Math 10B with Professor Stankova
Worksheet, Discussion #15; Thursday, 3/14/2019
GSI name: Roy Zhao

1 Variance and Covariance

1.1 Concepts

1.

Distribution	PMF	$E(X)$	Variance
Uniform	If $\#R(X) = n$, then $f(x) = \frac{1}{n}$ for all $x \in R(X)$.	$\sum_{i=1}^{n} \frac{x_i}{n}$	$\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{n}$
Bernoulli Trial	$f(0) = 1-p$, $f(1) = p$	p	$Var(X) = p(1-p)$
Binomial	$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geometric	$f(k) = (1-p)^k p$	$\frac{1-p}{p}$	$Var(X) = \frac{1-p}{p^2}$
Hyper-Geometric	$f(k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$	$\frac{nm}{N}$	$\frac{nm(N-m)(N-n)}{N^2(N-1)}$
Poisson	$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ

The **Variance** is defined as $Var(X) = E((X-\mu)^2)$. An easier form is $E(X^2) - E(X)^2$. It satisfies some properties:

- $Var(c) = 0$
- $Var(cX) = c^2 Var(X)$
- $Var(X+Y) = Var(X) + V(Y)$ for **independent** random variables.

The **Standard Error** is defined as $SE(X) = \sqrt{Var(X)}$. We use it to get the same units as X.

The **Covariance** is defined as $Cov(X,Y) = E[XY] - E[X]E[Y]$. It measures how "independent" two random variables are. For **independent** random variables, we have $Cov(X,Y) = 0$. Note that we can recover the definition of regular variance because the covariance of a random variable with itself is $Cov(X,X) = E[X^2] - E[X]^2 = Var(X)$. We can update the formula for the variance of the sum of two random variables as $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$ which holds for **all** random variables. Properties that hold for the random variable are:

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X, Y+Z) = Cov(X,Y) + Cov(X,Z)$
- $Cov(X, cY) = cCov(X,Y)$ for any constant c
- $Cov(X, c) = 0$ for any constant c

1.2 Examples

2. Let X be a uniform random variable with range $\{-1, 0, 1\}$. Let $Y = X^2$. Calculate $Cov(X, Y)$.

Solution: The PMF for X is

$$f(x) = \begin{cases} \frac{1}{3} & x = -1, 0, 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then the PMF for Y is

$$g(x) = \begin{cases} \frac{2}{3} & x = 0, 1 \\ 0 & \text{otherwise.} \end{cases}$$

Finally we calculate the PMF for $XY = X^3$ to be the same as the PMF for X. Now finally

$$\begin{aligned} Cov(X, Y) &= E[XY] - E[X]E[Y] \\ &= (1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{3}) - (1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{3})(1 \cdot \frac{2}{3} + 0 \cdot \frac{1}{3}) \\ &= 0 - 0 \cdot \frac{2}{3} = 0. \end{aligned}$$

1.3 Problems

3. True **FALSE** For independent random variables X, Y we have $Var(X - Y) = Var(X) - Var(Y)$.

Solution:

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

4. **TRUE** False The product of two Bernoulli trials is another Bernoulli trial.

Solution: A Bernoulli trial has range $\{0, 1\}$ and the product of any two numbers there are 0 or 1 so the product is still a Bernoulli trial.

5. **TRUE** False If c is a constant, then $Var(X + c) = Var(X)$.

Solution: Conceptually, this is telling you that if you shift a distribution left or right by c, it doesn't change how spread out it is. You can use the shortcut formula to also verify this.

6. True **FALSE** The covariance of two random variables is always ≥ 0.

Solution: The covariance of X with $-X$ is $Cov(X, -X) = E[-X^2] - E[X]E[-X] = -Var(X) \leq 0$.

7. **TRUE** False For random variables X, Y and constants c, d, we have $Cov(X + c, Y + d) = Cov(X, Y)$.

Solution: We can compute this out by plugging in $Cov(X + c, Y + d) = E[(X + c)(Y + d)] - E[X + c]E[Y + d]$ and using the fact that the expected value of a constant is the constant itself ($E[c] = c$) to simplify and get $E[XY] - E[X]E[Y] = Cov(X, Y)$.

8. While pulling out of a box of cookies, what is the expected number of cookies I have to pull out before I pull out an oatmeal raisin if 20% of cookies are oatmeal raisin? What is the variance?

Solution: This is a geometric distribution because I am counting the number of cookies I have to pull out before a success. The probability of success is $20\% = p = 1/5$. So the expected number of cookies I have to pull out is $\frac{1-p}{p} = 4$. The variance is $\frac{1-p}{p^2} = 4/(1/5) = 20$.

9. I flip a coin some number of times and I expected to see 90 heads with a standard deviation of 3 heads. What is the probability that I actually see 95 heads?

Solution: We are in a binomial distribution with $np = 90$ and $\sqrt{np(1-p)} = 3$ so $np(1-p) = 9$ and $1 - p = \frac{1}{10}$ so $p = \frac{9}{10}$ and $n = 100$. Thus, to get $k = 95$ heads, the probability is

$$f(95) = \binom{100}{95}(0.9)^{95}(0.1)^5.$$

10. I am at a casino and play a game and am expected to gain 10 cents per play with a variance of 1$^2 if I bet $10. What is the expected value and variance when I bet $100 instead?

Solution: We can think of this as $E[X]$ vs $E[10X]$. The expected value is multiplied by 10 so I expect to get $1 and the variance is multiplied by 10^2 so the variance is 100$^2.

11. Prove the short cut formula for variance from the definition of variance.

Solution:

$$\begin{aligned} Var[X] &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[XE[X]] + E[E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2. \end{aligned}$$

Where we use the fact that $E[X]$ is a constant so we can move it out of the expected value and the expected value of a constant is the constant itself.

12. Suppose that I flip a fair coin 10 times. Let T be the number of tails I get and H the number of heads. Calculate $E[T], E[H], Var[T], Var[H], Var[T + H]$. Now calculate $E[T - H]$ and $Var[T - H]$.

Solution: Both T, H are binomial distributions with $T + H = 10$ because there are 10 coin flips total. Thus, using the formula for the binomial distribution with $n = 10, p = \frac{1}{2}$, we get that

$$E[T] = E[H] = np = 5.$$

Then $Var[T] = Var[H] = np(1 - p) = 2.5$. Finally $Var[T + H] = Var[10] = 0$. And also $E[T - H] = E[T] - E[H] = 0$.

Now to calculate the variance, we cannot split it up since T, H are not independent. But, we know that $H + T = 10$ so $T - H = T - (10 - T) = 2T - 10$ and so $Var[T - H] = Var[2T - 10] = Var[2T] = 4Var[T] = 10$.

2 Average of Random Variables

2.1 Concepts

13. For X_i independent and identically distributed (i.i.d.) (e.g. rolling a die multiple times) with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$, then the average that we get (e.g. the average number that we roll) is **approximately** normal distributed with mean μ and standard deviation $\sigma/\sqrt{n}$. So
$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
is approximately normally distributed with $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \sigma^2/n$.

2.2 Examples

14. Show that the distribution of $\overline{X}$, the average of n i.i.d. random variables with mean μ and standard deviation σ has mean μ and standard deviation $\sigma/\sqrt{n}$.

Solution: First, we note that the mean is
$$E[\bar{X}] = E[\frac{X_1 + \cdots + X_n}{n}] = \frac{E[X_1] + \cdots + E[X_n]}{n} = \frac{\mu n}{n} = \mu.$$
Then, the variance adds and note that $Var(cX) = c^2 Var(X)$ to get that
$$Var(\bar{X}) = \frac{1}{n^2} Var(X_1 + \cdots + X_n) = \frac{Var(X_1) + \cdots + Var(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$
Therefore the standard error or standard deviation is $\sqrt{(\sigma^2)/n} = \sigma/\sqrt{n}$.

15. Suppose that the height of women is distributed with an average height of 63 inches and a standard deviation of 10 inches. Taking a sample of 100 women, what is the expected value and standard deviation for the average height?

Solution: The average height of 100 women will be approximately normally distributed with average 63 and standard deviation $10/\sqrt{100} = 1$.

2.3 Problems

16. **TRUE** False For a constant $c \geq 0$, we have that $SE(cX) = cSE(X)$.

Solution: This comes from the fact that $Var(cX) = c^2Var(X)$ and so $SE(cX) = \sqrt{Var(cX)} = \sqrt{c^2Var(X)} = c\sqrt{Var(X)} = cSE(X)$.

17. True **FALSE** Suppose that I roll a fair die 100 times. Then since the expected value of the average die roll is 3.5, I will roll a $1, 2$, or 3 50 times and a $4, 5$, or 6 50 times.

Solution: The average is 3.5 but I could roll all 6's for instance.

18. Suppose the weight of newborns is distributed with an average weight of 8 ounces and a standard deviation of 1 ounce. Today, there were 25 babies born at the Berkeley hospital. What is the expected value and variance of the average weight of these babies?

Solution: The average weight of these babies will be approximately normally distributed with mean 8 and standard deviation $1/\sqrt{25} = 0.2$ so variance of $0.2^2 = 0.04$.

19. Suppose that the average lifespan of a human is 75 years with a standard deviation of 10 years. What is the mean and standard error of the average lifespan of a class of 25 students?

Solution: The average lifespan of 25 students is approximately normally distributed with mean 75 and standard deviation $10/\sqrt{25} = 2$.

20. Suppose that in the 2012 election, 55% of people preferred Obama over Romney. If I sample 100 random people (assume that they are independently chosen), what is the expected value and variance for the percentage of the people sampled who support Obama?

Solution: Let X be a random variable that is 1 if the person prefers Obama and 0 otherwise. Then, we know that $E[X] = 0.55$ and $Var(X) = 0.55 \cdot (1 - 0.55)^2 + 0.45(0 - 0.55)^2 = 0.2475$ so $SE(X) \approx 0.5$. Let $\bar{X}$ be the average of asking 100 people, and hence $\bar{X}$ is normally distributed with mean 0.55 and standard deviation $0.5/\sqrt{100} = 0.05$.

21. The newest Berkeley quarterback throws an average of 0.75 TDs/game with a standard deviation of 1. What is his expected value and standard deviation for the number of TDs he throws next season (16 total games)?

Solution: In 16 games, he will average 0.75 TDs/game with a standard deviation of $1/\sqrt{16} = 0.25$. So multiplying by 16 gives us that he will average 12 TDs total with a standard deviation of $1/\sqrt{16} \cdot 16 = 4$ TDs.

22. Let $X_1, \ldots, X_4$ be i.i.d Bernoulli trials with $p = \frac{3}{4}$. Let $\bar{X}$ be the average of them. What is $Var[\bar{X}]$? Find $Cov(X_1, \bar{X})$ (Hint: Write $\bar{X} = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$).

Solution: Each of the X_i is Bernoulli so expected value of $p = \frac{3}{4}$ and variance of $p(1-p) = \frac{3}{16}$. Then the variance of $Var[\bar{X}] = \frac{1}{n} Var[X_1] = \frac{1}{4} \cdot \frac{3}{16} = \frac{3}{64}$. Finally, we have that

$$\begin{aligned} Cov(X_1, \bar{X}) &= Cov(X_1, \frac{1}{4}(X_1 + X_2 + X_3 + X_4)) \\ &= \frac{1}{4}(Cov(X_1, X_1) + Cov(X_1, X_2) + Cov(X_1, X_3) + Cov(X_1, X_4)) \\ &= \frac{1}{4}(Var(X_1) + 0 + 0 + 0) \\ &= \frac{1}{4}\frac{3}{16} = \frac{3}{64}. \end{aligned}$$