Confidence Intervals, Central Limit Theorem, and the Bootstrap Principle

1. Interpretation of a Confidence Interval

A confidence interval (CI) is used to estimate an unknown population parameter using sample data. The true population value is fixed but unknown; randomness comes from the sampling process.

A 95% confidence interval means that if the same sampling procedure were repeated many times, about 95% of the intervals constructed in this way would contain the true population value. Once a specific interval is calculated, the true value is either inside or outside the interval—-there is no probability attached to a fixed interval.

Example: A sample of 1,000 voters produces an approval rate of 58%. A 95% confidence interval is 54.8% to 61.2%. We state that we are 95% confident the true approval rate lies in this range.

---

2. Using the Central Limit Theorem for Confidence Intervals

A confidence interval provides a range of plausible values for a population parameter μ. Most confidence intervals are built around an estimate of μ, commonly a sample average or percentage.

When the estimate is an average or a percentage, the Central Limit Theorem (CLT) applies. Under the CLT, the sampling distribution of the estimate is approximately normal for large samples.

The general form of a confidence interval is:

Estimate ± z × SE

Typical z-values: - 90% confidence: z ≈ 1.65\

- 95% confidence: z ≈ 1.96 (often rounded to 2)\

- 99% confidence: $z \approx 2.58$

The standard error (SE) for an average or percentage is:

$SE = \sigma / \sqrt{n}$

Because the population standard deviation σ is usually unknown, it is replaced with a sample estimate using the bootstrap principle.

---

3. Estimating the Standard Error with the Bootstrap Principle

The bootstrap principle replaces unknown population quantities with their corresponding sample estimates.

For a population proportion:

$SE = \sqrt{[p(1 - p)]} / \sqrt{n}$

Since the true population proportion p is unknown, we substitute the sample proportion $\hat{p}$:

$SE \approx \sqrt{[\hat{p}(1 - \hat{p})]} / \sqrt{n}$

Example (Presidential approval poll): - Sample size: n = 1,000\

- Sample proportion: 58%\
- Estimated standard deviation: $\sqrt{[0.58(1 - 0.58)]} \approx 0.49$\
- 95% CI = 58% ± 2 × (0.49 / √1000)\
- Resulting CI: 54.9% to 61.1%

For physical measurements (such as estimating the speed of light), the sample standard deviation estimates the standard deviation of the measurement error. Adding a constant to all measurements does not change the standard deviation, validating the bootstrap principle. The principle extends to more complex estimation problems.

---

4. More About Confidence Intervals and the Square-Root Law

A confidence interval has the form:

Estimate ± z × SE

- The total width of a confidence interval is: 2 × z × SE\
- The quantity z × SE is called the margin of error

Because SE = σ / √n, the square-root law applies: - To reduce the width of the interval by half, the sample size must be quadrupled. - To make the width one-tenth as large, the sample size must increase by 100 times.

Another way to reduce interval width is to reduce the confidence level, which lowers z but also lowers certainty. There is always a trade-off between precision and confidence.

---

5. Rule of Thumb for a 95% Confidence Interval for a Proportion

For percentages, a fast approximation for a 95% confidence interval is:

$\hat{p} \pm 1 / \sqrt{n}$

This works because the maximum possible population standard deviation for a proportion is about 0.5, and with $z \approx 2$, the constants cancel out.

When confidence intervals are reported in the media without stating the confidence level, it is implicitly assumed to be 95%.

Key Takeaways

- A confidence interval estimates an unknown population parameter using sample data.
- The Central Limit Theorem justifies using a normal model for sample means and proportions.
- The standard error measures sampling variability and decreases as sample size increases.
- The bootstrap principle replaces unknown population parameters with sample estimates.
- Larger sample sizes produce narrower confidence intervals because of the square-root law.

- There is a fundamental trade-off between confidence level and precision.
- For proportions, a quick 95% confidence interval is approximately $\hat{p} \pm 1 / \sqrt{n}$.