CSE 312: Foundations of Computing II
Quiz Section #7: Zoo of 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

$$Var(X) = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

Notice that since this is an expectation of a _____________ random variable ($(X-\mu)^2$), variance is always _____________. 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

____

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 $X_1, \dots, X_n$ are i.i.d. (or iid) if they are ___________ and have the same ___________________________________.

Property of Variance: Let $a, b \in \mathbb{R}$ and X a random variable. Then,

$$Var(aX + b) = \_\_\_\_\_\_\_\_\_$$

Variance of Independent Variables: If $X \perp Y$, $Var(X+Y) = Var(X) + Var(Y)$. This 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 $X \perp Y$,

$$Var(aX + bY + c) = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

Continuous Random Variable: A r.v. which can take on an uncountably infinite number of values.

Probability Density Function (pdf or density): Let X be a continuous random variable. Then $f_X(x)$ is the density of X. Note that $f_X(x) \neq P(X = x)$, since $P(X = x) = 0$ for all x if X is continuous. However, the probability that X is close to x is proportional to $f_X(x)$: for small δ, $P\left(x - \frac{\delta}{2} < X < x + \frac{\delta}{2}\right) \approx \delta f_X(x)$.

Cumulative Distribution Function (cdf): For a continuous random variable, $P(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ and therefore $F_X'(x) = f_X(x)$.

Univariate: Discrete to Continuous:

	Discrete	**Continuous**
PMF/PDF	$p_X(x) = P(X = x)$	$f_X(x) \neq P(X = x)$

CDF	$F_X(x) = \sum_{t \le x} p_X(t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$
Normalization	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)dx = 1$
Expectation	$E[g(X)] = \sum_x g(x)p_X(x)$	$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$

Zoo of Discrete Random Variables

Uniform: $X \sim Unif(a, b)$ if X has the following probability mass function:

$$p_X(k) = \frac{1}{b-a+1}, \qquad k = a, \dots, b$$

$E[X] = \frac{a+b}{2}$ 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

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

Binomial: $X \sim Bin(n, p)$ if X is the sum of iid $Ber(p)$ random variables, and has pmf

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n$$

$E[X] = np$ 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 $X_1, \dots, X_n$ are independent Binomial r.v.'s, where $X_i \sim Bin(N_i, p)$, then $X = X_1 + \dots + X_n \sim Bin(N_1 + \dots + N_n, p)$.

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

$$p_X(k) = (1-p)^{k-1}p, \qquad k = 1, 2, \dots$$

$E[X] = \frac{1}{p}$ 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 $P(head) = p$.

Negative Binomial: $X \sim NegBin(r, p)$ if X is the sum of iid Geometric random variables, and has pmf

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \qquad k = r, r+1, \ldots$$

$E[X] = \frac{r}{p}$ 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 $P(head) = p$. If $X_1, \ldots, X_n$ are independent Negative Binomial r.v.'s, where $X_i \sim NegBin(r_i, p)$, then $X = X_1 + \cdots + X_n \sim NegBin(r_1 + \cdots + r_n, p)$.

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

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

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

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

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

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

Zoo of Continuous Random Variables

Uniform: $X \sim Unif(a, b)$ if X has the following probability density function:

$$f_X(x) = \frac{1}{b-a}, \qquad x \in [a, b]$$

$E[X] = \frac{a+b}{2}$ and $Var(X) = \frac{(b-a)^2}{12}$. This represents each real number from $[a, b]$ to be equally likely.

Exponential: $X \sim Exp(\lambda)$ if X has the following probability density function:

$$f_X(x) = \lambda e^{-\lambda x}, \qquad x \geq 0$$

$E[X] = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$. $F_X(x) = 1 - e^{-\lambda x}, x \geq 0$. The exponential random variable is the continuous analog to the geometric random variable: it represents the waiting time to the first success where $\lambda > 0$ is the average number of events per unit time. Note that the exponential measures how much time passes until the first success (any real number, continuous), where the Poisson measures how many events in a unit of time (nonnegative integer, discrete). X is memoryless:

for any $s, t \geq 0$, $P(X > s + t \mid X > s) = P(X > t)$.

The geometric r.v. also has this property.

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

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

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

d) whether or not my next fish is blue

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

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

2. Suppose I have $Y_1, \dots, Y_n$ 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 $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

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.

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 $p_U(u)$.

b) Find $E[U]$.

c) Find $E[Z]$.

d) Find $E[UV]$.

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

5. Suppose X has the following probability mass function:

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

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

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

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, $\sin^2(x) + \cos^2(x) = 1$).

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.

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.

7. 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?

8. Alex decided he wanted to create a "new" type of distribution that will be famous, but he needs some help. He knows he wants it to be continuous and have uniform density, but he needs help working out some of the details. We'll denote a random variable X having the "Uniform-2" distribution as $X \sim \text{Unif2}(a, b, c, d)$, where $a < b < c < d$. We want the density to be non-zero in $[a, b]$ and $[c, d]$, and zero everywhere else. Anywhere the density is non-zero, it must be equal to the same constant.

a) Find the probability density function, $f_X(x)$. Be sure to specify the values it takes on for every point in $(-\infty, \infty)$. (Hint: use a piece-wise definition).

b) Find the cumulative distribution function, $F_X(x)$. Be sure to specify the values it takes on for every point in $(-\infty, \infty)$. (Hint: use a piece-wise definition).

9. Suppose $X \sim \text{Unif}(0,1)$ and $Y = e^X$. Find $f_Y(y)$.

10. A single-stranded (1-dimensional) spider web, with length W centimeters, where W>4, is stretched taut between two fence posts. The homeowner (a spider) sits precisely at the midpoint of this web. Suppose that a fly gets caught at a random point on the strand, with each point being equally likely.
a) The spider is lazy, and it is only willing to walk over and eat the fly if the fly lands within 2 centimeters of where the spider sits. What is the probability that the spider eats the fly?

b) Let X be the random variable that represents the spider's distance from the fly's landing point. Calculate the CDF, PDF, expectation, and variance of X.

Cool puzzles from earlier topics

11. 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?

12. 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?

13. 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?