STA 131A: Introduction to Probability Theory

Lecture 26: Conclusion

Dogyoon Song

Spring 2026, UC Davis

Announcements

Final exam: Thu, June 11, 1:00–3:00 PM in Wellman Hall 226

- **Cumulative coverage:** Lectures 1–26
- **Arrive early:** The exam starts at 1:00 PM and ends at 3:00 PM sharp
- **Three hand-written cheat sheets:** Letter-size (8.5" $\times$ 11"), double-sided, formulas/notes
- **Calculator:** A simple non-graphing scientific calculator is allowed
- **No other materials:** No textbooks, notes, electronic devices, etc.
- **SDC accommodations:** Confirm your schedule with AES online ASAP

Preparation:

- Review lecture slides, homework, midterms, practice final, and solution keys
- Practice translating word problems into sample spaces, events, random variables, and distributions

Course evaluation: Please share your feedback comments by Thu, June 4

Office hours: 2:30–3:30 PM today at MSB 4220

Agenda

Goal: Reconnect the course as one probability-theory storyline and prepare for the final

First: spillover from Lecture 25

- CLT approximation for sums and averages
- Binomial normal approximation and continuity correction

Then: final review and synthesis

- Foundations: events, conditioning, Bayes, independence, counting
- Random variables: PMFs, PDFs, CDFs, named models
- Multiple variables: joint, marginal, conditional distributions
- Summary tools: expectation, variance, covariance, conditional variance
- Transformations, MGFs, random sums
- Deviation inequalities, WLLN, CLT

Exam preparation: Know definitions, workflows, computations, and interpretations

CLT approximation workflow

Let $X_1, \ldots, X_n$ be i.i.d. with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2, \qquad \sigma > 0.$$

For the sum $S_n = X_1 + \cdots + X_n$,

$$S_n \approx N(n\mu, n\sigma^2), \qquad \Longrightarrow \qquad P(S_n \le c) \approx \Phi\left(\frac{c - n\mu}{\sigma\sqrt{n}}\right).$$

For the sample average $\bar{X}_n = S_n/n$,

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right), \qquad \Longrightarrow \qquad P(\bar{X}_n \le a) \approx \Phi\left(\frac{a - \mu}{\sigma/\sqrt{n}}\right).$$

Workflow: identify μ, σ^2, n, then center, scale, and use Φ.

Example: Polling error and sample size (1/2)

Example (Polling; recall from Lecture 25)

We poll n voters and record the fraction M_n who support a candidate. If p is the true support probability, write

$$M_n = \frac{X_1 + \cdots + X_n}{n},$$

where the X_i's are independent $\mathrm{Bernoulli}(p)$. Then $\mathbb{E}[M_n] = p$ and $\mathrm{Var}(M_n) = \frac{p(1-p)}{n}$.
By the CLT,

$$M_n \approx G, \qquad G \sim N\left(p, \frac{p(1-p)}{n}\right).$$

For example, when $p = \frac{1}{2}$, $n = 100$, $\epsilon = 0.1$, the CLT gives

$$P(|M_n - p| \geq 0.1) \approx 2(1 - \Phi(2)) \approx 0.046.$$

By contrast, Chebyshev's inequality gives

$$P(|M_n - p| \geq 0.1) \leq \frac{0.25}{100(0.1)^2} = 0.25.$$

Example: Polling error and sample size (2/2)

Example (Polling sample size)

Question: How large should n be so that

$$P(|M_n - p| \leq 0.01) \approx 0.95?$$

Using the CLT, we want approximately

$$\frac{0.01\sqrt{n}}{\sqrt{p(1-p)}} \geq 1.96.$$

If p is unknown, use the worst-case bound $p(1-p) \leq 1/4$. Then it is sufficient to require

$$\frac{0.01\sqrt{n}}{1/2} \geq 1.96 \qquad \Longrightarrow \qquad n \geq 9604.$$

For comparison, using Chebyshev's inequality would give a more conservative conclusion:

$$\frac{1}{4n(0.01)^2} \leq 0.05, \qquad \Longrightarrow \qquad n \geq 50000.$$

De Moivre–Laplace approximation to the binomial

We consider a special case of CLT. Let

$$S_n \sim \mathrm{Binomial}(n, p).$$

We can write

$$S_n = X_1 + \cdots + X_n,$$

where $X_i \sim \mathrm{Bernoulli}(p)$ independently.

Since

$$\mathbb{E}[X_i] = p, \qquad \mathrm{Var}(X_i) = p(1-p),$$

the CLT gives the normal approximation

$$P(S_n \leq c) \approx \Phi\left(\frac{c - np}{\sqrt{np(1-p)}}\right).$$

This is called the **de Moivre–Laplace normal approximation** to the binomial.

Continuity correction

A binomial random variable is discrete, but the normal approximation is continuous.

To approximate probabilities involving integer values, use a **continuity correction**:

$$P(k \le S_n \le \ell) \approx P(k - 0.5 \le G \le \ell + 0.5),$$

where

$$G \sim N(np,\ np(1-p)).$$

For example,

$$P(10 \le S_n \le 20) \approx P(9.5 \le G \le 20.5).$$

For one-sided events:

$$P(S_n \le k) \approx P(G \le k + 0.5), \qquad P(S_n \ge k) \approx P(G \ge k - 0.5).$$

Message: the adjustment by 0.5 better aligns discrete bars with continuous area.

Example: Binomial normal approximation (1/2)

Example

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

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

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

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

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,

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

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 $G \sim N(40, 24)$.

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

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 accuracy of normal approximation.

STA 131A in one storyline

Theme: How do we model uncertainty, compute probabilities, and understand long-run behavior?

Stage	Main tools
Events	sample space, probability laws, set operations
Conditioning	conditional probability, total probability, Bayes' rule
Random variables	PMFs, PDFs, CDFs, named distributions
Multiple variables	joint, marginal, and conditional distributions
Summary quantities	expectation, variance, covariance, correlation
Transformations	derived distributions, MGFs, random sums
Large-sample behaviors	Markov, Chebyshev, WLLN, CLT

Exam mindset: Identify the structure first; then choose the right tool.

Probability fundamentals

Events are sets. Use unions, intersections, complements, and partitions carefully:

$$P(A^c) = 1 - P(A), \qquad P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Conditional probability

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

Law of total probability If $\{A_i\}$ is a partition, then

$$P(B) = \sum_i P(B \mid A_i) P(A_i).$$

Bayes' rule

$$P(A_j \mid B) = \frac{P(B \mid A_j) P(A_j)}{\sum_i P(B \mid A_i) P(A_i)}.$$

Key task: identify what is observed and what is hidden.

Independence: Events and random variables

Events:

$$A, B \text{ independent} \quad \iff \quad P(A \cap B) = P(A)P(B).$$

Independence is not the same as disjointness.

Random variables:

$$X, Y \text{ independent} \quad \iff \quad p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

in the discrete case, or

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

in the jointly continuous case.

Consequences of independence:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \qquad \mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Caution: independence implies zero covariance, but not vice versa.

Counting and discrete random variables

For finite equally likely outcomes,

$$P(A) = \frac{|A|}{|\Omega|}.$$

Counting tools

- Product rule: multiply choices across stages
- Permutations: order matters

$$\frac{n!}{(n-k)!}$$

- Combinations: order does not matter

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

  - Exactly k successes in n trials: choose the k success positions.

Discrete random variable:

$$p_X(x) = P(X = x), \qquad \sum_x p_X(x) = 1.$$

Common discrete models

Model	What it models	Key facts
Bernoulli(p)	one success/failure trial	$\mathbb{E}[X] = p, \ \mathrm{Var}(X) = p(1-p)$
Binomial(n, p)	number of successes in n independent trials	$\mathbb{E}[X] = np, \ \mathrm{Var}(X) = np(1-p)$
Geometric(p)	trial number of first success	$P(X = k) = (1-p)^{k-1}p$
Poisson(λ)	count in fixed time/space window	$\mathbb{E}[X] = \mathrm{Var}(X) = \lambda$

Key recognition skill: Identify whether the problem is about one trial, fixed number of trials, waiting time, or count in a window.

Continuous random variables: PDF and CDF

PDF

$$P(X \in A) = \int_A f_X(x)\,dx, \qquad f_X(x) \geq 0, \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

For continuous random variables,

$$P(X = x) = 0.$$

CDF

$$F_X(x) = P(X \leq x).$$

If X is continuous,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \qquad f_X(x) = F_X'(x)$$

where differentiable.

Key task: probabilities are areas, not point values of the PDF.

Common continuous models

Model	What it models	Key facts
Uniform(a, b)	equally likely on an interval	$\mathbb{E}[X] = \frac{a+b}{2}$, $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$
Exponential(λ)	continuous waiting time	$\mathbb{E}[X] = 1/\lambda$, $\mathrm{Var}(X) = 1/\lambda^2$, $P(X > t) = e^{-\lambda t}$
Normal(μ, σ^2)	aggregate noise, CLT approximations	standardize by $Z = (X - \mu)/\sigma$

Key recognition skill: interval model, waiting-time model, or normal approximation

Normal random variables and standardization

If $X \sim N(\mu, \sigma^2)$, then

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

Thus,

$$P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Useful identities

$$P(Z > z) = 1 - \Phi(z), \qquad \Phi(-z) = 1 - \Phi(z),$$

and

$$P(|Z| \leq z) = 2\Phi(z) - 1.$$

Common mistake: confusing variance σ^2 and standard deviation σ.

Joint distributions and conditioning

Joint distribution

$$p_{X,Y}(x,y) = P(X = x, Y = y) \qquad \text{or} \qquad P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy.$$

- **Geometry matters:** draw the support before choosing integration limits.

Marginals

$$p_X(x) = \sum_y p_{X,Y}(x,y), \qquad f_X(x) = \int f_{X,Y}(x,y)\,dy.$$

Conditional distribution

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

Key task: identify the slice and renormalize it.

Expectation, variance, and conditioning

Expectation

$$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x) \quad \text{or} \quad \mathbb{E}[g(X)] = \int g(x)f_X(x)\,dx.$$

Variance

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

Total expectation

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

Conditional expectation as prediction

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

Interpretation: $\mathbb{E}[X \mid Y]$ is the average value of X after observing Y.

Covariance, correlation, and total variance

Covariance

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

Correlation

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Variance of sums

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

Law of total variance

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

Interpretation:

total variability = average within-condition variability + between-condition variability.

Derived distributions

When $Y = g(X)$ or $Z = g(X, Y)$, find the distribution by tracking where probability goes.

Discrete case

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x).$$

Continuous case: CDF-first method is often safest

$$F_Y(y) = P(Y \leq y), \qquad f_Y(y) = F'_Y(y).$$

For joint transformations:

- Draw the support.
- Translate the event or transformation into inequalities.
- Integrate over the correct region.

Common mistake: writing integration limits before understanding the support.

MGFs, sums, and random sums

Moment generating function

$$M_X(t) = \mathbb{E}[e^{tX}].$$

- If M_X exists near 0, $M_X^{(k)}(0) = \mathbb{E}[X^k]$.
- For independent X, Y,

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Random sums: If

$$S = \sum_{i=1}^{N} X_i,$$

where N is independent of i.i.d. X_i's with mean μ and variance σ^2, then

$$\begin{aligned}
\mathbb{E}[S] &= \mathbb{E}[N]\mu, \\
\mathrm{Var}(S) &= \mathbb{E}[N]\sigma^2 + \mu^2 \mathrm{Var}(N), \\
M_S(t) &= M_N(\log M_X(t)).
\end{aligned}$$

Inequalities and limit theorems

Markov's inequality

$$X \geq 0 \quad \Longrightarrow \quad P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Chebyshev's inequality

$$P(|X - \mathbb{E}[X]| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}.$$

Weak law of large numbers If X_i are i.i.d. with mean μ, then

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

Central limit theorem

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \approx Z, \qquad Z \sim N(0,1) \qquad \text{for large } n.$$

Big picture: WLLN says averages stabilize; CLT describes the remaining fluctuations.

Final exam checklist

When reading a problem, ask:

- **Is it about events?** Use set operations, conditioning, total probability, Bayes, independence.
- **Is the sample space finite/equally likely?** Use counting.
- **Is a random variable discrete or continuous?** Use PMF/PDF/CDF accordingly.
- **Are there multiple random variables?** Use joint, marginal, conditional distributions.
- **Is it asking for a mean or variance?** Use expectation rules, covariance, total variance.
- **Is it a transformation?** Use CDF-first or support geometry.
- **Is it a sum or average?** Use MGFs, random-sum formulas, Chebyshev, WLLN, or CLT.

Strategy: identify the structure first, choose the right tool, and then compute carefully.

Textbook coverage: [BT08, Ch. 1.1–5.4]

How to use your cheat sheets

Your cheat sheets should help you recall **structure**, not just formulas.

Good items to include:

- Definitions: conditional probability, independence, PMF/PDF/CDF
- Core formulas: total probability, Bayes, expectation, variance, covariance
- Named distributions: PMFs/PDFs, means, variances, MGFs
- Workflows: normal standardization, joint density integration, CDF-first transformations
- Limit tools: Markov, Chebyshev, WLLN, CLT, continuity correction

Practice idea: Try solving a practice final once using only your cheat sheets. Then revise the sheets based on what you actually needed.

Final words

- **Probability is a language for uncertainty:** events, random variables, distributions, and expectations are its grammar.
- **Conditioning is the organizing principle:** many hard-looking problems become manageable once you condition on the right information.
- **Random variables turn outcomes into quantities:** PMFs, PDFs, CDFs, and joint distributions let us compute and reason systematically.
- **Dependence matters:** covariance, correlation, conditional variance, and independence describe how random quantities move together.
- **Large-sample theory explains why statistics works:** averages stabilize, and their remaining fluctuations become approximately normal.

Best of luck on the final exam!

References

Dimitri Bertsekas and John N Tsitsiklis.
Introduction to probability, volume 1.
Athena Scientific, 2nd edition, 2008.