Probability 2

Review

A **random variable** X is a function from a sample space S to the real numbers $\mathbb{R}$.

An **indicator** random variable for the event A, often notated $\mathbb{1}_A$, has value 1 for all $s \in A$ and value 0 for all $s \in \overline{A}$.

A **Bernoulli** random variable with parameter p has value 1 with probability p and value 0 with probability $1 - p$. Thus, an indicator random variable $\mathbb{1}_A$ is a Bernoulli($\mathbb{P}(A)$) random variable.

Given a random variable $X : S \to \mathbb{R}$, we use the shorthand "$X = x$" to denote the subset $\{s \in S : X(s) = x\}$, and the shorthand "$X \leq x$" to denote the subset $\{s \in S : X(s) \leq x\}$.

The **probability density function** (pdf) of a random variable X is the function $f_X : \mathbb{R} \to [0, 1]$ with $f_X(x) = \mathbb{P}(X = x)$.

The **cumulative distribution function** (cdf) of a random variable X is the function $F_X : \mathbb{R} \to [0, 1]$ with $F_X(x) = \mathbb{P}(X \leq x)$.

A **binomial experiment** consists of n independent trials, each of which realizes a Bernoulli random variable of some fixed probability p. The number of successes in this experiment is a **binomial(n, p) random variable**.

A binomial(n, p) random variable has distribution $f_x(k) = \binom{n}{k} p^k (1 - p)^{n-k}$ for each integer $0 \leq k \leq n$.

Two random variables X and Y are **independent** if for each $x, y \in \mathbb{R}$, the events $\{X = x\}$ and $\{Y = y\}$ are independent. The definitions of mutual and pairwise independence are analogous to those of events.

The **expected value**, **expectation**, or **mean** of a random variable X (denoted $\mathbb{E}(X)$) is the value that we would expect if we averaged a large number of independent samples from its distribution. Formally, we have,

$$\mathbb{E}(X) = \sum_{x \in \mathbb{R}} x \mathbb{P}(X = x).$$

Here, we are assuming that X is a discrete random variable, so this sum has only countably many non-zero terms.

If X is an indicator random variable for the event A with probability p (so a Bernoulli(p) random variable), then $\mathbb{E}(X) = \mathbb{P}(X = 1) = \mathbb{P}(A) = p$.

The main property that is used frequently in probability is the **linearity of expectation**. For any (not necessarily independent) random variables X and Y and any $a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$. This easily extends to any finite linear combination of random variables.

If X is a binomial(n, p) random variable, so a sum of n Bernoulli(p) random variables, linearity of expectation gives us that $\mathbb{E}(X) = np$.

If X and Y are independent random variables, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

The **variance** of a random variable X, denoted $\mathrm{Var}(X)$, is a measurement of how "spread out" its distribution is; that is how far a sample from its distribution will typically lie from its expected value. Formally,

$$\mathrm{Var}(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big) = \mathbb{E}\Big(X^2\Big) - \Big(\mathbb{E}(X)\Big)^2$$

The **standard deviation** of a random variable is the square root of its variance.

The **variance** of a Bernoulli(p) random variable, X is $\mathrm{Var}(X) = \mathbb{E}\Big(X^2\Big) - \Big(\mathbb{E}(X)\Big)^2 = p - p^2 = p(1-p)$.

Variance is not linear. Rather, $\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$ for any random variable X and any constant $c \in \mathbb{R}$. Moreover, given *independent* random variables X and Y, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. This latter property extends to any finite sum of mutually independent random variables, giving us that the variance of a binomial(n, p) random variable is $np(1-p)$.

Markov's Inequality and Chebyshev's Inequality are both used to upper bound the probability that a random variable takes on a value far from its mean.

Markov's Inequality Suppose that X is a nonnegative random variable and $\alpha > 0$ is a constant. Then,

$$\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}(X)}{\alpha}.$$

Chebyshev's Inequality Suppose that X is a random variable and $\beta > 0$ is a constant. Then,

$$\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq \beta\Big) \leq \frac{\mathrm{Var}(X)}{\beta^2}.$$

1. A prize wheel is divided into 50 evenly-sized segments.

- 30 segments are blank.
- 10 segments give a $10 gift card.
- 3 segments give a $50 gift card.
- 1 segment gives a $200 television.
- 1 segment gives a $20,000 car.

How much should the owner of the wheel charge per spin so that they make a profit of 1 dollar per spin, on average?

2. Consider an experiment where 2 fair 6-sided dice are rolled. Define random variable X as the sum of the dice rolls, and random variable Y for their difference.

(a) Determine the range, probability density function, and cumulative distribution function of X.

(b) Determine the range and probability density function of Y.

(c) Are X and Y independent? Explain your answer.

3. An urn contains 3 red balls and 5 yellow balls. Suppose that we randomly place 3 balls from the urn into a glass bowl.

(a) What is the sample space of the experiment?

(b) What is the probability that the bowl contains 3 red balls?

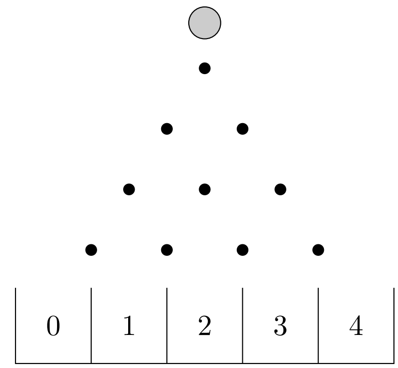
(c) Repeat your probability calculation from part (b) for the other possible numbers of red balls.

(d) What is the expected number of red balls in the glass bowl?

4.

The game of *Plinko* of size n includes a triangular array of n rows of pegs, above $n+1$ buckets, labeled 0 through n. The *Plinko* game of size 4 is shown on the right.

To play the game, a disk is dropped onto the top central peg. Then, it will deflect either left or right (each with probability $\frac{1}{2}$, landing squarely on one of the pegs in the second row. This process repeats until the disk lands in one of the buckets at the bottom. Assume that the direction the ball deflects off of lower pegs is independent of the previous path.



(a) Let B be the random variable representing the bucket number that the ball lands in. Describe the distribution of B. (**Hint:** How can you relate the bucket label with the directions that the ball bounced off of each of the pegs it hit?)

(b) Compute the probability density function of B for the $n = 4$ *Plinko* game.

(c) Suppose that when the game is set up, it is tipped slightly to the right. Now, the probability that the ball will deflect to the right is $\frac{3}{5}$. Describe the distribution of B in this case, and compute its probability density function in the case that $n = 4$.

5. Consider a group of 9 students. Suppose that each pair of students are friends with probability $\frac{1}{2}$ (independent of all other friendships).

(a) What is each student's expected number of friends?

(b) What is the probability that a student has exactly 3 friend?

(c) What is the expected number of students with exactly 3 friend?

6. A fair 6-sided dice is rolled 25 times.

(a) What is the expected value of a single dice roll?

(b) What is the expected sum of the dice rolls?

(c) What is the variance of a single dice roll?

(d) What is the variance of a the sum of the dice rolls?

(e) Use Markov's Inequality to bound the probability that the dice sum to at least 100.

(f) Use Chebyshev's Inequality to bound the probability that the dice sum to at least 100.

(g) Which inequality gave us a better bound?

7. A fair coin is flipped 30 times, and the results are recorded as a string of H's and T's, such as below:

TTHTTTHTTHTHTTTHTHTTHHHHTHHTHH

A *run* is a maximal sequence of identical flips. The above sequence has a run of 4 heads, and 2 runs of 3 tails. In this question, we'll explore some properties of runs in random sequences.

First, we'll compute the expected number of runs.

(a) For each $1 \leq i \leq 29$, let random variable D_i be an indicator variable for the event that flips i and $i+1$ were different. Describe the distribution of D_i.

(b) Express the number of runs, N, in terms of the D_i random variables.

(c) What is the expected number of runs?

(d) Compute the variance in the number of runs. (**Hint:** The D_i random variables are mutually independent. See if you can explain why.)

Note that these results give us a way to test whether a sequence written sequence of coin flips is likely to be random. Consider the sequence,

THTTHTHHTHTHTHHTTHTHTTHTHTTHTH

which has $N = 24$.

(e) Use Chebyshev's Inequality to bound the probability that such a sequence of 30 flips includes at least 24 runs.

Although this sequence "looks" random, our calculation shows that it is quite unlikely to be the result of a random experiment. The sequence shown at the top of the problem includes some longer runs, and has $N = 18$. This is very close to the expected value, so more likely to be random (in fact, this sequence did come from a random experiment).