CSE 312: Foundations of Computing II
Quiz Section #6: Variance, Independence, Zoo of Discrete Random Variables

Review/Mini-Lecture/Main Theorems and Concepts From Lecture

Variance: Let $X$ be a random variable and $\mu = E[X]$. The variance of $X$ is defined to be $Var(X) = E[(X - \mu)^2]$. 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 $Var(X) = E[X^2] - E^2[X]$.

Independence: Random variables $X$ and $Y$ are independent, written $X \perp Y$, iff
\[ \forall x, y, P(X = x \cap Y = y) = P(X = x)P(Y = y) \]
In this case, we have $E[XY] = E[X]E[Y]$ (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) if they are independent and have the same probability mass function.

Property of Variance: Let $a, b \in \mathbb{R}$ and $X$ a random variable. Then, $Var(ax + b) = a^2Var(X)$.

Linearity of Variance: If $X \perp Y$, $Var(X + Y) = Var(X) + Var(Y)$. Linearity of variance depends on independence, whereas linearity of expectation always holds. Note that this combined with the above show that $\forall a, b, c \in \mathbb{R}$ and if $X \perp Y$, $Var(ax + by + c) = a^2Var(X) + b^2Var(Y)$.

Zoo of Discrete Random Variables

Uniform: $X \sim Unif(a, b)$ if $X$ has the following probability mass function:
\[ p_X(k) = \frac{1}{b-a+1}, \quad k = a, \ldots, b \]
$E[X] = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)(b-a+2)}{12}$. 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): $X \sim Ber(p)$ if $X$ has the following probability mass function:
\[ p_X(k) = \begin{cases} p, & k = 1 \\ 1-p, & k = 0 \end{cases} \]
$E[X] = p$ and $Var(X) = p(1-p)$. An example of a Bernoulli r.v. is one flip of a coin with $P(head) = p$. By a clever trick, we can write
\[ p_X(k) = p^k(1-p)^{1-k}, \quad k = 0,1 \]

Binomial: $X \sim Bin(n, p)$ if $X$ is the sum of iid $Ber(p)$ random variables, and has pmf
\[ p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,1,\ldots,n \]
$E[X] = np$ and $Var(X) = np(1-p)$. An example of a Binomial r.v. is the number of heads in $n$ independent flips of a coin with $P(head) = p$. Note that $Bin(1, p) \equiv Ber(p)$.

As $n \to \infty$ and $p \to 0$, with $np = \lambda$, then $Bin(n, p) \to Poi(\lambda)$. If $X_1, \ldots, X_n$ are independent Binomial r.v.’s, where $X_i \sim Bin(N_i, p)$, then $X = X_1 + \cdots + X_n \sim Bin(N_1 + \cdots + N_n, p)$. 
Section #6 Review

Geometric: $X \sim \text{Geo}(p)$ if $X$ has the following probability mass function:

$$p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

$E[X] = \frac{1}{p}$ and $Var(X) = \frac{1-p}{p^2}$. 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: $X \sim \text{NegBin}(r, p)$ if $X$ is the sum of iid Geometric random variables, and has pmf

$$p_X(k) = \begin{pmatrix} r - 1 \\ k - 1 \end{pmatrix} p^r (1-p)^{k-r}, \quad k = r, r + 1, \ldots$$

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

Poisson: $X \sim \text{Poi} (\lambda )$ if $X$ has the following probability mass function:

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots$$

$E[X] = \lambda$ and $Var(X) = \lambda$. An example of a Poisson r.v. is the number of people being born in a minute, where $\lambda$ is the average rate per unit time. If $X_1, \ldots, X_n$ are independent Poisson r.v.’s, where $X_i \sim \text{Poi}(\lambda_i)$, then $X = X_1 + \cdots + X_n \sim \text{Poi}(\lambda_1 + \cdots + \lambda_n)$.

Hypergeometric: $X \sim \text{HypGeo}(N, K, n)$ if $X$ has the following probability mass function:

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

$E[X] = n \frac{K}{N}$. 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 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

   $$\text{Poi}(5r)$$

   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

   $$\text{Bin} \left( 10, \frac{B}{N - (B+G)} \right)$$
Section #6 Review

\textit{HypGeo}(N, B, 10)

f) how many fish I have to catch until I catch three red fish, if I catch and release \textit{NegBin}\left(3, \frac{R}{N}\right)

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

\[
E[Y] = E\left[\frac{1}{n} \sum_{i=1}^{n} iY_i\right] = \frac{1}{n} \sum_{i=1}^{n} iE[Y_i] = \frac{\mu}{n} \sum_{i=1}^{n} i = \frac{\mu n(n+1)}{2n} = \frac{\mu(n+1)}{2} \\
Var(Y) = Var\left(\frac{1}{n} \sum_{i=1}^{n} iY_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} i^2Var(Y_i) = \frac{\sigma^2 n(n+1)(2n+1)}{6n} = \frac{\sigma^2 (n+1)(2n+1)}{6n}
\]

3. Is the following statement true or false? If \(E[XY] = E[X]E[Y]\), 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 \(E[XY] = 0 = E[X]E[Y]\). But \(X\) and \(Y\) are not independent because \(P(Y = 1|X = 1) = 1 \neq \frac{2}{3} = P(Y = 1)\).

4. 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 \(p_U(u)\).

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

b) Find \(E[U]\).

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

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


d) Find $E[UV]$.


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

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

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

$$Var(U + V) = Var(X + Y) = Var(X) + Var(Y) = 2 \cdot 2 = 4$$

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

5. 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 $E[Y_1^2]$.

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

$$E[Y_1^2] = E[\sin^2(X)] = 1/4 \sin^2(0) + 1/2 \sin^2(\pi/2) + 1/4 \sin^2(\pi) = 1/2$$

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

$$E[Y_2^2] = E[\cos^2(X)] = 1/4 \cos^2(0) + 1/2 \cos^2(\pi/2) + 1/4 \cos^2(\pi) = 1/2$$

c) Suppose $Y = Y_1^2 + Y_2^2 = \sin^2(X) + \cos^2(X)$.  Before any calculation, what do you think $E[Y]$ should be?  Find $E[Y]$, 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.

$$E[Y] = E[Y_1^2 + Y_2^2] = E[Y_1^2] + E[Y_2^2] = 1/2 + 1/2 = 1$$
d) Let $W$ be any discrete random variable with probability mass function $p_W(w)$. Then, $E[\sin^2(W) + \cos^2(W)] = 1$. 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, $E[g(X)] = \sum_x g(x)p_X(x)$.

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

6. If electricity power failures occur according to a Poisson distribution with an average of 3 failures every twenty weeks, calculate the probability that there will be more than one failure during a particular week.

\[
X \sim \text{Poi}\left(\frac{3}{20}\right)
\]

$$P(X > 1) = 1 - p_X(0) - p_X(1) = 1 - e^{-0.15}0.15^0 - e^{-0.15}0.15^1 \approx 0.01$$

7. A company makes electric motors. The probability an electric motor is defective is 0.01, independent of other motors made. What is the probability that a sample of 300 electric motors will contain exactly 5 defective motors? Do it first exactly, then approximate it with the Poisson. How good was the approximation?

\[
X \sim \text{Bin}(300,0.01)
\]

$$p_X(5) = \binom{300}{5}(0.01)^5(0.99)^{295} \approx 0.10099$$

Average rate is $\lambda = np = 300 \times 0.01 = 3$

\[
X' \sim \text{Poi}(3)
\]

$$p_{X'}(5) = \frac{e^{-3}3^5}{5!} \approx 0.10082$$

8. An average page in a book contains one typo. What is the probability that there are exactly 8 typos in a given 10-page chapter, using the Poisson model?

\[
X \sim \text{Poi}(10)
\]

$$p_X(8) = \frac{e^{-10}10^8}{8!} \approx 0.113$$

Cool puzzles from earlier topics
9. A plane has 100 seats and 100 passengers. The first person to get on the plane lost his ticket and doesn't know his assigned seat, so he picks a seat uniformly at random to sit in. Every remaining person knows their seat, so if it is available they sit in it, and if it is unavailable they pick a uniform random remaining seat. What is the probability the last person to get on gets to sit in his own seat?

We will prove that only the first and last person's seat could possibly be available by the time the last person boards. Suppose for contradiction that the i-th person's seat, where 2 ≤ i ≤ 99, was available at the end. Then that i-th person's seat was available the whole time, and the i-th person would have sat there after boarding, which is a contradiction. So only the first or last person's seat could possibly be left when the last person boards.

One of those two seats must be occupied when the last person boards. Consider the person who took that seat, say passenger j, where 1 ≤ j ≤ 99. When j boards, both the first and last person’s seats are still available. Passenger j is equally likely to take either of those two seats. Therefore, the probability that the last person’s seat is available when the last person boards is ½.

10. Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?

It’s always better to switch doors. I like to think of this problem with 100 doors instead of 3 doors. When you pick one of them, the game show host opens all 98 other doors, so you’re left with the door you picked and the remaining door. What would you pick in this scenario? You had a 1/100 chance of being right on your first pick, and a 99/100 chance that the car was behind any of the other doors. Now that the 98 doors have been revealed, there is a 99/100 chance that the car is behind the last door.

11. You flip a fair coin independently and count the number of flips until the first tail, including that tail flip in the count. If the count is n, you receive $2^n$ dollars. What is the expected amount you will receive? How much would you be willing to pay at the start to play this game?

If you calculate the expected value of the amount you receive, you’ll see that it’s infinite. Would you still pay a lot of money to play this game? I mean, you are expected to receive an infinite amount of money in return, right?