CSCB63 Tutorial 10 — Expected Running Time, Stats, etc.

Recall the *indicator random variable*:

$$X_i = \begin{cases} 1 & \text{if an event } i \text{ occurs} \\ 0 & \text{if it does not} \end{cases}$$

One fact about indicator random variables that we use a lot is that

$$E[X] = Pr(X = 1)$$

Bernoulli distribution is the discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $q = 1 - p$. The expected value of a Bernoulli random variable X is $Pr(X = 1) = p$.

If we have n independent Bernoulli random variables, then the sum is the *binomial* distribution:

$$Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

is the probability of having k successes out of n total trials.

The *geometric* distribution gives the probability that the first occurrence of success requires k independent trials, each with success probability p. If the probability of success on each trial is p, then the probability that the k^{th} trial is the first success is

$$Pr(X = k) = (1-p)^{k-1} p$$

And the expected number of trials until the first success is

$$E(X) = 1/p$$

- Let's say we toss a fair coin that has a probability p of landing "heads" each time. Then the expected number of heads in n tosses is:

$$\begin{aligned} &E(\text{number of heads in } n \text{ tosses}) \\ =&Pr(1^{st} \text{ toss is head }) + \ldots + Pr(n^{th} \text{ toss is head}) \\ =&p + \ldots + p \\ =&n * p \end{aligned}$$

- (from CLRS) Suppose n customers leave their respective hats at coat check. Later the customers receive their hats back but randomly (uniformly) permuted. What is the expected number of customers who are lucky enough to get back their own hats?

  For each customer i, $1 \le i \le n$, let X_i be the indicator random variable

$$X_i = \begin{cases} 1 & \text{if customer } i \text{ gets their own hat} \\ 0 & \text{if customer } i \text{ gets someone else's hat} \end{cases}$$

  Then

$$\begin{aligned} &E[X_i] \\ =&Pr(\text{customer } i \text{ gets their own hat}) \\ =&(\text{number of outcomes when custmer } i \text{ gets their hat})/(\text{total number of outcomes}) \\ =&1 * (n-1)!/n! \\ =&1/n \end{aligned}$$

Then expected number of such customers is

$$\begin{aligned}E(X) =&E[X_1 + ... + X_n]\\=&E[X_1] + ... + E[X_n]\\=&(1/n) * n\\=&1\end{aligned}$$

- Consider the following piece of code. What is the expected final value of the variable x?

```
0. x := 0
1. do n times:
2.   if random(0,1) <= 3/4:
3.     x := x + 5
4.   else:
5.     x := x - 1
```

So, with probability 3/4 the value of x increases by 5 and with probability 1/4 the value of x decreases by 1, during each of n iterations.

Again, the first thing we need to do is define a random variable. Let $X_i be$ the random variable that corresponds to "during iteration i, the change to x is X_i". Then

$$X_i = \begin{cases} 5 & \text{if } random(0,1) \text{ returns a value} \leq 3/4 \\ -1 & \text{otherwise} \end{cases}$$

Then the expected value is

$$E[X_i] = 5 * (3/4) + (-1) * (1/4) = 7/2$$

Then after n iterations, the expected final value of x is $7n/2$.

- (from CLRS) Suppose we repeatedly toss a ball into one of b bins randomly. What is the expected number of tosses until every bin is non-empty?

We define a random variable N_i to be "the number of tosses required for the i^{th} bin to become non-empty, from the time $i-1$ bins are non-empty".

The scenario is then: we have $i-1$ non-empty bins, and we keep tossing a ball until it lands in an empty bin, thus making i bins non-empty. Since we have b bins total, and $i-1$ bins non-empty, then we have $b-i+1$ bins empty. Then the probability of a ball landing into one of them is $(b-i+1)/b$. How many tosses do we need until this happens? Note that we have a geometric distribution, and so

$$E(N_i) = 1/p = b/(b-i+1)$$

Finally, we calculate the sum:

$$\begin{aligned}E(N) =&E(N_1 + ... + N_b)\\=&E(N_1) + ... + E(N_b)\\=&b/b + b/(b-1) + ... + b/1\\=&b * \sum_{i=1}^{b} 1/i\\\leq&b * \log b\end{aligned}$$

- This one is more difficult. Consider the following piece of code:

```
c := 1
t := 0
while c <= n:
  t := t + 1
  if random(1, 2^c) = 1:
    c := c + 1
```

What is the expected final value of t? (Note that this amounts to the expected running time.)

Discussion: Using the definition of expected value will be too hard. Trying a sum of indicator/Bernoulli random variables, one per iteration, will raise the question of how many iterations there are.

Solution idea: similar to the one with the bins.

Let X be the random variable for the final t. This is the sum of the following simpler random variables:

For each i from 1 to n, let W_i be the random variable for how many times t increases until c changes from i to $i+1$.

In other words, at some point c's value becomes i. From that point, count the number iterations until c's value becomes $i+1$. That count is W_i.

When c's value is i, the code keeps trying $random(1, 2^i) = 1$ until it succeeds, and the number of trials is W_i. So W_i is a geometric random variable with success probability $1/2^i$. So

$$E(W_i) = 2^i$$

Finally,

$$E(X) = \sum_{i=1}^{n} E(W_i) = \sum_{i=1}^{n} 2^i = \left(\sum_{i=0}^{n-1} 2^i \right) + 2^n - 1 = \frac{1-2^n}{1-2} + 2^n - 1 = 2^{n+1} - 2$$