STA 130A: Mathematical Statistics: Brief Course

Lecture 22: From Random Samples to Estimation

Dogyoon Song

Fall 2026, UC Davis

Agenda: From normal approximation to estimation

First: Complete the CLT applications (spillover from Lecture 21)

- Normal approximation for sums and averages.
- Continuity correction for integer-valued counts.

Then: What if the population mean is unknown?

- Estimate population quantities from a random sample.
- Describe how estimates vary across samples.
- Distinguish individual variability from uncertainty in an estimate.

Our statistics module:

- Lectures 22–23: sampling distributions and standard errors
- Lecture 24: confidence intervals
- Lectures 25–26: constructing estimators

Recap: Normal approximation for sums and averages

Let $X_1, \ldots, X_n$ be i.i.d. with mean μ, variance σ^2, and $\sigma > 0$. For large n, the CLT gives

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \;\approx\; N(0,1), \qquad S_n = \sum_{i=1}^{n} X_i.$$

Quantity	Mean	Standard deviation
Sum S_n	$n\mu$	$\sigma\sqrt{n}$
Average $\bar{X}_n$	μ	$\sigma/\sqrt{n}$

Workflow: Identify the quantity $\longrightarrow$ center and scale $\longrightarrow$ use Φ and the correct tail.

$$P(\bar{X}_n > b) \approx 1 - \Phi\left(\frac{b-\mu}{\sigma/\sqrt{n}}\right).$$

Caution: Normal sampling gives an exact normal law for every n; otherwise this is an approximation, not a probability bound.

Example: Total package weight

Example

A plane carries 100 packages with independent weights $X_i \sim \mathrm{Uniform}(5, 50)$ in pounds.

Question: What is the approximate probability that the total weight exceeds 3000 pounds?

$$\mu = \mathbb{E}[X_i] = \frac{5+50}{2} = 27.5, \qquad \text{and} \qquad \sigma^2 = \mathrm{Var}(X_i) = \frac{(50-5)^2}{12} = \frac{2025}{12} = 168.75.$$

Let $S_{100} = X_1 + \cdots + X_{100}$. Then

$$S_{100} \approx N(100 \cdot 27.5,\ 100 \cdot 168.75).$$

Thus,

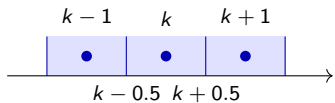
$$P(S_{100} > 3000) \approx 1 - \Phi\left(\frac{3000 - 2750}{\sqrt{16875}}\right) \approx 1 - \Phi(1.92) \approx 0.027.$$

Same event: $S_{100} > 3000$ exactly when $\bar{X}_{100} > 30$.

Binomial normal approximation: Continuity correction

For $S \sim \text{Binomial}(n, p)$, use $G \sim N(np, np(1-p))$ when a normal approximation is appropriate.

Discrete counts become intervals: The integer k corresponds to the cell $[k - 0.5, k + 0.5]$.



For integer cutoffs $k \leq \ell$,

$$P(k \leq S \leq \ell) \approx P(k - 0.5 \leq G \leq \ell + 0.5).$$

For one-sided events,

$$P(S \leq k) \approx P(G \leq k + 0.5), \qquad P(S \geq k) \approx P(G \geq k - 0.5).$$

Rule: Identify the included integers, then extend by half a unit before standardizing.

Example: Binomial normal approximation (1/2)

Example

Let $S \sim \text{Binomial}(100, 0.4)$.

Question: Approximate $P(35 \leq S \leq 45)$.

Since S is a sum of 100 independent Bernoulli(0.4) random variables,

$$\mathbb{E}[S] = np = 100(0.4) = 40, \qquad \text{Var}(S) = np(1-p) = 100(0.4)(0.6) = 24.$$

Thus, for probability calculations, we use $G \sim N(40, 24)$ as a normal approximation to S.

Without continuity correction, $P(35 \leq S \leq 45) \approx P(35 \leq G \leq 45)$. Therefore,

$$\begin{aligned} P(35 \leq S \leq 45) &\approx \Phi\left(\frac{45-40}{\sqrt{24}}\right) - \Phi\left(\frac{35-40}{\sqrt{24}}\right) \\ &\approx \Phi(1.02) - \Phi(-1.02) \\ &= 2\Phi(1.02) - 1 \approx 2(0.8461) - 1 = 0.6922. \end{aligned}$$

Example: Binomial normal approximation (2/2)

Example

With continuity correction, the integer event $35 \leq S \leq 45$ is approximated by the continuous interval $34.5 \leq G \leq 45.5$, where $G \sim N(40, 24)$.

$$\begin{aligned} P(35 \leq S \leq 45) &\approx P(34.5 \leq G \leq 45.5) \\ &= \Phi\left(\frac{45.5 - 40}{\sqrt{24}}\right) - \Phi\left(\frac{34.5 - 40}{\sqrt{24}}\right) \\ &\approx \Phi(1.12) - \Phi(-1.12) = 2\Phi(1.12) - 1 \approx 2(0.8686) - 1 = 0.7372. \end{aligned}$$

For comparison, the exact binomial probability is

$$P(35 \leq S \leq 45) = \sum_{k=35}^{45} \binom{100}{k} 0.4^k 0.6^{100-k} \approx 0.7386.$$

Takeaway: The continuity correction improves the normal approximation in this example.

Pop-up quiz: Continuity correction

Let

$$S \sim \text{Binomial}(100, 0.4).$$

Question: With continuity correction, which expression approximates $P(S \leq 45)$?

A) $\Phi\left(\dfrac{45 - 40}{\sqrt{24}}\right)$

B) $\Phi\left(\dfrac{45.5 - 40}{\sqrt{24}}\right)$

C) $1 - \Phi\left(\dfrac{45.5 - 40}{\sqrt{24}}\right)$

D) $\Phi\left(\dfrac{45 - 40}{24}\right)$

Answer: B.

The event $S \leq 45$ is approximated by the continuous event $G \leq 45.5$, where $G \sim N(40, 24)$.

Follow-up: What continuity correction would you use for $P(S \geq 45)$?

Motivation: What can a few package weights tell us?

A machine fills packages with a target weight of 500 grams. Its actual population mean μ and variance σ^2 are unknown.

Suppose we weigh 16 randomly selected packages and observe

$$\bar{x} = 501.2 \text{ grams.}$$

This is an illustrative dataset summary, not a report of actual measurements.

What do we want to learn?

- What is the machine's true average output μ?
- How much do individual package weights vary?
- How far might our sample average be from μ?

Key issue: Another sample would usually give a different answer.

Random sample versus observed data

Before observing the data: Model a random sample by

$$X_1, \ldots, X_n \quad \text{independent and identically distributed,} \qquad \mathbb{E}[X_i] = \mu, \quad \mathrm{Var}(X_i) = \sigma^2.$$

- Each X_i is random; the population quantities μ, σ^2 are fixed but unknown.
- The sampling assumptions are part of the model, not guaranteed by writing down a formula.

After observing the data: We have numerical values

$$x_1, \ldots, x_n.$$

For example, four observed weights might be

$$498, \quad 500, \quad 502, \quad 504 \quad \text{grams.}$$

Notation: Capitals for random quantities; lowercase for their observed values.

What is a statistic?

Definition (Statistic)

A **statistic** is a function of the sample whose value can be computed from the data, without knowing an unknown parameter:

$$T = T(X_1, \ldots, X_n).$$

Examples include

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad (n \geq 2).$$

- The rule is fixed; its output is random because the sample is random.
- If μ is unknown, $\bar{X} - \mu$ is a random variable but not a statistic: it cannot be computed from the sample alone.

Parameter, estimator, and estimate

	General notation	Mean-weight example
Parameter	θ: fixed unknown target	Population mean μ
Estimator	$\widehat{\theta} = T(X_1, \ldots, X_n)$	Sample mean $\bar{X}$
Estimate	$T(x_1, \ldots, x_n)$: observed value	$\bar{x} = 501.2$ grams

A statistic becomes an estimator when we use it to estimate a target.

- The estimator's distribution describes what would happen across repeated samples.
- The estimate is the number obtained from the particular sample in hand.

The same rule, different samples

Suppose we repeatedly take samples of four packages from the same machine. Apply the **same averaging rule** each time.

Sample	Observed weights (grams)	Observed mean
A	$498, 500, 502, 504$	501
B	$496, 498, 500, 502$	499
C	$499, 501, 503, 505$	502

These hypothetical samples illustrate variability, not a complete sampling distribution.

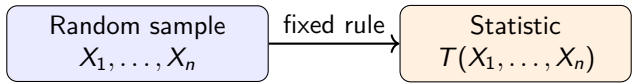
The population mean μ stays fixed; $\bar{X}$ varies from sample to sample.

Question: What distribution describes these possible sample means?

Sampling distribution of a statistic

Definition (Sampling distribution)

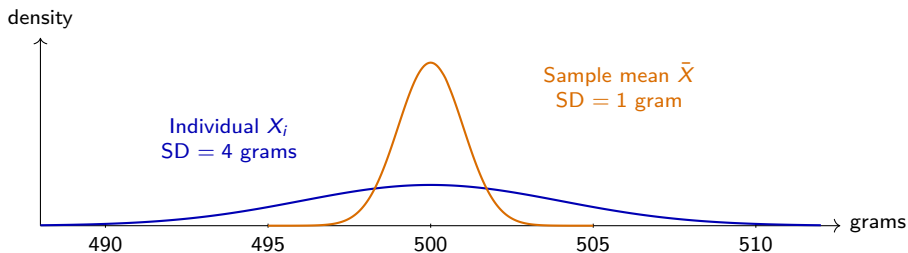
The **sampling distribution** is the probability distribution of $T(X_1, \ldots, X_n)$ under the sampling model.



- Imagine repeating the *entire sampling procedure*, not reusing the same observed data.
- A sampling distribution concerns the statistic, not the individual observations.
- Its shape and spread depend on the population model and the sample size.

Individual weights versus sample means

For illustration, suppose $X_i \sim N(500, 4^2)$ independently and $n = 16$. Then $\bar{X} \sim N(500, 1^2)$, exactly.



Same center, different spread. Averaging reduces sampling variability; it does not reduce the variability of the individual packages themselves.

Mean and variance of the sample mean

For i.i.d. observations with mean μ and variance σ^2,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

By linearity,

$$\mathbb{E}[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \mu.$$

By independence and variance scaling,

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

Statistical meaning: $\bar{X}$ is centered at its target μ, and its variability decreases as n grows.

- These moment identities do not require a normal population.

Standard deviation versus standard error

Population standard deviation:

$$\mathrm{SD}(X_i) = \sigma.$$

It measures the spread of individual observations.

Standard error of the sample mean:

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

It measures the spread of the estimator across repeated samples.

- A standard error is not the realized error $\bar{x} - \mu$.
- Population SD and standard error have the same units, but describe different variability.
- When σ is unknown, we will estimate this standard error using $S/\sqrt{n}$.

Example: What does a larger sample buy?

Example (Package weights)

Suppose the population standard deviation is $\sigma = 4$ grams.

Sample size	Individual SD	Standard error of $\bar{X}$
$n = 16$	4 grams	$4/\sqrt{16} = 1$ gram
$n = 64$	4 grams	$4/\sqrt{64} = 0.5$ gram

Question: How much more data is needed to halve the standard error?

Answer: Four times as many observations, because standard error decreases as $1/\sqrt{n}$.

Caution: More observations reduce sampling variability, not systematic sampling bias.

Exact normal sampling versus the CLT

Normal population: If $X_i \sim N(\mu, \sigma^2)$ independently, then

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{for every } n.$$

This is an exact distributional result, not an asymptotic approximation.

General i.i.d. population: The CLT gives, for large n,

$$P(\bar{X} \le c) \approx \Phi\left(\frac{c-\mu}{\sigma/\sqrt{n}}\right), \qquad \sigma > 0.$$

Approximation quality depends on the population and sample size.

In either case: $\mathbb{E}[\bar{X}] = \mu$ and $\mathrm{Var}(\bar{X}) = \sigma^2/n$. The individual observations do not become normal merely because n grows.

Pop-up quiz: What changes across samples?

Suppose we repeatedly draw samples of the same size from one unchanged population.

Question: Which statement is correct?

A) The population mean μ changes, but $\bar{X}$ is fixed.

B) Both μ and $\bar{X}$ must change.

C) The population mean μ is fixed, but $\bar{X}$ can change.

D) The standard error equals the observed error $\bar{x} - \mu$.

Answer: C. The estimator is random across samples; the population parameter stays fixed.

Follow-up: Once a sample has been observed, is its numerical mean $\bar{x}$ still an unknown outcome of that sampling experiment?

The sample variance

The population variance σ^2 is usually unknown. For $n \geq 2$, define

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \qquad S = \sqrt{S^2}.$$

- S^2 estimates the population variance; S estimates the population SD.
- After observing the data, compute s^2 and s using the same formulas.
- Variances have squared units; standard deviations have the original units.

Do not confuse:

$$\underbrace{S^2}_{\text{estimates } \sigma^2} \quad \text{and} \quad \underbrace{\frac{S^2}{n}}_{\text{estimates } \mathrm{Var}(\bar{X})=\sigma^2/n}.$$

Why divide by $n-1$?

We center the observations at the *sample* mean rather than the unknown population mean.

$$\sum_{i=1}^{n}(X_i - \bar{X}) = 0.$$

Once $n-1$ centered deviations are specified, the last is determined by this constraint.

The exact expectation calculation gives

$$\mathbb{E}\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = (n-1)\sigma^2.$$

Consequently,

$$\boxed{\mathbb{E}[S^2] = \sigma^2.}$$

Meaning: The sample variance is correct *on average* across samples (we call this *unbiasedness*). It need not equal σ^2 for a particular sample.

Self-study: Why S^2 is unbiased

The sum-of-squares identity is

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

The cross terms cancel because $\sum_i (X_i - \bar{X}) = 0$.

Taking expectations and using $\mathbb{E}[\bar{X}] = \mu$,

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] &= n\sigma^2 - n\mathrm{Var}(\bar{X}) \\ &= n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2.\end{aligned}$$

Thus $\mathbb{E}[S^2] = \sigma^2$.

No normality assumption is needed for this expectation identity. The exact distribution of S^2 is a different question, taken up next lecture.

Example: Compute and interpret three summaries

For the observed package weights $498, 500, 502, 504$ grams,

$$\bar{x} = \frac{498 + 500 + 502 + 504}{4} = 501 \text{ grams.}$$

The deviations from $\bar{x}$ are $-3, -1, 1, 3$, so

$$s^2 = \frac{9 + 1 + 1 + 9}{4 - 1} = \frac{20}{3} \text{ grams}^2, \qquad s \approx 2.58 \text{ grams.}$$

The estimated standard error of the mean is

$$\frac{s}{\sqrt{4}} = \sqrt{\frac{5}{3}} \approx 1.29 \text{ grams.}$$

- 501: our estimate of the population mean.
- 2.58: estimated spread of individual package weights.
- 1.29: estimated sampling spread of the mean, not its actual error.

Next question: What changes when σ is unknown?

For i.i.d. $N(\mu, \sigma^2)$ observations with $\sigma > 0$,

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

But σ is usually unknown; replacing it by S gives

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

- The denominator is now random too.
- This ratio is not exactly standard normal under normal sampling.
- We need the sampling distribution of S^2 and its relationship with $\bar{X}$.

Next lecture: Chi-square and Student's t arise from these sample quantities.

Wrap-up

Our statistical task:

$$\text{unknown parameter} \longrightarrow \text{random sample} \longrightarrow \text{estimator} \longrightarrow \text{observed estimate}.$$

Sampling distributions describe how estimators vary across samples.

$$\mathbb{E}[\bar{X}] = \mu, \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}, \qquad \mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

- S^2 estimates σ^2; $S/\sqrt{n}$ estimates the standard error of $\bar{X}$.
- Distinguish population variability, sampling variability, and the observed estimation error.
- Exact normal-sample results and CLT approximations are not interchangeable.

Suggested reading: Rice, Ch. 6 (sample mean and sample variance); Lectures 20–21.

Next lecture: Normal-sample theory, chi-square, and Student's t.

References


John A. Rice. *Mathematical Statistics and Data Analysis*, 3rd edition. Chapter 6: distributions derived from the normal distribution.


Dimitri P. Bertsekas and John N. Tsitsiklis. *Introduction to Probability*, 2nd edition. Chapter 5: limit theorems.