STATS 412

Tenth Class Note

In Son Zeng

10 October, 2018

My Office Hour:

My office hours are on **16:30 - 18:00 Tuesday** and **13:30 - 15:00 Friday**, at **USB 2165**. You may check the campus map to get to my office. I am prepared for your questions, so please feel free to come to my office hours. During the fall break, I have extra office hours on **14:00 - 16:00 Monday** and **16:30 - 18:00 Tuesday** before the midterm.

Calculus Review:

• Last week, I uploaded a book for probability to the Piazza. Particularly, I recommend you to read (Type the page number in PDF):

(1) Double integral: page 214 - 218

(2) Cumulative distribution function: page 224 - 229

(3) Expectation and Variance: page 235 - 244

(4) Joint Distributions: page 299 - 304

(5) Binomial Distribution: page 135 - 146

(6) Poisson Distribution: page

These are great practices to prepare you with essential integration skills and knowledge of distributions for the subsequent homework and the midterm.

Comments for Assignment 5

• Some problems are notable in this homework, including misplaced decimals in answer, not taking square root of variance for standard deviation and Check the Homework Comments 5 that I have uploaded through Piazza for detail.

• In question 5 and question 10, in part b), c), e) and f), I believe that you will try your very best to provide a reasonable interpretation/explanation to your decision. One notable idea may help your understanding and interpretation is what we will go through after the midterm:

p-value: The p-value is defined as the probability, under the null hypothesis (H_0, the hypothesis that we would like to find evidence to reject), of obtaining a result **equal to or more extreme than** what was actually observed.

Key Points during Lecture:

Poisson Distribution: It is clear in the lecture that Poisson distribution is an important and special distribution because its mean and variance are the same. That is, for a Poisson random variable $X \sim Poisson(\lambda)$:

$$\boxed{E(X) = V(X) = \lambda} \tag{1}$$

Test Reminder: When you encounter a question about normal distribution, you need to specify the probability correctly. For example, in lecture note page 12, the question says:

"What is the probability that the reaction time for a randomly selected person is **greater than 1.00 second?**"

Then you have to answer $P(X > 1.00)$. If the question becomes "**greater than or equal to 1.00 second**" or "**at least 1.00 second**", then you should answer $P(X \geq 1.00)$ to receive full credit.

You may argue that "well, does not it the two probabilities are the same because normal distribution is a continuous variable?" The answer is YES, the probability $P(X = 1.00) = 0$, and it is true for every single point. However, you should adhere to what the question asks you to do during the midterm examination. Good luck!

Symmetry of Standard Normal Distribution: During the class a question is raised: Why $P(Z > -2.72) = 1 - P(Z \leq -2.72) = P(Z \leq 2.72)$?

It seems intimidating at the first sight. There are several reasons why these relations are true. First of all, in the sense of CDF, $P(Z > -2.72)$ refers to the area where $Z > -2.72$, which is also the same as the total area subtracted by the part where $Z \leq 2.72$.

Another nice property of the standard normal distribution is that it is symmetric at $Z = 0$. Firstly, the PDF of standard normal distribution is an **even function** (see the proof below):

$$P_Z(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}, -\infty < z < \infty$$

$$P_Z(-z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(-z)^2}{2}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} = P_Z(z)$$

Therefore, the densities of the standard normal distribution are the same for $Z = 1$ and $Z = -1$, for example. Also, the probability of $P(Z \leq 0) = P(Z \geq 0) = \frac{1}{2}$, so we can safely use the nice equivalence relation:

$$P(Z \geq x) = P(Z \leq -x), -\infty < x < \infty$$

Issues about checking the standard normal Z table:

• If you are asked to characterize the lowest 2%, you need to look at the z-value corresponding to probability **0.02**, not 0.002, 0.0002, and not 0.2. Make sure you are clear about the translation of percentage to decimals before the midterm.

• Some people asked about using interpolation to find the z-value. For example, from the z-table, we will see that $P(Z \leq -2.57) = 0.0051, P(Z \leq -2.58) = 0.0049$. If you use linear interpolation to find the z-value corresponding to 0.005, what we will find is by specifying a line:

$$y - y_0 = k(x - x_0) \rightarrow -2.58 - (-2.57) = k(0.0049 - 0.0051) \rightarrow k = 50$$

Now we plug in $x = 0.005$:

$$y - (-2.57) = 50(0.005 - 0.0051) \rightarrow y = -2.57 - 0.005 = -2.575$$

Professor has mentioned that you can find the z-value by looking at the table and finding the one with the nearest probability. If you really like doing interpolation, you are free to do so as well.

Linear Functions of Normal Random Variables:

• If $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), ..., X_n \sim N(\mu_n, \sigma_n^2)$ are independent random variables, and $c_1, c_2, ..., c_n$ are constants where at least one of them is not 0, then we have

$$c_1X_1 + c_2X_2 + ... + c_nX_n \sim N\Big(\underbrace{c_1\mu_1 + c_2\mu_2 + ... + c_n\mu_n}_{=\text{mean}}, \underbrace{c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + ... + c_n^2\sigma_n^2}_{=\text{variance}}\Big)$$

Also, for example, if we have some subtractions, we have

$$c_1X_1 - c_2X_2 + c_3X_3 - c_4X_4 \sim N\Big(\underbrace{c_1\mu_1 - c_2\mu_2 + c_3\mu_3 - c_4\mu_4}_{=\text{mean}}, \underbrace{c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + c_3^2\sigma_3^2 + c_4^2\sigma_4^2}_{=\text{variance}}\Big)$$

Distribution of Sample mean in normal distribution:

• Since $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, we have

$$E(\bar{X}) = E\Big(\frac{1}{n}[X_1 + ... + X_n]\Big) = \frac{1}{n} \cdot (n\mu) = \mu$$

$$Var(\bar{X}) = Var\Big(\frac{1}{n}[X_1 + ... + X_n]\Big) = \frac{1}{n^2} \cdot (n\sigma^2) = \frac{1}{n}\sigma^2$$

Therefore, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. This result is important because it articulates the fundamental concept in statistical inference that when the sample size increases, the variability of the sample mean will go down (more accurate).

Last Comment:

Please inform me to fix the typos and grammatical mistakes if they exist. It is a great practice of writing and I appreciate your help!