CS 70 Discrete Mathematics and Probability Theory
Spring 2017 Rao
DIS 8b

1 Throwing Balls into a Depth-Limited Bin

Say you want to throw n balls into n bins with depth $k-1$ (they can fit $k-1$ balls, after that the bins overflow). Suppose that n is a large number and $k = 0.1n$. You throw the balls randomly into the bins, but you would like it if they don't overflow. You feel that you might expect not too many balls to land in each bin, but you're not sure, so you decide to investigate the probability of a bin overflowing.

(a) Focus on the first bin. Get an upper bound the number of ways that you can throw the balls into the bins such that this bin overflows. Try giving an argument about the following strategy: select k balls to put in the first bin, and then throw the remaining balls randomly. You should assume that the balls are distinguishable.

(b) Calculate an upper bound on the probability that the first bin will overflow.

(c) Upper bound the probability that some bin will overflow. [*Hint*: Use the union bound.]

(d) How does the above probability scale as n gets really large? [*Hint*: Use the union bound.]

Solution:

(a) We choose k of the balls to throw in the first bin and then throw the remaining $n-k$, giving us $\binom{n}{k}n^{n-k}$. Certainly any outcome of the ball-throwing that overflows the first bin is accounted for – we can simply choose the first k balls that land in the first bin and then simulate the rest of the outcome via random throwing. However, we are potentially overcounting: if $k+1$ balls go in the first bin, we have many choices for which k of them that could have been the "chosen" ones, and we count each one of these choices as distinct. However, they correspond to the same configuration, namely the one where $k+1$ balls are in the first bin. Hence we get an upper bound.

(b) We divide by the total number of ways the balls could have fallen into the bins, with order, so we get

$$\frac{\binom{n}{k}n^{n-k}}{n^n} = \frac{\binom{n}{k}}{n^k}.$$

(c) By symmetry, we can just upper bound this probability by n multiplied by the probability that a single bin (WLOG, the first bin) overflows. This gives about $n \cdot [\binom{n}{k}/n^k]$. This technique is called a *union bound*, where we upper bound the probability of the union of a bunch of events by the sums of the probabilities of the events.

(d) We get

$$n \cdot \frac{\binom{n}{k}}{n^k} = n \cdot \frac{n \cdot (n-1) \cdots (n-k+1)}{k! n^k} \leq n \cdot \frac{n^k}{k! n^k} = \frac{n}{k!} = \frac{n}{(0.1n) \cdot (k-1)!} = \frac{10}{(0.1n-1)!}.$$

Clearly, as n gets large this probability is going to 0. Note that this same analysis would work with $k = cn$ for any constant $0 < c < 1$. Hence, using some very coarse upper bounds, we can see that as the number of balls and bins grows, we have that it is very unlikely that we get a constant fraction of the balls in any single bin.

2 Telebears

Lydia has just started her Telebears appointment. She needs to register for a marine science class and CS 70. There are no waitlists, and she can attempt to enroll once per day in either class or both. The Telebears system is strange and picky, so the probability of enrolling in the marine science class is p_1 and the probability of enrolling in CS 70 is p_2. The probabilities are independent. Let M be the number of days it takes to enroll in the marine science class, and C be the number of days it takes to enroll in CS 70.

(a) What distribution do M and C follow? Are M and C independent?

(b) For some integer $k \geq 1$, what is $\Pr[C \geq k]$?

(c) For some integer $k \geq 1$, what is the probability that she is enrolled in both classes before day k?

Solution:

(a) $M \sim \text{Geom}(p_1)$, $C \sim \text{Geom}(p_2)$. Yes they are independent.

(b) We are looking for the probability that it takes at least k days to enroll in CS 70. Using the geometric distribution, this is $(1-p_2)^{k-1}$.

(c) Use independence. Let X be the number of days before she is enrolled in both.

$$\begin{aligned}\Pr[X < k] &= \Pr[M < k]\Pr[C < k] = (1 - \Pr[M \geq k])(1 - \Pr[C \geq k]) \\ &= (1 - (1-p_1)^{k-1})(1 - (1-p_2)^{k-1})\end{aligned}$$

3 Fishy Computations

Use the Poisson distribution to answer these questions:

(a) Suppose that on average, a fisherman catches 20 salmon per week. What is the probability that he will catch exactly 7 salmon this week?

(b) Suppose that on average, you go to Fisherman's Wharf twice a year. What is the probability that you will go at most once in 2018?

(c) Suppose that in March, on average, there are 5.7 boats that sail in Laguna Beach per day. What is the probability there will be *at least* 3 boats sailing throughout the *next two days* in Laguna?

Solution:

(a) $X \sim \text{Poiss}(20)$.

$$\Pr[X = 7] = \frac{20^7}{7!}e^{-20} \approx 5.23 \cdot 10^{-4}.$$

(b) $X \sim \text{Poiss}(2)$.

$$\Pr[X \leq 1] = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} \approx 0.41.$$

(c) Let Y be the number of boats that sail in the next two days. We can approximate Y as a Poisson distribution $Y \sim \text{Poiss}(\lambda = 11.4)$, where λ is the average number of boats that sail over two days. Now, we compute

$$\begin{aligned}
\Pr[Y \geq 3] &= 1 - \Pr[Y < 3] \\
&= 1 - \Pr[Y = 0 \cup Y = 1 \cup Y = 2] \\
&= 1 - (\Pr[Y = 0] + \Pr[Y = 1] + \Pr[Y = 2]) \\
&= 1 - \left(\frac{11.4^0}{0!}e^{-11.4} + \frac{11.4^1}{1!}e^{-11.4} + \frac{11.4^2}{2!}e^{-11.4}\right) \\
&\approx 0.999.
\end{aligned}$$

We can show what we did above formally with the following claim: the sum of two independent Poisson random variables is Poisson. We won't prove this, but from the above, you should intuitively know why this is true. Now, we can model sailing boats on day i as a Poisson distribution $X_i \sim \text{Poiss}(\lambda = 5.7)$. Now, let X_1 be the number of sailing boats on the next day, and X_2 be the number of sailing boats on the day after next. We are interested in $Y = X_1 + X_2$. Thus, we know $Y \sim \text{Poiss}(\lambda = 5.7 + 5.7 = 11.4)$.