STA255 Week-8

Shahriar Shams

02/02/2020

Review of Week-7

- Idea of Point Estimation
- Method of Moment Estimation
  - $\frac{1}{n}\sum_{i=1}^{n} X_i^k$ is an estimator of $E[X^k]$
- Mean Square Error (MSE): measuring accuracy of an estimator
  - $MSE[T] = var[T] + (Bias[T])^2$
- Bias and Underhandedness
  - $Bias[T] = E[T] - \theta$
- Maximum Likelihood Estimation
  - $L(\theta) = f(x_1, x_2, ..., x_n|\theta)$
  - If the obs are indep. $L(\theta) = f_\theta(x_1) * f_\theta(x_2) * ... * f_\theta(x_n)$
  - We look for θ that maximizes $L(\theta)$
  - Instead of maximizing $L(\theta)$, we maximize $logL(\theta)$

Learning goals

- Idea of interval estimation
- Definition of confidence interval
- Idea of Pivotal Quantity
- Confidence interval(CI) for μ with
  - population variance (σ^2) known
  - population variance (σ^2) unknown
- Two sided vs one sided CI
- Confidence interval for population proportion (will cover next week)
- Sample size calculation (will cover next week)
- Interpretation of confidence interval

Idea of interval estimation

- In **point estimation**,
  - we calculate an estimate (a single point on the number line) of an unknown parameter.
  - Suppose our unknown parameter is the average height of ALL UofT students (μ).
  - We take a sample of 100 students (say), calculate the sample mean and find the number to be 165.3cm
  - We say, 165.3cm is an estimate of μ

- It's very very likely(almost sure) that the value that we calculated from the sample (which is 163.3cm) is not the true value of mu.

- In **interval estimation**,
  - We try to find a range of values (also called an interval) that has a certain likelihood/probability of containing the true value of μ

Definition of confidence interval

- Suppose θ is our unknown parameter.
- An interval $(l(X_1, X_2, ..., X_n), u(X_1, X_2, ..., X_n))$ is a $100(1-\alpha)\%$- **confidence interval** for θ if

95% $\Rightarrow \alpha = 0.05$

$$P[l(X_1, X_2, ..., X_n) \leq \theta \leq u(X_1, X_2, ..., X_n)] = 1 - \alpha$$

where $1 - \alpha$ represents the confidence level of the interval.

- **In naive words,** we want "two numbers' which will have $1 - \alpha$ chance of containing the true parameter.
- We need to ensure two things here:
  - Need something that allows us connect the sample observations$(X_1, X_2, ..., X_n)$ to the parameter (θ)
  - We need a distribution so that we can calculate probability.

Idea of Pivotal Quantity

- It's a random variable
- Its expression involves the unknown parameter.
- But the distribution of it is free of the parameter.
- For example, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$
  - If you want to calculate a value of this you need to know μ.
  - But the distribution of this random variable does not depend on μ as we know

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Confidence Interval under Normal distribution for μ

Variance(σ^2) is known

- We know, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- Assuming $1-\alpha = 0.95$ we can write,

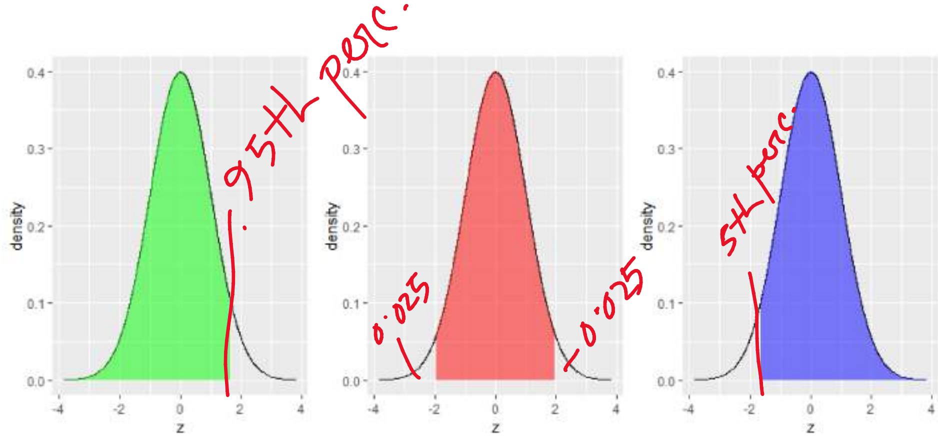
$$
\begin{aligned}
P\left[k_1 \le \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le k_2\right] &= 0.95 \\
\Longrightarrow P[k_1 * \frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le k_2 * \frac{\sigma}{\sqrt{n}}] &= 0.95 \\
\Longrightarrow P[\bar{X} - k_2 * \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} - k_1 * \frac{\sigma}{\sqrt{n}}] &= 0.95
\end{aligned}
$$

- k_1 and k_2 are quantiles of $N(0,1)$ distribution satisfying

$$P[k_1 \le Z \le k_2] = 0.95$$

  where Z is a standard Normal variable.

- **Question:** How do we decide what value of k_1 and k_2 to use?



- In green one, $k_1 = -\infty$ and $k_2 = 1.65 \Longleftarrow$ (0.95 quantile of a Standard Normal)
- In red one, $k_1 = -1.96$ and $k_2 = 1.96 \Longleftarrow$ (0.95 quantile of a Standard Normal)
- In blue one, $k_1 = -1.65$ and $k_2 = \infty$
- they all (along with infinitely many other) gives a total area of 0.95
- Simplest choice: pick the one with the shortest length of interval(which is the red one)

- In general for $100(1-\alpha)\%$ CI, $z_{\frac{\alpha}{2}}$ and $z_{1-\frac{\alpha}{2}}$ are preferred as the value of k_1 and k_2.
- Example: for $1-\alpha = 0.95 \implies \begin{cases} k_1 = z_{0.025} = -1.96 \\ k_2 = z_{0.975} = 1.96 \end{cases}$
- Finally, for $X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with σ^2 **known** we have the $100(1-\alpha)\%$-CI of μ as

$$\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\ ,\ \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

i = independently
i = identically
d = distributed

Example: Suppose \$(4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) $\overset{iid}{\sim} N(\mu, \sigma_0^2)$ with $\sigma_0^2 = 0.5$

Calculate the 95%-confidence interval for μ.

1. $n = 10$
2. $\bar{x} = \frac{1}{10}(4.7 + 5.5 + ... + 5.3) = 4.88$
3. $1 - \alpha = 0.95 \implies 1 - \frac{\alpha}{2} = 0.975$
4. using z-table or R [*qnorm(0.975)*], $z_{0.975} \approx 1.96$
5. 0.95-CI for μ:

$$4.88 \pm 1.96 * \frac{\sqrt{0.5}}{\sqrt{10}} = (4.442, 5.318)$$

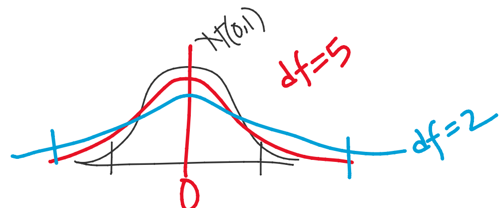
Variance(σ^2) is unknown

- When (σ^2) is unknown, we can't use $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ as a pivotal quantity anymore as we have two unknowns now.
- Since population variance is unknown, one intuitive solution that we can think of is replacing σ^2 by the variance calculated from the sample (S^2).
- Our Pivotal quantity becomes

$$\frac{\bar{X}-\mu}{S/\sqrt{n}}$$

- The distribution of this random variable is not $N(0,1)$ anymore.
- Rather it follows what's known as t-distribution with (n-1) degrees of freedom(df).

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(df=n-1)}$$



Some comments about t-distribution

- The term degrees of freedom(df) though have technical definition, but a naive way to look at it is that it specifies the parameter of the distribution.
- t-distribution looks similar to a Standard normal distribution with mean at zero and symmetric around the mean.
- t-distribution has a longer tail then standard normal.
- as the degrees of freedom increases, t-distribution more and more starts to look like a Standard Normal.

Now using the same idea used in the previous section,

- For $X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with σ^2 **unknown** we have the $100(1-\alpha)\%$-CI of μ as

$$\left(\bar{X} - t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\ ,\ \bar{X} + t_{1-\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}\right)$$

Example: (4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) $\overset{iid}{\sim} N(\mu, \sigma^2)$ with both μ and σ^2 unknown Calculate the 0.95-confidence interval for μ.

1. $n = 10$
2. $\bar{x} = \frac{1}{10}(4.7 + 5.5 + ... + 5.3) = 4.88$
3. $s = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}(\sum x_i^2 - n * (\bar{x})^2)} = 0.696$
4. $1 - \alpha = 0.95 \implies 1 - \frac{\alpha}{2} = 0.975$
5. using t-table or R [*qt(0.975,df=9)*], $t_{0.975,9} \approx 2.262$
6. 0.95-CI for μ:
$$4.88 \pm 2.262 * \frac{0.696}{\sqrt{10}} = (4.382, 5.378)$$

Two-sided vs. One sided CI

- So far in the two examples that we have done, we captured the middle $(1 - \alpha)$ area of the distribution.
- Which means we have discarded two ends of the distribution.
- We calculated both the lower and the upper bound.
- This is called two-sided CI

- In a on sided CI we start from the very left of the distribution or end at the very right.
- On the graph that's on page-4 of this document, the green and the blue CIs corresponds to one sided interval.
- So we only calculate either the lower bound or the upper bound.
- An $100(1 - \alpha)\%$ **upper confidence bound** for μ can be written as
  - When σ^2 is known,
  $$(\bar{X} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}})$$
  - When σ^2 is unknown,
  $$(\bar{X} + t_{1-\alpha,n-1}\frac{S}{\sqrt{n}})$$

- An $100(1-\alpha)\%$ **lower confidence bound** for μ can be written as
  - When σ^2 is known,
$$(\bar{X} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}})$$
  - When σ^2 is unknown,
$$(\bar{X} - t_{1-\alpha,n-1}\frac{S}{\sqrt{n}})$$

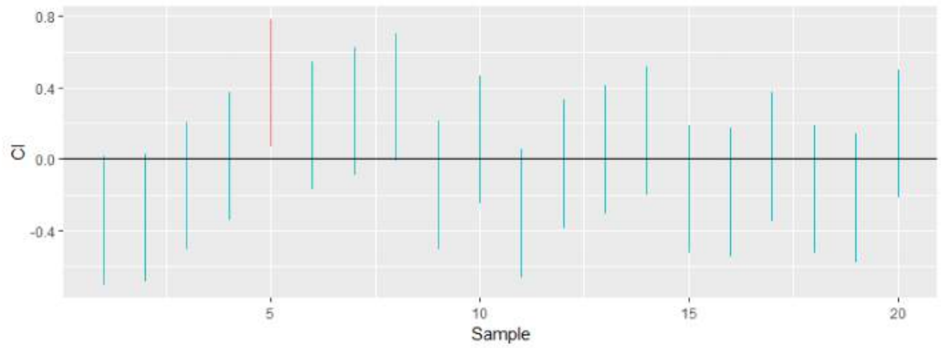
Interpretation of confidence interval

- We started with a goal in mind that we want to write a similar statement like the following
$$P[l() < \theta < u()] = 0.95$$
- Using example on page 4 of this document, we got the 95% CI for μ as (4.442, 5.318)
- Does it mean $P[4.442 < \mu < 5.318] = 0.95$?

- So far what we have learned in this course is based on the assumption that a parameter is a fixed constant (though unknown).
- Is this a valid statement if we believe μ is a constant?
$$P[4.442 < \mu < 5.318] = 0.95 \quad \leftarrow \text{invalid}$$

- We are trying to say μ, which is a constant has a 95% chance of being bounded by 4.442 and 5.318. Which is an invalid statement.

- Question: the two bounds that we have calculated, will we gt the same bounds if we take another 10 samples? Ans is No.
- So the two bounds are actually random variables that will change values from one sample to the other.
- The following graph descries this idea.
- I took a sample of 30 observations from $N(0,1)$ and calculated 95% CI for μ.
- That gave me the first vertical blue line.
- I repeated the task 20 times and that gave me the picture.



- 1 out of these 20 CIs missed the true mean ($\mu = 0$, the horizontal line)
- **Wrong interpretation:** There is 95% chance that μ is between 4.442 and 5.318
- **Correct interpretation:** If we keep taking samples (infinite times) and keep constructing 0.95-CIs, in 95% of the cases our CIs will capture the true value of the parameter.
- Question: The confidence interval that we calculated, does it include the true parameter? (In other words, the one that we calculated is it a red one or blue one in the graph)? - We don't know!

Chapter 8.4 and 8.5 not needed.

Homework

Chapter 8.1

1-3, 8

Chapter 8.2

12, 13, 17, 28

Chapter 8.3

33, 38