Section 5

Slides by Aman Thukral and Judy Tian

Administrivia

- Pset 4 due YESTERDAY
- Homework 5 due Wednesday (Feb. 15th) 11:59 pm PDT
    - It is highly recommended that you finish it within a week.
- Midterm on Monday (Feb. 13th) 9:30-10:20 am
    - See Anna's post on edstem

Review

1) **Uniform**: $X \sim \text{Uniform}(a,b)$ (Unif(a,b) for short), for integers $a \leqslant b$, iff X has the following probability mass function:

$$p_X(k) = \frac{1}{b-a+1}, \quad k = a, a+1, \ldots, b$$

$\mathbb{E}[X] = \frac{a+b}{2}$ and $\text{Var}(X) = \frac{(b-a)(b-a+2)}{12}$. This represents each integer from $[a,b]$ being equally likely. For example, a single roll of a fair die is Uniform$(1,6)$.

2) **Bernoulli (or indicator)**: $X \sim$ Bernoulli(p) (Ber(p) for short) iff X has the following probability mass function:

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1-p, & k = 0 \end{cases}$$

$\mathbb{E}[X] = p$ and $\text{Var}(X) = p(1-p)$. An example of a Bernoulli r.v. is one flip of a coin with $\mathbb{P}(\text{head}) = p$.

Review

3) Binomial: $X \sim$ Binomial(n, p) (Bin(n, p) for short) iff X is the sum of n iid Bernoulli(p) random variables. X has probability mass function

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

$\mathbb{E}[X] = np$ and $\mathrm{Var}(X) = np(1-p)$. An example of a Binomial r.v. is the number of heads in n independent flips of a coin with $\mathbb{P}(\text{head}) = p$. Note that Bin$(1, p) \equiv$ Ber(p). As $n \to \infty$ and $p \to 0$, with $np = \lambda$, then Bin$(n, p) \to$ Poi(λ). If $X_1, \ldots, X_n$ are independent Binomial r.v.'s, where $X_i \sim$ Bin(N_i, p), then $X = X_1 + \ldots + X_n \sim$ Bin$(N_1 + \ldots + N_n, p)$.

4) Geometric: $X \sim$ Geometric(p) (Geo(p) for short) iff X has the following probability mass function:

$$p_X(k) = (1-p)^{k-1} p, \quad k = 1, 2, \ldots$$

$\mathbb{E}[X] = \frac{1}{p}$ and $\mathrm{Var}(X) = \frac{1-p}{p^2}$. An example of a Geometric r.v. is the number of independent coin flips up to and including the first head, where $\mathbb{P}(\text{head}) = p$.

5) Poisson: $X \sim \text{Poisson}(\lambda)$ ($\text{Poi}(\lambda)$ for short) iff X has the following probability mass function:

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots$$

$\mathbb{E}[X] = \lambda$ and $\text{Var}(X) = \lambda$. An example of a Poisson r.v. is the number of people born during a particular minute, where λ is the average birth rate per minute. If $X_1, \ldots, X_n$ are independent Poisson r.v.'s, where $X_i \sim \text{Poi}(\lambda_i)$, then $X = X_1 + \ldots + X_n \sim \text{Poi}(\lambda_1 + \ldots + \lambda_n)$.

6) Hypergeometric: $X \sim \text{HyperGeometric}(N, K, n)$ ($\text{HypGeo}(N, K, n)$ for short) iff X has the following probability mass function:

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \qquad \text{where } n \leqslant N,\ k \leqslant \min(K, n) \text{ and } k \geqslant \max(0, n-(N-K)).$$

We have $\mathbb{E}[X] = n\frac{K}{N}$. ($\text{Var}(X) = n \cdot \frac{K(N-K)(N-n)}{N^2(2N-1)}$ which is not very memorable.) This represents the number of successes drawn, when n items are drawn from a bag with N items (K of which are successes, and $N-K$ failures) without replacement. If we did this with replacement, then this scenario would be represented as $\text{Bin}\left(n, \frac{K}{N}\right)$.

7) **Negative Binomial**: $X \sim \text{NegativeBinomial}(r, p)$ (NegBin(r, p) for short) iff X is the sum of r iid Geometric(p) random variables. X has probability mass function

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \ldots$$

$\mathbb{E}[X] = \frac{r}{p}$ and $\text{Var}(X) = \frac{r(1-p)}{p^2}$. An example of a Negative Binomial r.v. is the number of independent coin flips up to and including the r^{th} head, where $\mathbb{P}(\text{head}) = p$. If $X_1, \ldots, X_n$ are independent Negative Binomial r.v.'s, where $X_i \sim \text{NegBin}(r_i, p)$, then $X = X_1 + \ldots + X_n \sim \text{NegBin}(r_1 + \ldots + r_n, p)$.

Question 1: “Pond Fishing”

There are B blue fish, R red fish and G green fish in the pond, $B + R + G = N$. Identify the most appropriate distribution:

a) How many of the next 10 fish I catch are blue, if I catch and release.

Question 1: “Pond Fishing”

There are B blue fish, R red fish and G green fish in the pond, B + R+ G = N. Identify the most appropriate distribution:

a) How many of the next 10 fish I catch are blue, if I catch and release.

Bin(10, B/N)

Similar to “how many of my next 10 trials are a success”

Question 1: “Pond Fishing”

There are B blue fish, R red fish and G green fish in the pond, B + R+ G = N. Identify the most appropriate distribution:

b) How many fish I had to catch until my first green fish, if I catch and release

Question 1: “Pond Fishing”

There are B blue fish, R red fish and G green fish in the pond, B + R+ G = N. Identify the most appropriate distribution:

b) How many fish I had to catch until my first green fish, if I catch and release

Geo(G/N)

Similar to “how many trials until we see a success”

Question 1: “Pond Fishing”

There are B blue fish, R red fish and G green fish in the pond, B + R+ G = N. Identify the most appropriate distribution:

d) Whether or not the next fish is blue

Question 1: “Pond Fishing”

There are B blue fish, R red fish and G green fish in the pond, B + R+ G = N. Identify the most appropriate distribution:

d) Whether or not the next fish is blue

Ber(B/N)

One trial, similar to Binomial

Question 4: Memorylessness

We say that a random variable X is memoryless if $\mathbb{P}(X > k + i \mid X > k) = \mathbb{P}(X > i)$ for all non-negative integers k and i. The idea is that X does not *remember* its history. Let $X \sim Geo(p)$. Show that X is memoryless.

Question 4: Memorylessness

We say that a random variable X is memoryless if $\mathbb{P}(X > k + i \mid X > k) = \mathbb{P}(X > i)$ for all non-negative integers k and i. The idea is that X does not *remember* its history. Let $X \sim Geo(p)$. Show that X is memoryless.

Let's note that if $X \sim Geo(p)$, then $\mathbb{P}(X > k) = \mathbb{P}(\text{no successes in the first } k \text{ trials}) = (1-p)^k$.

Question 4: Memorylessness

We say that a random variable X is memoryless if $\mathbb{P}(X > k + i \mid X > k) = \mathbb{P}(X > i)$ for all non-negative integers k and i. The idea is that X does not *remember* its history. Let $X \sim Geo(p)$. Show that X is memoryless.

Let's note that if $X \sim Geo(p)$, then $\mathbb{P}(X > k) = \mathbb{P}(\text{no successes in the first } k \text{ trials}) = (1-p)^k$.

$$\mathbb{P}(X > k + i \mid X > k) = \frac{\mathbb{P}(X > k \mid X > k+i)\, \mathbb{P}(X > k+i)}{\mathbb{P}(X > k)} \qquad \text{[Bayes Theorem]}$$

Question 4: Memorylessness

We say that a random variable X is memoryless if $\mathbb{P}(X > k + i \mid X > k) = \mathbb{P}(X > i)$ for all non-negative integers k and i. The idea is that X does not *remember* its history. Let $X \sim Geo(p)$. Show that X is memoryless.

Let's note that if $X \sim Geo(p)$, then $\mathbb{P}(X > k) = \mathbb{P}(\text{no successes in the first } k \text{ trials}) = (1-p)^k$.

$$\mathbb{P}(X > k + i \mid X > k) = \frac{\mathbb{P}(X > k \mid X > k+i)\, \mathbb{P}(X > k+i)}{\mathbb{P}(X > k)} \qquad \text{[Bayes Theorem]}$$

$$= \frac{\mathbb{P}(X > k+i)}{\mathbb{P}(X > k)} \qquad [\mathbb{P}(X > k \mid X > k+i) = 1]$$

Question 4: Memorylessness

We say that a random variable X is memoryless if $\mathbb{P}(X > k + i \mid X > k) = \mathbb{P}(X > i)$ for all non-negative integers k and i. The idea is that X does not *remember* its history. Let $X \sim Geo(p)$. Show that X is memoryless.

Let's note that if $X \sim Geo(p)$, then $\mathbb{P}(X > k) = \mathbb{P}(\text{no successes in the first } k \text{ trials}) = (1-p)^k$.

$$\begin{aligned}
\mathbb{P}(X > k + i \mid X > k) &= \frac{\mathbb{P}(X > k \mid X > k+i)\, \mathbb{P}(X > k+i)}{\mathbb{P}(X > k)} && \text{[Bayes Theorem]} \\
&= \frac{\mathbb{P}(X > k+i)}{\mathbb{P}(X > k)} && [\mathbb{P}(X > k \mid X > k+i) = 1] \\
&= \frac{(1-p)^{k+i}}{(1-p)^k} && [\mathbb{P}(X > k) = (1-p)^k] \\
&= (1-p)^i \\
&= \mathbb{P}(X > i)
\end{aligned}$$

Midterm Practice

Very Important reminder for exam:

- Come early and find your room
- Print out cheatsheets
- Laptop should be fully charged
- Don’t open other tabs on your computer during exam!!!
- One time submission

Midterm Practice

- Go to the Canvas quizzes -> Practice Midterm 2023
- Open Wolfram Alpha on another tab on your computer https://www.wolframalpha.com/
- Try Question 2

Midterm Practice

Consider a Covid test. Suppose that:

The probability that a random person tests positive given that they have Covid is p.
The probability that a random person tests negative given that they do NOT have Covid is n.
The probability that a random person has Covid is c.
What is the probability that a random person has Covid given that they test positive?

Your answer should be correct to 4 decimal places. So for example, if your calculation shows that the answer is 0.02587329, then you should enter 0.0259.

Let C be the event that a random person has COVID and let $\bar{C}$ be the complement event. Let T^+ be the event that a random person tests positive and let T^- be the event that a random person tests negative.

Translating the statements given into probability notation, we have:

- $Pr(T^+|C) = p$
- $Pr(T^-|\bar{C}) = n$
- $Pr(C) = c$

The goal is to find the probability that a random person has Covid given that they test positive, ie:

$$Pr(C|T^+) = ???$$

First, we apply Bayes' Theorem:

$$Pr(C|T^+) = \frac{Pr(T^+|C)Pr(C)}{Pr(T^+)}$$

Then we expand the denominator with the Law of Total Probability, conditioning on the event that the test was positive given the person had Covid or not (ie if it's a false or a true positive):

$$Pr(C|T^+) = \frac{Pr(T^+|C)Pr(C)}{Pr(T^+|C)Pr(C) + Pr(T^+|\bar{C})Pr(\bar{C})}$$

We have all components here except for $Pr(T^+|\bar{C})$ and $Pr(\bar{C})$. Here we can use complementarity:

- $Pr(T^+|\bar{C}) = 1 - Pr(T^-|\bar{C}) = 1 - n$
- $Pr(\bar{C}) = 1 - Pr(C) = 1 - c$

Plugging all the values in we get our final solution:

$$Pr(C|T^+) = \frac{p * c}{p * c + (1 - n)(1 - c)}$$