STATS 412

Eleventh Class Note

In Son Zeng

23 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:

• After the exam 1, the requirements for integration will be lowered. However, to compute the Maximum Likelihood Estimator (MLE), you may encounter the difficulty for partial differentiation. If you have studied MATH 215 or the equivalent class before, you may review the notes. If you do not know how to perform partial differentiation, you may refer to the following websites for reference:

• http://tutorial.math.lamar.edu/Classes/CalcIII/PartialDerivsIntro.aspx

• https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivative-and-gradient-article a/introduction-to-partial-derivatives

These are great practices to prepare you with essential calculus skills and knowledge of distributions for the subsequent homework and the exam 2.

About Exam 1

• For the students who are satisfied about your exam 1 score, congratulation on the self-realization of your effort and I encourage you to continue on your good work so as to be successful.

• For the students who want more improvements, you are at the right time to work on your weaknesses. Researches show that people tend to retain longer memory about the details for unsatisfactory experiences than the satisfactory ones. You may raise questions for me to work on to help you improve your future exams and assignments. As discussed with Professor, the common misconceptions and unnecessary mistakes in exam 1 include, but not limited to, **0 otherwise, rounding expectation, misunderstanding about conditional and unconditional probability, integration**. Also, it is common that students may spend tremendous time on certain parts of the problem, while ignoring that some other parts may be easier to get to the answer. Therefore, work on concepts and time-management are two main big pictures to help you perform better in the next exam.

• Also, to be more successful in exams, it is better to pay efforts consistently and not wait until the last 5 days to realize that you are not familiar with a variety of concepts and cram for the review questions only.

• Finally, for the students who are not satisfied with your results, apart from my encouragement to work more systematically with the courseware and try organize the concepts better, I may try to identify your strengths and weaknesses (if you are willing to tell) so that I can develop different approaches of tutoring to accommodate different learning styles.

Future Homework Grading Policy:

Please include the final answer for each homework question. If the final answer is not included, you will risk 0.5 points for each missing part.

Key Points during Lecture:

Poisson Distribution Review: 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 $X \sim Poisson(\lambda)$:

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

$$\boxed{E(\bar{X}) = E(X) = \lambda, Var(\bar{X}) = \frac{\lambda}{n}} \tag{2}$$

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 (Review):

• 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).

p-value (Preview): There are multiple different definitions for the p-value from didderent probability and statistics books. From the Casella and Berger's book, 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.

Maximum Likelihood Estimate (MLE): As taught in the lecture, in the Poisson case, from the independence of $X_1, X_2, ...X_{30}$ we use the multiplicative rule to set up the **likelihood function**. The likelihood function is in the form:

$$\boxed{p(X_1 = x_1) \cdot p(X_2 = x_2) = ...p(X_{30} = x_{30}) = \prod_{i=1}^{30} \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!} = \frac{e^{-30\lambda}\lambda^{\sum_{i=1}^{30} x_i}}{\prod_{i=1}^{30} x_i!}} \tag{3}$$

The MLE can be obtained by finding the value of λ which maximizes this likelihood function. As taught in the lecture, we may take the derivative of the likelihood function with respect to λ.

However, this is not the whole story. In many case, it is difficult to take the derivative for a complicated likelihood function due to the complexity of probability distribution. One of the remedy is to take log to the likelihood function; such function is widely known as log-likelihood. Since log is a monotone increasing function, we can take the derivative of the log-likehood function to obtain the same MLE result.

• Maximum Likelihood Estimate (MLE) tends to be harder from the past experiences. Please ask questions during the office hours whenever you encounter trouble doing homework or revising the notes.

• I will include an extra question about MLE before the exam for revision. Again, I hope that students can help each other and post possible solution for the extra problems and I will definitely check.

The probability plots: Here is some comments about the probability plots. Firstly, if the observations fit perfectly to the $y = x$ line, it means that the observations (samples) are perfectly fit to the population values (the population may follow a variety of distributions different from the normal distribution, but normal distribution is prominent). Then we are confident to conclude the observed sample also follows roughly the same distribution.

Points shift up or shift down: It is not possible that all the points locate above the line. With the same logic, it is not possible that all the points locate below the line too. It may be the case that more points are located either above or below the line, but there should be at least one point located in the opposite direction of the line (you may also guess in which case it is right-skewed and in which case it is left-skewed, remember the definition).

Cauchy distribution (Optional): have longer-tails (or heavier tails) compared to the normal distribution. In other words, the Cauchy distribution has clear systematic departures from the normal distribution (the line). Look at the third plots in page , which looks approximately the Cauchy distribution.

Shapiro Wilk test (Optional): Shapiro Wilk test is a test for normality based on correlation between the data and the corresponding z-scores. The R code to perform the Shapiro Wilk test for one variable is shapiro.test(). Reference website: http://www.sthda.com/english/wiki/normality-test-in-r

Anderson Darling test (Optional): Another for normality for normality. The R code to perform the Shapiro Wilk test for one variable is as follows:

```r
install.packages("nortest")
library(nortest)
ad.test()
```

The data from the Anderson Darling test will inform us two results: A and p-value. The A value refers to the Anderson-Darling test statistics give the probability that the tested sample could have come from a normal distribution. Then, assuming that the null hypothesis (the process represents normally distributed data) is true, the p-value tells us the probability that our data follows the same distribution.

Reference website: https://rexplorations.wordpress.com/2015/08/11/normality-tests-in-r/

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!