$f_X(x) \geq 0 \ \forall x$, $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ $\qquad \Pr[X=x]=0$

$$f_X(x) = \begin{cases} 2x^2, & x\in[0,1/2] \\ 0, & o/w \end{cases}$$

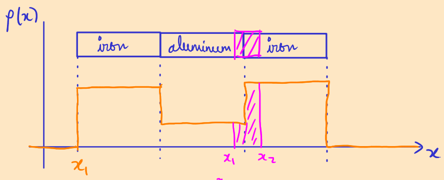
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \Pr[X\le x] = \Pr[X<x]$$

$$f_X(x) = \frac{d}{dx}F_X\Big|_{x=x} = \frac{d}{dt}\Pr[X\le t]\Big|_{t=x}$$

Prob $\longleftrightarrow$ mass

Prob density $\longleftrightarrow$ density

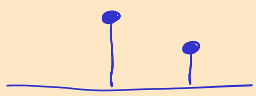
$f_X(x) \longleftrightarrow \rho(x)$



$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ $\qquad \int_{-\infty}^{\infty} x\rho(x)\,dx$ $\qquad \int_{x_1}^{x_2}\rho(x)\,dx$

$\int_{x_1}^{x_2} f_X(x)\,dx$

$f_X(x) = \begin{cases} 2 & x\in[0,0.5] \\ 0 & o/w \end{cases}$ $\qquad \int_{-\infty}^{\infty} x f_X(x)\,dx$



len(A)==n

E[# of swaps]

3 $\quad X \sim Geom(p_X)$, $Y \sim Geom(p_Y)$, $Z = \min\{X,Y\}$

CDF of Z

$$F_Z(z) = \Pr[Z\le z] = \Pr[\min\{X,Y\} \le z]$$

$$= 1 - \Pr[\min\{X,Y\} > z] = 1 - \underbrace{\Pr[X>z \cap Y>z]}_{\text{joint}} \overset{\text{indep}}{=} 1 - \underbrace{\Pr[X>z]\Pr[Y>z]}_{\text{marginals}} = 1-(1-p_X)^z(1-p_Y)^z$$

PMF of Z Show that Z is geometric

$$\Pr[Z=z] = \Pr[Z\le z] - \Pr[Z \le z-1]$$

$$= 1-(1-p_X)^z(1-p_Y)^z - \left(1-(1-p_X)^{z-1}(1-p_Y)^{z-1}\right)$$

$$= -(1-p_X)^z(1-p_Y)^z + (1-p_X)^{z-1}(1-p_Y)^{z-1}$$

$$= (1-p_X)^{z-1}(1-p_Y)^{z-1}\left(1-(1-p_X)(1-p_Y)\right)$$

$$= \left((1-p_X)(1-p_Y)\right)^{z-1}\left(1-(1-p_X)(1-p_Y)\right)$$

$$= (1-p_X+p_Y-p_Xp_Y)^{z-1}(p_X+p_Y-p_Xp_Y) \qquad p_Z = 1-\underbrace{(1-p_X)}_{\text{don't see H}}\underbrace{(1-p_Y)}_{\text{don't see 6}}$$

$$= p_Z(1-p_Z)^{z-1} \iff Z \sim Geom(p_Z)$$

$X \sim Geom(p)$

PMF $\quad \Pr[X=x] = (1-p)^{x-1}p$

$$\Pr[X>x] = \sum_{i=x+1}^{\infty}(1-p)^{i-1}p = p\sum_{j=x}^{\infty}(1-p)^j$$

$$p\sum_{j=0}^{\infty}(1-p)^j = \frac{1}{1-(1-p)} = \frac{1}{p} = 1$$

$$p\sum_{j=0}^{x-1}(1-p)^j = \frac{1-(1-p)^x}{1-(1-p)} = \frac{1-(1-p)^x}{p}$$

$$= 1-(1-p)^{x+1}$$

$$p\sum_{j=x}^{\infty}(1-p)^j = p\sum_{j=0}^{\infty}(1-p)^j - p\sum_{j=0}^{x}(1-p)^j$$

$$= \boxed{(1-p)^x}$$

X = waiting till H

Y = waiting till 6

$P[X \cup Y] = 1 - P[\bar{X} \cap \bar{Y}]$

$1-(1-p_X)(1-p_Y) = 1-p_X-p_Y+p_Xp_Y$

$p_X+p_Y-p_Xp_Y$

12

Geometric & Poisson Distributions and More

Computer Science Mentors 70

Nov 23 - Nov 27
Fall 2020

Prepared by: Aishani Sil, Austin Lei, Agnibho Roy, Debayan Bandyopadhyay, Abinav Routhu
Roast us here: *https://tinyurl.com/csm70-feedback20*

1 Geometric Distributions

1. **Intro**

   You are rolling a fair, six-sided die.

   (a) Compute the probability that it takes you 6 tries to roll a 6.

   Setup: Keep rolling till you see a 6. Let X be the # of rolls till you see a 6.

   $\Pr[X=6] = \left(\frac{5}{6}\right)^5\left(\frac{1}{6}\right) \approx 6.7\%$

   (b) Compute the expected number of times it takes you to roll a prime number. $\mathbb{E}[X] = 2$

   $X \sim \text{Geom}(1/2)$

2. **Free Food!**

   A cafe has a free sandwich menu, and each day gives you a random one of the 10 sandwiches on their menu for free. You want to try all of the available sandwiches, but you have two restrictions: you're allergic to fish, so you cannot have the 2 sandwiches that contain that ingredient, and you need all of your sandwiches toasted. The server doesn't care enough though, so you will still get a totally random sandwich that has a probability of $\frac{1}{2}$ of being toasted; you throw out the order if you cannot eat it. What is the expected number of days it takes you to try the 8 sandwiches you're able to have?

$$2\left(1+\frac{1}{2}+\cdots+\frac{1}{8}\right)$$

p-Coin toss

Binom(n,p) → how many succ?

Geom (p) → how long until 1^{st} succ

$$S = S_1 + \cdots + S_8$$

$$\mathbb{E}[S_1] = \frac{10}{8}\cdot 2$$

$$S_1 \sim \text{Geom}\left(\frac{8}{10}\cdot\frac{1}{2}\right)$$
$$S_2 \sim \text{Geom}\left(\frac{7}{10}\cdot\frac{1}{2}\right)$$
$$\vdots$$
$$S_i \sim \text{Geom}\left(\frac{9-i}{10}\cdot\frac{1}{2}\right)$$

$$\mathbb{E}[S] = \boxed{\sum_{i=1}^{8}\frac{20}{8+1-i}}$$

3. **Minimum of Geometric Random Variables**
   In this question, we will explore how the minimum, Z, of two geometric random variables, X and Y, is distributed geometrically as well. $Z = \min(X, Y)$ $\quad X \sim \text{Geom}(p_x),\ Y \sim \text{Geom}(p_Y)$

   (a) Find the CDF of Z in terms of F_X and F_Y, the CDFs of X and Y, respectively.

   (b) Let's assume that the probability parameters of X and Y are p_x and p_y, respectively. Using the previous part, explicitly calculate the CDF of Z.

   (c) Calculate the [redacted] of Z, show that it is geometrically distributed, and find its parameter, p_z.
   PMF
   ↑mass

4. **Throwing Darts**
   Alex and Bob are playing darts. They take turns throwing the dart and the first one to hit the center wins. Each turn, Alex hits the center with probability p and Bob hits the center with probability q. Alex gets to go first. We will explore two ways of finding out who is more likely to win.

   (a) What is the probability that Alex wins on the k^{th} turn?

   (b) Using the law of total probability, what is the probability that Alex wins?

   (c) We say that the game ended on turn k if either Alex or Bob hit the center on turn k. Let X be a random variable corresponding to the turn that the game ended on. What is the distribution of X (hint: what section are we on right now)?

(d) Using ONLY parts a and c, what is the probability that Alex wins given that the game ended on turn k.

(e) What does this tell us about the chances of Alex winning the game?

II Poisson Distributions

5. **Pancake Flips**
Leanne makes X pancakes, where X is a Poisson(λ) random variable. She flips each pancake one-handed, and each pancake has a probability p of being flipped properly, and probability $1 - p$ of flying off the pan, and the probability of each pancake getting flipped properly is independent of the other pancakes. Show that the number of pancakes that get flipped properly follows a Poisson($p\lambda$) distribution.

6. **Poisson as Limit of Binomial**
Let's explore merging 2 Poisson random variables, $A \sim Pois(\lambda_1)$ and $B \sim Pois(\lambda_2)$; that is, finding $E[A + B]$. It's totally possible to do this with tedious calculations, but let's do something different.

(a) The Poisson distribution is really a niche variant of the Binomial distribution for exceedingly rare events that have many, many chances of occurring (like car accidents, heart attacks, and sunny days in London). How can the Binomial distribution be used to model such an event? What would the expectation of this "coincidence" model be?

(b) Suppose that A corresponds to the number of seasons Berkeley beats Stanford in a football game and B corresponds to the number of seasons Berkeley beats USC. Using the "coincidence" model above, what is the probability of Berkeley beating Stanford in any one season? What about the probability of beating USC? And lastly, the probability of beating either?

(c) Using your solution from part (b), what can you say about $E[A+B]$, the number of seasons Berkeley beats Stanford or USC?

III Union Bound, Hashing, and Load Balancing

7. **Paper Triangle**
On a piece of paper, 70% of the paper is black and 30% is white. Show that an equilateral triangle can be drawn on the paper such that all of the vertices of the triangle are black. (Hint: Use the union bound to show that the probability of all the vertices being black is nonzero.)

8. **Hashing and Load Balancing Questions**
Imagine that we have m tasks and we are trying to distribute it amongst n identical processors. Imagine the task being distributed uniformly at random amongst the processors. We will discuss some metrics that we may be interested in.

(a) Using a balls and bins model, what is the probability that one of the processors does not end up with any tasks? What about the expected number of processors with no tasks (for this part, assume $n = m$ and approximate for a large n)? *Hint:* $\left(1 - \frac{1}{n}\right)^n \to e$ *as* $n \to \infty$

(b) We want to limit the chance that one or more of our processors become overloaded. What is the probability that any one processor gets exactly k tasks?

(c) What is an upper bound on the probability that any one processor gets *at least* k tasks?