Sampling

PSCI 1801 · Class Handout

Dr. Marc Trussler

This handout is designed as an in-class supplement to the textbook. The material contained on this handout does not represent the full set of information you need to know for the exams. Exclusively studying off of this handout will not be sufficient for the exams.

Law of Large Numbers & The Central Limit Theorem

The two foundational results of statistical inference.

Law of Large Numbers (LLN): As sample size n grows, the sample mean $\bar{X}_n$ converges to the population expected value $E[X]$.

Central Limit Theorem (CLT): For any random variable X with mean μ and finite variance σ^2, the sampling distribution of the sample mean is approximately normal for large n:

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

This is true regardless *of the shape of the underlying population — normal, skewed, uniform, whatever.*

Three principles

1. *The sample mean is unbiased:* $E[\bar{X}_n] = E[X]$
2. *The variance shrinks with* n*:* $V[\bar{X}_n] = V[X]/n$
3. *The sampling distribution is normal, centered on the population mean*

The Math (Brief)

Why do the three principles hold? We derive them from the expected value and variance operators.

Expected value of the sample mean

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

Using linearity of expectation ($E[cX] = cE[X]$ and $E[X+Y] = E[X]+E[Y]$):

$$E[\bar{X}_n] = \frac{1}{n} \cdot n \cdot E[X] = E[X]$$

Variance of the sample mean

The variance operator has different rules: $V[cX] = c^2V[X]$, and $V[X+Y] = V[X]+V[Y]$ only when X and Y are independent (which iid samples are):

$$V[\bar{X}_n] = \frac{1}{n^2}\sum V[X_i] = \frac{1}{n^2} \cdot n \cdot V[X] = \frac{V[X]}{n}$$

The standard error

The standard deviation of the sampling distribution is called the standard error *of the mean:*

$$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

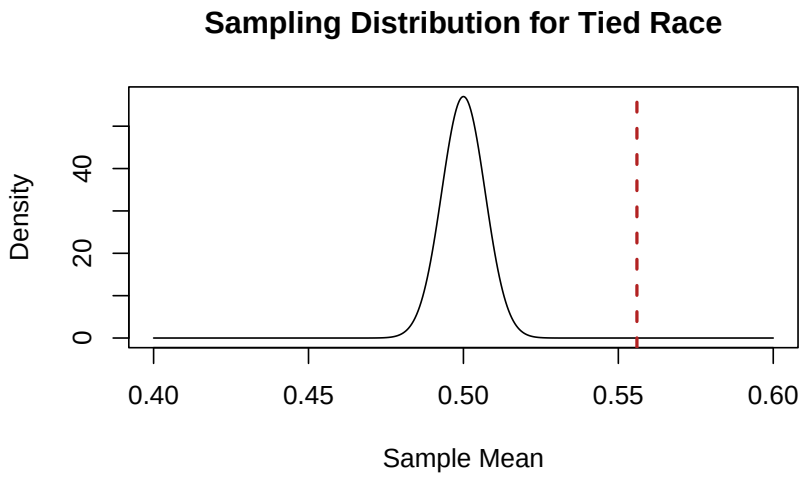
Sampling distributions get half *as wide when you* quadruple *the sample size.*

Building the Sampling Distribution

If we know (or hypothesize) μ and σ, we can draw the sampling distribution before *we sample.*

Example: If a race is truly tied, what does the sampling distribution of the proportion supporting a candidate look like with $n = 5000$?

- $\mu = 0.5$
- $\sigma = \sqrt{\pi(1-\pi)} = \sqrt{0.5 \cdot 0.5} = 0.5$ (Bernoulli variance from DiscreteRVs)
- $SE = 0.5/\sqrt{5000} \approx 0.0071$
- Sampling distribution: $N(0.5, 0.0071^2)$



If we actually observe a sample mean of 0.556, this is many standard errors away from 0.5 — very strong evidence against a tied race.

The 99% "expected range" bounds around the population mean:

$$\mu \pm 2.57 \cdot \frac{\sigma}{\sqrt{n}}$$

The Three Distributions

Whenever we do inferential statistics we are dealing with three *distinct distributions. Keep them straight — conflating them is the single most common mistake at this stage.*

1. **The population.** The random variable / DGP we are sampling from. Can be any shape.
2. **The sample.** The n numbers we actually drew. A rough approximation of the population, but never exact. Every sample looks a little different from every other.
3. **The sampling distribution of the sample mean.** The distribution of sample means we would get if we repeated the sampling process. Always (approximately) normal, centered on μ, with $SD = \sigma/\sqrt{n}$.

Statistics has no *clean way to answer "how much does my sample distribution differ from the population distribution?" But it* does *have a precise answer to "how much does my sample mean differ from the population mean?" — that's exactly what the CLT gives us.*

Because the sampling distribution of the sample mean is a normal distribution whose parameters we can write down before we sample, the sample mean is the only object in this picture whose behavior we can pin down with simple math. That's why we build inference around it.

What Makes a Good Estimator?

Language: population parameters *(fixed unknowns like μ) are estimated by* estimators *(procedures like "compute the mean of the sample"), which produce* estimates *(a specific number, like 0.556).*

Unbiasedness

$$E[\hat{\theta} - \theta] = 0$$

The average estimate across many samples equals the truth. The sample mean is unbiased — so is a wild alternative like "take only the first observation!"

Consistency

But unbiasedness alone isn't enough. We also want the estimator to get better *with more data. Combining low variance and unbiasedness gives us the* Mean Squared Error:

$$MSE = V[\hat{\theta}] + \Big(E[\hat{\theta} - \theta]\Big)^2 = \text{Variance} + \text{Bias}^2$$

A consistent *estimator is one whose* $MSE \to 0$ *as* $n \to \infty$*. The sample mean is consistent (variance shrinks as* $1/n$*). The "first observation" estimator is* not *(variance stays constant, no matter how large* n *gets).*

The RMSE (square root of MSE) is often easier to interpret because it's on the scale of the variable.

Loop-back: because the sample mean is unbiased, the bias term drops out and the MSE *reduces to* $V[X]/n$*. Its square root is* $sd(X)/\sqrt{n}$ *— which is exactly the standard error. So the* $1/\sqrt{n}$ *decay you'd see in a simulated RMSE curve is the same* $1/\sqrt{n}$ *decay that determined the width of every sampling distribution above. All one thing.*