MT2 Review

Math-UA 185 Prob Stat

Ed Chen

April 2, 2026

Content

Recent Material

- Law of Large Numbers (Strong, Weak, Empirical Prob.)
- Central Limit Theorem

Older Material

- Joint Distributions (PDF, CDF, Marginals)
- Covariance, Correlation
- Poisson Processes
- Markov, Chebyshev, Jensen Inequalities
- Moments, Exp/Var of Max/Min - not on this slideset

Topics are exhaustive, but this slideset is not!

- Review past exams, homework, quizzes.

Finite Moments

Having a **finite mean** (μ) and **finite variance** (σ^2) is equivalent to having finite first and second moments: $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[X^2] < \infty$. (both discrete and cont!)

Bounded RVs

If a random variable is strictly **bounded** (its support is completely within some finite interval $[a, b]$), all of its moments are finite. The converse is **not true**, for instance the exponential dist.

Weak Law of Large Numbers (WLLN)

For independent, identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$ with **finite mean** μ, the sample mean $\overline{X}_n$ converges in probability to μ.
For **any strictly positive fixed constant** $c > 0$:

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mu| > c) = 0$$

Weak Law of Large Numbers (WLLN)

For independent, identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$ with **finite mean** μ, the sample mean $\overline{X}_n$ converges in probability to μ.
For **any strictly positive fixed constant** $c > 0$:

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mu| > c) = 0$$

This is the limit of Chebyshev's inequality!

Strong Law of Large Numbers (SLLN)

For i.i.d. random variables with **finite mean** μ, the sample mean $\overline{X}_n$ converges to μ **almost surely** (with probability 1).

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1$$

LLN Sanity Check

Q: Let X_i be an i.i.d. sequence with mean μ. What does

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mu| < 0.001)$$

equal?

LLN Sanity Check

Q: Let X_i be an i.i.d. sequence with mean μ. What does

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| < 0.001)$$

equal?

A: It equals 1. Since

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > 0.001) = 0$$

by WLLN, the complement converges to 1.

LLN Sanity Check

Q: Let X_i be an i.i.d. sequence with mean μ. What does

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| < 0.001)$$

equal?

A: It equals 1. Since

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > 0.001) = 0$$

by WLLN, the complement converges to 1.

Q: Does the SLLN imply the WLLN?

LLN Sanity Check

Q: Let X_i be an i.i.d. sequence with mean μ. What does

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| < 0.001)$$

equal?

A: It equals 1. Since

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > 0.001) = 0$$

by WLLN, the complement converges to 1.

Q: Does the SLLN imply the WLLN?

A: Yes. Convergence in probability is weaker than almost sure convergence.

LLN Sanity Check

Q: Let X_i be an i.i.d. sequence with mean μ. What does

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| < 0.001)$$

equal?

A: It equals 1. Since

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > 0.001) = 0$$

by WLLN, the complement converges to 1.

Q: Does the SLLN imply the WLLN?

A: Yes. Convergence in probability is weaker than almost sure convergence.

Q: Does the WLLN imply the SLLN?

LLN Sanity Check

Q: Let X_i be an i.i.d. sequence with mean μ. What does

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| < 0.001)$$

equal?

A: It equals 1. Since

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > 0.001) = 0$$

by WLLN, the complement converges to 1.

Q: Does the SLLN imply the WLLN?

A: Yes. Convergence in probability is weaker than almost sure convergence.

Q: Does the WLLN imply the SLLN?

A: Not in general. Under standard i.i.d. assumptions with finite mean, they coincide.

Empirical Probabilities (Dekking 13.4)

If you define an indicator $Y_i = 1$ when an event occurs the sample mean $\overline{Y}_n$ represents the *empirical* fraction of times the event occurs.
By the LLN, this fraction converges to the theoretical probability:

$$\mathbb{E}[Y_i] = \mathbb{P}(\text{Event})$$

Q: You roll a fair 6-sided die independently 100,000 times. Let $Y_i = 1$ if the i-th roll is a 5 or a 6, and 0 otherwise. What value does $\overline{Y}_n$ converge to?

Empirical Probabilities (Dekking 13.4)

If you define an indicator $Y_i = 1$ when an event occurs the sample mean $\overline{Y}_n$ represents the *empirical* fraction of times the event occurs.
By the LLN, this fraction converges to the theoretical probability:

$$\mathbb{E}[Y_i] = \mathbb{P}(\text{Event})$$

Q: You roll a fair 6-sided die independently 100,000 times. Let $Y_i = 1$ if the i-th roll is a 5 or a 6, and 0 otherwise. What value does $\overline{Y}_n$ converge to?
A: By LLN, it converges to $\mathbb{E}[Y_i] = \mathbb{P}(\text{roll is 5 or 6})$:

$$\mathbb{P}(\text{roll is 5 or 6}) = \frac{2}{6} = \frac{1}{3}$$

Central Limit Theorem

If $X_1, X_2, \ldots$ are iid with **finite mean** μ and **finite variance** σ^2, the properly scaled difference between the sample mean and population mean converges in distribution to a standard Gaussian:

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

Q: Does CLT apply to an i.i.d. sequence where each variable is the number of heads in 10 coin flips?

Central Limit Theorem

If $X_1, X_2, \ldots$ are iid with **finite mean** μ and **finite variance** σ^2, the properly scaled difference between the sample mean and population mean converges in distribution to a standard Gaussian:

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

Q: Does CLT apply to an i.i.d. sequence where each variable is the number of heads in 10 coin flips?

A: Yes. Each variable is Binomial$(10, p)$, so it has finite mean and variance; CLT applies.

CLT, but for sample means

By the CLT, the sample mean itself is approximately normally distributed for large n:

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

(follows algebraically from CLT statement, rearrange and apply $Z \sim N(0,1)$ then $a + bZ \sim N(a, b^2)$)

Q: Let $X_i \sim \mathrm{Exp}(4)$. If

$$Y_n := \sqrt{n}\left(\overline{X}_n - \frac{1}{4}\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

what is σ^2?

CLT, but for sample means

By the CLT, the sample mean itself is approximately normally distributed for large n:

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

(follows algebraically from CLT statement, rearrange and apply $Z \sim N(0,1)$ then $a + bZ \sim N(a, b^2)$)

Q: Let $X_i \sim \mathrm{Exp}(4)$. If

$$Y_n := \sqrt{n}\left(\overline{X}_n - \frac{1}{4}\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

what is σ^2?
A: σ^2 is the variance of the underlying distribution:

$$\mathrm{Var}(\mathrm{Exp}(4)) = \frac{1}{4^2} = \frac{1}{16}.$$

Distribution of the Sum

If you sum n iid variables, the total $S_n \approx \mathcal{N}(n\mu, n\sigma^2)$.
Define the Z-score to standard normal: $Z = \frac{\text{value}-\mu}{\sigma}$
Notation ϕ and Φ denote the PDF and CDF of $N(0, 1)$

Q: Let $X_1, \ldots, X_{100}$ be i.i.d. $\text{Pois}(9)$ variables. Approximate $\mathbb{P}(S_{100} < 840)$.

Distribution of the Sum

If you sum n iid variables, the total $S_n \approx \mathcal{N}(n\mu, n\sigma^2)$.
Define the Z-score to standard normal: $Z = \frac{\text{value}-\mu}{\sigma}$
Notation ϕ and Φ denote the PDF and CDF of $N(0, 1)$

Q: Let $X_1, \ldots, X_{100}$ be i.i.d. $\mathrm{Pois}(9)$ variables. Approximate $\mathbb{P}(S_{100} < 840)$.
A: For S_{100}, mean is 900 and standard deviation is 30. Standardizing:

$$Z = \frac{840 - 900}{30} = -2$$

so $\mathbb{P}(S_{100} < 840) \approx \Phi(-2)$ (equivalently $1 - \Phi(2)$).

Joint & Marginal PDFs

For continuous RVs X, Y with joint PDF $f_{X,Y}(x,y)$, probabilities are found via double integration:
$\mathbb{P}((X,Y) \in A) = \iint_A f_{X,Y}(x,y)dxdy$.
Marginals are found by integrating out the other variable:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

Q: Let $f_{X,Y}(x,y) = \frac{3}{2}(x^2 + y)$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Find the marginal PDF $f_X(x)$.

Joint & Marginal PDFs

For continuous RVs X, Y with joint PDF $f_{X,Y}(x,y)$, probabilities are found via double integration:
$\mathbb{P}((X,Y) \in A) = \iint_A f_{X,Y}(x,y)dxdy$.
Marginals are found by integrating out the other variable:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

Q: Let $f_{X,Y}(x,y) = \frac{3}{2}(x^2 + y)$ for $0 \le x \le 1$ and $0 \le y \le 1$. Find the marginal PDF $f_X(x)$.
A: Integrate out y:

$$\begin{aligned} f_X(x) &= \int_0^1 \frac{3}{2}(x^2 + y)dy \\ &= \frac{3}{2}\left[x^2 y + \frac{y^2}{2}\right]_{y=0}^{y=1} = \frac{3}{2}\left(x^2 + \frac{1}{2}\right) \end{aligned}$$

for $0 \le x \le 1$.

Joint CDF Definition

The joint CDF evaluates the cumulative probability up to (x, y):

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)dvdu$$

Q: For $f_{X,Y}(x,y) = e^{-x-y}$ on $x > 0, y > 0$, compute $F_{X,Y}(x,y)$ for $x, y > 0$.

Joint CDF Definition

The joint CDF evaluates the cumulative probability up to (x, y):

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)dvdu$$

Q: For $f_{X,Y}(x, y) = e^{-x-y}$ on $x > 0, y > 0$, compute $F_{X,Y}(x, y)$ for $x, y > 0$.
A: Evaluate the double integral from 0 to the bounds:

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{0}^{x} \int_{0}^{y} e^{-u}e^{-v}dvdu \\ &= \left(\int_{0}^{x} e^{-u}du\right)\left(\int_{0}^{y} e^{-v}dv\right) \\ &= (1 - e^{-x})(1 - e^{-y}) \end{aligned}$$

Covariance & Correlation

Covariance measures the joint variability:

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

Correlation $\rho_{X,Y}$ is the normalized covariance, support in $[-1, 1]$:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Q: If X and Y are independent, what is their correlation? Is the converse always true?

Covariance & Correlation

Covariance measures the joint variability:

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

Correlation $\rho_{X,Y}$ is the normalized covariance, support in $[-1, 1]$:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Q: If X and Y are independent, what is their correlation? Is the converse always true?

A: Independence implies $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, so $\mathrm{Cov}(X, Y) = 0$ and $\rho_{X,Y} = 0$. The converse is **not** always true.

Poisson Processes

If events occur as a Poisson process with rate λ, the number of arrivals in time t is $N_t \sim \mathrm{Pois}(\lambda t)$.

Q: Customers arrive at rate $\lambda = 2$ per hour. Let T_3 be the time until the 3rd customer. Express $\mathbb{P}(T_3 > 1)$ via a Poisson probability.

Poisson Processes

If events occur as a Poisson process with rate λ, the number of arrivals in time t is $N_t \sim \mathrm{Pois}(\lambda t)$.

Q: Customers arrive at rate $\lambda = 2$ per hour. Let T_3 be the time until the 3rd customer. Express $\mathbb{P}(T_3 > 1)$ via a Poisson probability.
A: $T_3 > 1$ means fewer than 3 arrivals in the first hour, so

$$\mathbb{P}(T_3 > 1) = \mathbb{P}(N_1 \leq 2), \quad N_1 \sim \mathrm{Pois}(2).$$

Hence

$$\mathbb{P}(T_3 > 1) = e^{-2}\left(1 + 2 + \frac{2^2}{2!}\right) = 5e^{-2}.$$

Markov's Inequality

For any **non-negative** random variable $X \geq 0$ and any constant $a > 0$:

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

Q: Suppose exam scores are between 0 and 100 and the average score is 60. Give an upper bound for the proportion scoring at least 90.

Markov's Inequality

For any **non-negative** random variable $X \geq 0$ and any constant $a > 0$:

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

Q: Suppose exam scores are between 0 and 100 and the average score is 60. Give an upper bound for the proportion scoring at least 90.

A: By Markov with $a = 90$ and $\mathbb{E}[X] = 60$:

$$\mathbb{P}(X \geq 90) \leq \frac{60}{90} = \frac{2}{3}.$$

Chebyshev's Inequality

For any random variable X with finite mean μ and variance σ^2, and for any $k > 0$:

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Q: A factory has lifespan mean $\mu = 100$ hours and variance $\sigma^2 = 25$. Bound $\mathbb{P}(|X - 100| \geq 10)$.

Chebyshev's Inequality

For any random variable X with finite mean μ and variance σ^2, and for any $k > 0$:

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Q: A factory has lifespan mean $\mu = 100$ hours and variance $\sigma^2 = 25$. Bound $\mathbb{P}(|X - 100| \geq 10)$.

A: Using Chebyshev with $k = 10$:

$$\mathbb{P}(|X - 100| \geq 10) \leq \frac{25}{10^2} = \frac{1}{4}.$$

So at most 25% deviate by 10 hours or more.

Jensen's Inequality

For any random variable X and any **convex** function g (meaning $g''(x) \geq 0$):

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

Q: Use Jensen with the convex function $g(x) = x^2$ to show $\mathrm{Var}(X) \geq 0$.

Jensen's Inequality

For any random variable X and any **convex** function g (meaning $g''(x) \geq 0$):

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

Q: Use Jensen with the convex function $g(x) = x^2$ to show $\mathrm{Var}(X) \geq 0$.
A: Plugging $g(x) = x^2$ into Jensen gives:

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$$

Rearranging yields:

$$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$$

therefore $\mathrm{Var}(X) \geq 0$.