STA261: Probability and Statistics II

Shahriar Shams

Week 6 (Interval estimation: z, t, χ^2 and MLE based confidence intervals)

UNIVERSITY OF TORONTO

Winter 2020

Recap of Week 5

- Large sample property of MLE
  - $\hat{\theta}$ is the MLE of θ_0.
  - $nI(\theta_0)$ is the Fisher Information.
  - For $n \to \infty$
$$\hat{\theta} \xrightarrow{D} N(\theta_0, \frac{1}{nI(\theta_0)})$$

- Efficiency
  - Cramer Rao Lower Bound(CRLB) for variance of unbiased estimators.
$$var[T] \geq \frac{1}{nI(\theta_0)}$$

Learning goals for this week

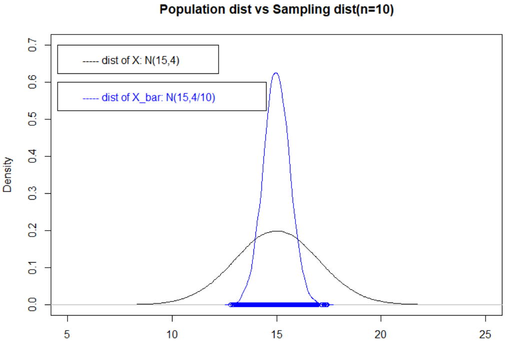
- Definition of Confidence Interval (CI)
- CI for parameters of Normal dist
    - CI for μ, (σ^2 known)
    - CI for μ, (σ^2 unknown)
    - CI for σ^2
- MLE based Confidence Intervals
- One-sided Confidence Intervals
- Few definitions related to CI and interpretation of CI

These are selected topics from
Evans and Rosenthal: chapter 6.3.2, 6.3.4, 6.5 and
John A. Rice: Chap 8.5.3

Section 1

Some revisions

Revisit: Population dist vs Sampling dist



- Each of the blue dots represents one value of $\bar{X}$ calculated based on one set of sample of size, n=10 from a $N(15, 4)$ distribution.
- If we increase the sample size (n) gradually, the blue density curve will get narrower and narrower. [Recall: $\bar{X} \sim N(\mu, \sigma^2/n)$]
- **Standard Error (SE):** the standard deviation of the blue curve
- $SE(\bar{X}) = \frac{2}{\sqrt{10}}$ for this example

Revisit some sampling distributions

- If $X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

- $X_1, X_2, ..., X_n$ *iid* from $f_{\theta_0}(x)$. Under some conditions and for $n \to \infty$

$$\frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{nI(\theta_0)}}} \xrightarrow{D} N(0,1)$$

Section 2

Definition of Confidence Interval

Definition of Confidence Interval (CI) (E&R - P326)

An interval $C(X_1, X_2, ..., X_n) = (l(X_1, X_2, ..., X_n), u(X_1, X_2, ..., X_n))$ is a γ-confidence interval for $\psi(\theta)$ if

$$P_\theta[\psi(\theta) \in C(X_1, X_2, ..., X_n)] \geq \gamma$$
$$\implies P_\theta[l(X_1, X_2, ..., X_n) \leq \psi(\theta) \leq u(X_1, X_2, ..., X_n)] \geq \gamma$$

for every $\theta \in \Omega$.

γ represents the confidence level of the interval.

In naive words, we want "two numbers" which will have at least γ chance of containing the true parameter.

Example explaining the definition of CI

- Assume the unknown parameter is μ
- Assume $\gamma = 0.95$
- We want an expression similar to this

$$P[l() \leq \mu \leq u()] \geq 0.95$$

- In most regular cases "$= 0.95$" interval is calculable.
- We need a tool that relates sample observations $(X_1, X_2, ..., X_n)$ to the parameter (μ) and finally allows calculating probability.
- This tool is called *Pivotal Quantity* or simply *Pivots*

Pivotal Quantity

Definition: A random variable defined in terms of the sample observations $X_1, X_2, ..., X_n$ is called a pivotal quantity

- if it involves the unknown parameters in it's expression
- but the distribution of this random variable does not depend on the parameters

The variables given on slide 6 are examples of Pivotal quantity.

Section 3

CI for parameters of Normal dist

Subsection 1

CI for μ, (σ^2 known)

CI for mean (μ) of Normal dist, σ^2 known

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

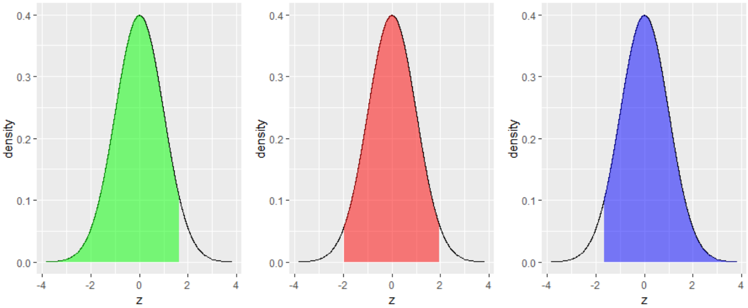
$$
\begin{aligned}
P\left[k_1 \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq k_2\right] &\geq 0.95 \\
\implies P[k_1 * \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq k_2 * \frac{\sigma}{\sqrt{n}}] &\geq 0.95 \\
\implies P[\bar{X} - k_2 * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - k_1 * \frac{\sigma}{\sqrt{n}}]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

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

$$P[k_1 \leq Z \leq k_2] \geq 0.95$$

where Z is a standard Normal variable.

choice of k_1 and k_2 assuming $\gamma = 0.95$



- In green one, $k_1 = -\infty$ and $k_2 = 1.65 \Longleftarrow$ (0.95 quantile)
- In red one, $k_1 = -1.96$ and $k_2 = 1.96$
- 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

Choice of k_1 and k_2 for any γ

- The sampling distribution is **unimodal and symmetric** around the mode, the middle γ part gives the shortest interval.
- $z_{(\frac{1-\gamma}{2})}$ and $z_{(\frac{1+\gamma}{2})}$ are preferred as the value of k_1 and k_2.
- Example: for $\gamma = 0.95 \implies \begin{cases} k_1 = z_{0.025} = -1.96 \\ k_2 = z_{0.975} = 1.96 \end{cases}$
- Finally, for $X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with σ^2 **known** we have the γ-CI of μ as

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

Example of CI for μ [Normal dist with known σ^2]

Exercise-6.3.1 (E&R):
$(4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) \overset{iid}{\sim} N(\mu, \sigma_0^2)$ with $\sigma_0^2 = 0.5$
Calculate the 0.95-confidence interval for μ.

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

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

Subsection 2

CI for μ, (σ^2 unknown)

CI for mean (μ) of Normal dist, σ^2 unknown

- When σ^2 is unknown, we use S^2 as an estimator of σ^2.
- Now we can't use $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ anymore.
- We use $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$
- We can use the same idea of slide 13-15
- For $X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ with σ^2 **unknown**
  we have the γ-CI of μ as

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

where, $t_{\frac{1+\gamma}{2}(n-1)}$ is the $\frac{1+\gamma}{2}$ quantile of a $t_{(n-1)}$ distribution.

Example of CI for μ [Normal dist with unknown σ^2]

Exercise-6.3.2 (E&R):
$(4.7, 5.5, 4.4, 3.3, 4.6, 5.3, 5.2, 4.8, 5.7, 5.3) \overset{iid}{\sim} N(\mu, \sigma^2)$ 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. $\gamma = 0.95 \implies \frac{1+\gamma}{2} = 0.975$
5. using t-table or R [*qt(0.975,df=9)*], $t_{0.975(9)} \approx 2.262$
6. 0.95-CI for μ:

$$4.88 \pm 2.262 * \frac{0.696}{\sqrt{10}} = (4.382, 5.378)$$

Subsection 3

CI for σ^2

CI for variance σ^2 of Normal distribution [E&R-P338]

- Recall, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$
- we can write,

$$P\left[\chi^2_{\frac{1-\gamma}{2}(n-1)} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\frac{1+\gamma}{2}(n-1)}\right] \geq \gamma$$

$$\implies P\left[\frac{\chi^2_{\frac{1-\gamma}{2}(n-1)}}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{\frac{1+\gamma}{2}(n-1)}}{(n-1)S^2}\right] \geq \gamma$$

$$\implies P\left[\frac{(n-1)S^2}{\chi^2_{\frac{1+\gamma}{2}(n-1)}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{1-\gamma}{2}(n-1)}}\right] \geq \gamma$$

- γ-level confidence interval:

$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{1+\gamma}{2}(n-1)}}, \frac{(n-1)S^2}{\chi^2_{\frac{1-\gamma}{2}(n-1)}}\right)$$

Example of CI of σ^2

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

1. $n = 10$
2. $\bar{x} = \frac{1}{10}(4.7 + 5.5 + ... + 5.3) = 4.88$
3. $(n-1)s^2 = \sum(x_i - \bar{x})^2 = \sum x_i^2 - n * (\bar{x})^2 = 4.356$
4. $\gamma = 0.95 \implies \frac{1-\gamma}{2} = 0.025$ and $\frac{1+\gamma}{2} = 0.975$
5. using χ^2-table or R, $\chi^2_{0.025(9)} \approx 2.7$ and $\chi^2_{0.975(9)} \approx 19.023$
6. 0.95-CI for μ:

$$\left(\frac{4.356}{19.023}, \frac{4.356}{2.7}\right) = (0.229, 1.613)$$

Comments on χ^2 based intervals

- χ^2 is not a symmetric distribution (at least for lower degrees of freedoms)
- It's shape depends on it's degrees of freedom.
- Using $\chi^2_{\frac{1-\gamma}{2}(n-1)}$ and $\chi^2_{\frac{1+\gamma}{2}(n-1)}$ as two ends may not result in the shortest length.

Section 4

MLE based CI for θ_0

CI for θ_0 using the asymptotic distribution of $\hat{\theta}$

Recall

For $n \to \infty$ we know $\frac{\hat{\theta}-\theta_0}{\sqrt{\frac{1}{nI(\theta_0)}}} \xrightarrow{D} N(0,1)$

- Using the same idea of slide 13-15, we can "write" that the γ-CI for θ_0 is

$$\left(\hat{\theta} - z_{\left(\frac{1+\gamma}{2}\right)}\sqrt{\frac{1}{nI(\theta_0)}}\,,\ \hat{\theta} + z_{\left(\frac{1+\gamma}{2}\right)}\sqrt{\frac{1}{nI(\theta_0)}}\right)$$

- **Question:** Can we use this for calculation? (why or why not?)

Estimate of the Fisher Information

- When Fisher Information involves the unknown parameter (θ_0) we can't use the expression on the previous page.
- We have two alternatives which give us an **estimate** of the Fisher information.

Plug-in estimate of Fisher Information

$nI(\hat{\theta}) = -E[\frac{\partial^2}{\partial\theta^2} \log f(X_1, X_2, ..., X_n|\theta)]\Big|_{\theta=\hat{\theta}}$

(In the expression of the Fisher information, replace θ by the mle, $\hat{\theta}$)

Observed Fisher Information (E&R page 364)

$$= -\frac{\partial^2}{\partial\theta^2} \log f(X_1, X_2, ..., X_n|\theta)\Big|_{\theta=\hat{\theta}}$$

(in the expression of the second-derivative of the negative log-likelihood replace θ by $\hat{\theta}$)

Estimate of the Fisher Information(cont...)

Though for the distributions that we have learned so far, both of these options produce same estimate(feel free to check, it will be a good practice), we will continue with the **plug-in estimate of Fisher Information**.

Using the *plug-in estimate* of Fisher Information, γ-level CI for θ_0 is

$$\left(\hat{\theta} - z_{\left(\frac{1+\gamma}{2}\right)}\sqrt{\frac{1}{nI(\hat{\theta})}}\,,\,\hat{\theta} + z_{\left(\frac{1+\gamma}{2}\right)}\sqrt{\frac{1}{nI(\hat{\theta})}}\right)$$

Example: CI for λ when data follows $Poisson(\lambda)$

- $\hat{\lambda} = \bar{X}$ is the MLE of λ
- Fisher Information, $nI(\lambda) = \frac{n}{\lambda}$
- Plug-in estimate, $nI(\hat{\lambda}) = \frac{n}{\bar{X}}$
- Finally, based on observed data the calculated γ-CI for λ is

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

$(4, 10, 10, 4, 6, 8, 8, 3, 4, 4) \overset{iid}{\sim} Pois(\lambda)$. Calculate 0.95-CI of λ

1. $\bar{x} = 6.1 \implies nI(\hat{\lambda}) = 10/6.1$
2. 0.95-CI of λ: $6.1 \pm 1.96 * \sqrt{6.1/10} \implies (4.569, 7.631)$

Section 5

One-sided Confidence Intervals

One-sided intervals

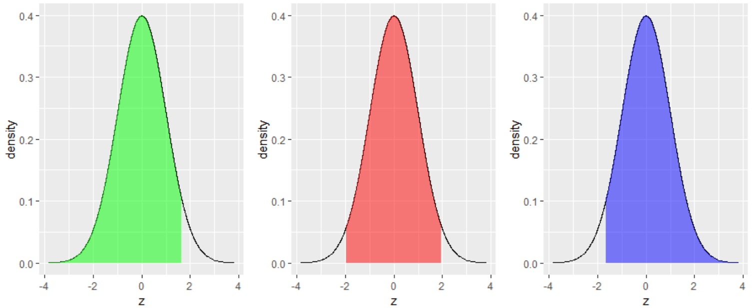
- Until now, all the intervals that have constructed some how represent the middle $\gamma * 100$ percent of the sampling distributions.
- These are called two-sided intervals (we are discarding both ends of the distribution)
- An one sided confidence interval looks like

$$P[-\infty \leq \psi(\theta) \leq u(X_1, X_2, ..., X_n)] \geq \gamma$$

or

$$P[l(X_1, X_2, ..., X_n) \leq \psi(\theta) \leq \infty] \geq \gamma$$

One-sided intervals (cont...)



- Left sided CI is represented by the green density.
- Right sided CI is represented by the blue density.

$(4, 10, 10, 4, 6, 8, 8, 3, 4, 4) \overset{iid}{\sim} Pois(\lambda)$.
Calculate **left sided** 0.95-CI of λ

1. $\bar{x} = 6.1 \implies nI(\hat{\lambda}) = 10/6.1$
2. left sided 0.95-CI of λ: $(-\infty, 6.1 + 1.65 * \sqrt{6.1/10} \implies (-\infty, 7.34)$

Section 6

Few definitions related to two-sided CI and interpreting CI

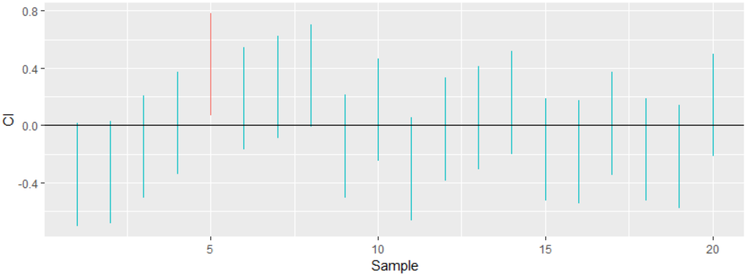
Few definitions related to CI (for two-sided z and t intervals)

1. For z and t interval, the sample mean $(\bar{x})$ is the midpoint of the lower and upper bound.
2. width of the interval = upper bound - lower bound.
3. Half of the width is known as the Margin of Error (ME).
4. CI: $[\bar{x} \pm \text{ME}]$
5. The width of the interval will increase as the confidence level (γ) increases. $(\gamma \uparrow \implies width \uparrow)$
6. The width of the interval will increase as the standard deviation (either σ or s) increases.
7. The width of the interval will decrease as the sample size (n) increases. $(n \uparrow \implies width \downarrow)$

Interpreting CI

- In slide 16, we got the 0.95-CI of μ as (4.442,5.318)
- Does it mean, $P[4.442 \leq \mu \leq 5.318] = 0.95$?
- Frequentist believe μ is a fixed number.
- Can we assign a probability statement to μ?

Interpreting CI (cont...)



- Generated 20 set of samples (each with size, n=30) from $N(0, 1)$
- Constructed the 0.95-CI for μ [just like slide 16, but 20 times]
- CIs are not fixed numbers rather random variables.
- 1 out of these 20 CIs missed the true mean ($\mu = 0$, the horizontal line)

Interpreting CI(cont...)

- **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 on slide 36)? - We don't know!
- In *Bayesian* school of thoughts, parameters are random variables. So assigning probabilities to a parameter is possible.

Assignment (Non-credit)

Evans and Rosenthal

Example: 6.3.7, 6.3.8, **6.3.16, 6.3.17**
Exercise(CI part of these ques): 6.3.1-6.3.4, 6.3.6, 6.3.8, 6.3.10, 6.3.12
R: 6.3.19, 6.3.21, 6.3.22
(CI part) 6.5.4, 6.5.5, 6.5.7, 6.5.8