CSE 312: Foundations of Computing II
Instructor: Alex Tsun
Date: 7/08/20
Lecture Topics: 3.5 Zoo of Discrete RVs II, 3.6 Zoo of Discrete RVs III

[**Tags:** PSet2 Q1, Zoo of Discrete RVs]

1. Match the following to the most appropriate distribution (from the Zoo of Discrete RVs), including parameters (e.g., your answer should be in the form like NegBin(30, 0.2), or Poi(100) for example). Distributions may be used more than once or not at all. Suppose there are B blue fish, R red fish, G green fish in a pond, where B + R + G = N. You do not need to show work for this problem.
   a. How many of the next 10 fish I catch are blue, if I catch and release.
   b. How many fish I had to catch until my first green fish, if I catch and release.
   c. How many red fish I catch in the next five minutes, if I catch on average r red fish per minute, if I catch and release.
   d. Whether or not my next fish is blue, if I catch and release.
   e. How many fish I had to catch until my third red fish, if I catch and release.
   f. How many red fish I caught in one scoop of a net containing M fish.

Solution: Watch lecture ☺

[**Tags:** Zoo of Discrete RVs]

2. Suppose you are working at Amazon, and you are unfortunately on-call for your team the entire year (that means, you are the person that they may ping in the middle of the night to debug issues). There are 5 SWE's on your team (including yourself), and each person independently introduces on average 0.1 bugs per work-week (Mon-Fri).
   a. What is the probability of having a bug-free work-week?
   b. What is the probability of having a bug-free day? What's the relationship between your answer to this part and the previous part?
   c. What is the probability that in a (52-week) year, that there are at least 40 bug-free weeks?
   d. Suppose it's the first Monday of the year. When would you expect the first day where you had to debug (at least) one issue (in number of work-days from today)?
   e. Suppose it's the first Monday of the year. What is the probability that your tenth day of debugging happens in February or later (> 20 work-days from now)?

Solution:

a. The number of bugs in a week in total is $X \sim Poi(0.5)$ since we add 5 independent $Poi(0.1)$ rvs. Then,

$$P(X = 0) = e^{-0.5}\frac{0.5^0}{0!} = e^{-0.5} \approx 0.60653$$

b. The number of bugs in a day in total is $Y \sim Poi(0.1)$ since we add 5 independent $Poi(0.02)$ rvs. Then,

$$P(Y = 0) = e^{-0.1}\frac{0.1^0}{0!} \approx 0.90484$$

c. The number of bug-free weeks in a year is $Z \sim Bin(52, 0.60653)$. So

$$P(Z \geq 40) = \sum_{k=40}^{52}\binom{52}{k}0.60653^k(1-0.60653)^{52-k}$$

d. The days until the first bug is $W \sim Geo(p = 1 - 0.90484)$. Hence, $E[W] = \frac{1}{p} = \frac{1}{0.09516} \approx 10.508$.

e. The days until the tenth bug is $W \sim NegBin(10, 0.09516)$. Hence,

$$P(W > 20) = \sum_{k=21}^{\infty}\binom{k-1}{10-1}0.09516^{10}(1-0.09516)^{k-10}$$

Alternatively and equivalently, we can ask the probability that we had < 10 bug-free days in the first 20 days. The number of days with bugs in the first 20 days is $V \sim Bin(20, 0.09516)$, so

$$P(V < 10) = \sum_{k=0}^{9}\binom{20}{k}0.09516^k(1-0.09516)^{20-k}$$

[**Tags:** Zoo of Discrete RVs]

3. Suppose we have a hash function $h: \mathcal{U} \to \{0,1,\ldots,m-1\}$ which maps from a universe $\mathcal{U}$ of strings (with length < 100) into m buckets, with each string independently and equally likely to be hashed into any bucket. We want to insert n strings $s_1, \ldots, s_n$ into our hash table.
    a. Let $X_1 = h(s_1)$ be the index of the bucket that string s_1 hashes into. What distribution does X_1 have from our zoo?
    b. What is the probability that two particular strings s_1 and s_2 hash to the same bucket?
    c. If Y_1 is the number of strings in the first bucket after inserting all n strings, what distribution does Y_1 have from our zoo? What is the probability that the first bucket is empty?
    d. What is the expected number of empty buckets?

Solution:

a. $X_1 \sim Unif(0, m-1)$.
b. $P(X_1 = X_2) = \frac{1}{m}$ since the first string hashes to some bucket, and the probability that the second string also hashes to that bucket is just $\frac{1}{m}$.
c. $Y_1 \sim Bin\left(n, \frac{1}{m}\right)$, so $P(Y_1 = 0) = \left(1 - \frac{1}{m}\right)^n$.
d. Let $Z_0, \ldots, Z_{m-1}$ be indicator rvs which are 1 if the i^{th} bucket is empty and 0 otherwise. Notes that $P(Z_i = 1) = P(Y_i = 0)$, since the probability that a bucket is empty is the same as the probability it has zero strings in it. Then,
$$E[Z_i] = P(Z_i = 1) = \left(1 - \frac{1}{m}\right)^n$$
Hence, if $Z = \sum_{i=0}^{m-1} Z_i$ is the number of empty buckets, then
$$E[Z] = \sum_{i=0}^{m-1} E[Z_i] = m\left(1 - \frac{1}{m}\right)^n$$