**Math 10B with Professor Stankova**
**Worksheet, Discussion #16; Tuesday, 3/19/2019**
**GSI name: Roy Zhao**

# 1 Z-Scores, CLT, LoLN

## 1.1 Concepts

1. 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, $z(1.5) = P(0 \leq Z \leq 1.5)$, where $Z$ is the standard normal distribution; the bell curve with mean $\mu = 0$ and standard deviation $\sigma = 1$. To find how many standard deviations a value $a$ is away from the mean, you can use the formula $\frac{|a-\mu|}{\sigma}$.

   One key area these pop up in is when taking the average of a bunch of trials. For $X_i$ independent and identically distributed (i.i.d.) (e.g. rolling a die multiple times) with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$, the **Central Limit Theorem (CLT)** tells us that the average that we get (e.g. the average number that we roll)

   - is **approximately** normal distributed (we can approximate probabilities with $z$-scores)
   - $\bar{\mu} = E[\bar{X}] = \mu$,
   - $\bar{\sigma} = SE(\bar{X}) = \sigma/\sqrt{n}$.

   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 general, you can think of it in this table:

| | |
|---|---|
| $X$ | The test score of a single student in $10B$. |
| $\mu = E[X]$ | The average test score of the students in $10B$. |
| $\sigma = SE(X) = \sqrt{Var(X)}$ | The likelihood of a random student having a test score far away from the average $\mu$. |
| $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ | The average test score in a particular discussion section. |
| $\bar{\mu} = E[\bar{X}]$ | The average over all sections of the average test score of the section. |
| $\bar{\sigma} = SE(\bar{X}) = \sqrt{Var(\bar{X})}$ | The likelihood of a random discussion section having an average test score far away from the average $\bar{\mu}$. |

Note that putting a bar over the constants $\mu, \sigma, X$ just tells us that we are considering the $\mu, \sigma$ for the averaged random variable $\bar{X}$. Now the reason that $\mu = \bar{\mu}$ is because if you think about it, we are still taking the average over the same population (everyone in $10B$). The reason that $\bar{\sigma} \leq \sigma$ is because if we take the average test score of a particular section, if there are students with high scores, it is likely their score will cancel with students with lower scores. So, it is less likely that a section average will be far away from the class average.

In order to compute probabilities, we compute the $z$ score. Given a normal distribution with mean $\mu$ and standard deviation $\sigma$, 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 $\mu$. Namely, $E[\bar{X}] \to \mu$ and the probability that we are far away from the mean goes to 0. Mathematically, it says that for any fixed $\epsilon > 0$, we have

$$\lim_{n \to \infty} P(|\bar{X} - \mu| > \epsilon) = 0.$$

## 1.2 Examples

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

**Solution:** We have $P(-1 \leq X \leq 1) = P(-2 \leq X \leq 1) - P(-2 \leq X \leq -1)$ and we calculate the $z$ scores. To calculate the $z$ score value of $a$, we take $\frac{|a-\mu|}{\sigma}$. The first is $\frac{|1-(-2)|}{4} = \frac{3}{4}$ and the second is $\frac{|-1-(-2)|}{4} = \frac{1}{4}$. Then looking pictorially, we see that we need to subtract these two probabilities. Thus, the probability is $z(0.75) - z(0.25)$.

3. Suppose you ask a group of 9 voters whether they voted for Romney or Obama and suppose that they are equally likely to have voted for either. Calculate the exact probability that less than a third of them voted for Romney. Use the CLT to approximate the probability that less than a third of them voted for Romney.

**Solution:** We can think of this as a Binomial distribution with a success being someone voting for Romney. The range of $\bar{X}$ or the percentage of people voting for Romney is $Range(\bar{X}) = \{0, 1/9, 2/9, \ldots, 9/9 = 1\}$. Then, this is a binomial distribution so

$$f(k/9) = \binom{9}{k}(0.5)^k(1-0.5)^{9-k} = \frac{\binom{9}{k}}{2^9}.$$

So, the probability that less than a third of them voted for Romney is if $0, 1$, or 2 of them voted for him so

$$f(0) + f(1/9) + f(2/9) = \frac{\binom{9}{0}}{2^9} + \frac{\binom{9}{1}}{2^9} + \frac{\binom{9}{2}}{2^9} = \frac{46}{2^9}.$$

Now to approximate the probability, we want to use $z$-scores to calculate the probability of the normal distribution. The CLT tells us that $\bar{X}$, which is an average of Bernoulli trials, is approximately normally distributed with $\bar{\mu} = \mu = p = \frac{1}{2}$. Then $\bar{\sigma} = \sigma/\sqrt{n} = \sqrt{p(1-p)}/\sqrt{9} = \frac{1}{6}$. We want to calculate

$$\begin{aligned} P(\bar{X} \le 1/3) &= P(\bar{X} \le 1/2) - P(1/3 \le \bar{X} \le 1/2) \\ &= 1/2 - z(\frac{|1/3 - 1/2|}{1/6}) \\ &= 1/2 - z(1). \end{aligned}$$

## 1.3 Problems

4. **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.

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

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

6. **TRUE** False You can use the Central Limit Theorem to prove the Law of Large Numbers.

**Solution:** We know that $\overline{X}$ has distribution with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$ and hence as $n \to \infty$, the standard deviation $\to 0$ which means that all of the probability gets concentrated at $\mu$, which is exactly what the law of large numbers says.

7. True **FALSE** Suppose I calculate that probability that in a sample of $10,000$ men, their average height is less than 66 inches is 99.999%. Then all but one or two men in a sample of $10,000$ men will have a height of less than 66 inches.

**Solution:** The probability that the average is less than 66 inches is very high but that does not mean that same percentage will actually have height less than 66 inches. (Average vs individual values)

8. 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 $0.5 - z(0.5)$.

9. 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 $z(0.25) + z(0.75)$.

10. 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 $0.5 - z(|66 - 60|/4) = 0.5 - z(1.5)$.

11. 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 $P(\bar{X} \geq 1) = 0.5 - P(0.9 \leq \bar{X} \leq 1) = 0.5 - z(\frac{|1-0.9|}{0.25}) = 0.5 - z(0.4)$.

12. Suppose that the height of NBA players is distributed with an average height of 83 inches and a standard deviation of 10 inches. Taking a sample of 100 players, approximate the probability that the average of the heights of these 100 players is between 82 and 84 inches.

**Solution:** The average height of 100 players will be approximately normally distributed with average 83 and standard deviation $10/\sqrt{100} = 1$. Therefore, $P(82 \leq \bar{X} \leq 84) = P(82 \leq \bar{X} \leq 83) + P(83 \leq \bar{X} \leq 84) = z(1) + z(1) = 2z(1)$.