ECON 120A: Synthesized Lecture Notes
Probability, Inference, and Core Derivations

Compiled from ECON 120A lecture sequence (Winter 2022)

Contents

1 Scope and Notation

These notes synthesize the main technical content of ECON 120A into one coherent document. The emphasis is on:

1. probability foundations,
2. sampling distributions and asymptotics,
3. estimation, confidence intervals, and hypothesis testing,
4. dependence between multiple random variables.

Throughout, unless stated otherwise, $X_1, \ldots, X_n$ are i.i.d. random variables with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Sample mean and sample variance are

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

2 Data, Populations, and Sampling

2.1 Population vs Sample

A *population* is the complete collection of units of interest; a *sample* is a subset used for inference. In most empirical contexts, full census measurement is impossible due to cost, time, or feasibility constraints.

2.2 Sampling Design and Representativeness

The quality of inference depends on sampling design.

- **Simple random sampling (SRS):** each size-n sample has equal selection probability.
- **Systematic sampling:** every kth unit after a random start.
- **Stratified sampling:** sample within strata to preserve key composition.
- **Cluster sampling:** sample clusters, then units within selected clusters.

Nonrandom sampling introduces selection bias and can invalidate standard inferential formulas.

2.3 Descriptive Statistics

For a sample $x_1, \ldots, x_n$, common summaries are:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

and for a binary variable $d_i \in \{0, 1\}$,

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} d_i.$$

These descriptive objects are also estimators once sampling uncertainty is introduced.

3 Probability Foundations

3.1 Events and Axioms

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition 3.1 (Probability measure)**.** A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure if:

1. $\mathbb{P}(\Omega) = 1$,
2. $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$,
3. for pairwise disjoint $A_1, A_2, \dots$,

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Proposition 3.2 (Addition rule)**.** *For any events* A, B*,*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Proof. Decompose into disjoint parts:

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B).$$

Using additivity and rewriting $\mathbb{P}(A)$ and $\mathbb{P}(B)$ by disjoint decomposition yields the formula. □

3.2 Conditional Probability and Bayes Rule

Definition 3.3. For $\mathbb{P}(B) > 0$, the conditional probability of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Proposition 3.4 (Multiplication rule)**.**

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A).$$

Proposition 3.5 (Bayes theorem)**.** *If* $B_1, \dots, B_m$ *partition* Ω *and* $\mathbb{P}(B_j) > 0$*, then*

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_{j=1}^{m} \mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}.$$

3.3 Independence

Events A, B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Random variables X, Y are independent if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \text{for all } x, y,$$

equivalently $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ where densities exist.

4 Random Variables, Distributions, and Moments

4.1 PMF, PDF, CDF

For discrete X, PMF is $p_X(x) = \mathbb{P}(X = x)$. For continuous X, PDF f_X satisfies

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(t)\,dt.$$

CDF is always defined:

$$F_X(x) = \mathbb{P}(X \le x).$$

4.2 Expectation and Variance

$$\mathbb{E}[X] = \begin{cases} \sum_x x p_X(x), & X \text{ discrete,} \\ \int_{-\infty}^{\infty} x f_X(x)\,dx, & X \text{ continuous,} \end{cases}$$

$$\operatorname{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Proposition 4.1 (Linearity and variance algebra)**.** *For constants a, b,*

$$\mathbb{E}[a + bX] = a + b\mathbb{E}[X], \qquad \operatorname{Var}(a + bX) = b^2 \operatorname{Var}(X).$$

For random variables X, Y,

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y).$$

4.3 Moment Generating Functions

The MGF of X is $M_X(t) = \mathbb{E}[e^{tX}]$ when finite near $t = 0$. If MGFs exist in a neighborhood of zero, they identify distributions and simplify derivations of sums of independent variables:

$$M_{X+Y}(t) = M_X(t)M_Y(t) \quad \text{if } X \perp Y.$$

5 Canonical Distributions

5.1 Bernoulli and Binomial

If $X \sim \text{Bernoulli}(p)$, then

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p, \quad \mathbb{E}[X] = p, \quad \operatorname{Var}(X) = p(1 - p).$$

If $X \sim \text{Binomial}(n, p)$,

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

with

$$\mathbb{E}[X] = np, \qquad \operatorname{Var}(X) = np(1 - p).$$

Derivation of binomial moments. Write $X = \sum_{i=1}^n B_i$ with i.i.d. $B_i \sim \text{Bernoulli}(p)$. Then

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[B_i] = np,$$

$$\operatorname{Var}(X) = \sum_{i=1}^n \operatorname{Var}(B_i) = np(1 - p),$$

using independence. □

5.2 Normal Distribution

$X \sim N(\mu, \sigma^2)$ has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Standardization gives

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1),$$

which is used for tail probabilities and inference cutoffs.

6 Joint Distributions and Dependence

6.1 Marginal and Conditional Laws

For joint PMF $p_{X,Y}(x,y)$,

$$p_X(x) = \sum_y p_{X,Y}(x,y), \qquad p_Y(y) = \sum_x p_{X,Y}(x,y).$$

For joint density $f_{X,Y}(x,y)$,

$$f_X(x) = \int f_{X,Y}(x,y)\,dy, \qquad f_Y(y) = \int f_{X,Y}(x,y)\,dx.$$

Conditional density:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

6.2 Conditional Expectation and Iterated Expectation

Theorem 6.1 (Law of iterated expectations)**.** *If* $\mathbb{E}[|Y|] < \infty$*, then*

$$\mathbb{E}[Y] = \mathbb{E}\big[\mathbb{E}[Y \mid X]\big].$$

Proof. For discrete variables,

$$\mathbb{E}\big[\mathbb{E}[Y \mid X]\big] = \sum_x \left(\sum_y y\, p_{Y|X}(y \mid x)\right) p_X(x) = \sum_x \sum_y y\, p_{X,Y}(x,y) = \sum_y y\, p_Y(y) = \mathbb{E}[Y].$$

The continuous case follows by replacing sums with integrals. □

Theorem 6.2 (Law of total variance)**.**

$$\mathrm{Var}(Y) = \mathbb{E}\big[\,\mathrm{Var}(Y \mid X)\big] + \mathrm{Var}\left(\mathbb{E}[Y \mid X]\right).$$

6.3 Covariance and Correlation

$$\operatorname{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])], \qquad \operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

Proposition 6.3 (Correlation bounds)**.** *If variances are positive, then*

$$-1 \le \operatorname{Corr}(X,Y) \le 1.$$

Proof. Apply Cauchy–Schwarz to centered variables:

$$|\operatorname{Cov}(X,Y)| = |\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]| \le \sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}.$$

Divide both sides by $\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}$. □

Remark 6.4. Correlation captures linear association, not causality.

7 Sampling Distributions

7.1 Sample Mean as a Random Variable

Because the sample changes across repeated draws, $\bar{X}_n$ is random and has its own distribution.

Proposition 7.1 (Moments of the sample mean)**.** *For i.i.d. X_i with mean μ and variance σ^2,*

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

Proof.

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

Also,

$$\operatorname{Var}(\bar{X}_n) = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) = \frac{\sigma^2}{n},$$

using independence. □

7.2 Finite Population Correction

If sampling without replacement from finite population size N, then

$$\operatorname{Var}(\bar{X}_n) = \frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}.$$

When n/N is small (common rule: below 0.05), the correction is near 1.

7.3 Sample Proportion

For Bernoulli data $D_i \in \{0,1\}$ with $\mathbb{P}(D_i = 1) = p$,

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n}D_i, \qquad \mathbb{E}[\hat{p}] = p, \qquad \operatorname{Var}(\hat{p}) = \frac{p(1-p)}{n}.$$

8 Law of Large Numbers

8.1 Statement

Theorem 8.1 (Weak law of large numbers)**.** *If $X_1, X_2, \dots$ are i.i.d. with finite mean μ and finite variance σ^2, then*

$$\bar{X}_n \xrightarrow{p} \mu.$$

8.2 Proof via Chebyshev

Proof. For any $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \le \frac{\operatorname{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

Hence $\bar{X}_n \to \mu$ in probability. □

Remark 8.2. This theorem justifies replacing population means with sample averages in large samples.

9 Central Limit Theorem

9.1 Lindeberg–Levy CLT

Theorem 9.1 (Classical CLT)**.** *For i.i.d. X_i with mean μ and variance $\sigma^2 \in (0, \infty)$,*

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1).$$

9.2 Proof Sketch by Characteristic Functions

Let $Y_i = (X_i - \mu)/\sigma$ so that $\mathbb{E}[Y_i] = 0$, $\operatorname{Var}(Y_i) = 1$, and define

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i.$$

Its characteristic function is

$$\phi_{Z_n}(t) = \left(\phi_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Using a second-order expansion around zero,

$$\phi_Y(u) = 1 - \frac{u^2}{2} + o(u^2), \qquad u \to 0,$$

so

$$\phi_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-t^2/2},$$

which is the characteristic function of $N(0, 1)$. Levy's continuity theorem yields convergence in distribution.

9.3 Practical Consequence

For large n,

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

which is the basis for large-sample confidence intervals and tests.

10 Estimation Theory

10.1 Point Estimators

An estimator $\hat{\theta}_n$ maps sample data into a parameter estimate. Desirable properties:

- **Unbiasedness:** $\mathbb{E}[\hat{\theta}_n] = \theta$.
- **Efficiency:** smaller variance among unbiased estimators.
- **Consistency:** $\hat{\theta}_n \xrightarrow{p} \theta$.

10.2 Bias–Variance Decomposition

Proposition 10.1. *For any estimator $\hat{\theta}$,*

$$MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathrm{Var}(\hat{\theta}) + \left(Bias(\hat{\theta})\right)^2.$$

Proof. Write

$$\hat{\theta} - \theta = (\hat{\theta} - \mathbb{E}[\hat{\theta}]) + (\mathbb{E}[\hat{\theta}] - \theta).$$

Square and take expectation; cross-term vanishes because $\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]] = 0$. □

10.3 Unbiasedness of the Sample Mean

Proposition 10.2. *$\bar{X}_n$ is unbiased for μ.*

Proof. Immediate from linearity:

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

□

10.4 Unbiasedness of Sample Variance

Theorem 10.3. *For i.i.d. data with variance σ^2,*

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

Proof. Use identity

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

Take expectations:

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

Divide by $n - 1$. □

10.5 Consistency of Mean and Variance Estimators

Proposition 10.4. *$\bar{X}_n \xrightarrow{p} \mu$ and $S_n^2 \xrightarrow{p} \sigma^2$ under finite fourth moment.*

Idea. First result is LLN. For S_n^2, rewrite

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2\right),$$

and apply LLN to X_i and X_i^2, then continuous mapping. □

11 Confidence Intervals

11.1 General Form

A two-sided $(1-\alpha)$ confidence interval is

$$\hat{\theta} \pm c_{1-\alpha/2} \cdot SE(\hat{\theta}).$$

Larger confidence level increases interval width.

11.2 Mean with Known Variance

If population variance is known and either normality holds or n is large,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{(exact under normality; approx by CLT otherwise).}$$

Hence

$$\mu \in \left[\bar{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

11.3 Mean with Unknown Variance

If $X_i \sim N(\mu, \sigma^2)$ and σ unknown,

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}.$$

Thus an exact $(1-\alpha)$ CI is

$$\left[\bar{X}_n - t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}},\ \bar{X}_n + t_{n-1,1-\alpha/2}\frac{S_n}{\sqrt{n}}\right].$$

11.4 Proportion CI

For large samples,

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right),$$

so a plug-in CI is

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

11.5 Sample Size for Margin of Error

For target half-width E in a mean CI with known σ:

$$n \geq \left(\frac{z_{1-\alpha/2}\sigma}{E}\right)^2.$$

For proportion (using pilot value p_0):

$$n \geq \frac{z_{1-\alpha/2}^2 p_0(1-p_0)}{E^2}.$$

Conservative choice $p_0 = 0.5$ maximizes required n.

12 Hypothesis Testing

12.1 Framework

1. State H_0 and H_a.
2. Choose significance level α.
3. Compute test statistic under H_0.
4. Compute critical value or p-value.
5. Reject or fail to reject H_0.

12.2 Type I/II Errors and Power

- Type I error: reject true H_0 (probability α).
- Type II error: fail to reject false H_0 (probability $\beta(\theta)$).
- Power: $1 - \beta(\theta)$.

Increasing sample size increases power for fixed α.

12.3 Tests for Mean

For $H_0 : \mu = \mu_0$:

- Known σ:
$$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}.$$
- Unknown σ with normal population:
$$T = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} \sim t_{n-1}.$$

Adjust rejection region for two-sided, left-sided, or right-sided alternatives.

12.4 Test for Proportion

For $H_0 : p = p_0$:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}},$$

with normal approximation valid when np_0 and $n(1 - p_0)$ are sufficiently large.

12.5 p-value Interpretation

The p-value is the probability, under H_0, of observing a test statistic at least as extreme as the realized one. Small p-values indicate incompatibility with H_0.

Proposition 12.1 (CI-test duality)**.** *For two-sided tests in regular one-parameter settings, reject $H_0 : \theta = \theta_0$ at level α if and only if θ_0 lies outside the $(1 - \alpha)$ confidence interval.*

Proof. Both procedures are based on the same pivot inequality

$$\left| \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})} \right| \le c_{1-\alpha/2}.$$

Rearranging this inequality in terms of θ_0 yields the CI inclusion condition. □

13 Normal Approximation to Binomial

If $X \sim \text{Bin}(n, p)$ with large n, then

$$\frac{X - np}{\sqrt{np(1 - p)}} \approx N(0, 1).$$

A continuity correction improves approximation:

$$\mathbb{P}(X \le k) \approx \Phi \left(\frac{k + 0.5 - np}{\sqrt{np(1 - p)}} \right).$$

14 Multiple Random Variables: Compact Toolkit

14.1 Joint and Marginal Objects

Given joint density/PMF, obtain marginals by integrating/summing out the other variable, then obtain conditional laws by normalization.

14.2 Conditional Means and Prediction

$\mathbb{E}[Y \mid X]$ is the mean-squared optimal predictor of Y based on X. This is the first conceptual bridge toward regression in ECON 120B and 120C.

14.3 Dependence Summary

- Independence $\Rightarrow$ zero covariance.
- Zero covariance does not generally imply independence.
- Under joint normality, zero covariance and independence coincide.

15 Worked Derivations and Examples

15.1 Example A: CLT-Based Probability for the Sample Mean

Suppose i.i.d. observations have mean $\mu = 50$, standard deviation $\sigma = 12$, and $n = 64$. Approximate

$$\mathbb{P}(48 \le \bar{X}_n \le 52).$$

By CLT,

$$\bar{X}_n \approx N\left(50, \frac{12^2}{64}\right) = N(50, 2.25),$$

so standard error is 1.5 and

$$\mathbb{P}(48 \le \bar{X}_n \le 52) \approx \mathbb{P}\left(\frac{-2}{1.5} \le Z \le \frac{2}{1.5}\right) = \mathbb{P}(-1.333 \le Z \le 1.333) \approx 0.817.$$

15.2 Example B: One-Sample z Test

Observed: $n = 100$, known $\sigma = 12$, sample mean $\bar{x} = 450$. Test

$$H_0 : \mu = 454, \qquad H_a : \mu \ne 454,$$

at $\alpha = 0.05$.

$$z = \frac{450 - 454}{12/\sqrt{100}} = \frac{-4}{1.2} = -3.33.$$

Critical value for two-sided 5% test is 1.96. Since $|z| > 1.96$, reject H_0.

15.3 Example C: Test for Proportion

Suppose $n = 125$, observed count $x = 46$, so $\hat{p} = 0.368$. Test

$$H_0 : p = 0.292, \qquad H_a : p > 0.292.$$

Statistic:

$$z = \frac{0.368 - 0.292}{\sqrt{0.292(1 - 0.292)/125}} \approx 1.85.$$

At $\alpha = 0.10$, right-tail critical value is 1.282, so reject H_0. At $\alpha = 0.05$, critical value 1.645, still reject (borderline but above threshold).

15.4 Example D: Confidence Interval for Mean (Unknown Variance)

Given $n = 30$, $\bar{x} = 599.5$, $s = 84.84$, construct a 95% CI for μ:

$$\bar{x} \pm t_{29,0.975} \frac{s}{\sqrt{30}} \approx 599.5 \pm 2.045 \cdot 15.49 \approx 599.5 \pm 31.7.$$

So CI is approximately $(567.8, 631.2)$.

15.5 Example E: MSE Comparison

Suppose two estimators of the same scalar parameter satisfy:

$$\text{Bias}(\hat{\theta}_1) = 0, \quad \text{Var}(\hat{\theta}_1) = 4,$$

$$\text{Bias}(\hat{\theta}_2) = 1, \quad \text{Var}(\hat{\theta}_2) = 2.$$

Then

$$\text{MSE}(\hat{\theta}_1) = 4, \qquad \text{MSE}(\hat{\theta}_2) = 2 + 1^2 = 3.$$

Even with bias, $\hat{\theta}_2$ is closer to the target in mean squared error.

15.6 Example F: Correlation from Sample Moments

Given data pairs $(70, 60), (80, 65), (90, 85)$:

$$\bar{x} = 80, \quad \bar{y} = 70,$$

$$\widehat{\text{Cov}}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 125,$$

with sample standard deviations $s_x = 10$, $s_y \approx 13.229$, so

$$\hat{\rho} = \frac{125}{10 \cdot 13.229} \approx 0.945.$$

Strong positive linear association, but no causal claim follows from this alone.

16 Proof Appendix

16.1 Markov Inequality

Theorem 16.1. *If $Y \geq 0$ and $a > 0$, then*

$$\mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}[Y]}{a}.$$

Proof. Since $Y \geq a\mathbf{1}\{Y \geq a\}$,

$$\mathbb{E}[Y] \geq \mathbb{E}[a\mathbf{1}\{Y \geq a\}] = a\mathbb{P}(Y \geq a).$$

Divide by a. □

16.2 Chebyshev Inequality

Theorem 16.2. *For random variable X with finite mean μ and variance σ^2,*

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}, \qquad t > 0.$$

Proof. Apply Markov to $Y = (X - \mu)^2$:

$$\mathbb{P}\left((X - \mu)^2 \geq t^2\right) \leq \frac{\mathbb{E}[(X - \mu)^2]}{t^2} = \frac{\sigma^2}{t^2}.$$

□

16.3 Consistency from Bias and Variance Conditions

Proposition 16.3. *If Bias*$(\hat{\theta}_n) \to 0$ *and* $\mathrm{Var}(\hat{\theta}_n) \to 0$*, then* $\hat{\theta}_n \xrightarrow{p} \theta$*.*

Proof. By MSE decomposition,

$$\mathbb{E}[(\hat{\theta}_n - \theta)^2] = \mathrm{Var}(\hat{\theta}_n) + \mathrm{Bias}(\hat{\theta}_n)^2 \to 0.$$

Chebyshev implies

$$\mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{\mathbb{E}[(\hat{\theta}_n - \theta)^2]}{\varepsilon^2} \to 0.$$

□

17 Distribution Toolkit Used in 120A Inference

17.1 Hypergeometric Distribution (Finite Population Sampling)

Suppose a finite population has size N with K successes and $N-K$ failures. Draw n units without replacement and let X be the number of successes drawn. Then

$$\mathbb{P}(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}.$$

This is the exact finite-population model underlying the finite population correction in sampling variance formulas.

17.2 Poisson Distribution as a Binomial Limit

If $X_n \sim \mathrm{Bin}(n, p_n)$ and $np_n \to \lambda > 0$, then for fixed k,

$$\mathbb{P}(X_n = k) \to e^{-\lambda}\frac{\lambda^k}{k!}.$$

So the Poisson distribution appears as the rare-event limit of Bernoulli repetition.

Sketch. Write

$$\mathbb{P}(X_n = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Use

$$\binom{n}{k}\left(\frac{1}{n}\right)^k \to \frac{1}{k!}, \qquad \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}, \qquad \left(1 - \frac{\lambda}{n}\right)^{-k} \to 1.$$

Multiply limits. □

17.3 Exponential and Gamma Family

The exponential distribution with rate λ has density

$$f(x) = \lambda e^{-\lambda x}\mathbf{1}\{x \geq 0\},$$

mean $1/\lambda$, variance $1/\lambda^2$, and memoryless property.

If $X_1, \ldots, X_m$ are i.i.d. exponential(λ), then

$$\sum_{i=1}^{m} X_i \sim \mathrm{Gamma}(m, \lambda).$$

This is often used for waiting-time and count-process modeling.

17.4 Chi-square, t, and F Distributions

Let $Z_1, \dots, Z_\nu$ be i.i.d. $N(0,1)$. Then

$$Q = \sum_{j=1}^{\nu} Z_j^2 \sim \chi^2_\nu.$$

If $Z \sim N(0,1)$ and $Q \sim \chi^2_\nu$ independent, then

$$T = \frac{Z}{\sqrt{Q/\nu}} \sim t_\nu.$$

If $Q_1 \sim \chi^2_{\nu_1}$ and $Q_2 \sim \chi^2_{\nu_2}$ independent, then

$$F = \frac{(Q_1/\nu_1)}{(Q_2/\nu_2)} \sim F_{\nu_1,\nu_2}.$$

These distributions are central for exact finite-sample inference in Gaussian models.

Proposition 17.1 (Exact t pivot for Gaussian sample mean)**.** *If* $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$*, then*

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}.$$

Sketch. Under normality,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1), \qquad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1},$$

and these two statistics are independent. Their ratio has a t_{n-1} law by definition. □

18 Asymptotic Toolkit Beyond LLN and CLT

18.1 Modes of Convergence

Key modes:

- $X_n \xrightarrow{p} X$ (in probability),
- $X_n \xrightarrow{d} X$ (in distribution),
- $X_n \xrightarrow{a.s.} X$ (almost surely).

18.2 Continuous Mapping Theorem

Theorem 18.1. *If* $X_n \xrightarrow{p} X$ *and* g *is continuous at points in the support of* X*, then*

$$g(X_n) \xrightarrow{p} g(X).$$

If $X_n \xrightarrow{d} X$*, then* $g(X_n) \xrightarrow{d} g(X)$*.*

18.3 Slutsky's Theorem

Theorem 18.2. *If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ for constant c, then*

$$X_n + Y_n \xrightarrow{d} X + c, \quad X_n Y_n \xrightarrow{d} cX, \quad \frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c} \quad (c \neq 0).$$

18.4 Delta Method

Theorem 18.3 (Univariate delta method)**.** *If*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, V),$$

and g is differentiable at θ_0 with $g'(\theta_0) \neq 0$, then

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{d} N(0,\, [g'(\theta_0)]^2 V).$$

Sketch. First-order Taylor expansion:

$$g(\hat{\theta}_n) - g(\theta_0) = g'(\theta_0)(\hat{\theta}_n - \theta_0) + r_n,$$

where $r_n = o_p(n^{-1/2})$. Multiply by $\sqrt{n}$ and apply Slutsky. □

Example 18.4 (Variance stabilization for proportions)**.** If $\hat{p} \xrightarrow{p} p$ and $\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, p(1-p))$, then with

$$g(p) = \arcsin(\sqrt{p}), \qquad g'(p) = \frac{1}{2\sqrt{p(1-p)}},$$

the delta method gives an asymptotic variance close to $1/4$ after scaling, useful for certain interval constructions.

19 Power, Size, and Sample Size in Hypothesis Testing

19.1 Power Function for z Test (Known Variance)

Consider testing

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_a : \mu > \mu_0$$

with rejection rule

$$\bar{X}_n > \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}.$$

At true value μ_1, power is

$$\pi(\mu_1) = \mathbb{P}_{\mu_1}\left(\bar{X}_n > \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}\right) = 1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma}\right).$$

19.2 Sample Size from Target Power

Given desired size α, power $1 - \beta$, and effect $\Delta = \mu_1 - \mu_0 > 0$, solve

$$1 - \beta = 1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{n}\Delta}{\sigma}\right),$$

which implies

$$n \geq \left(\frac{(z_{1-\alpha} + z_{1-\beta})\sigma}{\Delta}\right)^2.$$

This is the standard one-sided planning formula in Gaussian designs.

19.3 Two-sided Designs

For two-sided tests, replace $z_{1-\alpha}$ by $z_{1-\alpha/2}$ in the expression above.

20 Extended Worked Derivations

20.1 Problem 1: Derive the Variance of the Sample Proportion

Let $D_i \in \{0, 1\}$ be i.i.d. with $\mathbb{P}(D_i = 1) = p$. Then

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} D_i.$$

Derive $E[\hat{p}]$ and $\mathrm{Var}(\hat{p})$.

Solution. Since $E[D_i] = p$,

$$E[\hat{p}] = \frac{1}{n} \sum_{i=1}^{n} E[D_i] = p.$$

Also $\mathrm{Var}(D_i) = p(1-p)$ and independence gives

$$\mathrm{Var}(\hat{p}) = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} D_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(D_i) = \frac{p(1-p)}{n}.$$

□

20.2 Problem 2: CLT-based CI for a Mean

Suppose $n = 64$, $\bar{x} = 102$, known $\sigma = 20$. Build a 95% CI for μ.

$$102 \pm 1.96 \cdot \frac{20}{8} = 102 \pm 4.9.$$

Therefore the interval is $(97.1, 106.9)$.

20.3 Problem 3: CI Width and Sample Size

Target a 95% CI half-width of 2 with known $\sigma = 15$.

$$n \geq \left(\frac{1.96 \cdot 15}{2}\right)^2 = 216.09.$$

So minimum integer sample size is $n = 217$.

20.4 Problem 4: Two-sided z Test and p-value

Observed $z = 2.21$ in a two-sided test. Then

$$p\text{-value} = 2\big(1 - \Phi(2.21)\big) \approx 0.027.$$

Reject at 5%, not at 1%.

20.5 Problem 5: One-sided t Test

Given $n = 16$, $\bar{x} = 8.5$, $s = 2$, test $H_0 : \mu = 8$ vs $H_a : \mu > 8$.

$$t = \frac{8.5 - 8}{2/\sqrt{16}} = 1.0.$$

With $df = 15$, one-sided 5% critical value is about 1.753. Since $1.0 < 1.753$, fail to reject.

20.6 Problem 6: CI-Test Duality in Practice

If a 95% CI for μ is $(4.2, 6.1)$, then:

- test of $H_0 : \mu = 5$ at 5% two-sided fails to reject,
- test of $H_0 : \mu = 7$ at 5% two-sided rejects.

No extra calculation is needed because interval inversion already encodes the test.

20.7 Problem 7: Correlation and Linear Transformation

If $Y = a + bX$ with $b > 0$, then $\mathrm{Corr}(X, Y) = 1$. If $b < 0$, then $\mathrm{Corr}(X, Y) = -1$.

Proof.

$$\mathrm{Cov}(X, a + bX) = b\,\mathrm{Var}(X), \qquad \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(a + bX)} = \sqrt{\mathrm{Var}(X)\,b^2\,\mathrm{Var}(X)} = |b|\,\mathrm{Var}(X).$$

Hence

$$\mathrm{Corr}(X, Y) = \frac{b\,\mathrm{Var}(X)}{|b|\,\mathrm{Var}(X)} = \mathrm{sign}(b).$$

□

20.8 Problem 8: Chebyshev Bound vs Normal Approximation

Suppose $\mu = 100$, $\sigma = 20$, $n = 100$.

$$\mathrm{Var}(\bar{X}) = \frac{20^2}{100} = 4, \quad sd(\bar{X}) = 2.$$

Chebyshev for $|\bar{X} - \mu| \geq 4$:

$$\mathbb{P}(|\bar{X} - \mu| \geq 4) \leq \frac{4}{16} = 0.25.$$

CLT approximation:

$$\mathbb{P}(|\bar{X} - \mu| \geq 4) \approx 2(1 - \Phi(2)) \approx 0.0455.$$

Chebyshev is general but conservative.

20.9 Problem 9: Approximate Distribution of a Nonlinear Estimator

Let $\hat{\mu}_n = \bar{X}_n$ with

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

For $g(\mu) = \mu^2$,

$$\sqrt{n}(\hat{\mu}_n^2 - \mu^2) \xrightarrow{d} N\left(0,\, 4\mu^2\sigma^2\right)$$

by the delta method.

20.10 Problem 10: Finite Population Correction Impact

Take SRS without replacement from $N = 1000$ and $n = 400$.

$$\text{FPC} = \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{600}{999}} \approx 0.775.$$

Ignoring FPC would overstate the standard error by about $1/0.775 - 1 \approx 29\%$.

21 Derivation and Proof Checklist

Before finalizing an inference statement, verify:

1. sampling assumptions (randomness, independence, replacement conditions),
2. moment conditions (finite variance for LLN/CLT versions used),
3. finite-sample vs asymptotic justification (exact t vs approximate z),
4. correct standard error formula (including finite population correction when needed),
5. alternative hypothesis direction and matching rejection region,
6. interpretation language ("fail to reject" is not "accept as true"),
7. distinction between dependence and causality in multivariate summaries.

22 Additional Proof Notes

22.1 Variance of the Sample Mean with Correlation Terms

For general (not necessarily independent) observations $X_1, \ldots, X_n$,

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) + \frac{2}{n^2}\sum_{1\le i<j\le n}\text{Cov}(X_i, X_j).$$

Under independence, covariance terms vanish. This derivation clarifies why dependence can severely inflate standard errors.

22.2 Deriving the Wald Statistic

Suppose

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, V),$$

and a consistent variance estimator $\hat{V}_n \xrightarrow{p} V$ is available. Then

$$W_n = \frac{(\hat{\theta}_n - \theta_0)^2}{\hat{V}_n/n} = \left(\frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{\sqrt{\hat{V}_n}}\right)^2 \xrightarrow{d} \chi_1^2.$$

This is the one-parameter Wald test foundation used in many econometric settings.

22.3 CLT for Sample Proportions from Bernoulli CLT

Let $D_i \sim \text{Bernoulli}(p)$ i.i.d. Then

$$\sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} = \frac{\sum_{i=1}^{n}(D_i - p)}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

Hence large-sample proportion tests and CIs are direct consequences of the CLT.

22.4 Asymptotic Normality of Standardized Mean with Estimated Variance

By CLT,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1).$$

If $S_n^2 \xrightarrow{p} \sigma^2$, then by Slutsky,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1).$$

This gives asymptotic justification for replacing unknown σ with S_n outside exact normal finite-sample settings.

22.5 One-sided vs Two-sided Logic

Two-sided tests allocate type-I error to both tails:

$$\text{reject if } |Z| > z_{1-\alpha/2}.$$

Right-sided tests concentrate rejection probability on one tail:

$$\text{reject if } Z > z_{1-\alpha}.$$

The same observed statistic can be significant one-sided but not two-sided. Direction must be chosen from economic question before looking at the sample.

23 Comprehensive Practice Set (With Solution Sketches)

23.1 Q1: Bernoulli and Binomial Setup

Let $X \sim \text{Bin}(n,p)$.

1. Show $X = \sum_{i=1}^{n} B_i$ with i.i.d. Bernoulli(p) indicators.
2. Use this representation to derive $\mathbb{E}[X]$ and $\text{Var}(X)$.

Solution sketch: linearity yields $\mathbb{E}[X] = np$. Independence yields $\text{Var}(X) = np(1-p)$.

23.2 Q2: Computing an Exact Binomial Probability

Suppose $n = 20$, $p = 0.4$. Compute $\mathbb{P}(X = 8)$:

$$\mathbb{P}(X = 8) = \binom{20}{8}(0.4)^8(0.6)^{12} \approx 0.1797.$$

The point is procedural fluency in mapping a word problem to (n,p,x).

23.3 Q3: Approximate Binomial Tail Using Normal Approximation

With $n = 200$, $p = 0.5$, approximate $\mathbb{P}(X \geq 115)$. Using continuity correction:

$$\mathbb{P}(X \geq 115) \approx \mathbb{P}\Big(Z \geq \frac{114.5 - 100}{\sqrt{50}}\Big) = \mathbb{P}(Z \geq 2.051) \approx 0.0201.$$

23.4 Q4: LLN Interpretation

Suppose daily returns have mean 0.1 and finite variance. Interpret

$$\bar{R}_n \xrightarrow{p} 0.1.$$

Solution sketch: long-run average return stabilizes near 0.1; realized sample means fluctuate less as n grows, but finite-sample noise remains.

23.5 Q5: CLT Standardization

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

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}.$$

Explain why Z_n is approximately standard normal for large n. **Solution sketch:** this is exactly the CLT statement; approximation quality increases with sample size and depends on tail behavior.

23.6 Q6: Building a 99% Mean CI

Given $n = 49$, $\bar{x} = 75$, known $\sigma = 14$, build a 99% CI:

$$75 \pm 2.576 \cdot \frac{14}{7} = 75 \pm 5.152,$$

so CI is $(69.848, 80.152)$.

23.7 Q7: Unknown Variance CI

Given $n = 12$, $\bar{x} = 20$, $s = 3$, build a 95% CI. Use $t_{11,0.975} \approx 2.201$:

$$20 \pm 2.201 \cdot \frac{3}{\sqrt{12}} = 20 \pm 1.906.$$

Interval: $(18.094, 21.906)$.

23.8 Q8: Testing a Mean (Right-tail)

Test $H_0 : \mu = 50$ vs $H_a : \mu > 50$ with known $\sigma = 10$, $n = 36$, $\bar{x} = 53$.

$$z = \frac{53 - 50}{10/6} = 1.8.$$

At 5%, critical value is 1.645, so reject H_0. At 1%, critical value is 2.326, fail to reject.

23.9 Q9: Testing a Proportion

Suppose $n = 400$, observed support rate $\hat{p} = 0.56$. Test $H_0 : p = 0.5$ vs $H_a : p \neq 0.5$.

$$z = \frac{0.56 - 0.5}{\sqrt{0.5 \cdot 0.5/400}} = \frac{0.06}{0.025} = 2.4.$$

Two-sided p-value is about 0.016, so reject at 5%.

23.10 Q10: Type I and Type II Tradeoff

Why does lowering α generally reduce power for fixed n? **Solution sketch:** lower α moves critical values farther into tails, shrinking rejection region under both null and alternative distributions unless sample size or effect size compensates.

23.11 Q11: Correlation Pitfall

Suppose ice-cream sales and drownings are positively correlated. Does this imply causality? **Solution sketch:** no. A third variable (temperature) can move both. Correlation is a descriptive dependence measure, not a causal identification strategy.

23.12 Q12: Finite Population Correction Decision

When should FPC be used? **Solution sketch:** sampling without replacement from a finite population where sampling fraction n/N is non-negligible (often above 5%). Otherwise the correction is close to one and can be ignored.

23.13 Q13: MSE Ranking Exercise

Estimator A: bias 0.2, variance 1.0. Estimator B: bias 0, variance 1.3.

$$\text{MSE}(A) = 1.0 + 0.2^2 = 1.04, \qquad \text{MSE}(B) = 1.3.$$

A is preferred by MSE despite bias.

23.14 Q14: CI Width and Confidence Level

Holding data fixed, compare 90%, 95%, 99% CIs. **Solution sketch:** widths are proportional to critical values:

$$1.645 < 1.96 < 2.576.$$

Higher confidence means wider interval.

23.15 Q15: Delta Method Mechanics

Given asymptotic normality of $\hat{\theta}_n$ and smooth g, list the three steps.

1. write first-order Taylor expansion around θ_0,
2. multiply by $\sqrt{n}$ and isolate leading term,
3. apply Slutsky to replace random slope by constant derivative.

23.16 Q16: Interpreting "Fail to Reject"

If a test fails to reject H_0, what does that mean? **Solution sketch:** sample evidence is not strong enough against H_0 at the chosen significance level; it does not prove the null is true.

23.17 Q17: One-sided p-value Conversion

If two-sided p-value is 0.04 and estimate sign matches right-tail alternative, one-sided p-value is approximately 0.02. If sign is opposite, one-sided p-value is close to 0.98.

23.18 Q18: Standard Error Intuition

For sample mean,

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

Doubling precision (halving SE) requires quadrupling sample size.

23.19 Q19: Independence vs Zero Correlation

Provide an example where $\mathrm{Cov}(X, Y) = 0$ but dependence exists. **Solution sketch:** let $X \sim N(0, 1)$ and $Y = X^2$. Then $\mathrm{Cov}(X, Y) = E[X^3] - E[X]E[X^2] = 0$, but Y is fully determined by X, so dependence is strong.

23.20 Q20: Inference Workflow

Minimum rigorous sequence:

1. define parameter and sampling context,
2. derive estimator and standard error,
3. choose exact or asymptotic distributional justification,
4. report estimate, interval, and test with clear assumptions,
5. state limits: sampling bias risk, model assumptions, and interpretation scope.

24 Assumption Diagnostics and Robustness Notes

24.1 When Independence is Dubious

Many formulas above use independent sampling. If observations are clustered, serially correlated, or network-dependent, naive standard errors are too small. A practical implication is to replace i.i.d.-based standard errors with dependence-robust alternatives once dependence structure is identified.

24.2 Heavy Tails and CLT Speed

CLT may converge slowly under heavy tails or strong skewness. In such cases:

1. increase sample size requirements,
2. inspect tail behavior empirically,

3. prefer robust summaries (medians, trimmed means) when outliers dominate means,
4. use bootstrap methods where analytic approximations are fragile.

24.3 Finite-Sample vs Asymptotic Inference

Finite-sample exact pivots (for normal models) are preferred when available. Asymptotic methods are justified when:

- no exact finite-sample law is tractable,
- sample size is sufficiently large for approximation quality,
- regularity conditions are credible in context.

Inference writeups should explicitly state which regime is being used.

24.4 Interpretation Discipline

Statistical significance is not economic significance. A coherent report should always include:

1. effect magnitude in natural units,
2. uncertainty quantification (CI or standard error),
3. practical relevance in application context,
4. assumptions required for interpretation.

24.5 Checklist for Empirical Credibility in 120A-Level Work

1. Data provenance documented.
2. Sampling mechanism explained.
3. Parameter and estimator clearly matched.
4. Distributional approximations justified.
5. Test direction chosen before examining outcomes.
6. Null and alternative statements economically interpretable.
7. Final conclusions separate statistical and substantive claims.

25 Proof Prompts for Mastery

25.1 Prompt 1: Show that Uncorrelatedness Does Not Imply Independence

Construct random variables with zero covariance but dependence.

One construction. Let X take values $-1, 0, 1$ each with probability $1/3$, and define $Y = X^2$. Then

$$E[X] = 0, \quad E[Y] = \frac{2}{3}, \quad E[XY] = E[X^3] = 0.$$

So

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.$$

But Y is a deterministic function of X, so they are dependent. □

25.2 Prompt 2: Verify That Sample Mean is Minimum-Variance Linear Unbiased under i.i.d.

Consider linear estimators of μ of form

$$\tilde{\mu} = \sum_{i=1}^{n} w_i X_i, \qquad \sum_{i=1}^{n} w_i = 1.$$

Then

$$\mathrm{Var}(\tilde{\mu}) = \sigma^2 \sum_{i=1}^{n} w_i^2.$$

By Cauchy–Schwarz,

$$\left(\sum_{i=1}^{n} w_i\right)^2 \le n \sum_{i=1}^{n} w_i^2 \quad \Rightarrow \quad \sum_{i=1}^{n} w_i^2 \ge \frac{1}{n},$$

with equality at $w_i = 1/n$. Hence $\bar{X}_n$ minimizes variance among unbiased linear estimators.

25.3 Prompt 3: Derive One-Sided p-value Formula

For right-tail z test with observed z_{obs},

$$p\text{-value} = \mathbb{P}_{H_0}(Z \ge z_{obs}) = 1 - \Phi(z_{obs}).$$

For left-tail tests use $\Phi(z_{obs})$. For two-sided symmetric tests:

$$p\text{-value} = 2 \min\{\Phi(z_{obs}), 1 - \Phi(z_{obs})\} = 2(1 - \Phi(|z_{obs}|)).$$

25.4 Prompt 4: Show CI Width Shrinks at Rate $n^{-1/2}$

In z-CI for mean with known variance,

$$\text{half-width} = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

Thus doubling n multiplies width by $1/\sqrt{2}$, and reducing width by half requires quadrupling n.

25.5 Prompt 5: Invert a Two-Sided z Test to Recover the CI

Reject $H_0 : \mu = \mu_0$ if

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{1-\alpha/2}.$$

Equivalently do not reject if

$$\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \le \mu_0 \le \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

The non-rejected set of null values is exactly the confidence interval.

26 Compact Summary for Exam Review

1. **Probability core:** axioms, conditional probability, Bayes, independence.
2. **Distribution core:** Bernoulli/Binomial/Normal, PMF-PDF-CDF relationships.
3. **Sampling core:** $\mathbb{E}[\bar{X}] = \mu$, $\text{Var}(\bar{X}) = \sigma^2/n$.
4. **Asymptotics:** LLN justifies consistency; CLT justifies normal approximations.
5. **Estimator quality:** bias, variance, MSE, consistency.
6. **Inference:** CI and hypothesis tests are two views of the same pivot logic.
7. **Multivariate basics:** marginals, conditionals, covariance, correlation, limits of interpretation.

27 Next-Step Bridge

This 120A toolkit is exactly what is needed for 120B/120C topics: regression estimators, causal inference design, and asymptotic theory for econometric models.