Math 10B with Professor Stankova
Worksheet, Discussion #14; Tuesday, 3/6/2018
GSI name: Roy Zhao

1 Covariance

1.1 Concepts

1. 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 Example

2. If $Var(X) = 1$ and $Cov(X, Y) = 0$, what is $Cov(X, X + Y)$?

Solution: Following the formula, we have that

$$Cov(X, X+Y) = E[X(X+Y)] - E[X]E[X+Y] = E[X^2] + E[XY] - E[X](E[X] + E[Y])$$

$$= E[X^2] - E[X]^2 + E[XY] - E[X]E[Y] = Var(X) + Cov(X, Y) = 1 + 0 = 1.$$

1.3 Problems

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

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

4. **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 $E[XY] - E[X]E[Y] = Cov(X, Y)$.

5. If $Var(X) = 4$ and $Var(Y) = 1$ and $Cov(X, Y) = -2$, what is $Var(X + 2Y)$?

Solution: Computing and plugging $X + 2Y$ into the formula for variance, we get that

$$Var(X+2Y) = E[(X+2Y)^2] - E[X+2Y]^2 = E[X^2+4XY+4Y^2] - (E[X]^2+4E[X]E[Y]+4E[Y]^2)$$

$$= E[X^2] - E[X]^2 + 4E[XY] - 4E[X]E[Y] + 4E[Y^2] - 4E[Y]^2$$
$$= Var(X) + 4Cov(X, Y) + 4Var(Y) = 4 - 8 + 4 = 0.$$

Another way to do this is use the formula above to get

$$Var(X+2Y) = Var(X)+Var(2Y)+2Cov(X, 2Y) = Var(X)+4Var(Y)+4Cov(X, Y) = 0.$$

1.4 Extra Problems

6. If $Var(X) = 1$ and $Var(Y) = 1$ and $Cov(X, Y) = 3$, what is $Var(2X + Y)$?

Solution: Use the formula above to get

$$Var(2X+Y) = Var(2X)+Var(Y)+2Cov(2X, Y) = 4Var(X)+Var(Y)+4Cov(X, Y) = 17.$$

2 Z-Scores, CLT, LoLN

2.1 Concepts

7. In order to compute the probability $P(a \leq X \leq b)$ for a normal distribution, we need to take an integral $\int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/\sigma^2}$ and this integral is almost impossible to do without a calculator. So, what we do is have a table of values for this integral and look up the value that we need. Given a z score such as 1.5, when we look it up in the table,

$z(1.5) = P(0 \le Z \le 1.5)$, where Z is the standard normal distribution; the bell curve with mean $\mu = 0$ and standard deviation $\sigma = 1$.

One key area these pop up in is when taking the average of a bunch of trials. The **Central Limit Theorem (CLT)** tells us that 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 $Var(\bar{X}) = \sigma^2/n$.

In order to compute probabilities, we compute the z score. Given a normal distribution with mean μ and standard deviation σ, the z score of a value a is $\frac{|a-\mu|}{\sigma}$. Then we look up this value in a table.

The **Law of Large Numbers** is a weaker statement that just says that as we take averages and let $n \to \infty$, then the sample mean becomes closer and closer to the actual mean μ. Namely, $E[\bar{X}] \to \mu$ and the probability that we are far away from the mean goes to 0.

2.2 Examples

8. Let f be normally distributed with mean -2 and standard deviation 4. Calculate the probability $P(-1 \le X \le 1)$.

Solution: We have $P(-1 \le X \le 1) = P(-2 \le X \le 1) - P(-2 \le X \le -1)$ and we calculate the z scores. The first is $\frac{|1-(-2)|}{4} = \frac{3}{4}$ and the second is $\frac{|-1-(-2)|}{4} = \frac{1}{4}$. Thus, the probability is $z(0.75) - z(0.25)$.

9. 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 probability that the average of the heights of these 100 women is between 62 and 64 inches?

Solution: The average height of 100 women will be approximately normally distributed with average 63 and standard deviation $10/\sqrt{100} = 1$. Therefore, $P(62 \le \bar{X} \le 64) = P(62 \le \bar{X} \le 63) + P(63 \le \bar{X} \le 64) = z(1) + z(1) = 2z(1)$.

2.3 Problems

10. **TRUE** False We can only use the z score to calculate probabilities of normal distributions (bell curves).

Solution: The table only applies for probability of normal distributions.

11. True **FALSE** The normal distribution with positive mean can only take on positive values. ($P(X \leq 0) = 0$)

Solution: The normal distribution can take on any real value.

12. Let f be normally distributed with mean 1 and standard deviation 4. Calculate the probability $P(X \geq 3)$.

Solution: In order to calculate this probability, we need the probability area to touch the median so we have $P(X \geq 3) = P(X \geq 1) - P(1 \leq X \leq 3)$. The first probability is $\frac{1}{2}$ and the second has the z score $\frac{|3-1|}{4} = \frac{1}{2}$. So the answer is $0.5 - z(0.5)$.

13. Let f be normally distributed with mean -2 and standard deviation 4. Calculate the probability $P(-3 \leq X \leq 1)$.

Solution: We have $P(-3 \leq X \leq 1) = P(-3 \leq X \leq -2) + P(-2 \leq X \leq 1)$ and we calculate the z scores. The first is $\frac{|-3-(-2)|}{4} = \frac{1}{4}$ and the second is $\frac{|1-(-2)|}{4} = \frac{3}{4}$. Thus, the probability is $z(0.25) + z(0.75)$.

14. Let f be normally distributed with mean 5 and standard deviation 2. Calculate the probability $P(X \leq 3)$.

Solution: We have $P(X \leq 3) = P(X \leq 5) - P(3 \leq X \leq 5)$ and we calculate the z scores. The first is just $\frac{1}{2}$ and the second is $\frac{|3-5|}{2} = 1$. Thus, the probability is $0.5 - z(1)$.

15. Let f be normally distributed with mean 0 and standard deviation 5. Calculate the probability $P(-2 \leq X \leq -1)$.

Solution: We have $P(-2 \leq X \leq -1) = P(-2 \leq X \leq 0) - P(-1 \leq X \leq 0)$ and we calculate the z scores. The first is $\frac{|-2-0|}{5} = \frac{2}{5}$ and the second is $\frac{|-1-0|}{5} = \frac{1}{5}$. Thus, the probability is $z(2/5) - z(1/5)$.

16. 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 probability that the average weight of these newborns is less than 7.5 ounces?

> **Solution:** The average weight of these babies will be approximately normally distributed with mean 8 and standard deviation $1/\sqrt{25} = 0.2$. The probability is $P(\bar{X} \leq 7.5) = 0.5 - P(7.5 \leq \bar{X} \leq 8) = 0.5 - z(2.5)$.

17. Suppose that the average lifespan of a human is 75 years with a standard deviation of 10 years. What is the probability that in a class of 25 students, they will on average live longer than 80 years?

> **Solution:** The average lifespan of 25 students is approximately normally distributed with mean 75 and standard deviation $10/\sqrt{25} = 2$. Thus $P(\bar{X} \geq 80) = 0.5 - z(|80 - 75|/2) = 0.5 - z(2.5)$.

2.4 Extra Problems

18. Let f be normally distributed with mean 3 and standard deviation 5. Calculate the probability $P(X \geq 0)$.

> **Solution:** We have $P(X \geq 0) = P(0 \leq X \leq 3) + P(3 \leq X)$ and we calculate the z scores. The second is just $\frac{1}{2}$ and the first is $\frac{|0-3|}{5} = 0.6$. Thus, the probability is $0.5 + z(0.6)$.

19. Let f be normally distributed with mean 2 and standard deviation 1. Calculate the probability $P(X \leq 0)$.

> **Solution:** We have $P(X \leq 0) = P(X \leq 2) - P(0 \leq X \leq 2)$ and we calculate the z scores. The first is just $\frac{1}{2}$ and the second is $\frac{|0-2|}{1} = 2$. Thus, the probability is $0.5 - z(2)$.

20. The newest Berkeley quarterback throws an average of 0.9 TDs/game with a standard deviation of 1. What is the probability that he averages at least 1 TD/game next season (16 total games)?

Solution: In 16 games, he will average 0.9 TDs/game with a standard deviation of $1/\sqrt{16} = 0.25$. So the probability that he averages at least 1 TD/game is $P(\bar{X} \geq 1) = 0.5 - P(0.9 \leq \bar{X} \leq 1) = 0.5 - z(\frac{|1-0.9|}{0.25}) = 0.5 - z(0.4)$.

21. Suppose that the average shopper spends 100 dollars during Black Friday, with a standard deviation of 50 dollars. What is the approximate probability that a random sample of 25 shoppers will have spent more than \$3000?

Solution: In a sample of 25 shoppers, the average shopper will spend 100 dollars with a standard deviation of $50/\sqrt{25} = 10$. Thus, the probability that a random sample will spend more than 3000 dollars is the probability that a random sample will average more than $3000/25 = 120$ dollars per person. This probability is $0.5 - z(|120 - 100|/10) = 0.5 - z(2)$.

22. Suppose that on the most recent midterm, the average was 60 and the standard deviation 20. What is the approximate probability that a class of 25 had an average score of at least 66?

Solution: In a class of 25, the average score will be distributed with mean 60 and standard deviation $20/\sqrt{25} = 4$. The probability that they had an average score of at least 66 is $0.5 - z(|66 - 60|/4) = 0.5 - z(1.5)$.