CSE 312 Section 4

Random Variables and Expectation

Administrivia

Announcements & Reminders

- HW2
  - Grades released on Gradescope – check your submission to read comments
  - Regrade requests open ~24 hours after grades are released and close after a week

- HW3
  - Was due yesterday, Wednesday 7/12 @ 11pm
  - Late deadline Friday 7/14 @ 11pm
  - **This is the last homework before the midterm!**

Review & Questions

Review of Main Concepts

- **Random Variable (rv):** A numeric function $X : \Omega \to \mathbb{R}$ of the outcome.
- **Range/Support:** The support/range of a random variable X, denoted Ω_X, is the set of all possible values that X can take on.
- **Discrete Random Variable (drv):** A random variable taking on a countable (either finite or countably infinite) number of possible values.
- **Probability Mass Function (pmf)** for a discrete random variable X: a function $p_X : \Omega_X \to [0,1]$ with $p_X(x) = \mathbb{P}(X = x)$ that maps possible values of a discrete random variable to the probability of that value happening, such that $\sum_x p_X(x) = 1$.
- **Cumulative Distribution Function (CDF)** for a random variable X: a function $F_X : \mathbb{R} \to \mathbb{R}$ with $F_X(x) = \mathbb{P}(X \leq x)$

Review of Main Concepts

- **Expectation (expected value, mean, or average):** The expectation of a discrete random variable is defined to be

$$\mathbb{E}[X] = \sum_x x\, p_X(x) = \sum_x x\, \mathbb{P}(X = x)$$

The expectation of a function of a discrete random variable $g(X)$ is

$$\mathbb{E}[g(X)] = \sum_x g(x)\, p_X(x)$$

- **Linearity of Expectation:** Let X and Y be random variables, and $a, b, c \in \mathbb{R}$. Then, $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$. Also, for any random variables $X_1, \dots, X_n$,

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]$$

Review of Main Concepts

- **Variance:** Let X be a random variable and $\mu = E[X]$. The variance of X is defined to be

$$Var(X) = \mathbb{E}[(X-\mu)^2]$$

  Notice that since this is an expectation of a non-negative random variable ($(X-\mu)^2$), variance is always nonnegative. With some algebra, we can simplify this to

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- **Standard Deviation:** Let X be a random variable. We define the standard deviation of X to be the square root of the variance, and denote it $\sigma = \sqrt{Var(X)}$

- **Property of Variance:** Let $a, b \in \mathbb{R}$ and let X be a random variable. Then,

$$Var(aX+b) = a^2 Var(X)$$

Review of Main Concepts

- **Independence:** Random variables X and Y are independent iff

$$\forall x \forall y, \mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\, \mathbb{P}(Y = y)$$

  In this case, we have $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ (the converse is not necessarily true).

- **i.i.d. (independent and identically distributed):** Random variables $X_1, \ldots, X_n$ are i.i.d. (or iid) iff they are independent and have the same probability mass function..

- **Variance of Independent Variables:** If X is independent of Y,

$$Var(X + Y) = Var(X) + Var(Y)$$

  This depends on independence, whereas linearity of expectation always holds. Note that this combined with the above shows that $\forall a, b, c \in \mathbb{R}$ and if X is independent of Y,

$$Var(aX + bY + c) = a^2 Var(X) + b^2 Var(Y)$$

Review of Main Concepts – Zoo of RVs

Distribution	Parameters	Possible Description	Range Ω_X	$\mathbb{E}[X]$	Var(X)	PMF ($p_X(k)$ for $k \in \Omega_X$)
Uniform (disc)	$X \sim \text{Unif}(a,b)$ for $a,b \in \mathbb{Z}$ and $a \leq b$	Equally likely to be any *integer* in $[a,b]$	$\{a,\dots,b\}$	$\frac{a+b}{2}$	$\frac{(b-a)(b-a+2)}{12}$	$\frac{1}{b-a+1}$
Bernoulli	$X \sim \text{Ber}(p)$ for $p \in [0,1]$	Takes value 1 with prob p and 0 with prob $1-p$	$\{0,1\}$	p	$p(1-p)$	$p^k(1-p)^{1-k}$
Binomial	$X \sim \text{Bin}(n,p)$ for $n \in \mathbb{N}$, and $p \in [0,1]$	Sum of n iid $Ber(p)$ rvs. # of heads in n independent coin flips with $P(head) = p$.	$\{0,1,\dots,n\}$	np	$np(1-p)$	$\binom{n}{k}p^k(1-p)^{n-k}$
Poisson	$X \sim \text{Poi}(\lambda)$ for $\lambda > 0$	# of events that occur in **one** unit of time independently with rate λ per unit time	$\{0,1,\dots\}$	λ	λ	$e^{-\lambda}\frac{\lambda^k}{k!}$
Geometric	$X \sim \text{Geo}(p)$ for $p \in [0,1]$	# of independent Bernoulli trials with parameter p *up to and **including** first success*	$\{1,2,\dots\}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$(1-p)^{k-1}p$
Hypergeometric	$X \sim \text{HypGeo}(N,K,n)$ for $n, K \leq N$ and $n,K,N \in \mathbb{N}$	# of successes in n draws (w/o replacement) from N items that contain K successes in total	$\{\max(0, n+K-N), \dots, \min(n,K)\}$	$n\frac{K}{N}$	$n\frac{K(N-K)(N-n)}{N^2(N-1)}$	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$
Negative Binomial	$X \sim \text{NegBin}(r,p)$ for $r \in \mathbb{N}, p \in [0,1]$	Sum of r iid $Geo(p)$ rvs. # of independent flips until r^{th} head with $P(head) = p$	$\{r, r+1, \dots\}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\binom{k-1}{r-1}p^r(1-p)^{k-r}$

Source: https://courses.cs.washington.edu/courses/cse312/23su/resources/distributions.pdf

Problem 3 – Kit Kats Again

3 – Kit Kats Again

Suppose we have N candies in a jar, K of which are kit kats. Suppose we draw (without replacement) until we have (exactly) k kit kats, $k \leq K \leq N$. Let X be the number of draws until the kth kit kat. What is Ω_X, the range of X? What is $p_X(n) = \mathbb{P}(X = n)$? (We say X is a "negative hypergeometric" random variable).

Work on this problem with the people around you, and then we'll go over it together!

3 – Kit Kats Again

Suppose we have N candies in a jar, K of which are kit kats. Suppose we draw (without replacement) until we have (exactly) k kit kats, $k \leq K \leq N$. Let X be the number of draws until the kth kit kat. What is Ω_X, the range of X? What is $p_X(n) = \mathbb{P}(X = n)$? (We say X is a "negative hypergeometric" random variable).

3 – Kit Kats Again

Suppose we have N candies in a jar, K of which are kit kats. Suppose we draw (without replacement) until we have (exactly) k kit kats, $k \leq K \leq N$. Let X be the number of draws until the kth kit kat. What is Ω_X, the range of X? What is $p_X(n) = \mathbb{P}(X = n)$? (We say X is a "negative hypergeometric" random variable).

Consider all N candies to be arranged randomly in a row. We will treat the first n candies to be the "chosen" candies. We know that the nth candy is a kitkat (the kth kitkat) to be chosen. This nth candy divides the row into 2 sections. On the left, are the $n - 1$ candies that were chosen (all candies chosen except the last kitkat). On the right, are the candies remaining in the jar. So the first term in the numerator is choosing the $k - 1$ spots where kitkats will be chosen from the $n - 1$ spots. The second term is choosing the $K - k$ spots for the kitkats that remain in the jar among the remaining $N - n$ candies. The denominator is the total number of possible ways to place the K kitkats among N candies.

$$\Omega_X = \{k, k+1, \dots, N - K + k\}$$

$$p_X(n) = \mathbb{P}(X = x) = \frac{\binom{n-1}{k-1}\binom{N-n}{K-k}}{\binom{N}{K}}$$

Problem 4 – Hungry Washing Machine

4 – Hungry Washing Machine

You have 10 pairs of socks (so 20 socks in total), with each pair being a different color. You put them in the washing machine, but the washing machine eats 4 of the socks chosen at random. Every subset of 4 socks is equally probable to be the subset that gets eaten. Let X be the number of complete pairs of socks that you have left.

a) What is the range of X, Ω_X (the set of possible values it can take on)? What is the probability mass function of X?

b) Find $\mathbb{E}[X]$ from the definition of expectation.

Work on this problem with the people around you, and then we'll go over it together!

4 – Hungry Washing Machine

a) What is the range of X, Ω_X (the set of possible values it can take on)? What is the probability mass function of X?

4 – Hungry Washing Machine

a) What is the range of X, Ω_X (the set of possible values it can take on)? What is the probability mass function of X?

The washing machine eats 4 socks every time. It can either eat a single sock from 4 pairs of socks, leaving us with 6 complete pairs, or a single sock from 2 pairs and a matching pair, leaving us with 7 complete pairs, or 2 pairs of matching socks, leaving us with 8 complete pairs.

$$\Omega_X = \{6,7,8\}$$

4 – Hungry Washing Machine

a) What is the range of X, Ω_X (the set of possible values it can take on)? What is the probability mass function of X?

The washing machine eats 4 socks every time. It can either eat a single sock from 4 pairs of socks, leaving us with 6 complete pairs, or a single sock from 2 pairs and a matching pair, leaving us with 7 complete pairs, or 2 pairs of matching socks, leaving us with 8 complete pairs.

$$\Omega_X = \{6,7,8\}$$

We are dealing with a sample space with equally likely outcomes. As such, we can compute use the formula $\mathbb{P}(E) = \frac{|E|}{|\Omega|}$. We know that $|\Omega| = \binom{20}{4}$ because the washing machine picks a set of 4 socks out of 20 possible socks.

4 – Hungry Washing Machine

a) What is the range of X, Ω_X (the set of possible values it can take on)? What is the probability mass function of X?

To define the pmf of X, we consider each value in the range of X.

For $k = 6$, we first pick 4 out of 10 pairs of socks from which we will eat a single sock ($\binom{10}{4}$ ways), and for each of these 4 pairs we have two socks to pick from ($\binom{2}{1}^4$ ways). Using the product rule, we get $|X = 6| = \binom{10}{4}2^4$

For $k = 7$, we first pick 1 out of 10 pairs of socks to eat in its entirety ($\binom{10}{1}$ ways), and then 2 out of the 9 remaining pairs from which we will eat a single sock ($\binom{9}{2}$ ways), and for each of these 2 pairs we have two socks to pick from ($\binom{2}{1}^2$ ways). Using the product rule, we get $|X = 7| = 10\binom{9}{4}2^2$

For $k = 8$, we pick 2 out of 10 pairs of socks to eat ($\binom{10}{2}$ ways). We get $|X = 8| = \binom{10}{2}$

4 – Hungry Washing Machine

a) What is the range of X, Ω_X (the set of possible values it can take on)? What is the probability mass function of X?

To define the pmf of X, we consider each value in the range of X.

For $k = 6$, we first pick 4 out of 10 pairs of socks from which we will eat a single sock ($\binom{10}{4}$ ways), and for each of these 4 pairs we have two socks to pick from ($\binom{2}{1}^4$ ways). Using the product rule, we get $|X = 6| = \binom{10}{4}2^4$

For $k = 7$, we first pick 1 out of 10 pairs of socks to eat in its entirety ($\binom{10}{1}$ ways), and then 2 out of the 9 remaining pairs from which we will eat a single sock ($\binom{9}{2}$ ways), and for each of these 2 pairs we have two socks to pick from ($\binom{2}{1}^2$ ways). Using the product rule, we get $|X = 7| = 10\binom{9}{4}2^2$

For $k = 8$, we pick 2 out of 10 pairs of socks to eat ($\binom{10}{2}$ ways). We get $|X = 8| = \binom{10}{2}$

$$p_X = \begin{cases} \dfrac{\binom{10}{4}2^4}{\binom{20}{4}} & k = 6 \\ \dfrac{10\binom{9}{4}2^2}{\binom{20}{4}} & k = 7 \\ \dfrac{\binom{10}{2}}{\binom{20}{4}} & k = 8 \end{cases}$$

4 – Hungry Washing Machine

b) Find $\mathbb{E}[X]$ from the definition of expectation.

4 – Hungry Washing Machine

b) Find $\mathbb{E}[X]$ from the definition of expectation.

$$\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot p_X(k) = 6 \cdot \frac{\binom{10}{4} 2^4}{\binom{20}{4}} + 7 \cdot \frac{10\binom{9}{2} 2^2}{\binom{20}{4}} + 8 \cdot \frac{\binom{10}{2}}{\binom{20}{4}} = \frac{120}{19}$$

Problem 6 – Frogger

6 – Frogger

A frog starts on a 1-dimensional number line at 0. At each second, independently, the frog takes a unit step right with probability p_1, to the left with probability p_2, and doesn't move with probability p_3, where $p_1 + p_2 + p_3 = 1$. After 2 seconds, let X be the location of the frog.

a) Find $p_X(k)$, the probability mass function for X.

b) Compute $\mathbb{E}[X]$ from the definition.

Work on this problem with the people around you, and then we'll go over it together!

6 – Frogger

A frog starts on a 1-dimensional number line at 0. At each second, independently, the frog takes a unit step right with probability p_1, to the left with probability p_2, and doesn't move with probability p_3, where $p_1 + p_2 + p_3 = 1$. After 2 seconds, let X be the location of the frog.

a) Find $p_X(k)$, the probability mass function for X.

6 – Frogger

A frog starts on a 1-dimensional number line at 0. At each second, independently, the frog takes a unit step right with probability p_1, to the left with probability p_2, and doesn't move with probability p_3, where $p_1 + p_2 + p_3 = 1$. After 2 seconds, let X be the location of the frog.

a) Find $p_X(k)$, the probability mass function for X.

Let L be a left step, R be a right step, and N be no step.

The range of X is $\{-2, -1, 0, 1, 2\}$. We can compute:

$p_X(-2) = \mathbb{P}(X = -2) = \mathbb{P}(LL) = p_2^2$

$p_X(-1) = \mathbb{P}(X = -1) = \mathbb{P}(LN \cup NL) = 2p_2p_3$

$p_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(NN \cup LR \cup RL) = p_3^2 + 2p_1p_2$

$p_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(RN \cup NR) = 2p_1p_3$

$p_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(RR) = p_1^2$

6 – Frogger

A frog starts on a 1-dimensional number line at 0. At each second, independently, the frog takes a unit step right with probability p_1, to the left with probability p_2, and doesn't move with probability p_3, where $p_1 + p_2 + p_3 = 1$. After 2 seconds, let X be the location of the frog.

a) Find $p_X(k)$, the probability mass function for X.

Let L be a left step, R be a right step, and N be no step.

The range of X is $\{-2, -1, 0, 1, 2\}$. We can compute:

$p_X(-2) = \mathbb{P}(X = -2) = \mathbb{P}(LL) = p_2^2$

$p_X(-1) = \mathbb{P}(X = -1) = \mathbb{P}(LN \cup NL) = 2p_2p_3$

$p_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(NN \cup LR \cup RL) = p_3^2 + 2p_1p_2$

$p_X(1) = \mathbb{P}(X = 1) = \mathbb{P}(RN \cup NR) = 2p_1p_3$

$p_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(RR) = p_1^2$

$$p_X = \begin{cases} p_2^2 & k = -2 \\ 2p_2p_3 & k = -1 \\ p_3^2 + 2p_1p_2 & k = 0 \\ 2p_1p_3 & k = 1 \\ p_1^2 & k = 2 \end{cases}$$

6 – Frogger

A frog starts on a 1-dimensional number line at 0. At each second, independently, the frog takes a unit step right with probability p_1, to the left with probability p_2, and doesn't move with probability p_3, where $p_1 + p_2 + p_3 = 1$. After 2 seconds, let X be the location of the frog.

b) Compute $\mathbb{E}[X]$ from the definition.

6 – Frogger

A frog starts on a 1-dimensional number line at 0. At each second, independently, the frog takes a unit step right with probability p_1, to the left with probability p_2, and doesn't move with probability p_3, where $p_1 + p_2 + p_3 = 1$. After 2 seconds, let X be the location of the frog.

b) Compute $\mathbb{E}[X]$ from the definition.

$$\mathbb{E}[X] = (-2)(p_2^2) + (-1)(2p_2p_3) + (0)(p_3^2 + 2p_1p_2) + (1)(2p_1p_3) + (2)(p_1^2) = 2(p_1 - p_2)$$

Problem 13 – Pond Fishing

13 – Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$.
For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

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

d) whether or not my next fish is blue

13 – Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$.
For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

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

Same as: how many of next 10 trials (fish) are a success (blue)

- Binomial distribution

If I catch and release

- $p = \frac{B}{N}$
- Each trial is independent

$$\text{Bin}\left(10, \frac{B}{N}\right)$$

13 – Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$.
For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

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

Same as: how many trials (fish) until we see a success (green)
- Geometric distribution

If I catch and release
- $p = \frac{G}{N}$

$$\text{Geo}\left(\frac{G}{N}\right)$$

13 – Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$.
For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

c) how many red fish I catch in the next five minutes, if I catch on average r red fish per minute

Same as: how many occurrences of an event given an average rate

- Poisson distribution

In the next five minutes -- this is our time unit
Average rate needs to match the time unit

- $5r$

$$\mathrm{Poi}(5r)$$

13 – Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$.
For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

d) whether or not my next fish is blue

Same as: how many of next 1 trial (fish) are a success (blue)

- Bernoulli distribution

$$\mathrm{Ber}\left(\frac{B}{N}\right)$$

13 – Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$.
For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

e) how many of the next 10 fish I catch are blue, if I do not release the fish back to the pond after each catch

f) how many fish I have to catch until I catch three red fish, if I catch and release

13 – Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$.
For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

e) how many of the next 10 fish I catch are blue, if I do not release the fish back to the pond after each catch

Same as: number of successes in 10 draws (without replacement)

- Hypergeometric distribution

In this case, we have 10 draws (without replacement because we do not catch and release), and out of the N fish, B are blue (a success).

$$\text{HypGeo}(N, B, 10)$$

13 – Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B + R + G = N$.
For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

f) how many fish I have to catch until I catch three red fish, if I catch and release

Same as: number of trials until 3 successes (red fish)

- Negative binomial distribution

In this case, our trials are caught fish (with replacement this time) and our success is if the fish are red, which happens with probability $\frac{R}{N}$.

$$\mathrm{NegBin}\left(3, \frac{R}{N}\right)$$

That's All, Folks!

Thanks for coming to section this week!
Any questions?