Variance, Covariance, Concentration Inequalities, and WLLN

Computer Science Mentors 70

Amogh Gupta, Sylvia Jin, Aekus Bhathal, Abinav Routhu, Debayan Bandyopadhyay

Roast us here: `https://tinyurl.com/csm70-feedback20`

1 Variance

1. **Introduction to Variance**

We are making paper boats out of square sheets of paper. Suppose the size of the sail of the boat is same as the length of the side of the paper. We plan to go to "The Random Sheet Shop" to buy the sheet. This shop have 6 sheets of varying side lengths.

(a) The shop has sheets of side length 1 unit to 6 units. Find the mean and variance for the size of the sail of the boat we could make out of these sheets if the shop gives us a sheet uniformly at random.

Solution: The probability of getting any sheet is

$$\frac{1}{6}$$

Then 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) Now suppose we go to a shop which lists sizes a little differently and gives a sheets with a different probability distribution. The shop lists the sheet as difference from mean but never tells you the mean of side but rather tells the mean of the area of the sheet is $\frac{101}{6}$. Find the mean and variance for the size of the sail of the boat we could make out of these sheets if the shop gives us a sheet with distribution $(-\frac{5}{2} : \frac{1}{3}, -\frac{3}{2} : \frac{1}{12}, -\frac{1}{2} : \frac{1}{12}, +\frac{1}{2} : \frac{1}{12}, -\frac{3}{2} : \frac{1}{12}, +\frac{5}{2} : \frac{1}{3})$

Solution: The variance value is:

$$\text{Var}(X) = E[(X - E[X])^2] = \frac{2}{12}(\frac{1}{4} + \frac{9}{4}) + \frac{2}{3}(\frac{25}{4}) = \frac{5}{12} + \frac{50}{12} = \frac{55}{12}$$

The the expected value is:

$$E[X] = \sqrt{E[X^2] - \text{Var}(X)} = \sqrt{\frac{101}{6} - \frac{55}{12}} = \sqrt{\frac{49}{4}} = \frac{7}{2}$$

(c) Give an intuitive explanation for the difference between the variance and mean 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 side length 1 and 6 are higher in part (b) than part (a).

2. **Chaotic Santa**

(Fall '17 Disc) A delivery guy is out delivering n packages to n customers, where $n \in \mathbb{N}$, $n > 1$. Not only does he hand a random package to each customer, he opens the package before delivering it with probability $\frac{1}{2}$.

(a) What is the expected number of customers that get their own package unopened?

Solution: Define an indicator variable C_i to be 1 if customer i gets their own package unopened. Since it is an indicator and follows a Bernoulli distribution,

$$E[C_i] = P(C_i) = \frac{1}{2} \cdot \frac{1}{n}.$$

Let C = number of customers that get their own package unopened;

$$C = C_1 + C_2 + \ldots + C_n = n \cdot \frac{1}{2} \cdot \frac{1}{n} = \frac{1}{2}.$$

(b) What is the variance for the random variable above?

Solution:
Remember that $\text{Var}(C) = E[C^2] - E[C]^2$. We can get the second term by squaring part (a), so we must still find $E[C^2]$.

$$E[C^2] = E[(C_1 + C_2 + \ldots + C_n)^2] = E[\sum_{i,j} C_i C_j] = \sum_{i,j} E[C_i C_j].$$

For this kind of setup, we split up the summation into cases where $i = j$ and cases where $i \neq j$, so our expression is

$$E[C^2] = \sum_i E[C_i^2] + \sum_{i \neq j} E[C_i C_j]$$
$$E[C_i^2] = E[C_i] = \frac{1}{2n}\,;\, E[C_i C_j] = P(C_i C_j = 1)$$

We must now find the probability that the two selected people both get their own package unopened. Note that we cannot use linearity of expectation here, so we need to count this as one event.

$$P(C_i C_j = 1) =$$
$$\frac{1}{2n} \text{ (first person gets their package unopened)}$$
$$\cdot$$
$$\frac{1}{2(n-1)} \text{ (second person gets their package unopened after the first person's stuff arrives)}$$

So, we add up n instances of the first part of the summation, and $n(n-1)$ instances of the second part for each combination where $i \neq j$...

$$E[C^2] = n\left(\frac{1}{2n}\right) + n(n-1)\frac{1}{2n \cdot 2(n-1)} = \frac{3}{4}$$
$$\text{Var}(C) = E[C^2] - E[C]^2 = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}.$$

3. **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 Covariance and Correlation

1. **One D -> Two D**

Recall from long, long ago, the binary truth operators OR, AND, XOR, and so forth. In this problem we analyze these operators as probability distributions.

(a) Suppose X, Y are discrete random variables taking on values in $\{0, 1, 2, 3\}$ and let $\mathbb{P}(X = x, Y = y) = c(x \oplus y)$. $\oplus$ is the XOR operator, and can be expressed in terms of elementary logic operators as $P \oplus Q = (P \vee Q) \wedge \neg(P \wedge Q)$.

Since X and Y are not constrained to 0 or 1, in this case $\oplus$ is applied bitwise (e.g. $3 \oplus 2 = 11_2 \oplus 10_2 = 01_2 = 1$).

Find c, and express the joint distribution of X and Y with a probability table, and find the marginal distributions for X and Y.

Solution: First, we write the table in terms of c: (Notice it doesn't matter what axis is X or Y since the distribution is symmetric).

	0	1	2	3
0	0	c	2c	3c
1	c	0	3c	2c
2	2c	3c	0	c
3	3c	2c	c	0

The sum over all values for x and y must be 1, so for this to be a valid probability distribution,

$$\sum_{x=0}^{3} \sum_{y=0}^{3} \mathbb{P}(X = x, Y = y) = 24c = 1 \tag{12}$$

so $c = \frac{1}{24}$. Now, notice that each row and each column sum to $\frac{1}{4}$, hence the marginal distribution is

$$P(X = x) = P(Y = y) = \frac{1}{4} \tag{13}$$

(b) Using the same definitions from part (a), find the covariance and the correlation of X, Y [5cm]

Solution: Using the definition of covariance,

$$Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{14}$$

X and Y are uniform and identically distributed so $\mathbb{E}[X] = \mathbb{E}[Y] = \frac{3}{2}$

To find $\mathbb{E}[XY]$, we sum over all values for X, Y and find the weighted average (if this formula doesn't make sense to you, you can also define a new variable $Z = XY$, and take the expectation of that):

$$\mathbb{E}[XY] = \sum_{x=0}^{3} \sum_{y=0}^{3} xy\mathbb{P}(X = x, Y = y) \tag{15}$$

$$= 2 \cdot \frac{3}{24} + 3 \cdot \frac{2}{24} + 6 \cdot \frac{1}{24} + 2 \cdot \frac{3}{24} + 3 \cdot \frac{2}{24} + 6 \cdot \frac{1}{24} \tag{16}$$

$$= \frac{3}{2} \tag{17}$$

Hence, we have

$$Cov(X,Y) = \frac{3}{2} - (\frac{3}{2})^2 \tag{18}$$
$$= -\frac{3}{4} \tag{19}$$
$$\tag{20}$$

$\sigma_X = \sigma_Y = \sqrt{Var(X)} = \sqrt{\frac{4^2-1}{12}} = \sqrt{\frac{5}{4}}$ using the formula for variance of a uniform distribution. Hence, using the formula for correlation,

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = -\frac{3}{5} \tag{21}$$

This makes sense since if you plot the graph, you will notice a negative correlation.

(c) Find $\mathbb{E}[X+Y]$ and $Var[X+Y]$

Solution: By linearity of expectation, $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 3$.

By bilinearity of covariance,

$$Var(X+Y) = Cov(X+Y, X+Y) \tag{22}$$
$$= Cov(X,X) + 2Cov(X,Y) + Cov(Y,Y) \tag{23}$$
$$= Var(X) + 2Cov(X,Y) + Var(Y) \tag{24}$$
$$= \frac{5}{4} - \frac{6}{4} + \frac{5}{4} \tag{25}$$
$$= 1 \tag{26}$$

3 Concentration Inequalities

1. Let X be the sum of 20 i.i.d. Poisson random variables $X_1, \ldots, X_{20}$ with $E(X_i) = 1$. Find an upper bound of $\Pr[X \geq 26]$ using,

(a) Markov's inequality:

Solution:

$$\begin{aligned} \Pr[X \geq a] &\leq \frac{E(X)}{a} \text{ for all } a > 0 \\ \Pr[X \geq 26] &\leq \frac{20}{26} \\ &\approx 0.769 \end{aligned}$$

(b) Chebyshev's inequality:

Solution:

$$\begin{aligned} \Pr[|X - E(X)| \geq c] &\leq \frac{\sigma_X^2}{c^2} \\ \Pr[|X - 20| \geq 6] &\leq \frac{20}{36} \\ &\approx 0.5556 \end{aligned}$$

2. Suppose we have a sequence of iid random variables $X_1, X_2, ..., X_n$
Let $A_n = \frac{X_1+X_2+...+X_n}{n}$ be the sample mean.
Show that the true mean of $X_i = \mu$ is within the interval $[\mu - 4.5\frac{\sigma}{\sqrt{n}}, \mu + 4.5\frac{\sigma}{\sqrt{n}}]$ with 95% probability.

Solution: Using Chebyshev's Inequality, we have:
$P(|A_n - \mu| > 4.5\frac{\sigma}{\sqrt{n}}) \leq \frac{var(A_n)}{(4.5\frac{\sigma}{\sqrt{n}})^2}$
$var(A_n) = var(\frac{X_1+X_2+...X_n}{n}) = \frac{1}{n^2}var(X_1 + X_2 + ... + X_n) = \frac{1}{n^2} * n * var(X_1) = \frac{\sigma^2}{n}$
$P(|A_n - \mu| > 4.5\frac{\sigma}{\sqrt{n}}) \leq \frac{\frac{\sigma^2}{n}}{20\frac{\sigma^2}{n}} = \frac{1}{20}$.
The probability that you are outside of the confidence interval is 0.05, so therefore, the probability that the true mean lies within the interval is $1 - 0.05 = 0.95$.

4 Weak Law of Large Numbers

Introduction to LLN

Leanne has a weighted coin that shows up heads with probability $\frac{4}{5}$ and tails with probability $\frac{1}{5}$. Leanne flips the coin 100 times, and computes X, the average number of coins that show up heads.

(a) What is $E[X]$?

Solution: Let I_i be the indicator variable for the event that the ith coin flip is heads. We can express $X = \frac{I_1+\cdots+I_{100}}{100}$. Note that $E[I_i] = \frac{4}{5}$ for all i, so by Linearity of Expectation we have that

$$E[X] = E\left[\frac{I_1 + \cdots + I_{100}}{100}\right] = \frac{1}{100}E[I_1] + \cdots + \frac{1}{100}E[I_{100}] = \frac{1}{100} \cdot 100 \cdot \frac{4}{5} = \frac{4}{5}.$$

(b) What is $\text{Var}(X)$?

Solution: Note that for any i, I_i is a Bernoulli random variable with parameter $p = \frac{4}{5}$, so $\text{Var} I_i = \frac{4}{5}\left(1 - \frac{4}{5}\right) = \frac{4}{25}$. Since we can express $X = \frac{I_1+\cdots+I_{100}}{100}$, and the I_i are independent, we have that

$$\begin{aligned}
\text{Var}(X) &= \text{Var}\left(\frac{I_1 + \cdots + I_{100}}{100}\right) \\
&= \left(\frac{1}{100}\right)^2 \text{Var}\left(I_1 + \cdots + I_{100}\right) \\
&= \left(\frac{1}{100}\right)^2 \text{Var}\left(I_1 + \cdots + I_{100}\right) \\
&= \left(\frac{1}{100}\right)^2 (\text{Var}(I_1) + \cdots + \text{Var}(I_{100})) \\
&= \frac{100 \cdot \frac{4}{25}}{10000} \\
&= \frac{1}{625}.
\end{aligned}$$

(c) Suppose Leanne flips n coins instead of 100. What does the LLN tell us about X?

Solution: The LLN tells us that for any $\varepsilon > 0$, the probability that X is within ε of 0.8 goes to 1 as $n \to \infty$. This can be seen for $n = 100$; the variance is small, so it is unlikely for X to be very far from 0.8.