Dis 8: Intro to Estimation

Related textbook chapter: 10
Ch 10 handout and solution offered by Dr. Pac can be accessed here: Handout Solution
This handout incorporates reviews with all exercises from the handout given by Dr. Pac.

1 Motivation

- In our last discussion before Spring break, we talked about sampling distributions, which describe how sample statistics are distributed.
- Recall that our goal is to perform statistical inference: use sample statistic to draw conclusion on population parameter.
- We are finally going to connect the pieces:
  - The sample statistics of interest from last week are **point estimators**. A point estimator takes a best guess at the true value of an underlying population parameter.
  - Sometimes, one might instead want to estimate a range that's likely to include the true population parameter. The estimator that provides such a range is called an **interval estimator**.
- Sorting through some terminologies:
  - An estimator (point or interval) tries to estimate (a point value or a range of) the corresponding true population parameter.
  - An estimator follows a sampling distribution.
  - A population parameter follows a probability distribution.

2 Point Estimator

- Definition: a point estimator takes a (single) best guess at the true value of an underlying population parameter.
- Examples of point estimator:
  - $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a point estimator used to estimate population mean μ_X.
  - X_1 (the first sample observation) is also a point estimator, which can be used to estimate population mean μ_X.
- Begs the question: how do you evaluate if an estimator is "good"?

  $\Rightarrow$ use the following three criteria:

  1. **Unbiased**: an estimator is unbiased if

  $$E[\text{estimator}] = \text{population parameter}$$

  2. **Relatively efficient**: an estimator is relatively efficient if, compared to another estiamtor with the same amount of bias, it has lower variance. That is, if

  $$E[\text{estimator}_a] = E[\text{estimator}_b]$$

Then estimator$_a$ is relatively efficient if

$$V(\text{estimator}_a) < V(\text{estimator}_b)$$

3. **Consistent**: an estimator is consistent if the following two hold:
   (a) Asymptotically unbiased: $E[\text{estimator}] \to$ population parameter as $n \to \infty$, and
   (b) $V(\text{estimator}) \to 0$ as $n \to \infty$

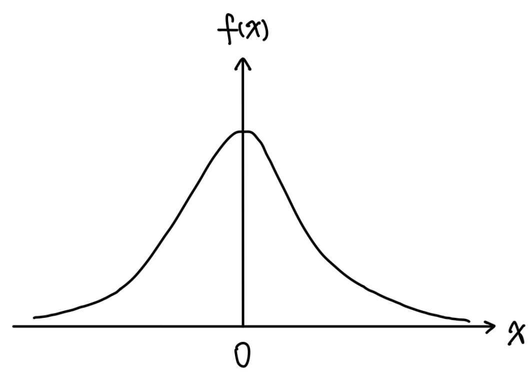
[Go to Exercise 1 and 2]

3 Interval Estimator: Confidence Interval

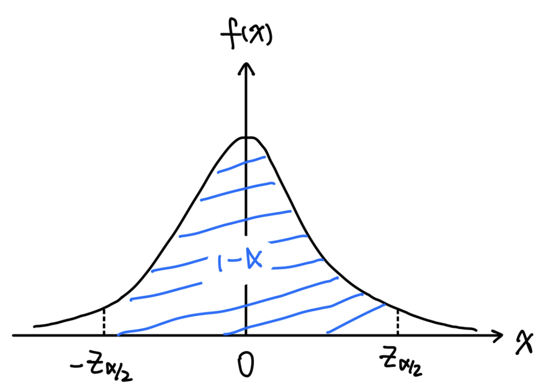
- Definition: an interval estimator estimates a range that's likely to include the true population parameter.
- One interval estimator that we often look at: **confidence interval**

3.1 Construct a confidence interval

- Think about a standard normal distribution:



- If we want to cover $(1-\alpha)$ portion of this standard normal distribution, then



- We can think about this standard normal distribution as the sampling distribution of the mean, where the sample mean estimator has been standardized:

$$\begin{aligned}
& P\left(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\right) && = 1-\alpha \\
\Leftrightarrow\quad & P\left(-Z_{\alpha/2} \leq \frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \leq Z_{\alpha/2}\right) && = 1-\alpha \\
\Leftrightarrow\quad & P\left(-Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \leq \bar{X}-\mu_X \leq Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) && = 1-\alpha \\
\Leftrightarrow\quad & P\left(\bar{X}-Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \leq \mu_X \leq \bar{X}+Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) && = 1-\alpha
\end{aligned}$$

  Thus, for $(1-\alpha)$ portion of area covered, the confidence interval constructed is

$$\left[\bar{X}-Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{X}+Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right]$$

  We call $(1-\alpha)$ the **confidence level** for the above interval.

- What are some common confidence level and the associated Z score ($Z_{\alpha/2}$)?

Confidence level	α	$Z_{\alpha/2}$
90%	0.1	1.645
95%	0.05	1.96
99%	0.01	2.575

- **Interpretation**:

  Say that, for example, a 95% confidence interval of the mean of X, using a sample of size 70, is estimated to be $[4, 8]$. The following are some examples of correct interpretation of this confidence interval constructed.

  - **Correct version 1**: There's a 5% probability that the population mean of X lies outside of the confidence interval <u>estimator</u>. For this sample of size 70, we <u>estimate</u> the confidence interval to be $[4, 8]$.
  - **Correct version 2**: If random sample of size 70 were <u>repeatedly selected</u>, then in the long run, 95% of the confidence intervals formed would contain the true mean of X, which <u>in this case</u> is between 4 and 8.

[Go to Exercise 3]

3.2 Sample size needed given a already constructed confidence interval and confidence level

- We just saw that a confidence interval with $(1-\alpha)$ confidence level is constructed to be

$$\left[\bar{X}-Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{X}+Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right]$$

In other words, the lower and upper bound of this confidence interval is calculated to be

$$\bar{X} \mp Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}$$

- Say that instead, we want to specify how tight the confidence interval is. Usually, we do this by specifying a bound (B), which is the value that is subtracted from or added to the $\bar{X}$. That is, we want the lower and upper bound of a confidence interval to be calculated as

$$\bar{X} \mp B$$

- This implies that

$$B = Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}$$

- Using this expression, we have chosen what B is. Often times, people also have in mind of what they want the confidence level to be (i.e. α is chosen), and σ_X is given. Thus, in order to set the bound as B, one can specify the sample size n:

$$\begin{aligned} B &= Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \\ \sqrt{n} &= \frac{Z_{\alpha/2} \cdot \sigma_X}{B} \\ n &= \left(\frac{Z_{\alpha/2} \cdot \sigma_X}{B}\right)^2 \end{aligned}$$

- The sample size n obtained in this way is, more appropriately speaking, a lower bound, since a bigger n always shrinks the variance of the sample mean, meaning that the bound can be even tighter if needed.

  Thus, in order to achieve bound B under some α and σ_X, one needs sample size

$$n \geq \left(\frac{Z_{\alpha/2} \cdot \sigma_X}{B}\right)^2$$

[Go to Exercise 4]

4 Exercises

1. Let $\{X_i, i = 1, \ldots, n\}$ be a simple random sample drawn from a population with mean μ and variance σ^2.

   (a) Is $\tilde{X} = \frac{1}{n+1}\sum_{i=1}^{n} X_i$ a biased or unbiased estimator of μ?

(b) Can we conclude that $\tilde{X}$ is relatively efficient compared to the sample mean $\bar{X}$?

(c) Is $\tilde{X}$ a consistent estimator?

2. Suppose you have a sample of size n drawn from a population with mean μ and variance σ^2. Which estimator for μ is more efficient: X_1 (the first observation), or $\bar{X}$?

3. Suppose you draw a sample from a population with a standard deviation of 25. You draw 50 observations and end up with a sample mean of 100.

   (a) Estimate a 90% confidence interval for the population mean

   (b) Estimate a 95% confidence interval for the population mean

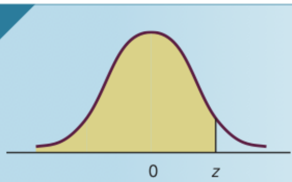
   (c) Estimate a 99% confidence interval for the population mean

   (d) What effect does increasing the confidence level have on the resulting confidence interval?

(e) Carefully interpret your confidence interval from part (a)

4. You would like to produce a confidence interval for the mean output of a new telephone production line. Output has a population standard deviation of 15 phones, and you would like to estimate the confidence interval to within plus or minus 2 phones (with 95% confidence). How many observations do you need?

TABLE **3** ***(Continued)***



$P(-\infty < Z < z)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990