IEDA 2540 Statistics for Engineers
Interval Estimation

Wei YOU

香港科技大學
THE HONG KONG UNIVERSITY OF
SCIENCE AND TECHNOLOGY

Spring, 2025

Introduction

Previously, we talked about

- possible ways to evaluate an estimator: error, bias, variance, and MSE, etc.; and
- two general methods for obtaining **point estimates**: MoM and MLE.

Despite the ease of use, the main problem of point estimators is that they don't convey information about the *uncertainty* or *reliability* of the estimate.

- A point estimator does not reflect the level of "confidence" we have.

Point Estimators May Not Equal to the Parameter with High Probability

No matter how small MSE is, there is no reason to expect a point estimate to be exactly the same as the parameter it estimates. (Because of the randomness in the random sample and thus the estimator.)

Example: Let p be the probability of a head from a coin toss:

- Suppose we observe 2 flips, then the MLE is $\widehat{p} = X/2$.
- Note that $\mathbb{P}(\widehat{p} = 0) = 1/4$, $\mathbb{P}(\widehat{p} = 1/2) = 1/2$, and $\mathbb{P}(\widehat{p} = 1) = 1/4$.
- What is the probability that $\widehat{p} = p$?
  - If the unknown p happens to be one of 0, $1/2$, or 1, then the probability $\mathbb{P}(\widehat{p} = p)$ is positive. Yet, if we take one more observation, $\widehat{p}$ only take values 0, $1/3$, $2/3$, or 1.
  - What if p does not equal 0, $1/2$, or 1? Then the probability is 0!
  - We cannot expect $\mathbb{P}(\widehat{p} = p)$ to stay positive for all sample sizes!

Point Estimators Does Not Reflect the Level of Confidence

How "confident" are we in a point estimate?

Example: Let p be the probability of a head from a coin toss:

- Case 1: Observe 3 flips, 2 heads and 1 tail.
- Case 2: Observe 300 flips, 200 heads and 100 tails.
- In both cases, the point estimate is $\widehat{p} = 2/3$.

Would you be more confident in $\widehat{p} = 2/3$ under Case 1 or Case 2?

Introduction

Suppose that, instead of a single value (point estimate), we report **an interval** within which we *expect to find the true parameter with high probability*.

- Such an interval is called a **confidence interval (CI)** or an **interval estimator**.
- Compared with a single value, an interval has a much higher chance to contain the true parameter.

Example: Let p be the probability of a head from a coin toss:

- Observe 3 flips, 2 heads and 1 tail.
- Can you propose an interval that you believe p will fall in with high probability?

Example: Bernoulli

For example, you may propose the interval $[0, 1]$. However,

- Although you are certain that this interval will cover the true p,
- $[0, 1]$ provides no additional information regarding the value of p beyond what we already know even without the data.

Based on what we will learn in this topic, we have:

- **Example:** **(Case 1):** Observe 3 flips, 2 heads and 1 tail. We are "95% sure" that the probability of getting a head is within $[0.133, 1.120]$.
- **Example:** **(Case 2):** Observe 300 flips, 200 heads and 100 tails. We are "95% sure" that the probability of getting a head is within $[0.613, 0.720]$.

What exactly do we mean by "**95% sure**"? How can we construct such an interval based on our observations?

Example: Normal

Suppose we take four observations from a $\mathcal{N}(\mu, 1)$ population, say, $\boldsymbol{X} = X_1, X_2, X_3, X_4$, and we wish to estimate μ. A good point estimator for μ is the sample mean $\bar{X}$. An intuitive way to construct an interval estimate is to *provide a range of values around the point estimator*.

- For example, we may propose the interval $\bar{X} - 1,\ \bar{X} + 1$.

Consider the following cases:

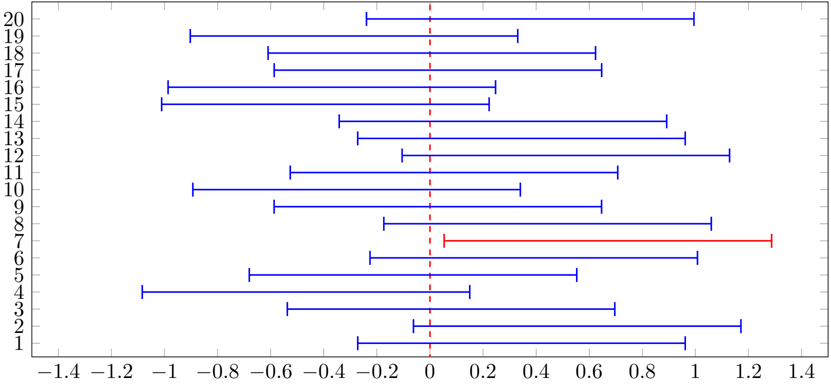
- Suppose the observations are $x = -0.492, 0.732, 1.180, -0.395$. Then, $\bar{x} = 0.256$ and the interval is $\bar{x} - 1,\ \bar{x} + 1 = [-0.744,\ 1.256]$.
- Suppose the observations are $x = -0.444, -0.901, -0.006, -0.040$. Then, $\bar{x} = -0.348$ and the interval is $\bar{x} - 1,\ \bar{x} + 1 = [-1.348,\ 0.652]$.

Intervals such as $[-0.744,\ 1.256]$ and $[-1.348,\ 0.652]$ (with specific numeric values) are called the **observed confidence intervals**.

Example: Normal, cont'd

For each repetition of the experiment, we calculate the observed CI from the data.

For each observed CI, we can check whether the interval contains the true mean μ.



Measuring the reliability of the CI

To measure the reliability of the interval, we look at the success rate (i.e., the proportion of intervals that cover μ) of the CI.

- This success rate is called the **confidence level** of a CI.
- **Reliability**: if we repeat the experiment many times, $\sim$95% of the resulting intervals would contain the true μ, while only $\sim$5% would fail to cover μ.

Purpose of Interval Estimation

- **Measure of uncertainty:** Confidence intervals provide a range around the point estimate that likely contains the true population parameter. This range reflects the uncertainty associated with the point estimate, which is crucial for assessing its reliability.
- **Indication of precision:** The width of a confidence interval gives insight into the precision of the estimate.
  - Narrower intervals suggest a more precise estimate.
  - Wider intervals indicate less precision.

Purpose of Interval Estimation

- **Transparent reporting:** Reporting confidence intervals alongside point estimates promotes transparency in research and data analysis. It provides a fuller picture of the findings, allowing others to assess the reliability and applicability of the results.
  Example:
  - If you propose a financial portfolio and report an annual return of 20%, is this sustainable? Or is it simply due to a bullish market?
  - Suppose you also report a confidence interval of $[-50\%, 40\%]$ with a 0.95 confidence level; then we know that a positive return is not guaranteed.
  - Suppose you report a confidence interval of $[15\%, 27\%]$ with a 0.95 confidence level; then we know that the portfolio is performing well.

Purpose of Interval Estimation

- **Guidance for decision making:** In practical applications, knowing just the point estimate might not be enough for making informed decisions. Confidence intervals provide a range of plausible values for the parameter, which can be crucial for decision-making in fields like medicine, policy-making, and business.
  Example: If we estimate the hourly number of arrivals to an Emergency Department to be $[10, 20]$, then we may hire doctors based on the upper confidence bound, 20, to ensure fast service.

Confidence Interval

Definition (Confidence interval)

An interval estimate of a real-valued parameter θ is any *pair of statistics*, L and U, of a random sample that satisfies $L \leq U$. The random interval $[L, U]$ is called an **interval estimator**.

- Being statistics, the interval is calculated from the data and is therefore random. To be precise, we can write $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$.
- Once $\boldsymbol{X}$ is observed, we compute $l = L(\boldsymbol{X})$ and $u = U(\boldsymbol{X})$. Then the inference $l \leq \theta \leq u$ is made.

The endpoints l and u are called the **lower-** and **upper-confidence limits (bounds)**.

The goal is to provide $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$ such that:

- $L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})$ occurs with high probability (reliable!),
- and the length of the interval, $U(\boldsymbol{X}) - L(\boldsymbol{X})$, is as short as possible (precise!).

Confidence Level – The Reliability of a CI

Definition (Confidence level/coverage Probability)

The **coverage probability** of an interval estimate is the probability that the random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ covers the true parameter θ.

$$\mathbb{P}_\theta\Big(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})]\Big)$$

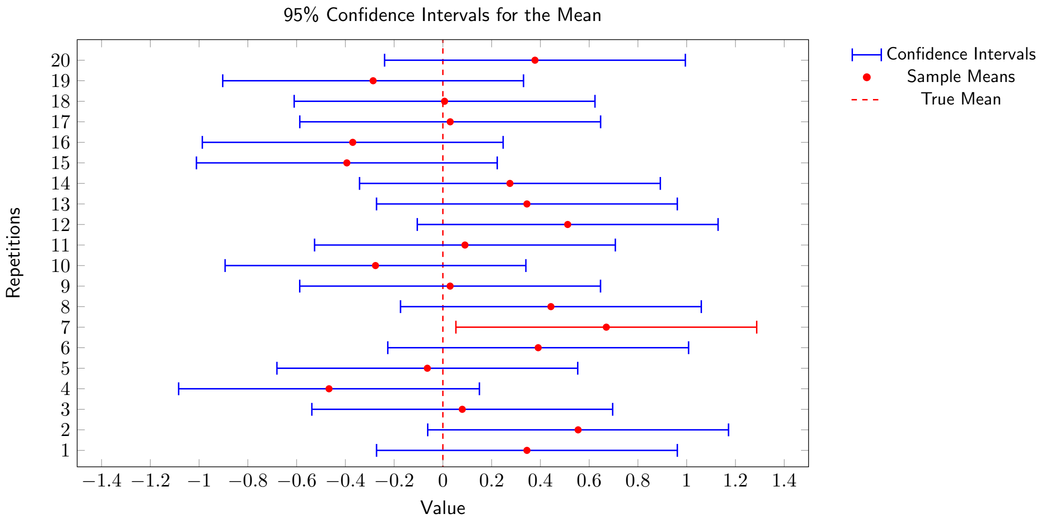
This is also called the **confidence level**, usually written as $(1-\alpha) \times 100\%$.

$(1-\alpha) \times 100\%$ Confidence Interval

The interval estimator with a $(1-\alpha) \times 100\%$ confidence level is called a $(1-\alpha) \times 100\%$ confidence interval.

- The parameter θ is considered unknown and deterministic.
- The CI (the pair of statistics $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$) is random because it depends on the sample $\boldsymbol{X}$.
- We can think of the confidence level as the probability that the deterministic parameter θ falls into the random interval.

Example: A 95% confidence interval means that if we collect 20 random samples from the population, roughly one sample would yield an interval that does NOT cover the unknown parameter.


95% Confidence Intervals for the Mean
Confidence Intervals
Sample Means
True Mean
Repetitions
Value
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
−1.4 −1.2 −1 −0.8 −0.6 −0.4 −0.2 0 0.2 0.4 0.6 0.8 1 1.2 1.4

Interval Estimator

- If both $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$ are finite-valued, the interval called two-sided.
- In some cases, we only care about bounds on one side.
  Example: For light bulbs, quality control department care about the lower bound of the lifespan, $[L(\boldsymbol{X}), \infty)$. I.e., with 95% confidence, the mean lifespan of the light bulb can burn is at least 5 years.
  Example: In clinical trials, we wish to understand the upper bound on the average time that a vaccine takes effect, $(-\infty, U(\boldsymbol{X})]$.
- These are called one-sided confidence intervals, $(-\infty, U(\boldsymbol{X})]$ or $[L(\boldsymbol{X}), \infty)$.

Example: Normal

For a sample $\{X_1, X_2, X_3, X_4\}$ from $N(\mu, 1)$, we wish to estimate μ and propose the interval

$$[\bar{X} - 1, \bar{X} + 1]$$

as an interval estimator.

Question: Is this interval reasonable?

To assess the reliability of this interval we calculate its confidence level:

$$\begin{aligned}
\mathbb{P}\Big(L(\boldsymbol{X}) \le \mu \le U(\boldsymbol{X})\Big) &= \mathbb{P}\Big(\bar{X} - 1 \le \mu \le \bar{X} + 1\Big) \\
&= \mathbb{P}\Big(-1 \le \bar{X} - \mu \le 1\Big) = \mathbb{P}\Big(\frac{-1}{\sqrt{\sigma^2/n}} \le \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \le \frac{1}{\sqrt{\sigma^2/n}}\Big) \\
&= \mathbb{P}\Big(\frac{-1}{\sqrt{1/4}} \le Z \le \frac{1}{\sqrt{\sigma^2/n}}\Big) \approx 0.9544.
\end{aligned}$$

Example: Uniform

For a sample $X_1, \ldots, X_n$ from $\mathrm{Unif}(0, \theta)$. We wish to construct a CI for θ.

- **An intuitive way to construct an interval estimation is to provide a range of values around a good point estimator.**
- The MLE of θ is $Y = X_{(n)} = \max_i X_i$.

We can use $[aY, bY]$ for some constants $1 \le a \le b$ as a interval estimator for θ. Then the confidence coefficient is

$$\mathbb{P}_\theta(\theta \in [aY, bY]) = \mathbb{P}_\theta\left(\frac{1}{b} \le \frac{Y}{\theta} \le \frac{1}{a}\right) = \frac{1}{a^n} - \frac{1}{b^n}.$$

Let $a = 1$ and b be a constant slightly larger than 1. Then the confidence coefficient is close to 1 when the sample size n is large.

When the Coverage Probability depends on θ

Example: Consider the same sample $X_1, \ldots, X_n$ from $\text{Unif}(0, \theta)$. We can also use $[Y + c, Y + d]$ for some positive constants $c \leq d$ as an interval estimator for θ.

$$\mathbb{P}_\theta(\theta \in [Y + c, Y + d]) = \mathbb{P}_\theta\left(1 - \frac{d}{\theta} \leq \frac{Y}{\theta} \leq 1 - \frac{c}{\theta}\right) = \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n$$

*Above holds when $\theta \geq d$.

Observations

- The coverage probability can depend on the parameter θ.
- As θ gets large, the coverage probability decreases to 0!

Reference: Distribution of $X_{(n)}$

- Distribution of $X_{(n)}$

$$F_{X_{(n)}}(x) = \mathbb{P}\Big(X_{(n)} \leq x\Big) = \mathbb{P}\Big(\max\{X_1, X_2, \dots, X_n\} \leq x\Big).$$

- Since the X_i are independent,

$$F_{X_{(n)}}(x) = \mathbb{P}(X_1 \leq x,\ X_2 \leq x,\ \dots,\ X_n \leq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) = \Big[F_\theta(x)\Big]^n.$$

- Here, $F_\theta(x)$ is the CDF of the Uniform$(0,\theta)$ distribution:

$$F_\theta(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x}{\theta}, & 0 \leq x \leq \theta, \\ 1, & x > \theta. \end{cases}$$

When the Coverage Probability depends on θ

When the coverage probability depends on the parameter θ, we use a more robust concept of the confidence coefficient, to *guard against worst case confidence level over the possible range of parameters*:

Definition (Confidence Coefficient)

The confidence coefficient of an interval estimate is is the infimum of the coverage probabilities.

$$1 - \alpha = \inf_{\theta \in \Theta} \mathbb{P}_\theta(\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})])$$

α here is called the **significance level**.

Example: Consider the same sample $X_1, \ldots, X_n$ from $\mathrm{Unif}(0, \theta)$. The confidence level of $[Y + c, Y + d]$ is 0%!

Case 1: CI for Normal Mean μ with Known Variance

Let $X_1, \ldots, X_n$ be i.i.d. random variables from $\mathcal{N}(\mu, \sigma^2)$.

- Suppose σ^2 is known, so that it can be used in your estimators.
- We are interested in estimating μ.
- The goal is to find a $(1-\alpha) \times 100\%$ confidence interval.

How to find such an interval?

- The core idea is to start with a good point estimator and expand it to a range.
- Suppose we propose the interval $\bar{X} - c, \bar{X} + c$, where $\bar{X}$ is the sample mean and c is a constant *to be determined*.
- We seek to find a constant c such that

$$\mathbb{P}\Big(\bar{X} - c \leq \mu \leq \bar{X} + c\Big) = 1 - \alpha.$$

Case 1: CI for Normal Mean μ with Known Variance

To calculate the coverage probability $\mathbb{P}\Big(\bar{X} - c \leq \mu \leq \bar{X} + c\Big)$, we need to know the distribution of the sample mean $\bar{X}$.

- **Distribution of the Sample Mean:**

$$\bar{X} \sim \mathcal{N}\Big(\mu, \frac{\sigma^2}{n}\Big), \quad \text{hence} \quad Z \equiv \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

- **Coverage Probability:** We express the probability that μ lies in the interval $\bar{X} - c, \bar{X} + c$ as

$$\mathbb{P}\Big(\bar{X} - c \leq \mu \leq \bar{X} + c\Big) = \mathbb{P}\Big(-c \leq \bar{X} - \mu \leq c\Big) = \mathbb{P}\Big(-\frac{c}{\sigma/\sqrt{n}} \leq Z \leq \frac{c}{\sigma/\sqrt{n}}\Big)$$
$$= \Phi\Big(\frac{c}{\sigma/\sqrt{n}}\Big) - \Phi\Big(-\frac{c}{\sigma/\sqrt{n}}\Big)$$

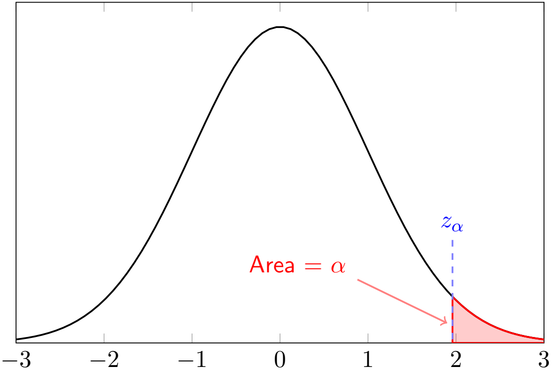
Critical Value for Standard Normal

Critical value/z-score

Let Z be a standard normal random variable. We define z_α to be the value such that

$$\mathbb{P}(Z > z_\alpha) = \alpha.$$

- An α proportion of observations fall **above** z_α.



Example: $z_{0.05} = 1.645,\ z_{0.025} = 1.96,\ z_{0.005} = 2.576.$

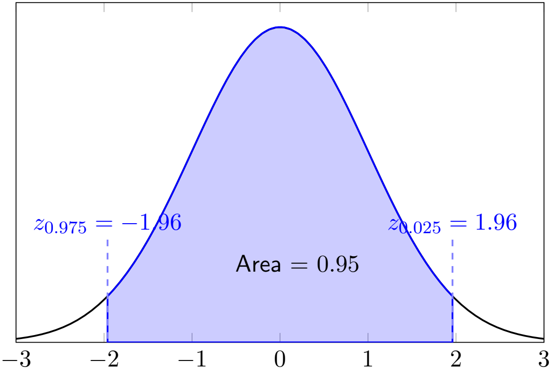
Determining the Constant c

We wish to find a constant c such that

$$\mathbb{P}\left(-\frac{c}{\sigma/\sqrt{n}} \leq Z \leq \frac{c}{\sigma/\sqrt{n}}\right) = 1 - \alpha.$$

The relevant critical values from the standard normal distribution are $z_{\alpha/2}$ and $z_{1-\alpha/2}$. By symmetry, we have

$$z_{\alpha/2} = -z_{1-\alpha/2}.$$



Hence, we set

$$\frac{c}{\sigma/\sqrt{n}} = z_{\alpha/2} \quad \Rightarrow \quad c = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

CI for Normal Population Mean with Known Variance

CI for μ with known σ

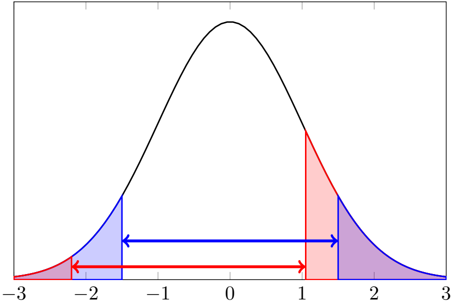
If $\bar{x}$ is the *observed* sample mean of a random sample with size n and *known* variance σ^2, a $(1-\alpha)\times 100\%$ confidence interval on μ is given by

$$\left[\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}\right]$$

- Notice that we used a lower-case $\bar{x}$ to denote the realized value, instead of the random statistic.
- The CI requires knowledge of the variance σ^2. We will see later how to deal with the unknown variance.
- We constructed this CI by finding a symmetric interval around the sample mean $\bar{X}$ such that the coverage probability is $1-\alpha$. **Why a symmetric interval?**

Why Symmetric Interval?

- The goal is to provide CI with a guaranteed coverage probability (i.e., the interval is reliable).
- At the same time, we want the interval length, $U(\boldsymbol{X}) - L(\boldsymbol{X})$, to be as short as possible (i.e., the information is more precise).
- Given a fixed coverage probability $1 - \alpha$, a symmetric interval around the sample mean yields the shortest possible interval.



Notice that $\text{Red Area} = \text{Blue Area} = \alpha$, but **the symmetrical blue interval is shorter**, hence is preferred over the red.

Discussion: the Width of the CI

The width of the confidence interval is given by

$$2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

- σ: Assumed to be known. A larger variance implies a wider CI.
- α: Related to the confidence level (usually given). A smaller significance level α (i.e., a higher coverage confidence level $1-\alpha$) implies a wider CI.
- n: Sample size, chosen by the experimenter. A larger sample size implies a narrower CI.

Example:

Suppose we send a signal of value μ from A to B. The destination B receives $\mu + Z$, where $Z \sim \mathcal{N}(0, 4)$. The signal is sent 9 times and we receive:

$$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5.$$

Then, $\bar{X} = 9, \sigma^2 = 4$, $n = 9$, and $z_{0.05} = 1.645, nz_{0.025} = 1.96, z_{0.005} = 2.576$.

- The 99% confidence interval for μ is:

$$\bar{x} \pm z_{0.005} \cdot \sigma^2/\sqrt{n} = 9 \pm 2.576 \cdot \frac{2}{3} \approx (9 - 1.717, 9 + 1.717) \approx (7.28,\ 10.72).$$

- The 95% confidence interval for μ is:

$$\bar{x} \pm z_{0.025} \cdot \sigma^2/\sqrt{n} = 9 \pm 1.96 \cdot \frac{2}{3} \approx (9 - 1.307, 9 + 1.307) \approx (7.69,\ 10.31).$$

- The 90% confidence interval for μ is:

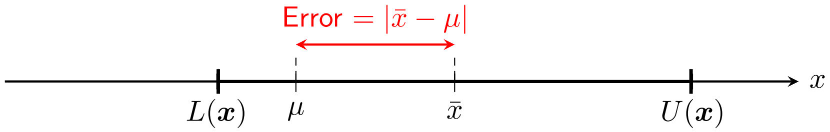
$$\bar{x} \pm z_{0.05} \cdot \sigma^2/\sqrt{n} = 9 \pm 1.645 \cdot \frac{2}{3} \approx (9 - 1.097, 9 + 1.097) \approx (7.91,\ 10.09).$$

Controlling the Width of the Confidence Interval

The width of the CI is given by $2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. To reduce the width of the CI, we can:

- Decrease σ (i.e., reduce variability in the data), which is not always possible.
- Increase the sample size n.

We say that the sample mean estimate has an error less than $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ with $(1-\alpha)\times 100\%$ confidence. **This is because the error is less than the half-width of the CI if and only if μ is covered by the CI.**



Choice of Sample Size

Question

How many observations do we need so that the estimation error is smaller than some tolerance ε with $1-\alpha$ confidence?

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \varepsilon \Rightarrow n \geq \left\lceil \left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2 \right\rceil$$

where $\lceil x \rceil$ is the ceiling function, i.e., the smallest integer that is $\geq x$.

- As confidence level increase (or as α decrease), we need more sample.
- As error tolerance decrease, we need more sample.
- As variance increase, we need more sample.
- The sample size does not depend on the center of the data, i.e. $\bar{x}$.

Example: Mean Weight of Salmon

Suppose the weight of salmons from a farm is normally distributed with a known standard deviation $\sigma = 0.3$ pounds. We wish to estimate the mean μ from a sample.

If we want to be 95% confident that our estimate is correct to within ± 0.1 pound, how large a sample is needed?

- Since we want to be 95% confident, we have $1 - \alpha = 0.95 \Rightarrow \alpha/2 = 0.025$.
- The critical value is $z_{0.025} = 1.96$.
- The margin of error is given by $z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 0.1$.
- Solving for n, we obtain:

$$\sqrt{n} \geq \frac{1.96 \times 0.3}{0.1} \Rightarrow n \geq \left(\frac{1.96 \times 0.3}{0.1}\right)^2 \approx 34.57.$$

- Thus, we need $n \geq 35$.

One-sided Confidence Interval

In the previous case, we constructed two-sided confidence intervals, i.e., both $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$ are finite.

- If we care only about one side, we may also construct a one-sided confidence interval, e.g., $[L(\boldsymbol{X}), \infty)$ or $(-\infty, U(\boldsymbol{X})]$.

Example: (Confidence lower bound) We consider an intuitive choice $L(\boldsymbol{X}) = \bar{X} - c$.

- Given a confidence level $1 - \alpha$, we choose c such that

$$\mathbb{P}\Big(\mu \in [\bar{X} - c, \infty)\Big) = 1 - \alpha.$$

- Note that

$$\mathbb{P}\Big(\mu \in [\bar{X} - c, \infty)\Big) = \mathbb{P}\Big(\bar{X} - \mu \leq c\Big) = \mathbb{P}\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{c}{\sigma/\sqrt{n}}\right).$$

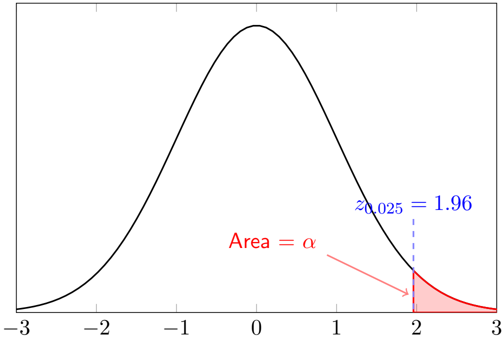
- In contrast to the two-sided CI, here we only care about one side of the inequality.

We wish to choose c so that

$$\mathbb{P}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{c}{\sigma/\sqrt{n}}\right) = 1-\alpha \quad \Leftrightarrow \quad \mathbb{P}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{c}{\sigma/\sqrt{n}}\right) = \alpha.$$

We then set

$$\frac{c}{\sigma/\sqrt{n}} = z_\alpha \quad \Leftrightarrow \quad c = z_\alpha \frac{\sigma}{\sqrt{n}}.$$



One-sided CI for μ with known σ

If $\bar{x}$ is the observed sample mean of a random sample with size n and known variance σ^2, a $(1-\alpha) \times 100\%$ **one-sided confidence lower bound** on μ is given by

$$\left[\bar{x} - z_{\alpha}\sigma/\sqrt{n}, \infty\right).$$

A $(1-\alpha) \times 100\%$ **one-sided confidence upper bound** on μ is given by

$$\left(-\infty, \bar{x} + z_{\alpha}\sigma/\sqrt{n}\right].$$

Case 2: Normal Population Mean with Unknown Variance

Previously, we constructed CIs under the assumption of known variance, because our CIs (e.g., $\left[\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}\right]$) require σ^2.

- If the variance is unknown, we can no longer compute these CIs.
- **A natural idea:** estimate the variance using the sample variance S^2 and plug it into our previous CI. For example, one might consider

$$\left[\bar{X} - c\frac{S}{\sqrt{n}}, \bar{X} + c\frac{S}{\sqrt{n}}\right].$$

However, the key step in constructing the CI is to calculate the exact coverage probability, i.e., we need to determine

$$\mathbb{P}\left(\bar{X} - c\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + c\frac{S}{\sqrt{n}}\right) = \mathbb{P}\left(-c \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq c\right).$$

- **Question:** What is the distribution of $\frac{\bar{X}-\mu}{S/\sqrt{n}}$?

The t-Statistic

Let us standardizes the sample mean using the sample standard deviation:

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

In contrast, when variance is known, we have

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

- Although T differs from Z, *they should not be too different*. Since S^2 is a consistent estimator of σ^2, it will converge to σ^2 as the sample size increases, making T and Z nearly indistinguishable for large samples.
- With a large sample size, we may consider $T \approx Z$.
- The main differences appear in moderate to small samples.

It turns out that T follows exactly the so-called Student's t-distribution with degrees of freedom $\nu = n - 1$.

Student's t distribution

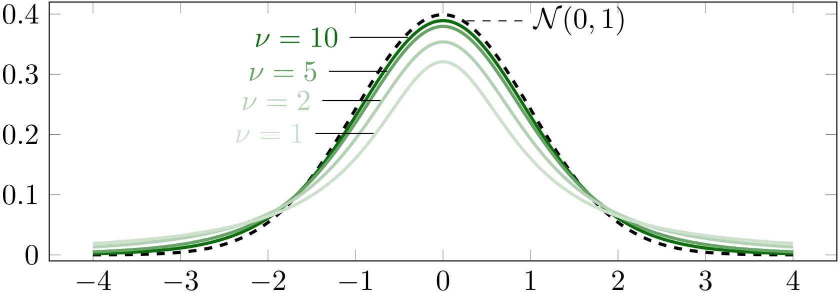
Let

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

Then, T has a **Student's t distribution with degree of freedom** $n - 1$.

- A t-distribution has one parameter, $\nu = n - 1$, and is typically denoted as t_ν.
- The distribution is named after William Sealy Gosset's 1908 paper in *Biometrika*, published under the pseudonym "Student".

PDF of the t distribution

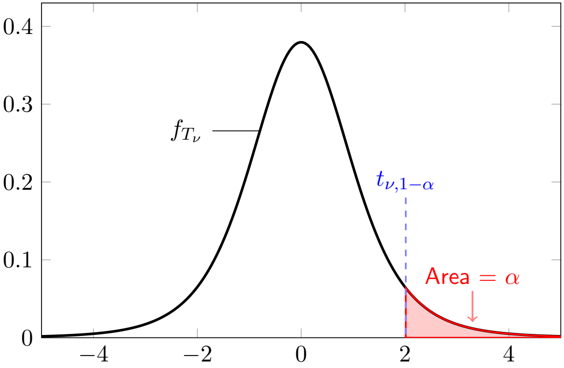


- Its pdf has a "bell" shape, similar to normal.
- As the sample size n grows to ∞, the t distribution is closer and closer to a standard normal distribution. Why?
- It has "heavier tail" than normal distributions, so it will more frequently produce outliers. This is due to the additional randomness introduced by the need to use sample variance.

Critical value for t distribution

Let T be a Student's t distribution with degree of freedom n. We define the critical value (also called the t-score) associated with the desired significance level α to be the value $t_{n,\alpha}$ such that

$$\mathbb{P}(T > t_{n,\alpha}) = \alpha.$$



- An α proportion of observations fall **above** $t_{\nu,1-\alpha}$.
- To find out the value of $t_{n,\alpha}$, use `scipy.stats.t.ppf(1-alpha,DoF)` in Python.

Case 2: Normal Population Mean with Unknown Variance

We can write the coverage probability of the CI $\left[\bar{X} - c\frac{S}{\sqrt{n}}, \bar{X} + c\frac{S}{\sqrt{n}}\right]$ with $c = t_{n-1,\alpha/2}$ as

$$\mathbb{P}\left(\bar{X} - c\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + c\frac{S}{\sqrt{n}}\right) = \mathbb{P}\left(-t_{n-1,\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{n-1,\alpha/2}\right)$$
$$= \mathbb{P}(-t_{n-1,\alpha/2} \leq T \leq t_{n-1,\alpha/2}) = \alpha.$$

CI for μ with unknown σ

If $\bar{x}$ is the observed sample mean of a random sample with size n and unknown variance, a $(1-\alpha) \times 100\%$ confidence interval on μ is given by

$$\left[\bar{x} - t_{n-1,\alpha/2} s/\sqrt{n}, \bar{x} + t_{n-1,\alpha/2} s/\sqrt{n}\right]$$

Notice that we used a lower-case $\bar{x}$ and s to denote the realized values, instead of the random statistics.

Example on Slide 29

- **Data:** $\bar{X} = 9$, $S^2 = 9.5$, and $n = 9$.
- **Known Variance:** Assume $\sigma^2 = 4$ (so $\sigma = 2$). Then, the 95% CI for μ is

$$9 \pm 1.96 \frac{2}{\sqrt{9}} = 9 \pm 1.307 \quad \Longrightarrow \quad (7.69, 10.31).$$

- **Unknown Variance:** Using the sample variance and $t_{0.025,8} = 2.306$, the 95% CI for μ is

$$9 \pm 2.306 \frac{\sqrt{9.5}}{\sqrt{9}} \approx 9 \pm 2.370 \quad \Longrightarrow \quad (6.63, 11.37).$$

- **Other Confidence Levels (Unknown Variance):**
  - 99% CI: $9 \pm 3.335 \frac{\sqrt{9.5}}{\sqrt{9}} \approx (7.28, 10.72)$.
  - 90% CI: $9 \pm 1.859 \frac{\sqrt{9.5}}{\sqrt{9}} \approx (7.91, 10.09)$.
- **Observation:** When σ^2 is unknown, the CI becomes wider due to the additional uncertainty in estimating the variance.

Choice of Sample Size

Recall the CI with unknown variance is given by

$$[\bar{x} - t_{n-1,\alpha/2} s/\sqrt{n}, \bar{x} + t_{n-1,\alpha/2} s/\sqrt{n}]$$

Question

How many observations do we need so that the estimation error is smaller than some tolerance ε?

We need

$$t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \leq \varepsilon \Rightarrow n \geq \left\lceil \left(\frac{t_{n-1,\alpha/2} s}{\varepsilon} \right)^2 \right\rceil$$

One-sided Confidence Interval

If we care only about one side of the value, we construct *one-sided confidence interval*.

- In this case, we assign probability only to one side of the sample mean.

One-sided CI for normal population mean with unknown variance

A $(1-\alpha) \times 100\%$ **upper-confidence bound** is

$$\mu \le \bar{x} + t_{n-1,\alpha} s/\sqrt{n}.$$

A $(1-\alpha) \times 100\%$ **lower-confidence bound** is

$$\mu \ge \bar{x} - t_{n-1,\alpha} s/\sqrt{n}.$$

Case 3: Confidence Interval for the Population Variance

Let $X_1, \ldots, X_n$ be i.i.d. $\mathcal{N}(\mu, \sigma^2)$.

- Suppose both the mean μ and the variance σ^2 are unknown.
- We want to construct a 95% CI for the unknown population variance σ^2.
- We know that the sample variance S^2 is a good point estimator for σ^2.

Question: What kind of distribution does the sample variance S^2 follow?

Chi-squared Distribution

Let

$$X^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 .$$

- Think of this as a standardized version of S^2.

χ^2 distribution

Let $X^2 = \frac{(n-1)S^2}{\sigma^2}$, then X^2 follows the **chi-squared distribution with degree of freedom** $\nu = n - 1$, usually denoted as χ^2_{n-1}.

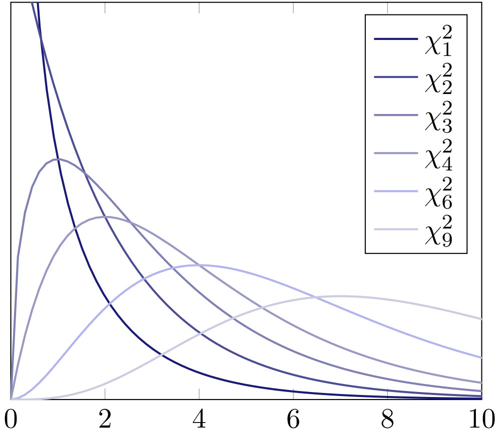
Notice that X^2 is NOT a statistic because it depends on the unknown parameter σ^2.

- The chi-squared distribution with ν degrees of freedom can be written as

$$\chi^2_\nu = \sum_{i=1}^{\nu} Z_i^2,$$

where the Z_i are independent standard normal random variables.
- χ^2_{n-1} is always non-negative.
- This distribution is asymmetric and skewed to the right.
- $\mathbb{E}[\chi^2_\nu] = \nu, \quad \mathrm{Var}(\chi^2_\nu) = 2\nu.$

PDF of the χ^2_ν distribution



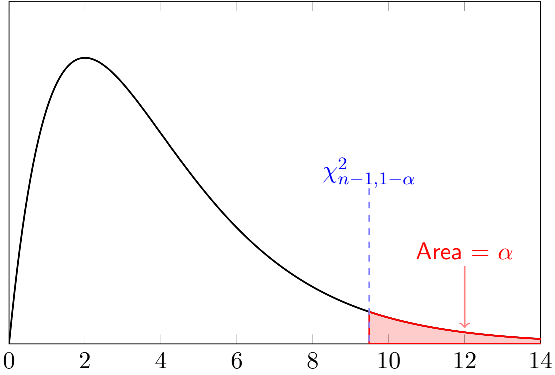
χ^2 critical value

Let X^2 be a χ^2_{n-1} random variable. We define $\chi^2_{n-1,\alpha}$ to be the value such that

$$\mathbb{P}(X^2 > \chi^2_{n-1,\alpha}) = \alpha.$$

- An α proportion of observations fall **above** $\chi^2_{n-1,\alpha}$.

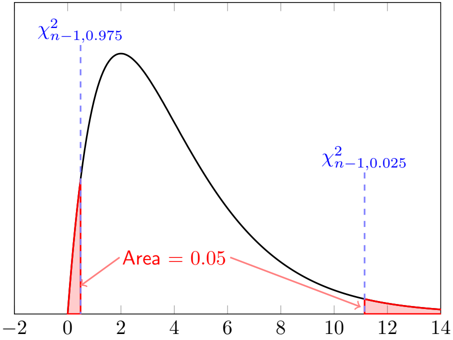
PDF of the χ^2_{n-1} distribution



Case 3: Confidence Interval for the Population Variance

Steps to construct a confidence interval for the population variance σ^2:

1. Find a statistic S^2 informative about the parameter.
2. Find the distribution of the statistic:
$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$
3. Write down the probability equality
$$\mathbb{P}(\chi^2_{n-1,1-\alpha/2} \leq X^2 \leq \chi^2_{n-1,\alpha/2}) = 1 - \alpha.$$
4. Re-arrange terms to get the coverage probability
$$\mathbb{P}\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right) = 1 - \alpha.$$



Two-sided CI for population variance

A two-sided $(1-\alpha) \times 100\%$ confidence interval for the population variance σ^2 is

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \right].$$

Notice that we used a lower-case s to denote the realized value.

Remarks

- One-sided CI
$$\left(-\infty, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha}} \right] \quad \text{and} \quad \left[\frac{(n-1)s^2}{\chi^2_{n-1,\alpha}}, \infty \right)$$
- Symmetric CI? In fact, this is not optimal in terms of minimizing the length. Beyond the scope of this course.

Case 4: Large-Sample Confidence Interval for Population Proportion

It is often necessary to construct confidence intervals on a population proportion.

Example: A random sample has been taken from a large population and that X observations in this sample belong to a class of interest. Then the sample proportion, $\widehat{P} = X/n$, is a point estimator for the population proportion of that class.

Example: (Bernoulli) Observe 300 flips, 200 heads and 100 tails. We are "95% sure" that the probability of getting a head is within $[0.613, 0.720]$.

- How to construct a CI for population proportion?

Steps:

- Statistic: $\widehat{P} = X/n$.
- Distribution of the statistic: (for large-sample $n \geq 30$)

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\widehat{P} - p}{\sqrt{p(1-p)/n}} \approx N(0,1).$$

- Probability equality

$$\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \approx 1 - \alpha.$$

- Re-arrange terms

$$\mathbb{P}\left(\widehat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \widehat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) \approx 1 - \alpha.$$

We have a problem!

- CI as statistics cannot have unknown parameters.

Solution: it is often satisfactory if we simply replace p by $\widehat{P}$, when the sample size n is suitably large $(n \geq 30)$.

Large-Sample CI for population proportion

A $(1-\alpha) \times 100\%$ confidence interval for the population proportion p is

$$\left[\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right].$$

Notice that we used a lower-case $\widehat{p}$ to denote the realized value.

Example: Observe 300 coin flips, 200 heads and 100 tails. Calculate a 95% CI:

$$\left[\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]$$

$$= \left[\frac{2}{3} - 1.96 * \sqrt{\frac{\frac{2}{3}(1-\frac{2}{3})}{300}}, \frac{2}{3} + 1.96 * \sqrt{\frac{\frac{2}{3}(1-\frac{2}{3})}{300}}\right]$$

$$= [0.613, 0.720].$$

Example: Observe 3 coin flips, 2 heads and 1 tails. Calculate a 95% CI:

$$\left[\frac{2}{3} - 1.96 * \sqrt{\frac{\frac{2}{3}(1-\frac{2}{3})}{3}}, \frac{2}{3} + 1.96 * \sqrt{\frac{\frac{2}{3}(1-\frac{2}{3})}{3}}\right] = [0.133, 1.120].$$

Example: Inferring Sample Size from a Poll

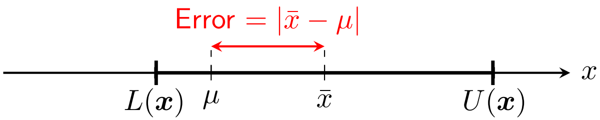
On October 14, 2003, the New York Times reported that a poll indicated that 52% of the population was in favor of President Bush's job performance with 95% confidence and a margin of error of $\pm 4\%$. What does this mean? Can we infer how many people were questioned?

Interpretation:

- With 95% confidence, the true proportion p lies in the interval

$$[0.52 - 0.04, 0.52 + 0.04] = [0.48, 0.56].$$

- The margin of error (half-width) of the 95% CI is given by $z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where $\hat{p} = 0.52$ and $z_{0.025} = 1.96$.



Example: Inferring Sample Size from a Poll

Calculations:

$$1.96\sqrt{\frac{0.52 \times 0.48}{n}} = 0.04.$$

Solving for n:

$$\sqrt{\frac{0.52 \times 0.48}{n}} = \frac{0.04}{1.96} \quad \Longrightarrow \quad n = \frac{0.52 \times 0.48}{\left(\frac{0.04}{1.96}\right)^2} \approx 599.29.$$

Thus, approximately 600 people were surveyed.

Case 5: Large-Sample Confidence Interval

For real-world problems, the population is not normally distributed.

For population mean μ

- By CLT, for large sample $(n \geq 30)$ we have $\frac{\bar{X}-\mu}{\sigma\sqrt{n}} \approx N(0,1)$.
- So

$$[\bar{x} + z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{1-\alpha/2}\sigma/\sqrt{n}]$$

  is still an effective $(1-\alpha) \times 100\%$ CI.
- Similarly, when σ is unknown, replace σ by s.
- For one-sided CI, we have

$$(-\infty, \bar{x} + z_\alpha s/\sqrt{n}] \quad \text{or} \quad [\bar{x} - z_\alpha s/\sqrt{n}, +\infty).$$

Similarly, we can construct CI for σ just as we did for normal sample.

General Steps for Constructing a $(1-\alpha) \times 100\%$ CI

1. Use **only** the **samples**, the **known parameter(s)** and the **parameter of interest** to construct a quantity Y, such that its distribution is known, and the distribution does not depend on θ.
   - Usually starts with a point estimator of the target parameter.
2. Compute a interval $[A, B]$ such that the following (approximately) holds

$$P(Y \in [A, B]) = 1 - \alpha.$$

   - For one-sided case, set A to $-\infty$ or B to $+\infty$ accordingly;
   - For two-sided case, find A, B such that they assign equal probability $\alpha/2$ to each side.
3. Rearrange the interval $A \leqslant Y \leqslant B$ using the expression of Y obtained in Step 1, and obtain a interval that covers the population parameter of interest. This final interval will be the desired CI.

Extended Reading

- Chapter 8 of Douglas C. Montgomery and George C. Runger, *Applied Statistics and Probability for Engineers*, 7th Ed.
- (**Advanced**) More on t distribution and χ^2 distribution. Section 4.6 of this book:
  https://www.utstat.toronto.edu/mikevans/jeffrosenthal/chapt4.pdf