, so:

$$k \; log \; k = \Theta(\frac{log \; n}{log \; log \; n} * log \; log \; n) = \Theta(log \; n)$$

.

Then, since $n^2(\frac{e}{k})^k \leq 1$:

$$E[max_i X_i] \leq k*Pr(max_i X_i \leq k)+n*Pr(max_i X_i \geq k) \leq n^2(\frac{e}{k})^k+k = 1+O(\frac{log \; n}{log \; log \; n}) = O(\frac{log \; n}{log \; log \; n})$$