Expected Value and Indicators

TFs: , Timothy Kang (tkang01@college), Credits to Timothy Kang

Practice Problems

Example 1. Dice collector.
What is the expected number of times a die must be rolled until the numbers 1 through 6 have all shown up at least once?

Solution
Let X be the number of times the die is rolled. Let's express X as the sum of simpler r.v.s:

$$X = T_1 + T_2 + \ldots + T_6$$

where T_1 is the number of rolls until the 1^{st} unique number shows up, T_2 is the number of *additional* rolls until the 2^{nd} unique number, and so on. Note that T_1 always equals 1. Using the story of the Geometric, $T_2 \sim 1 + \text{Geom}\left(\frac{5}{6}\right)$, $T_3 \sim 1 + \text{Geom}\left(\frac{4}{6}\right)$, etc.

By linearity of expectation:

$$\begin{aligned} E(X) &= E(T_1) + E(T_2) + \ldots + E(T_6) \\ &= 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} \\ &= \boxed{14.7} \end{aligned}$$

Example 2. Mutual Friends.
Alice and Bob have just met, and wonder whether they have a mutual friend. Each has 50 friends, out of 1000 other people who live in their town. They think that it's unlikely that they have a friend in common, saying each of us is only friends with 5% of the people here, so it would be very unlikely that our two 5%'s overlap.

Assume that Alice's 50 friends are a random sample of the 1000 people (equally likely to be any 50 of the 1000), and similarly for Bob. Also assume that knowing who Alice's friends are gives no information about who Bob's friends are. Let X be the number of mutual friends they have.

(a) Compute $E(X)$

(b) Find the PMF of X.

(c) Is the distribution of X one of the important distributions we have looked at? If so, which?

Solution

(a) Let I_k be an indicator random variable representing whether the kth person is a mutual friend of Alice and Bob. Then, we can write

$$\begin{aligned} E(X) &= E\left(\sum_{k=1}^{1000} I_k\right) \\ &= \sum_{k=1}^{1000} E(I_k) \\ &= 1000 \cdot P(I_k = 1) \\ &= 1000 \cdot \left(\frac{50}{1000}\right)^2 = \boxed{2.5} \end{aligned}$$

Here, we used linearity of expectation and then the fundamental bridge. The probability that any particular person is a mutual friend of Alice and Bob is $(50/1000)^2$ because being a friend of Alice and a friend of Bob are independent events.

(b) Suppose that Alice is friends with a fixed group of 50 people. We need to calculate the probability that out of those 50 people, exactly k of them are also Bob's friends. This would be the probability $P(X = k)$.

First, we need to pick k of these 50 of Alice's friends to be Bob's friends as well. Then, we need to choose $50 - k$ other friends for Bob from the 950 remaining non-Alice friends. We multiply those two together to get the total number of ways Bob can have 50 friends. Finally, we divide by the number of ways Bob could have any 50 friends out of 1000. We get:

$$P(X = k) = \frac{\binom{50}{k}\binom{950}{50-k}}{\binom{1000}{50}}$$

(c) Yes, this is the PMF of a hypergeometric distribution. We can think of the problem as follows: Bob is choosing 50 balls out of 50 white balls (Alice's friends) and 950 black balls (not Alice's friends). The number of white balls he ends up choosing is distributed $\text{HGeom}(50, 950, 50)$.

Example 3. Min and Max.
Let $X \sim \text{Bin}\left(n, \frac{1}{2}\right)$ and $Y \sim \text{Bin}\left(n+1, \frac{1}{2}\right)$ independently.

(a) Let V = $\min(X, Y)$ be the smaller of X and Y, and let $W = \max(X, Y)$. So if X crystalizes to x and Y crystalizes to y, then V crystalizes to $\min(x, y)$ and W crystalizes to $\max(x, y)$. Find $E(V) + E(W)$.

(b) Show that $E|X - Y| = E(W) - E(V)$, with notation as in (a).

(c) Compute $\text{Var}(n - X)$.

Solution

(a) Note that $V + W = X + Y$ because adding the larger and smaller of two numbers is the same as adding both numbers). Therefore, by linearity of expectation, we get:

$$E(V) + E(W) = E(V + W) = E(X + Y) = E(X) + E(Y) = \boxed{\frac{2n+1}{2}}$$

(b) Note that $|X - Y| = W - V$ because the absolute difference between two numbers is equal to the larger minus the smaller. Therefore,

$$E|X - Y| = E(W - V) = E(W) - E(V)$$

(c) We can use the property of variances, which says that $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. We get

$$\mathrm{Var}(n - X) = \mathrm{Var}(n) + \mathrm{Var}(-X) = 0 + \mathrm{Var}(X) = \boxed{n/4}$$

.

Example 4. Expected Number and Variance of Matches.
Suppose 100 people, each with a hat. We mix the hats and hand them out randomly to each person.

(a) What is the expected number of people who get their own hat?

(b) What is the variance of the number of people who get their own hat?

Solution
Let I_k be an indicator random variable representing whether the kth person gets his own hat.

(a) We have $E(I_k) = \frac{1}{100}$ since the it is equally probable for the kth person to get any of the 100 hats. Hence,

$$\begin{aligned} E(X) &= E\left(\sum_{k=1}^{100} I_k\right) \\ &= \sum_{k=1}^{100} E(I_k) \\ &= 100 \cdot \frac{1}{100} \\ &= \boxed{1} \end{aligned}$$

(b) By the definition of variance,

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = E(X^2) - 1$$

using the result in (a). To find $E(X^2)$, we can expand the indicator r.v. representation and use properties of indicators:

$$\begin{aligned}
E(X^2) &= E\left[(I_1 + I_2 + \ldots + I_{100})^2\right] \\
&= E(I_1^2 + I_2^2 + \ldots + I_{100}^2 + 2\sum_{i<j} I_i I_j) \\
&= E(I_1 + I_2 + \ldots + I_{100} + 2\sum_{i<j} I_i I_j) \\
&= 100E(I_1) + 2\binom{100}{2} E(I_1 I_2)
\end{aligned}$$

Since $I_1 I_2$ is the indicator that the 1st and 2nd people both get their own hats, $E(I_1 I_2) = \frac{1}{100}\frac{1}{99}$. Thus,

$$E(X^2) = 100 \cdot \frac{1}{100} + 2\binom{100}{2} \cdot \frac{1}{100} \cdot \frac{1}{99} = 2$$

and

$$\text{Var}(X) = E(X^2) - 1 = 2 - 1 = \boxed{1}$$

Example 5. Prizes.
There are n prizes, with values \$1, \$2 … \$$n$. You get to choose k random prizes, without replacement. What is the expected total value of the prizes you get?

Solution
Let I_j be indicator random variables representing whether the prize of \$$j$ is received. We know that $E(I_j) = P(\text{Gift of \$}j \text{ is selected}) = \frac{k}{n}$ (because there are $\binom{n-1}{k-1}$ for the item to be selected, and $\binom{n}{k}$ total ways to select k items, and their ratio is $\frac{k}{n}$).

Let V be the total value of the k prizes you receive. Then, we can write:

$$V = I_1 + 2I_2 + 3I_3 + \ldots + n \cdot I_n$$

Finding $E(V)$ and using the fundamental bridge, we get:

$$\begin{aligned}
E(V) &= E(I_1) + E(2I_2) + E(3I_3) + \ldots E(nI_n) \\
&= E(I_1) + 2E(I_2) + 3E(I_3) + \ldots nE(I_n) \\
&= \frac{k}{n}(1 + 2 + \ldots + n) \\
&= \frac{k}{n} \cdot \frac{n(n+1)}{2} \\
&= \boxed{\frac{k(n+1)}{2}}
\end{aligned}$$

Example 6. Coin Runs.
A coin with probability p of Heads is flipped n times. The sequence of outcomes canbe divided

into runs (blocks of H's or blocks of T's), e.g., $HHHTTHTTTH$ becomes $\boxed{HHH}\boxed{TT}\boxed{H}\boxed{TTT}\boxed{H}$, which has 5 runs. Find the expected number of runs.

Hint: Start by finding the expected number of tosses (other than the first) where the outcome is different from the previous one.

Solution
Let I_j be an indicator random variable for whether or not the jth flip is different from the $j-1$st, for $j = 2, 3, \ldots n$. The probability that the jth flip is different from the $j-1$st is equal to $p(1-p) + (1-p)p = 2p(1-p)$, meaning $E(I_j) = 2p(1-p)$. Let X be the number of flips such that that flip is different from the previous one. Then $X = I_2 + \ldots + I_n$ and

$$\begin{aligned} E(X) &= (n-1) \cdot E(I_j) \\ &= 2(n-1)(p)(1-p) \end{aligned}$$

The number of flips for which this is true is one less than the number of runs, so the expected number of runs is: $\boxed{1 + 2(n-1)(p)(1-p)}$.

Example 7. True/False.

(a) If X and Y have the same CDF, they have the same expectation.

(b) If X and Y have the same expectation, they have the same CDF.

(c) If X and Y have the same CDF, they must be dependent.

(d) If X and Y have the same CDF, they must be independent.

(e) If X and Y have the same distribution, $P(X < Y)$ is at most $1/2$.

(f) If X and Y are independent and have the same distribution, $P(X < Y)$ is at most $1/2$.

Solution

(a) **True**. If X and Y have the same CDF, then they would have the same distribution, and hence the same expectation.

(b) **False**. Consider the expected values and CDFs of $\text{Bern}(1)$ and $\text{Bin}(2, 0.5)$.

(c) **False**. Having the same CDF (and equivalently distribution) does not indicate anything about the dependency of the two random variables.

(d) **False**. Same as above.

(e) **False**. Think of a clock with two hands, whereby one of the hands is always one unit faster than the other hand.

(f) **True**. This means that X and Y are i.i.d., and hence by symmetry, $P(X < Y) = 1/2 \cdot (1 - P(X = Y))$ which is at most $\frac{1}{2}$.