CSE 312: Foundations of Computing II

Section 5: Variance, Important Discrete Distributions Solutions

0. Review of Main Concepts

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

(b) **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 $\sigma = \sqrt{Var(X)}$.

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

(d) **Independence**: Random variables X and Y are independent iff

$$\forall x \forall y, \quad \mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

In this case, we have $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ (the converse is not necessarily true).

(e) **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.

(f) **Variance of Independent Variables**: If X is independent of 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 is independent of Y, $Var(aX+bY+c) = a^2 Var(X) + b^2 Var(Y)$.

1. Zoo of Discrete Random Variables

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

(b) **Bernoulli (or indicator)**: $X \sim \text{Bernoulli}(p)$ ($\text{Ber}(p)$ for short) 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 $Var(X) = p(1-p)$. An example of a Bernoulli r.v. is one flip of a coin with $\mathbb{P}(\text{head}) = p$.

(c) **Binomial**: $X \sim \text{Binomial}(n, p)$ ($\text{Bin}(n, p)$ for short) iff X is the sum of n iid $\text{Bernoulli}(p)$ 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 $Var(X) = np(1-p)$. An example of a Binomial r.v. is the number of heads in n independent flips of a coin with $\mathbb{P}(\text{head}) = p$. Note that $\text{Bin}(1, p) \equiv \text{Ber}(p)$. As $n \to \infty$ and $p \to 0$, with $np = \lambda$, then $\text{Bin}(n, p) \to \text{Poi}(\lambda)$. If $X_1, \ldots, X_n$ are independent Binomial r.v.'s, where $X_i \sim \text{Bin}(N_i, p)$, then $X = X_1 + \ldots + X_n \sim \text{Bin}(N_1 + \ldots + N_n, p)$.

(d) **Geometric:** $X \sim \text{Geometric}(p)$ ($\text{Geo}(p)$ for short) 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 $Var(X) = \frac{1-p}{p^2}$. An example of a Geometric r.v. is the number of independent coin flips up to and including the first head, where $\mathbb{P}(\text{head}) = p$.

(e) **Poisson**: $X \sim$ Poisson(λ) (Poi(λ) for short) 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 $Var(X) = \lambda$. 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(λ_i), then $X = X_1 + \ldots + X_n \sim$ Poi$(\lambda_1 + \ldots + \lambda_n)$.

2. Pond Fishing

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

Solution:

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

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

Solution:

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

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

Solution:

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

(d) whether or not my next fish is blue

Solution:

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

3. Best Coach Ever!!

You are a hardworking boxer. Your coach tells you that the probability of your winning a boxing match is 0.2 independently of every other match.

(a) How many matches do you expect to fight until you win 10 times ?

Solution:

The number of matches you have to fight until you win 10 times can be modeled by $\sum_{i=1}^{10} X_i$ where $X_i \sim$ Geometric(0.2) is the number of matches you have to fight to win the i^{th} time. Recall $\mathbb{E}[X_i] = \frac{1}{0.2} = 5$. $\mathbb{E}\left[\sum_{i=1}^{10} X_i\right] = \sum_{i=1}^{10} \mathbb{E}[X_i] = \sum_{i}^{10} \frac{1}{0.2} = 10 \cdot 5 = 50$.

(b) You only get to play 12 matches every year. To win a spot in the Annual Boxing Championship, a boxer needs to win at least 10 matches in a year. What is the probability that you will go to the Championship this year?

Solution:

You can go to the championship if you win more than or equal to 10 times this year. Let Y be the number of matches you win out of the 12 matches. Note that $Y \sim \text{Binomial}(12, 0.2)$. We are interested in

$$\mathbb{P}(Y = 10) + \mathbb{P}(Y = 11) + \mathbb{P}(Y = 12) = \sum_{i=10}^{12} \binom{12}{i} 0.2^i (1 - 0.2)^{12-i}$$

(c) Let p be your answer to part (b). How many times can you expect to go to the Championship in your 20 year career?

Solution:

The number of times you go to the championship can be modeled by $Y \sim \text{Binomial}(20, p)$. So, $E[Y] = 20 \cdot p$.

4. Variance of a Product

Let X, Y, Z be independent random variables with means μ_X, μ_Y, μ_Z and variances $\sigma_X^2, \sigma_Y^2, \sigma_Z^2$, respectively. Find $Var(XY - Z)$.

Solution:

First notice that $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \implies \mathbb{E}[X^2] = Var(X) + \mathbb{E}[X]^2 = \sigma_X^2 + \mu_X^2$, and same for Y.

$$Var(XY) = \mathbb{E}[X^2Y^2] - \mathbb{E}[XY]^2 \text{ (by theorem in class)}$$

$$= \mathbb{E}[X^2]\,\mathbb{E}[Y^2] - \mathbb{E}[X]^2\,\mathbb{E}[Y]^2 \text{ (by independence)}$$

$$= (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) - \mu_X^2\mu_Y^2$$

By independence,

$$Var(XY - Z) = Var(XY) + Var(Z)$$

$$= (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) - \mu_X^2\mu_Y^2 + \sigma_Z^2$$

5. True or False?

Identify the following statements as true or false (true means always true). Justify your answer.

(a) For any random variable X, we have $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$.

Solution:

True, since $0 \leq Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

(b) Let X, Y be random variables. Then, X and Y are independent if and only if $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

Solution:

False. The forward implication is true, but the reverse is not. For example, if $X \sim \text{Uniform}(-1, 1)$ (equally likely to be in $\{-1, 0, 1\}$), and $Y = X^2$, we have $\mathbb{E}[X] = 0$, so $\mathbb{E}[X]\,\mathbb{E}[Y] = 0$. However, since $X = X^3$ (why?), $\mathbb{E}[XY] = \mathbb{E}[XX^2] = \mathbb{E}[X^3] = \mathbb{E}[X] = 0$, we have that $\mathbb{E}[X]\,\mathbb{E}[Y] = 0 = \mathbb{E}[XY]$. However, X and Y are not independent; indeed, $\mathbb{P}(Y = 0|X = 0) = 1 \neq \frac{1}{3} = \mathbb{P}(Y = 0)$.

(c) Let $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, p)$ be independent. Then, $X + Y \sim \text{Binomial}(n + m, p)$.

Solution:

True. X is the sum of n independent Bernoulli trials, and Y is the sum of m. So $X+Y$ is the sum of $n+m$ independent Bernoulli trials, so $X+Y \sim$ Binomial$(n+m, p)$.

(d) Let $X_1, ..., X_{n+1}$ be independent Bernoulli(p) random variables. Then, $\mathbb{E}[\sum_{i=1}^{n} X_i X_{i+1}] = np^2$.

Solution:

True. Notice that $X_i X_{i+1}$ is also Bernoulli (only takes on 0 and 1), but is 1 iff both are 1, so $X_i X_{i+1} \sim$ Bernoulli(p^2). The statement holds by linearity, since $\mathbb{E}[X_i X_{i+1}] = p^2$.

(e) Let $X_1, ..., X_{n+1}$ be independent Bernoulli(p) random variables. Then, $Y = \sum_{i=1}^{n} X_i X_{i+1} \sim$ Binomial(n, p^2).

Solution:

False. They are all Bernoulli p^2 as determined in the previous part, but they are not independent. Indeed, $\mathbb{P}(X_1 X_2 = 1 | X_2 X_3 = 1) = \mathbb{P}(X_1 = 1) = p \neq p^2 = \mathbb{P}(X_1 X_2 = 1)$.

(f) If $X \sim$ Bernoulli(p), then $nX \sim$ Binomial(n, p).

Solution:

False. The range of X is $\{0, 1\}$, so the range of nX is $\{0, n\}$. nX cannot be Bin(n, p), otherwise its range would be $\{0, 1, ..., n\}$.

(g) If $X \sim$ Binomial(n, p), then $\frac{X}{n} \sim$ Bernoulli(p).

Solution:

False. Again, the range of X is $\{0, 1, ..., n\}$, so the range of $\frac{X}{n}$ is $\{0, \frac{1}{n}, \frac{2}{n}, ..., 1\}$. Hence it cannot be Ber$(p)$, otherwise its range would be $\{0, 1\}$.

(h) For any two independent random variables X, Y, we have $Var(X - Y) = Var(X) - Var(Y)$.

Solution:

False. $Var(X - Y) = Var(X + (-Y)) = Var(X) + (-1)^2 Var(Y) = Var(X) + Var(Y)$.