Section 5

Review

- **Variance.** Var$(X) = \mathbb{E}[(X - \mathbb{E}[\mathbb{X}])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ Var$(aX + b) = a^2$Var(X).

  Notice that since this is an expectation of a non-negative random variable ($(X - \mu)^2$), variance is always non-negative.

- **Independence.** Two random variables X and Y are **independent** if $\forall x \in \Omega_X, \forall y \in \Omega_Y$, the following holds true: $\mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$.

  When two random variables are independent, we have $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ (the converse is not necessarily true).

- **Variance and Independence.** For any two independent random variables X and Y, Var$(X + Y) =$ ____________________

  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, $\mathrm{Var}(aX + bY + c) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$.

- **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.

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

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

- **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$(N_1 + \ldots + N_n, p)$.

- **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 $\mathbb{P}(\text{head}) = p$.

- **Poisson**: $X \sim$ Poisson(λ) (Poi(λ) 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 $\mathrm{Var}(X) = \lambda$. 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 $X_i \sim$ Poi(λ_i), then $X = X_1 + \ldots + X_n \sim$ Poi($\lambda_1 + \ldots + \lambda_n$).

- **Hypergeometric**: $X \sim$ HyperGeometric(N, K, n) (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 $\mathbb{E}[X] = n\frac{K}{N}$. ($\mathrm{Var}(X) = n \cdot \frac{K(N-K)(N-n)}{N^2(2N-1)}$ 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 Bin $\left(n, \frac{K}{N}\right)$.

- **Negative Binomial**: $X \sim$ 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 $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$. 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 $X_i \sim$ NegBin(r_i, p), then $X = X_1 + \ldots + X_n \sim$ NegBin($r_1 + \ldots + r_n, p$).

Task 1 – Content Review Questions

a) True or false: $Var(A+B) = Var(A) + Var(B)$ for **any** two random variables A and B

b) What is $Var(3A+4)$?

- ◯ $3Var(A) + 4$
- ◯ $3Var(A)$
- ◯ $9Var(A)$
- ◯ $Var(A)$

c) What is $P(X = 4)$ if X is a **continuous** random variable?

- ◯ 1
- ◯ 0
- ◯ not enough information

d) The cumulative distribution function for a continuous random variable X is $F_X(k) =$

- ◯ $\int_{-\infty}^{k} f_X(x)dx$
- ◯ $\int_{-\infty}^{\infty} f_X(x)dx$
- ◯ $\int_{k}^{\infty} f_X(x)dx$
- ◯ $\frac{d}{dk}f_X(k)$

e) The probability density function for a continuous random variable X is $f_X(k) =$

- ◯ $\int_{-\infty}^{k} f_X(x)dx$
- ◯ $\frac{d}{dk}F_X(k)$

f) **True** or **False**. If X is a continuous random variable, $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$

g) **True** or **False**. If X is a continuous random variable, $Var(X) = E[X^2] - (E[X])^2$

Task 2 – 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

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

Task 3 – Best Coach Ever!!

You are a hardworking boxer. Your coach tells you that the probability of your winning a boxing match is 0.2 independently of every other match.

a) How many matches do you expect to fight until you win 10 times and what kind of random variable is this?

b) You only get to play 12 matches every year. To win a spot in the Annual Boxing Championship, a boxer needs to win at least 10 matches in a year. What is the probability that you will go to the Championship this year and what kind of random variable is the number of matches you win out of the 12?

c) Let p be your answer to part (b). How many times can you expect to go to the Championship in your 20 year career?

Task 4 – True or False?

Identify the following statements as true or false (true means always true). Justify your answer.

a) For any random variable X, we have $\mathbb{E}[X^2] \geqslant \mathbb{E}[X]^2$.

b) Let X, Y be random variables. Then, X and Y are independent if and only if $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

c) Let $X \sim$ Binomial(n, p) and $Y \sim$ Binomial(m, p) be independent. Then, $X + Y \sim$ Binomial$(n + m, p)$.

d) Let $X_1, ..., X_{n+1}$ be independent Bernoulli(p) random variables. Then, $\mathbb{E}[\sum_{i=1}^{n} X_i X_{i+1}] = np^2$.

e) Let $X_1, ..., X_{n+1}$ be independent Bernoulli(p) random variables. Then, $Y = \sum_{i=1}^{n} X_i X_{i+1} \sim$ Binomial(n, p^2).

f) If $X \sim$ Bernoulli(p), then $nX \sim$ Binomial(n, p).

g) If $X \sim$ Binomial(n, p), then $\frac{X}{n} \sim$ Bernoulli(p).

h) For any two independent random variables X, Y, we have $\mathrm{Var}(X - Y) = \mathrm{Var}(X) - \mathrm{Var}(Y)$.

Task 5 – 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.

Task 6 – Fun with Poissons

Let $X \sim Poisson(\lambda_1)$ and $Y \sim Poisson(\lambda_2)$, where X and Y are independent.

a) Show that $X + Y \sim Poisson(\lambda_1 + \lambda_2)$. To show that a random variable is distributed according to a particular distribution, we must show that they have the same PMF. Thus, we are trying to show that $P(X + Y = n) = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}$

b) Show that $P(X = k \mid X + Y = n) = P(W = k)$ where $W \sim Bin(n, \frac{\lambda_1}{\lambda_1+\lambda_2})$

Task 7 – Hat Check

At a reception, n people give their hats to a hat-check person. When they leave, the hat-check person gives each of them a hat chosen at random from the hats that remain. What is the expected number of people who get their own hats back? (Notice that the hats returned to two people are not independent events: if a certain hat is returned to one person, it cannot also be returned to the other person.)

Task 8 – Balls and Bins

Throw n balls into m bins, where m and n are positive integers. Let X be the number of bins with exactly one ball. Compute $\mathrm{Var}(X)$.

Task 9 – Continuous r.v. example

Suppose that X is a random variable with pdf

$$f_X(x) = \begin{cases} 2C(2x - x^2) & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

where C is an appropriately chosen constant.

a) What must the constant C be for this to be a valid pdf?

b) Using this C, what is $\mathbb{P}(X > 1)$?

Task 10 – Throwing a dart

Consider the closed unit circle of radius r, i.e., $S = \{(x, y) : x^2 + y^2 \leqslant r^2\}$. Suppose we throw a dart onto this circle and are guaranteed to hit it, but the dart is equally likely to land anywhere in S. Concretely this means that the probability that the dart lands in any particular area of size A (that is entirely inside the circle of radius R), is equal to $\frac{\text{A}}{\text{Area of whole circle}}$. The density outside the circle of radius r is 0.

Let X be the distance the dart lands from the center. What is the CDF and pdf of X? What is $\mathbb{E}[X]$ and $\mathrm{Var}(X)$?

Task 11 – A square dartboard?

You throw a dart at an $s \times s$ square dartboard. The goal of this game is to get the dart to land as close to the lower left corner of the dartboard as possible. However, your aim is such that the dart is equally likely to land at any point on the dartboard. Let random variable X be the length of the side of the smallest *square* B in the lower left corner of the dartboard that contains the point where the dart lands. That is, the lower left corner of B must be the same point as the lower left corner of the dartboard, and the dart lands somewhere along the upper or right edge of B. For random variable X, find the CDF, PDF, $\mathbb{E}[X]$, and $\mathrm{Var}(X)$.