CS 70 Discrete Mathematics and Probability Theory
Spring 2026 Sinclair, Song **DIS 11A**

Random Variables Intro

Note 16

Random Variable: A random variable X is a function from $\Omega \to \mathbb{R}$, mapping the possible outcomes to real numbers. Note that this function itself is not random; the *outcomes* are random. We define

$$\mathbb{P}[X = k] = \mathbb{P}[\{\omega \in \Omega : X(\omega) = k\}].$$

Distribution of a random variable: the set of all $(k, \mathbb{P}[X = k])$, describing the probability of attaining each value of the random variable.

Bernoulli Distribution: $X \sim \text{Bernoulli}(p)$; X represents the outcome of a biased coin flip. X is often called an *indicator random variable* of an event with probability p. The distribution is described as

$$\mathbb{P}[X = k] = \begin{cases} p & \text{if } k = 1 \\ 1 - p & \text{if } k = 0 \end{cases}$$

Binomial Distribution: $X \sim \text{Binomial}(n, p)$; X represents the number of successes in n independent trials, where p is the probability of success in each trial.

Joint Distribution of two RVs X, Y is $\mathbb{P}[X = x, Y = y]$. The *marginal distributions* are

$$\mathbb{P}[X = x] = \sum_y \mathbb{P}[X = x, Y = y]$$

$$\mathbb{P}[Y = y] = \sum_x \mathbb{P}[X = x, Y = y]$$

The conditional probability with two random variables is defined as

$$\mathbb{P}[X = x \mid Y = y] = \frac{\mathbb{P}[X = x, Y = y]}{\mathbb{P}[Y = y]}.$$

Two random variables are *independent* if and only if for all x, y:

$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x]\,\mathbb{P}[Y = y].$$

This is equivalent to saying that for all x, y:

$$\mathbb{P}[X = x \mid Y = y] = \mathbb{P}[X = x].$$

Expectation: just like a weighted average; we weight the values that X can take on by the probabilities of getting those values. Expectation is defined as

$$\mathbb{E}[X] = \sum_k k \cdot \mathbb{P}[X = k].$$

Linearity of Expectation: for two random variables X, Y (which could be dependent) and constants a, b:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

1 Head Count

Note 16

Consider a coin with $\mathbb{P}[\text{Heads}] = 2/5$. Suppose you flip the coin 20 times, and define X to be the number of heads.

(a) What is $\mathbb{P}[X = k]$, for some $0 \leq k \leq 20$? Express your answer in terms of k. (Do not just copy down a formula—re-derive it yourself!)

(b) What is the name of the distribution of X, and what are its parameters?

(c) What is $\mathbb{P}[12 \leq X \leq 14]$?

(d) What is $\mathbb{P}[X \geq 1]$? *Hint: You should be able to do this without a summation.*

(e) Now consider a second coin also with $\mathbb{P}[\text{Heads}] = 2/5$. Suppose you flip this second coin 30 times, and define Y to be the number of heads. What is the distribution of the *total* number of heads among these two coins, i.e. what is the distribution of $X + Y$?

Solution:

(a) There are a total of $\binom{20}{k}$ ways to select k coins to be heads. The probability that the selected k coins to be heads is $(\frac{2}{5})^k$, and the probability that the rest are tails is $(\frac{3}{5})^{20-k}$. Putting this together, we have

$$\mathbb{P}[X = k] = \binom{20}{k}\left(\frac{2}{5}\right)^k\left(\frac{3}{5}\right)^{20-k}.$$

(b) Since we have 20 independent trials, with each trial having a probability $2/5$ of success, we can write $X \sim \text{Binomial}(20, \frac{2}{5})$.

(c) The only way to write out this expression is as a sum of 3 different probabilities:

$$\begin{aligned}\mathbb{P}[12 \leq X \leq 14] &= \mathbb{P}[X = 12] + \mathbb{P}[X = 13] + \mathbb{P}[X = 14] \\ &= \binom{20}{12}\left(\frac{2}{5}\right)^{12}\left(\frac{3}{5}\right)^{8} + \binom{20}{13}\left(\frac{2}{5}\right)^{13}\left(\frac{3}{5}\right)^{7} + \binom{20}{14}\left(\frac{2}{5}\right)^{14}\left(\frac{3}{5}\right)^{6}.\end{aligned}$$

(d) Note that the probability that there is at least one head is the complement to the probability that there are zero heads. This means that

$$\mathbb{P}[X \geq 1] = 1 - \mathbb{P}[X = 0] = 1 - \left(\frac{3}{5}\right)^{20}.$$

(e) Since these two coins have the exact same behavior (i.e. the same bias), we can treat this as an experiment where we flip just one coin 50 times. Here, this means that $X+Y \sim \text{Binomial}(50, \frac{2}{5})$.

Note that more generally, for two independent binomial random variables $X \sim \text{Binomial}(n_1, p)$ and $Y \sim \text{Binomial}(n_2, p)$ with identical success probabilities, we have $X+Y \sim \text{Binomial}(n_1 + n_2, p)$, since we're essentially just increasing the number of trials we're considering.

2 Family Planning

Note 16

Mr. and Mrs. Johnson decide to continue having children until they either have their first girl or until they have three children. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let G denote the numbers of girls that the Johnsons have. Let C be the total number of children they have.

(a) Determine the sample space, along with the probability of each sample point.

(b) Compute the joint distribution of G and C. Fill in the table below.

	$C=1$	$C=2$	$C=3$
$G=0$			
$G=1$			

(c) Use the joint distribution to compute the marginal distributions of G and C and confirm that the values are as you'd expect. Fill in the tables below.

$\mathbb{P}[G=0]$	
$\mathbb{P}[G=1]$	

$\mathbb{P}[C=1]$	$\mathbb{P}[C=2]$	$\mathbb{P}[C=3]$

(d) Are G and C independent?

(e) What is the expected number of girls the Johnsons will have? What is the expected number of children that the Johnsons will have?

Solution:

(a) The sample space is the set of all possible sequences of children that the Johnsons can have: $\Omega = \{g, bg, bbg, bbb\}$. The probabilities of these sample points are:

$$\mathbb{P}[g] = \frac{1}{2}$$
$$\mathbb{P}[bg] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$
$$\mathbb{P}[bbg] = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$
$$\mathbb{P}[bbb] = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

(b)

	$C=1$	$C=2$	$C=3$
$G=0$	0	0	$\mathbb{P}[bbb]=1/8$
$G=1$	$\mathbb{P}[g]=1/2$	$\mathbb{P}[bg]=1/4$	$\mathbb{P}[bbg]=1/8$

(c) Marginal distribution for G:

$$\mathbb{P}[G=0]=0+0+\frac{1}{8}=\frac{1}{8}$$
$$\mathbb{P}[G=1]=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}=\frac{7}{8}$$

Marginal distribution for C:

$$\mathbb{P}[C=1]=0+\frac{1}{2}=\frac{1}{2}$$
$$\mathbb{P}[C=2]=0+\frac{1}{4}=\frac{1}{4}$$
$$\mathbb{P}[C=3]=\frac{1}{8}+\frac{1}{8}=\frac{1}{4}$$

(d) No, G and C are not independent. If two random variables are independent, then

$$\mathbb{P}[X=x,Y=y]=\mathbb{P}[X=x]\,\mathbb{P}[Y=y].$$

To show this dependence, consider an entry in the joint distribution table, such as $\mathbb{P}[G=0,C=3]=1/8$. This is not equal to $\mathbb{P}[G=0]\,\mathbb{P}[C=3]=(1/8)\cdot(1/4)=1/32$, so the random variables are not independent.

(e) We can apply the definition of expectation directly for this problem, since we've computed the marginal distribution for both random variables.

$$\mathbb{E}[G]=0\cdot\mathbb{P}[G=0]+1\cdot\mathbb{P}[G=1]=1\cdot\frac{7}{8}=\frac{7}{8}$$
$$\mathbb{E}[C]=1\cdot\mathbb{P}[C=1]+2\cdot\mathbb{P}[C=2]+3\cdot\mathbb{P}[C=3]=1\cdot\frac{1}{2}+2\cdot\frac{1}{4}+3\cdot\frac{1}{4}=\frac{7}{4}$$

3 Pullout Balls

Note 16

Suppose you have a bag containing four balls numbered $1,2,3,4$.

(a) You perform the following experiment: pull out a single ball and record its number. What is the expected value of the number that you record?

(b) You repeat the experiment from part (a), except this time you pull out two balls together and record the product of their numbers. What is the expected value of the total that you record?

Solution:

(a) Let X be the number that you record. Each ball is equally likely to be chosen, so

$$\mathbb{E}[X] = \sum_x x \cdot \mathbb{P}[X = x] = 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} + 4 \times \frac{1}{4} = 2.5.$$

As demonstrated here, the expected value of a random variable need not, and often is not, a feasible value of that random variable (there is no outcome ω for which $X(\omega) = 2.5$).

(b) Let Y be the product of two numbers that you pull out. Then

$$\mathbb{E}[Y] = \frac{1}{\binom{4}{2}}(1 \cdot 2 + 1 \cdot 3 + 1 \cdot 4 + 2 \cdot 3 + 2 \cdot 4 + 3 \cdot 4) = \frac{2+3+4+6+8+12}{6} = \frac{35}{6}.$$

4 Linearity

Note 16

Solve each of the following problems using linearity of expectation. Explain your methods clearly.

(a) In an arcade, you play game A 10 times and game B 20 times. Each time you play game A, you win with probability $1/3$ (independently of the other times), and if you win you get 3 tickets (redeemable for prizes), and if you lose you get 0 tickets. Game B is similar, but you win with probability $1/5$, and if you win you get 4 tickets. What is the expected total number of tickets you receive?

(b) A monkey types at a 26-letter keyboard with one key corresponding to each of the lower-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possibilities. If the monkey types 1 million letters, what is the expected number of times the sequence "book" appears? (*Hint*: Consider where the sequence "book" can appear in the string.)

Solution:

(a) Let T be the random variable for the total number of tickets you receive. Then, we can express T as

$$\begin{aligned} T &= 3 \cdot (\text{number of wins in game A}) + 4 \cdot (\text{number of wins in game B}) \\ &= 3 \cdot (A_1 + A_2 + \cdots + A_{10}) + 4 \cdot (B_1 + B_2 + \cdots + B_{20}). \end{aligned}$$

where A_i is the indicator variable that is 1 if you win on the i-th play of game A, and B_i is the indicator variable that is 1 if you win on the i-th play of game B. The expected value of A_i and B_i are

$$\begin{aligned} \mathbb{E}[A_i] &= 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} = \frac{1}{3}, \\ \mathbb{E}[B_i] &= 1 \cdot \frac{1}{5} + 0 \cdot \frac{4}{5} = \frac{1}{5}. \end{aligned}$$

Then by linearity of expectation,

$$\begin{aligned} E[T] &= 3\,\mathbb{E}[A_1] + \cdots + 3\,\mathbb{E}[A_{10}] + 4\,\mathbb{E}[B_1] + \cdots + 4\,\mathbb{E}[B_{20}] \\ &= 10\left(3 \cdot \frac{1}{3}\right) + 20\left(4 \cdot \frac{1}{5}\right) \\ &= 26. \end{aligned}$$

Note that $10\left(3 \cdot \frac{1}{3}\right)$ and $20\left(4 \cdot \frac{1}{5}\right)$ matches the expression directly gotten using the expected value of a binomial random variable.

(b) There are $1,000,000 - 4 + 1 = 999,997$ places where "book" can appear, each with a (non-independent) probability of $1/26^4$ of happening. If A is the random variable that tells how many times "book" appears, and A_i is the indicator variable that is 1 if "book" appears starting at the ith letter, then

$$\begin{aligned} \mathbb{E}[A] &= \mathbb{E}[A_1 + \cdots + A_{999,997}] \\ &= \mathbb{E}[A_1] + \cdots + \mathbb{E}[A_{999,997}] \\ &= \frac{999,997}{26^4} \approx 2.19. \end{aligned}$$