CS 70	Discrete Mathematics and Probability Theory	
Spring 2015	Vazirani	Discussion 12M

1. Sanity Check!

a. Define X to be the sum of n standard six-sided dice. What is $E[X]$?

Answer: We define X_i to be the value of the i^{th} dice. Then, by linearity of expectation $\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = n \cdot \frac{1}{6}(1+2+3+4+5+6) = 3.5n$.

b. Suppose we have a biased coin that comes up heads with probability p. After n tosses, what is the expected number of occurences of the subsequence HTH? (For example, the sequence $HTHHTHTTH$ has two occurences of HTH.)

Answer: Let X_i be the indicator variable for the event that we have subsequence HTH starting at position i. Then, we are after the expectation of $X = \sum_{i=1}^{n-2} X_i$. By linearity of expectation,

$$\mathbf{E}[X] = \sum_{i=1}^{n-2} \mathbf{E}[X_i] = (n-2)p(1-p)p = (n-2)(1-p)p^2.$$

2. Bernoulli and Binomial Distribution

A random variable X is called a Bernoulli random variable with parameter p if $X = 1$ with probability p and $X = 0$ with probability $1-p$.

a. Calculate $\mathrm{E}[X]$ and $\mathrm{Var}[X]$.

Answer:

$$\mathrm{E}[X] = p \cdot 1 + (1-p) \cdot 0 = p$$

$$\mathrm{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = p \cdot 1^2 + (1-p) \cdot 0 - p^2 = p(1-p)$$

b. A Binomial random variable with parameters n and p is defined to be the sum of n independent, identically distributed Bernoulli random variables with parameter p. If Z is a Binomial random variable with parameters n and p, what are $\mathrm{E}[Z]$ and $\mathrm{Var}[Z]$?

Answer: Let X_i be i.i.d. Bernoulli random variables with parameter p. By linearity of expectation,

$$\mathrm{E}[Z] = \sum_{i=1}^{n} \mathrm{E}[X_i] = np$$

Moreover, since X_i's are independent,

$$\mathrm{Var}[Z] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = np(1-p).$$

3. Chopping up DNA

In a certain biological experiment, a piece of DNA consisting of a linear sequence (or string) of 4000 nucleotides is subjected to bombardment by various enzymes. The effect of the bombardment is to randomly cut the string between pairs of adjacent nucleotides: each of the 3999 possible cuts occurs independently and with probability $1/500$.

a. What is the expected number of pieces into which the string is cut?

Answer: Let X_i be the indicator variable for the event that a cut occurs at position i. Then, the number of pieces into which the string is cut is $X = 1 + \sum_{i=1}^{3999} X_i$ (the number of pieces is one greater than the number of cuts). By linearity of expectation,

$$\mathrm{E}[X] = 1 + \sum_{i=1}^{3999} \frac{1}{500} = 1 + 3999/500.$$

b. What is the variance of the above quantity? (Hint: use problem 2.)

Answer: Note that

$$\mathrm{Var}[X] = \mathrm{Var}\left[1 + \sum_{i=1}^{3999} X_i\right] = \mathrm{Var}[1] + \mathrm{Var}\left[\sum_{i=1}^{3999} X_i\right] = \mathrm{Var}\left[\sum_{i=1}^{3999} X_i\right].$$

Note that $Z = \sum_{i=1}^{3999} X_i$ is a Binomial random variable with parameters $n = 3999$ and $p = 1/500$. Hence, the variance is $np(1-p) = 3999 \cdot \frac{1}{500} \cdot \frac{499}{500}$.

c. Suppose that the cuts are no longer independent, but highly correlated: when a cut occurs in a particular location, nearby locations are much more likely to be cut as well. The probability of each individual cut remains $1/500$. Does the expected number of pieces increase, decrease, or stay the same?

Answer: Since $\mathrm{E}[X_i]$ is still $1/500$ for each i, the expectation $\mathrm{E}[X] = 1 + \sum_{i=1}^{3999} \mathrm{E}[X_i]$ will stay the same.

4. Will I Get My Package?

A sneaky delivery guy of some company is out delivering n packages to n customers. Not only does he hand a random package to each customer, he tends to open a package before delivering with probability $\frac{1}{2}$ (independently of the choice of the package). Let X be the number of customers who receive their own packages unopened.

a. Compute the expectation $\mathrm{E}(X)$.

Answer: Define $X_i = \begin{cases} 1 & \text{if the } i\text{-th customer gets his/her own package unopened} \\ 0 & \text{otherwise.} \end{cases}$

Then, $\mathrm{E}(X_i) = \Pr[X_i = 1] = \frac{1}{2n}$, since the i-th customer will get his/her own package with probability $\frac{1}{n}$ and it will be unopened with probability $\frac{1}{2}$.

By linearity of expectation, $\mathrm{E}[X] = \mathrm{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{E}[X_i] = n \cdot \frac{1}{2n} = \frac{1}{2}$.

b. What is the probability that customers i and j both receive their own packages unopened?

Answer: The probability that both customers receive their own packages is $\frac{(n-2)!}{n!}$, and the probability that both customers receive unopened packages is $\frac{1}{4}$. Since the two events are independent, the probability that both customers receive their own packages unopened is $\frac{1}{4}\frac{(n-2)!}{n!} = \frac{1}{4n(n-1)}$.

c. Compute the variance Var(X).

Answer: Since $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$, we need to first compute $\mathrm{E}(X^2)$.
Note that $\mathrm{E}(X^2) = \mathrm{E}\big((X_1 + X_2 + \ldots + X_n)^2\big) = \mathrm{E}(\sum_{i,j} X_i X_j) = \sum_{i,j} \mathrm{E}(X_i X_j)$, where the last equality follows from linearity of expectation.
If $i = j$, then $\mathrm{E}(X_i X_j) = \mathrm{E}(X_i^2) = \frac{1}{2n}$.
If $i \neq j$, then $\mathrm{E}(X_i X_j) = \Pr(X_i X_j = 1) \cdot 1 + \Pr(X_i X_j = 0) \cdot 0 = \frac{1}{4n(n-1)}$.
Hence, $\mathrm{E}(X^2) = \sum_{i,j} \mathrm{E}(X_i X_j) = \sum_i \mathrm{E}(X_i^2) + \sum_{i \neq j} \mathrm{E}(X_i X_j) = n \cdot \dfrac{1}{2n} + n(n-1) \cdot \dfrac{1}{4n(n-1)} = \dfrac{1}{2} + \dfrac{1}{4} = \dfrac{3}{4}$.
It follows that $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$.