10 Variance, (Joint) Distributions, Independence, Expectation

Computer Science Mentors 70

Nov 6 - Nov 10
Fall 2023

Amogh Gupta, Sylvia Jin, Aekus Bhathal, Abinav Routhu, Debayan Bandyopadhyay, Thomas Chen, Dylan Xu

1 Expectation and Variance

1. **Introduction to Expectation**

Imagine that Sylvia has two, 3-sided loaded (non-uniform probability) dice, which are represented by the random variables X and Y, respectively. X and Y are distributed as follows:

$$P(X=1)=\frac{1}{2} \qquad P(Y=1)=\frac{1}{6}$$
$$P(X=2)=\frac{1}{4} \qquad P(Y=2)=\frac{1}{6}$$
$$P(X=3)=\frac{1}{4} \qquad P(Y=3)=\frac{2}{3}$$

(a) Sylvia rolls the first die, represented by random variable X.What is the expected value of the roll of the first die. What is the probability it will roll the expected value?

Solution: We first find the expected value of the first die, random variable X

$$E[X]=\frac{1}{2}(1)+\frac{1}{4}(2)+\frac{1}{4}(3)=\frac{7}{4}$$

As the expected value is not a number on the die, the probability that the die will role it is zero

(b) What is expected value of a roll of the second die, represented by random variable Y?

Solution:

$$E[Y]=\frac{1}{6}(1)+\frac{1}{6}(2)+\frac{2}{3}(3)=\frac{5}{2}$$

(c) What is the expected value of the product of the two dice?

Solution: We take the expectation of each possible product:

$$E[XY]=\frac{1}{2}\cdot\frac{1}{6}(1)(1)+\frac{1}{2}\cdot\frac{1}{6}(1)(2)+\frac{1}{2}\cdot\frac{2}{3}(1)(3)+\frac{1}{4}\cdot\frac{1}{6}(2)(1)+\frac{1}{4}\cdot\frac{1}{6}(2)(2)$$
$$+\frac{1}{4}\cdot\frac{2}{3}(2)(3)+\frac{1}{4}\cdot\frac{1}{6}(3)(1)+\frac{1}{4}\cdot\frac{1}{6}(3)(2)+\frac{1}{4}\cdot\frac{2}{3}(3)(3)=4.375$$

2. **Introduction to Variance**

Let us say that we are dealing with a biased die and we want to know how often my roll varies from turn to turn. Consider the following 6-length tuple assigning probabilities to each of the 6 rolls on the die, whose distribution we name X:

$$(p_1,p_2,p_3,p_4,p_5,p_6)$$

(a) Find the mean and variance for the die with distribution $(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6})$

Solution: The expected value is:

$$E[X] = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} = 3.5$$

The variance is:

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{1}{6}(1+4+9+16+25+36) - 3.5^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$$

(b) Find the mean and variance for the die with distribution $(\frac{1}{3}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{3})$

Solution: The expected value is:

$$E[X] = \frac{1}{12}(2+3+4+5) + \frac{1}{3}(1+6) = \frac{7}{2} = 3.5$$

The variance is:

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{1}{12}(4+9+16+25) + \frac{1}{3}(1+36) - 3.5^2 = \frac{101}{6} - \frac{49}{4} = \frac{55}{12}$$

(c) Give an intuitive explanation for the difference between the results in part(a) and part (b)

Solution: Although the means of the two die are the same, we expect there to be a higher variance because the probabilities of a 1 and 6 are higher in part (b) than part (a).

3. **Money Balls**

We are drawing dollar bills out of a bag consisting of R red balls and B blue balls. We get 1 dollar for every blue ball, but whenever we draw a red ball, the game stops and we retrieve the amount of money we have made so far.

(a) Find the expected value of the amount of money made from this game.

Solution: We can create an indicator, X_i, which represents the indicator that the i^{th} blue ball appears before the first red ball. The total number of balls that we draw before we draw this blue ball represents the total amount of money made in the game, which can be represented as:

$$X = \sum_{i=1}^{B} X_i$$

The probability that the event represented by the indicator occurs is $\frac{1}{R+1}$. Consider drawing all balls out of the bag. Out of the $R+1$ choices (in between all red balls) that the i^{th} blue ball can be placed, it must be placed before all of them. Using the linearity of expectation:

$$X = \sum_{i=1}^{B} E[X_i] = B * E[X_1] = \frac{B}{R+1}$$

(b) Find the variance of the amount of money made in this game.

Solution: To compute the variance, we have: $\text{Var}(X) = E[X^2] - E[X]^2$. We first find:

$$\begin{aligned} E[X^2] &= E[(X_1 + X_2 + \cdots + X_B)^2] \\ &= \sum_{i=1}^{B} E[X_i^2] + \sum_{i \neq j} E[X_i X_j] \\ &= BE[X_i^2] + B(B-1)E[X_1 X_2] \\ &= B(\frac{1}{R+1}) + B(B-1)\frac{2}{(R+1)(R+2)} \\ &= \frac{B(R+2) + 2B(B-1)}{(R+1)(R+2)} \end{aligned}$$

Keep in mind that the square of an indicator is the same as the indicator itself, which is the simplification in line 3 of the above simplification. Furthermore, $E[X_i X_j] = \frac{2}{(R+1)(R+2)}$ because after the i^{th} blue ball, we have a total of $(R+1)+1$ places to place the jth blue ball, and it can either go at the beginning or after the i^{th} blue ball, leading to 2 spots. We simplify the summation using the linearity of expectation. Combining everything, we have:

$$\text{Var}(X) = \frac{B(R+2) + 2B(B-1)}{(R+1)(R+2)} - \left(\frac{B}{R+1}\right)^2$$

4. **Jensen's for Special Polynomials**

Jensen's inequality says that for any convex function f, $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$ (You don't need to know this for CS 70). In this problem, we will prove that Jensen's inequality holds for a subclass of convex functions called "special polynomials" (this is a made up name). We define a special polynomial as any function that can be written as

$$f(x) = a_n x^{(2^{n-1})} + a_{n-1} x^{(2^{n-2})} + \ldots + a_1 x$$

for some $n \in \mathbb{N}$ and $\forall 1 \leq i \leq n$, $a_i \geq 0$

(a) Prove that $\mathbb{E}[X^2] \geq E[X]^2$ (*Hint: use the definition of variance*)

Solution: Recall from the definition of Variance that $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, implying that

$$\mathbb{E}[X^2] = Var(X) + \mathbb{E}[X]^2 \tag{1}$$

Now let $Y = (X - \mu)^2$. Then, $Var(X) = E(Y)$. Since Y is a non-negative random variable (a square of a real-valued function can never be negative), we must have that $E(Y) = Var(X) \geq 0$, since the weighted average of non-negative values must be non-negative. Thus,

$$\mathbb{E}[X^2] \geq \mathbb{E}[X]^2 \tag{2}$$

(b) Use part (a) to prove that $\mathbb{E}[X^{(2^k)}] \geq E[X]^{(2^k)}$ for some $k \in \mathbb{N}$

Solution: From part (a), we have proven that the statements holds when $k = 1$. We use induction to prove it holds for arbitrary k.

Suppose the statement holds for $k-1$. Then,

$$\begin{aligned}
\mathbb{E}[X^{(2^k)}] &= \mathbb{E}[(X^{(2^{k-1})})^2] && (3)\\
&\geq \mathbb{E}[X^{(2^{k-1})}]^2 && (4)\\
&\geq (\mathbb{E}[X]^{2^{k-1}})^2 && (5)\\
&= \mathbb{E}[X]^{(2^k)} && (6)
\end{aligned}$$

A lot happened here. Let's break it down line by line.

On line 3, we rearranged exponents to isolate $k-1$.

Then, from line 3 to line 4, we applied the inequality in part (a).

From line 4 to line 5, we applied the inductive hypothesis, and finally we rearranged exponents from line 5 to line 6. This completes the proof!

(c) Use part (b) and properties of expectation to prove that $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$.

Solution: Most of the heavy lifting is done. We now apply the function definition from the problem statement and use linearity of expectation to isolate each term. Then, we use the inequality from part (b) on each term

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[a_n X^{(2^{n-1})} + a_{n-1} x^{(2^{n-2})} + \ldots + a_1 x] && (7)\\
&= \mathbb{E}[a_n X^{(2^{n-1})}] + \mathbb{E}[a_{n-1} X^{(2^{n-2})}] + \ldots + \mathbb{E}[a_1 X] && (8)\\
&= a_n \mathbb{E}[X^{(2^{n-1})}] + a_{n-1}\mathbb{E}[X^{(2^{n-2})}] + \ldots + a_1 \mathbb{E}[X] && (9)\\
&\geq a_n \mathbb{E}[X]^{(2^{n-1})} + a_{n-1}\mathbb{E}[X]^{(2^{n-2})} + \ldots + a_1 \mathbb{E}[X] && (10)\\
&= f(\mathbb{E}[X]) && (11)
\end{aligned}$$

Boom!

2 Independence

1. Assume we have 10 coins, each with a different bias towards heads. The first coin has $p = 0.1$ of flipping heads, the second has $p = 0.2$, etc up to the 10th coin which has $p = 1$. What is the expected number of heads from flipping all 10 coins at once?

Solution: Since we know that the expectation of a single Bernoulli distribution is $E(X) = p$, and we have 10 coins, we can write the following (by Linearity of Expectation):

$$E(X) = E(X_1 + X_2 + X_3 + \ldots + X_{10}) = E(X_1) + E(X_2) + E(X_3) + \ldots + E(X_{10})$$

$$E(X) = \sum_{i=1}^{10} p = \sum_{i=1}^{10} \frac{i}{10} = 5.5$$

2. **The Coupon Collector Returns**
The staff of CS 70 are selling CS Seventy Cereal boxes, with each box containing one card of a notable Berkeley CS alum. There exist n distinct cards in total, each of which has $1/n$ chance of being in any given box. Our card collector Catherine wants (any) k of those distinct cards, where $k \leq n$.

(a) How many boxes should she expect to need to buy before she gets k distinct cards?

Solution: Like in the original coupon collector problem, we denote X_i as the number of boxes Catherine buys when trying to get the i-th new card. Let S_k be the total number of boxes needed to collect k boxes; then:

$$S_k = X_1 + X_2 + \cdots + X_k$$

We can model X_i as a geometric distribution with $p = \frac{n-i+1}{n}$. Thus, $\mathbb{E}[X_i] = 1/p = \frac{n}{n-i+1}$. Then, by linearity of expectation,

$$\mathbb{E}[S_k] = \sum_{i=1}^{k} \mathbb{E}[X_i] = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{n-k+1} = n \sum_{i=n-k+1}^{n} \frac{1}{i}$$

(b) Now suppose that Catherine wants a *specific* subset of k distinct cards from the n total. How many boxes should she expect to need to buy?

Solution: Let X_i be the same as in part a). Every X_i is still a geometric distribution, but X_1 has probability $p = \frac{k}{n}$, since we have to draw one of the k cards that Catherine wants. Similarly, X_2 has probability $p = \frac{k-1}{n}$ and X_i has probability $p = \frac{k-i+1}{n}$. Thus, $\mathbb{E}[x_i] = \frac{n}{k-i+1}$, meaning that

$$\mathbb{E}[S_k] = \sum_{i=1}^{k} \mathbb{E}[X_i] = \frac{n}{k} + \frac{n}{k-1} + \cdots + \frac{n}{1} = n \sum_{i=1}^{k} \frac{1}{i}$$

(As a sanity check, since $k \leq n$, the answer for part b) is $\geq$ the answer for part a), which makes sense as adding a condition should increase the average boxes Catherine needs.)

3 Distributions

1. **Rolling Chopsticks**

   The content mentors were trying to eat noodles in a new way. Rather than eating noodles by chopsticks directly, they tried eating noodles by rolling one noodle on the chopstick and eat it. This is seemingly a hard way to eat noodles so the probability they successfully eat a noodle on each attempt is p.

   (a) Suppose they attempt to eat a noodle, and eat the noodle on the attempt X. What is the distribution of X? What is the distribution of unsuccessful attempts to eat that noodle, X', in terms of X?

   Solution: We notice that X is modeling how long it takes until the first successful attempt hence $X \sim \text{Geometric}(p)$. Now if $X = k$ we know the last one is the successful attempt so there are $k-1$ unsuccessful attempts. Then $X' = X - 1$. So the distribution of unsuccessful attempts to eat that noodle $X' \sim X - 1$. That is $\mathbb{P}[X' = k] = \mathbb{P}[X = k+1]$

   (b) Let Y be the number of unsuccessful attempts in trying to eat 2 noodles. What is the distribution of Y?

   Solution: We can model Y as $Y = X_1' + X_2'$, where X_1' and X_2' are identical and independent random variables over the same distribution as X'.

   For the distribution of Y, if $Y = k$ then there were k unsuccessful attempts between trying to eat both noodles. We observe that if we made w unsuccessful attempts before 1^{st} successful attempt, then we made $k - w$ unsuccessful attempts before the 2^{nd} successful event. There are $k + 1$ different possibilities from $w = 0$ to $w = k$.
   Also we know the distribution of X' only in terms of X, so we'll need to substitute $X' = X - 1 \implies X = X' + 1$.

$$
\begin{aligned}
\mathbb{P}[Y = k] &= \mathbb{P}[X_1' + X_2' = k] \\
&= \sum_{w=0}^{k} \mathbb{P}[X_1' = w \cap X_2' = k - w] \\
&= \sum_{w=0}^{k} \mathbb{P}[X_1 = w + 1 \cap X_2 = k - w + 1] \\
&= \sum_{w=0}^{k} (1-p)^w p \cdot (1-p)^{k-w} p \\
&= (1-p)^k p^2 \cdot \sum_{w=0}^{k} 1 \\
&= (k+1)(1-p)^k p^2
\end{aligned}
$$

   (c) Not content with their distribution Y and eating 2 noodles, the content mentors want to find the distribution Z for the total unsuccessful attempts of eating the whole bowl of R noodles. They were planning to proceed as part b) but then Aekus, a random variable distribution enthusiast, suggested to use $P(Z = k) = \binom{r+k-1}{k}(1-p)^k p^r$ where $r = R$.

   The distribution Z is defined by 2 parameters: 1) r - the number of successful attempts and 2) p - the probability of a successful attempt so we will write Z as $Z(r, p)$.

   Show by induction on r with base case $r = 1$ that Aekus's suggestion is correction; Z is the sum of independent random variables drawing from the distribution of X'. (*Hint*: Remember the "Hockey stick" identity $\sum_{i=0}^{k-1} \binom{n+i}{i} = \binom{n+k}{k-1}$)

Solution: For intuition let's try to put small values for r.

$$\begin{aligned}\mathbb{P}[Z(1,p)=k] &= \binom{1+k-1}{k}(1-p)^k p^1 \\ &= (1-p)^k p^1 \\ &= \mathbb{P}(X=k+1) = \mathbb{P}(X'=k)\end{aligned}$$

When $r=1$ we get that $Z(1,p) \sim X'$ and when $r=2$ we get $Z(2,p) \sim Y$. Now this helps us to see that we could use our old friend induction to prove this.

Base Case: The base case is when $r=1$ and is true as we saw above.

Hypothesis: Assume that $Z(r,p)$ is the sum of r X_i', where $X_i' = X_i - 1$ for IID $X_i \sim$ Geometric(p).

Step : Now let $W = Z + X'$. We show that W is the Z distribution with $r_W = r+1$.

$$\begin{aligned}\mathbb{P}[W=w] &= \mathbb{P}[X'+Z=w] \\ &= \sum_{k=0}^{w} \mathbb{P}[(Z=k)\cap(X'=w-k)] \\ &= \sum_{k=0}^{w} \mathbb{P}[(Z=k)\cap(X=w-k+1)] && \text{Using the substitution } X'=X-1 \\ &= \sum_{k=0}^{w} \binom{r+k-1}{k} p^r(1-p)^k \cdot p(1-p)^{w-k} \\ &= \sum_{k=0}^{w} \binom{r+k-1}{k} p^{r+1}(1-p)^w \\ &= p^{r+1}(1-p)^w \sum_{k=0}^{w} \binom{r+k-1}{k} \\ &= \binom{r+w}{w} p^{r+1}(1-p)^w && \text{Using "Hockey Stick" with } n=r-1, k=w+1.\end{aligned}$$

This shows that W is the Z distribution with $r_W = r+1$ and probability of success as p, $Z(r+1,p)$ and it is now evident that if $X_i' = X_i - 1$ where $X \sim$ Geometric(p) for $i = 1, 2, \ldots, n$ are IID, then $\sum_i^r X_i' = Z(r,p)$

The distribution Z has a name, it's called the Negative Binomial Distribution, just a fun fact.

(d) What is the expected value of total unsuccessful attempts of eating the whole bowl of R noodles, the random variable Z?.

Solution: From part c) we proved that Z is the sum of r independent X' random variables. Then

$$\begin{aligned}Z &= rX' \\ E[Z] &= E[rX'] \\ &= rE[X'] && \text{Using Linearity of expectation} \\ &= rE[X-1] && \text{Using the substitution } X'=X-1 \\ &= r(E[X]-1) \\ &= \frac{r}{p} - r && \text{As } X \sim \text{Geometric}(p)\text{, then } E[X]=1/p \\ &= \frac{r(1-p)}{p}\end{aligned}$$

2. Mr. and Mrs. Brown decide to continue having children until they either have their first girl or until they have three children. Assume that each child is equally likely to be a boy or a girl, independent of all other children, and that there are no multiple births. Let G denote the numbers of girls that the Browns have. Let C be the total number of children they have.

(a) Determine the sample space, along with the probability of each sample point.

Solution: The sample space is the set of all possible sequences of children that the Browns can have: $\omega = \{g, bg, bbg, bbbg\}$. The probabilities of these sample points are:

$$\P(g) = \frac{1}{2}$$
$$\P(bg) = \frac{1}{2} + \frac{1}{2} = \frac{1}{4}$$
$$\P(bbg) = \frac{1}{2}^3 = \frac{1}{8}$$
$$\P(bbb) = \frac{1}{2}^3 = \frac{1}{8}$$

(b) Compute the joint distribution of G and C. Fill in the table below.

	C = 1	C = 2	C = 3
G = 0			
G = 1			

Solution:

	C = 1	C = 2	C = 3
G = 0	0	0	¶(bbb) = 1/8
G = 1	¶(g) = 1/2	¶(bg) = 1/4	¶(g) = 1/8

(c) Use the joint distribution to compute the marginal distributions of G and C and confirm that the values are as you'd expect. Fill in the tables below.

P(G = 0)	
P(G = 1)	

P(C = 1)	P(C = 2)	P(C = 3)

Solution: Marginal distribution for G:

$$\P(G = 0) = 0 + 0 + \frac{1}{8} = \frac{1}{8}$$
$$\P(G = 1) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$$

Marginal distribution for C:

$$\P(C = 1) = 0 + \frac{1}{2} = \frac{1}{2}$$
$$\P(C = 2) = 0 + \frac{1}{4} = \frac{1}{4}$$
$$\P(C = 3) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

(d) Are G and C independent?

Solution: No, G and C are not independent. If two random variables are independent, then

$$\P(X = x, Y = y) = \P(X = x)\P(Y = y)$$

To show this dependence, consider an entry in the joint distribution table, such as $\P(G = 0, C = 3) = \frac{1}{8}$. This is not equal to $\P(G = 0)\P(C = 3) = \frac{1}{8} * \frac{1}{4} = \frac{1}{32}$, so the random variables are not independent.

(e) What is the expected number of girls the Browns will have? What is the expected number of children that the Browns will have?

Solution: We can apply the definition of expectation directly for this problem, since we?ve computed the marginal distribution for both random variables.

$$\mathsf{E}(G) = 0 * \P(G = 0) + 1 * \P(G = 1) = \frac{7}{8}$$

$$\mathsf{E}(C) = 1 * \P(C = 1) + 2 * \P(C = 2) + 3 * \P(C = 3) = 1 * \frac{1}{2} + 2 * \frac{1}{4} + 3 * \frac{1}{4} = \frac{7}{4}$$