CSE 312: Foundations of Computing II
Quiz Section #6: Discrete RV's, Conditional Expectation, Tail Bounds
(solutions)

Review: Main Theorems and Concepts

Variance: Let X be a random variable and $\mu = \mathbb{E}[X]$. The variance of X is defined to be $\text{Var}(X) = \mathbb{E}[(X-\mu)^2]$. Notice that since this is an expectation of a nonnegative random variable ($(X-\mu)^2$), variance is always nonnegative. With some algebra, we can simplify this to $\text{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X]$.

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, $\text{Var}(aX + b) = a^2 Var(X)$.

Independence: Random variables X and Y are independent, written $X \perp Y$, 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 $X \perp Y$, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$. 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 $X \perp Y$, $\text{Var}(aX + bY + c) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)$.

Conditional Expectation: Let X be a random variable, and E be an event. Then, $\mathbb{E}[X \mid E] = \sum_x x \cdot \mathbb{P}(X = x \mid E)$.

Law of Total Expectation: Let X be a random variable, and $E_1, \ldots, E_n$ a partition of the sample space. Then, $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X \mid E_i] \cdot \mathbb{P}(E_i)$. In particular, if Y is a random variable, then $\mathbb{E}[X] = \sum_y \mathbb{E}[X \mid Y = y] \cdot \mathbb{P}(Y = y)$, since the events where $\{Y = y\}$ form a partition.

Markov's Inequality: Let X be a non-negative random variable, and $\alpha > 0$. Then, $\mathbb{P}(X \geq \alpha) \leq \frac{\mathbb{E}[X]}{\alpha}$.

Chebyshev's Inequality: Suppose Y is a random variable with $\mathbb{E}[Y] = \mu$ and $\text{Var}(Y) = \sigma^2$. Then, for any $\alpha > 0$, $P(|Y - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$.

Zoo of Discrete Random Variables

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

Bernoulli (or indicator): $X \sim Ber(p)$ 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 $\text{Var}(X) = p(1-p)$. An example of a Bernoulli r.v. is one flip of a coin with $P(\text{head}) = p$. By a clever trick, we can write

$$p_X(k) = p^k (1-p)^{1-k}, \quad k = 0, 1$$

Binomial: $X \sim Bin(n, p)$ iff X is the sum of n iid $Ber(p)$ 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 $\text{Var}(X) = np(1-p)$. An example of a Binomial r.v. is the number of heads in n independent flips of a coin with $P(\text{head}) = p$. Note that $Bin(1, p) \equiv Ber(p)$. As $n \to \infty$ and $p \to 0$, with $np = \lambda$, then $\text{Bin}(n, p) \to Poi(\lambda)$. 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: $X \sim Geo(p)$ 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 $\text{Var}(X) = \frac{1-p}{p^2}$. An example of a Geometric r.v. is the number of independent coin flips up to and including the first head, where $P(\text{head}) = p$.

Negative Binomial: $X \sim NegBin(r, p)$ iff X is the sum of r iid $Geo(p)$ 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 $\text{Var}(X) = \frac{r(1-p)}{p^2}$. An example of a Negative Binomial r.v. is the number of independent coin flips up to and including the r^{th} head, where $P(\text{head}) = p$. If $X_1, \ldots, X_n$ are independent Negative Binomial r.v.'s, where $X_i \sim NegBin(r_i, p)$, then $X = X_1 + \ldots + X_n \sim NegBin(r_1 + \ldots + r_n, p)$.

Poisson: $X \sim Poi(\lambda)$ 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 $\text{Var}(X) = \lambda$. 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 $X_i \sim Poi(\lambda_i)$, then $X = X_1 + \ldots + X_n \sim Poi(\lambda_1 + \ldots + \lambda_n)$.

Hypergeometric: $X \sim HypGeo(N, K, n)$ iff X has the following probability mass function:

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad k = max\{0, n+K-N\}, \ldots, \min\{K, n\}$$

$\mathbb{E}[X] = n\frac{K}{N}$. 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 $\text{Bin}\left(n, \frac{K}{N}\right)$.

Exercises

1. Suppose we roll two fair 5-sided dice independently. Let X be the value of the first die, Y be the value of the second die, $Z = X + Y$ be their sum, $U = \min\{X, Y\}$ and $V = \max\{X, Y\}$.

   (a) Find $p_U(u)$.

$$p_U(u) = \begin{cases} \frac{9}{25}, & u = 1 \\ \frac{7}{25}, & u = 2 \\ \frac{5}{25}, & u = 3 \\ \frac{3}{25}, & u = 4 \\ \frac{1}{25}, & u = 5 \end{cases}$$

   (b) Find $\mathbb{E}[U]$.

$$\mathbb{E}[U] = 1 \cdot \frac{9}{25} + 2 \cdot \frac{7}{25} + 3 \cdot \frac{5}{25} + 4 \cdot \frac{3}{25} + 5 \cdot \frac{1}{25} = \frac{55}{25} = 2.2$$

   (c) Find $\mathbb{E}[Z]$.

   We know $X, Y \sim Unif(1, 5)$, so $\mathbb{E}[X] = \mathbb{E}[Y] = \frac{1+5}{2} = 3$.

$$\mathbb{E}[Z] = \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 3 + 3 = 6$$

   (d) Find $\mathbb{E}[UV]$.

$$\mathbb{E}[\text{UV}] = \mathbb{E}[\text{XY}] = \mathbb{E}[X]\,\mathbb{E}[Y] = 3^2 = 9$$

   Since $UV = XY$, and then X, Y are independent.

   (e) Find $Var(U + V)$.

   Since $X, Y \sim Unif(1, 5)$, $\text{Var}(X) = Var(Y) = \frac{(5-1)(5-1+2)}{12} = 2$.

$$\text{Var}(U+V) = Var(X+Y) = Var(X) + Var(Y) = 2 + 2 = 4$$

Since $U + V = X + Y$, and then X, Y are independent.

2. Suppose X has the following probability mass function:

$$p_X(x) = \begin{cases} c, & x = 0 \\ 2c, & x = \frac{\pi}{2} \\ c, & x = \pi \\ 0, & \text{otherwise} \end{cases}$$

(a) Suppose $Y_1 = \sin(X)$. Find $\mathbb{E}[Y_1^2]$.

Probabilities must sum to 1, so $c = 1/4$.

$$\mathbb{E}\left[Y_1^2\right] = \mathbb{E}\left[\sin^2(X)\right] = \frac{1}{4}\sin^2(0) + \frac{1}{2}\sin^2\left(\frac{\pi}{2}\right) + \frac{1}{4}\sin^2(\pi) = \frac{1}{2}$$

(b) Suppose $Y_2 = \cos(X)$. Find $\mathbb{E}[Y_2^2]$.

$$\mathbb{E}\left[Y_2^2\right] = \mathbb{E}\left[\cos^2(X)\right] = \frac{1}{4}\cos^2(0) + \frac{1}{2}\cos^2\left(\frac{\pi}{2}\right) + \frac{1}{4}\cos^2(\pi) = \frac{1}{2}$$

(c) Suppose $Y = Y_1^2 + Y_2^2 = \sin^2(X) + \cos^2(X)$. Before any calculation, what do you think $\mathbb{E}[Y]$ should be? Find $\mathbb{E}[Y]$, and see if your hypothesis was correct. (Recall for any real number x, $\sin^2(x) + \cos^2(x) = 1$).

I expect the answer to be 1, since for any real number x, $\sin^2(x) + \cos^2(x) = 1$, but I'm not sure since these are random variables and not real numbers.

$$\mathbb{E}[Y] = \mathbb{E}\left[Y_1^2 + Y_2^2\right] = \mathbb{E}\left[Y_1^2\right] + \mathbb{E}\left[Y_2^2\right] = \frac{1}{2} + \frac{1}{2} = 1$$

(d) Let W be any discrete random variable with probability mass function $p_W(w)$. Then, $\mathbb{E}[\sin^2(W) + \cos^2(W)] = 1$. Is this statement always true? If so, prove it. If not, give a counterexample by giving a probability mass function for a discrete random variable W for which the statement is false.

This is true. Recall for a discrete random variable, $\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$.

$$\mathbb{E}[\sin^2(W) + \cos^2(W)] = \sum_w (\sin^2(w) + \cos^2(w))p_W(w) = \sum_w p_W(w) = 1$$

3. Consider the following scenarios:

(a) Let $X_1, ..., X_n$ be iid (independent and identically distributed) random variables with mean μ and variance σ^2. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean. Compute $\mathbb{E}[\overline{X}_n]$ and $Var(\overline{X}_n)$.

$$\mathbb{E}[\overline{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n}\cdot n\mu = \mu$$

$$Var(\overline{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

(b) Suppose n students take a CSE 312 exam with scores ranging in $\{0, 1, ..., 100\}$, mean 50. Give an upper bound on the probability that a student gets over 80.

Using Markov's inequality, $\mathbb{P}(X \geq 80) \leq \frac{\mathbb{E}[X]}{80} = \frac{50}{80} = \frac{5}{8}$

(c) Continuing from the previous part, suppose you also know the variance of scores is 25. Give an upper bound on the probability that a student gets over 80.

Using Chebyshev's inequality,

$$\mathbb{P}(X \geq 80) \leq \mathbb{P}(X \geq 80 \text{ or } X \leq 20) = \mathbb{P}(|X - 50| \geq 30) \leq \frac{Var(X)}{30^2} = \frac{25}{30^2} = \frac{1}{36}$$

(d) How large should n be such that the sample average is farther away from 50 by 10 with probability at most 0.01?

Let $X_1, ..., X_n$ be the student's scores. Then, $\mathbb{E}[\overline{X}_n] = 50$ and $Var(\overline{X}_n) = \frac{25}{n}$ by part (a). By Chebyshev's inequality, we have

$$\mathbb{P}(|\overline{X}_n - 50| \geq 10) \leq \frac{Var(\overline{X}_n)}{10^2} = \frac{25}{100n} = \frac{1}{4n}$$

We want this ≤ 0.01, so we solve $\frac{1}{4n} \leq 0.01$ to get that $n \geq 25$.

4. Suppose I run a lemonade stand outside, which costs me \$100 a day to operate. I sell a drink of lemonade for \$20. Each person who walks by my stand either buys no drink, or exactly 1 drink. If it rains, only n_1 people walk by my stand, and each buy a drink independently with probability p_1. If it doesn't rain, n_2 people walk by my stand, and each buy a drink independently with probability p_2. It rains every day with probability p_3, independently of each other day. Let X be my **profit** over the next week. What is $\mathbb{E}[X]$?

   Let R be the event it rains, and X_i be how many drinks I sell on day i for $i \in [7]$. Then,

$$X = \sum_{i=1}^{7} (20X_i - 100)$$

   We can see that

$$(X_i \mid R) \sim Bin(n_1, p_1), \quad \mathbb{E}[X_i \mid R] = n_1 p_1$$
$$(X_i \mid R^C) \sim Bin(n_2, p_2), \quad \mathbb{E}[X_i \mid R^C] = n_2 p_2$$

   By the law of total expectation,

$$\mathbb{E}[X_i] = \mathbb{E}[X_i \mid R]\mathbb{P}(R) + \mathbb{E}[X_i \mid R^C]\mathbb{P}(R^C) = n_1 p_1 p_3 + n_2 p_2 (1 - p_3)$$

   So

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{7} (20X_i - 100)\right] = 20 \sum_{i=1}^{7} \mathbb{E}[X_i] - 700 = 140[n_1 p_1 p_3 + n_2 p_2 (1 - p_3)] - 700$$

5. Let N be a random variable which can take on only nonnegative integer values, which has mean γ. Let $X_1, ..., X_N$ be a **random** number of iid random variables with common mean μ, such that each X_i is independent of N. Define $X = \sum_{i=1}^{N} X_i$. What is $\mathbb{E}[X]$?

   First, notice that

$$\mathbb{E}[X \mid N = n] = \mathbb{E}\left[\sum_{i=1}^{N} X_i \mid N = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu$$

   By the law of total expectation,

$$\mathbb{E}[X] = \sum_{n} \mathbb{E}[X \mid N = n]\mathbb{P}(N = n) = \sum_{n} n\mu \cdot \mathbb{P}(N = n) = \mu \sum_{n} n \cdot \mathbb{P}(N = n) = \mu \cdot \mathbb{E}[N] = \gamma\mu$$

6. 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

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

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

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

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

$$Poi(5r)$$

(d) whether or not my next fish is blue

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

(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

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

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

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

7. Suppose $Y_1, \ldots, Y_n$ are iid with $\mathbb{E}[Y_i] = \mu$ and $\text{Var}(Y_i) = \sigma^2$, and let $Y = \frac{1}{n}\sum_{i=1}^n iY_i$. What is $\mathbb{E}[Y]$ and $\text{Var}(Y)$? Recall that $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ and $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$.

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n iY_i\right] = \frac{1}{n}\sum_{i=1}^n i\mathbb{E}[Y_i] = \frac{\mu}{n}\sum_{i=1}^n i = \frac{\mu}{n}\frac{n(n+1)}{2} = \frac{\mu(n+1)}{2}$$

$$\text{Var}(Y) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^n iY_i\right) = \frac{1}{n^2}\sum_{i=1}^n i^2\text{Var}(Y_i) = \frac{\sigma^2}{n^2}\frac{n(n+1)(2n+1)}{6} = \sigma^2\frac{(n+1)(2n+1)}{6n}$$

8. Is the following statement true or false? If $\mathbb{E}[\mathrm{XY}] = \mathbb{E}[X]\,\mathbb{E}[Y]$, then $X \perp Y$. If it is true, prove it. If not, provide a counterexample.

As mentioned in the review section, this is false.

Let $X \sim Unif(-1, 1)$ and $Y = X^2$. Notice that $XY = X^3 = X$. Then $\mathbb{E}[\mathrm{XY}] = 0 = \mathbb{E}[X]\,\mathbb{E}[Y]$. But X and Y are not independent because $P(Y = 1|X = 1) = 1 \neq \frac{2}{3} = P(Y = 1)$.