STATS 412

Thirteenth Class Note

In Son Zeng

29 October, 2018

My Office Hour:

I have adjustments of my office hours this week. The office hours this week are on **16:30 - 18:00 Tuesday** and **17:30 - 19:00 Thursday**, 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.

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.

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.

Homework 7 Reminder:

• Make sure you are applying the correct formula for the MSE: $MSE(x) = [Bias(x)]^2 + Var(x)$. Additionally, there is always a tradeoff between increasing the bias while reducing the variance, or decreasing the bias while increasing the variance. In the realm of statistics, depending on the context, we may select an estimator which minimizes the variance, or is unbiased, or minimizes the MSE.

• In question 2 and 3, be sure to take the derivative with the correct parameter and check that the MLE is indeed the maximum by taking the second derivative.

• For question 4 - 8, be careful about the difference of the sample mean $\bar{X}$ and S_n; they approximate to normal distribution by CLT with different variance.

• For the quantiles, you may use the table provided by your book, or use statistical softwares to find out the corresponding z-score. The R code for deriving z-score is qnorm, for example, if you want to find 15-quantile, 25-quantile, 40-quantile, 80-quantile, 99.5-quantile, the corresponding z-scores are:

```
qnorm(0.15)
```

```
## [1] -1.036433
```

```
qnorm(0.25)
```

```
## [1] -0.6744898
```

```
qnorm(0.40)
```

```
## [1] -0.2533471
```

```
qnorm(0.80)
```

```
## [1] 0.8416212
```

```
qnorm(0.995)
```

```
## [1] 2.575829
```

Key Points during Lecture:

Test Reminder: There will be the standard normal cumulative distribution function table provided for reference during the test.

Large Sample Size: Regarding the notion of large sample size, in our class we use $n \geq 30$ as criteria.

Continuity correction: For a sample from Binomial distribution, as we perform the binomial approximation to normal, we will employ continuity correction to derive the probabilities.

• For example, if we role 100 dices, what if the probability that 20 or more are 6? For this one, since the sample size is $n = 100 \geq 30$, we have a large sample so that we can use Binomial approximation to Normal and calculate the mean as $\mu = np = 100 \cdot \left(\frac{1}{6}\right) = \frac{50}{3}$ and the standard deviation as $\sigma = \sqrt{np(1-p)} = \sqrt{100 \cdot \frac{1}{6}\frac{5}{6}} = \sqrt{13.89} = 3.727$. Here we have to be particularly careful, $P(X \geq 20) \approx P(X > 19.5) = P\Big(Z > \frac{19.5-\frac{50}{3}}{3.727}\Big) = P(Z > 1.506) = 1 - 0.9339 = 0.0661$. Here, $P(X \geq n) \to P(X > n - 0.5)$, and we also express $P(X = n) \to P(n - 0.5 < X < n + 0.5)$

Reference: https://www.statisticshowto.datasciencecentral.com/what-is-the-continuity-correction-factor/

p-value: 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.

MLE for normal distribution (Review): If $X_1, X_2, ...X_n$ are independent normally distributed random variables, the joint PDF (likelihood function) is $f(x_1, x_2, ..., x_n; \mu, \sigma^2) = (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}} \cdot exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\Big)$. Take the log we obtain the log-likelihood function:

$$\boxed{ln(f(x_1, x_2, ..., x_n; \mu, \sigma^2) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{1}$$

To find out the MLE estimator for the mean, we take the first derivative with respect to μ (I skipped the verification part which requires you to take the second derivative of μ and conclude that the second derivative of the log-likelihood function is negative for all x), and set the derivative as 0 (if you are having trouble this step, please also revise the chain rule in calculus:

$$\boxed{\frac{\partial ln(f(x_1, x_2, ..., x_n; \mu, \sigma^2)}{\partial \mu} = -\frac{1}{2\sigma^2} \cdot 2 \cdot (-1) \cdot \sum_{i=1}^{n}(x_i - \mu) =^{set} 0 \to \sum_{i=1}^{n} -n\mu = 0 \to \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}} \tag{2}$$

To find out the MLE estimator for the variance, we take the first derivative with respect to σ^2 (Again, I skipped the verification part which requires you to take the second derivative of σ^2 and find the result $\frac{\partial^2 ln()}{\partial(\sigma^2)^2} = -\frac{n}{2}$ that the second derivative of the log-likelihood function is negative for all x), and set the derivative as 0 (if you are having trouble this step, please also revise the chain rule in calculus:

$$\boxed{\frac{\partial ln(f(x_1, x_2, ..., x_n; \mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \cdot \sum_{i=1}^{n}(x_i - \mu)^2 =^{set} 0 \to \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 = n \to \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}} \tag{3}$$

I added a hat on the top of the μ and σ^2 to distinguish that they are, in fact, an estimator, In addition, for the variance, if we consider the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, we the MLE of the variance can be rewritten as $\hat{\sigma}^2 = \frac{n-1}{n}S^2$.

Properties of MLE (Review):

• Since the MLE does not necessarily exist and unique, the derivation of MLE may not work.

• Invariance Principle: For example, $\hat{\sigma}^2 = \frac{n-1}{n}S^2$ implies $\hat{\sigma} = \sqrt{\frac{n-1}{n}S^2}$.

• Since MLE approximates the minimum variance (MVUE), it does not necessary achieve the minimum variance. Students do not need to revise whether MLE achieves the minumum variance for the exam 2.

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!