18.175: Lecture 8

DeMoivre-Laplace and weak convergence

Scott Sheffield

MIT

Outline

Kolmogorov zero-one law and three-series theorem

Large deviations

DeMoivre-Laplace limit theorem

Weak convergence

Characteristic functions

Outline

Kolmogorov zero-one law and three-series theorem

Large deviations

DeMoivre-Laplace limit theorem

Weak convergence

Characteristic functions

Kolmogorov zero-one law

- Consider sequence of random variables X_n on some probability space. Write $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \ldots)$ and $\mathcal{T} = \cap_n \mathcal{F}'_n$.

Kolmogorov zero-one law

- Consider sequence of random variables X_n on some probability space. Write $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \ldots)$ and $\mathcal{T} = \cap_n \mathcal{F}'_n$.
- $\mathcal{T}$ is called the **tail σ-algebra**. It contains the information you can observe by looking only at stuff arbitrarily far into the future. Intuitively, membership in tail event doesn't change when finitely many X_n are changed.

Kolmogorov zero-one law

- Consider sequence of random variables X_n on some probability space. Write $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \ldots)$ and $\mathcal{T} = \cap_n \mathcal{F}'_n$.
- $\mathcal{T}$ is called the **tail σ-algebra**. It contains the information you can observe by looking only at stuff arbitrarily far into the future. Intuitively, membership in tail event doesn't change when finitely many X_n are changed.
- Event that X_n converge to a limit is example of a tail event. Other examples?

Kolmogorov zero-one law

- Consider sequence of random variables X_n on some probability space. Write $\mathcal{F}'_n = \sigma(X_n, X_{n+1}, \ldots)$ and $\mathcal{T} = \cap_n \mathcal{F}'_n$.
- $\mathcal{T}$ is called the **tail σ-algebra**. It contains the information you can observe by looking only at stuff arbitrarily far into the future. Intuitively, membership in tail event doesn't change when finitely many X_n are changed.
- Event that X_n converge to a limit is example of a tail event. Other examples?
- **Theorem:** If $X_1, X_2, \ldots$ are independent and $A \in \mathcal{T}$ then $P(A) \in \{0, 1\}$.

Kolmogorov zero-one law proof idea

- **Theorem:** If $X_1, X_2, \ldots$ are independent and $A \in \mathcal{T}$ then $P(A) \in \{0, 1\}$.

Kolmogorov zero-one law proof idea

- **Theorem:** If $X_1, X_2, \ldots$ are independent and $A \in \mathcal{T}$ then $P(A) \in \{0, 1\}$.
- **Main idea of proof:** Statement is equivalent to saying that A is independent of itself, i.e., $P(A) = P(A \cap A) = P(A)^2$. How do we prove that?

Kolmogorov zero-one law proof idea

- **Theorem:** If $X_1, X_2, \dots$ are independent and $A \in \mathcal{T}$ then $P(A) \in \{0, 1\}$.
- **Main idea of proof:** Statement is equivalent to saying that A is independent of itself, i.e., $P(A) = P(A \cap A) = P(A)^2$. How do we prove that?
- Recall theorem that if $\mathcal{A}_i$ are independent π-systems, then σA_i are independent.

Kolmogorov zero-one law proof idea

- **Theorem:** If $X_1, X_2, \dots$ are independent and $A \in \mathcal{T}$ then $P(A) \in \{0, 1\}$.
- **Main idea of proof:** Statement is equivalent to saying that A is independent of itself, i.e., $P(A) = P(A \cap A) = P(A)^2$. How do we prove that?
- Recall theorem that if $\mathcal{A}_i$ are independent π-systems, then σA_i are independent.
- Deduce that $\sigma(X_1, X_2, \dots, X_n)$ and $\sigma(X_{n+1}, X_{n+2}, \dots)$ are independent. Then deduce that $\sigma(X_1, X_2, \dots)$ and $\mathcal{T}$ are independent, using fact that $\cup_k \sigma(X_1, \dots, X_k)$ and $\mathcal{T}$ are π-systems.

Kolmogorov maximal inequality

- **Theorem:** Suppose X_i are independent with mean zero and finite variances, and $S_n = \sum_{i=1}^n X_n$. Then

$$P(\max_{1 \leq k \leq n} |S_k| \geq x) \leq x^{-2}\mathrm{Var}(S_n) = x^{-2}E|S_n|^2.$$

Kolmogorov maximal inequality

- **Theorem:** Suppose X_i are independent with mean zero and finite variances, and $S_n = \sum_{i=1}^n X_n$. Then

$$P(\max_{1\le k\le n} |S_k| \ge x) \le x^{-2}\mathrm{Var}(S_n) = x^{-2}E|S_n|^2.$$

- **Main idea of proof:** Consider first time maximum is exceeded. Bound below the expected square sum on that event.

Kolmogorov three-series theorem

- **Theorem:** Let $X_1, X_2, \ldots$ be independent and fix $A > 0$. Write $Y_i = X_i 1_{(|X_i| \leq A)}$. Then $\sum X_i$ converges a.s. if and only if the following are all true:

Kolmogorov three-series theorem

- **Theorem:** Let $X_1, X_2, \dots$ be independent and fix $A > 0$. Write $Y_i = X_i 1_{(|X_i| \le A)}$. Then $\sum X_i$ converges a.s. if and only if the following are all true:
  - $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$

Kolmogorov three-series theorem

- **Theorem:** Let $X_1, X_2, \ldots$ be independent and fix $A > 0$. Write $Y_i = X_i 1_{(|X_i| \le A)}$. Then $\sum X_i$ converges a.s. if and only if the following are all true:
  - $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$
  - $\sum_{n=1}^{\infty} EY_n$ converges

Kolmogorov three-series theorem

- **Theorem:** Let $X_1, X_2, \dots$ be independent and fix $A > 0$. Write $Y_i = X_i 1_{(|X_i| \le A)}$. Then $\sum X_i$ converges a.s. if and only if the following are all true:
  - $\sum_{n=1}^\infty P(|X_n| > A) < \infty$
  - $\sum_{n=1}^\infty EY_n$ converges
  - $\sum_{n=1}^\infty \mathrm{Var}(Y_n) < \infty$

Kolmogorov three-series theorem

- **Theorem:** Let $X_1, X_2, \ldots$ be independent and fix $A > 0$. Write $Y_i = X_i 1_{(|X_i| \leq A)}$. Then $\sum X_i$ converges a.s. if and only if the following are all true:
  - $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$
  - $\sum_{n=1}^{\infty} EY_n$ converges
  - $\sum_{n=1}^{\infty} \mathrm{Var}(Y_n) < \infty$
- **Main ideas behind the proof:** Kolmogorov zero-one law implies that $\sum X_i$ converges with probability $p \in \{0, 1\}$. We just have to show that $p = 1$ when all hypotheses are satisfied (sufficiency of conditions) and $p = 0$ if any one of them fails (necessity).

Kolmogorov three-series theorem

- **Theorem:** Let $X_1, X_2, \ldots$ be independent and fix $A > 0$. Write $Y_i = X_i 1_{(|X_i| \le A)}$. Then $\sum X_i$ converges a.s. if and only if the following are all true:
  - $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$
  - $\sum_{n=1}^{\infty} EY_n$ converges
  - $\sum_{n=1}^{\infty} \mathrm{Var}(Y_n) < \infty$
- **Main ideas behind the proof:** Kolmogorov zero-one law implies that $\sum X_i$ converges with probability $p \in \{0, 1\}$. We just have to show that $p = 1$ when all hypotheses are satisfied (sufficiency of conditions) and $p = 0$ if any one of them fails (necessity).
- To prove sufficiency, apply Borel-Cantelli to see that probability that $X_n \ne Y_n$ i.o. is zero. Subtract means from Y_n, reduce to case that each Y_n has mean zero. Apply Kolmogorov maximal inequality.

Outline

Outline

Recall: moment generating functions

- Let X be a random variable.

Recall: moment generating functions

- Let X be a random variable.
- The **moment generating function** of X is defined by $M(t) = M_X(t) := E[e^{tX}]$.

Recall: moment generating functions

- Let X be a random variable.
- The **moment generating function** of X is defined by $M(t) = M_X(t) := E[e^{tX}]$.

Recall: moment generating functions

- Let X be a random variable.
- The **moment generating function** of X is defined by $M(t) = M_X(t) := E[e^{tX}]$.
- When X is discrete, can write $M(t) = \sum_x e^{tx} p_X(x)$. So $M(t)$ is a weighted average of countably many exponential functions.

Recall: moment generating functions

- Let X be a random variable.
- The **moment generating function** of X is defined by $M(t) = M_X(t) := E[e^{tX}]$.
- When X is discrete, can write $M(t) = \sum_x e^{tx} p_X(x)$. So $M(t)$ is a weighted average of countably many exponential functions.
- When X is continuous, can write $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$. So $M(t)$ is a weighted average of a continuum of exponential functions.

Recall: moment generating functions

- Let X be a random variable.
- The **moment generating function** of X is defined by $M(t) = M_X(t) := E[e^{tX}]$.
- When X is discrete, can write $M(t) = \sum_x e^{tx} p_X(x)$. So $M(t)$ is a weighted average of countably many exponential functions.
- When X is continuous, can write $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$. So $M(t)$ is a weighted average of a continuum of exponential functions.
- We always have $M(0) = 1$.

Recall: moment generating functions

- Let X be a random variable.
- The **moment generating function** of X is defined by $M(t) = M_X(t) := E[e^{tX}]$.
- When X is discrete, can write $M(t) = \sum_x e^{tx} p_X(x)$. So $M(t)$ is a weighted average of countably many exponential functions.
- When X is continuous, can write $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$. So $M(t)$ is a weighted average of a continuum of exponential functions.
- We always have $M(0) = 1$.
- If $b > 0$ and $t > 0$ then $E[e^{tX}] \geq E[e^{t \min\{X,b\}}] \geq P\{X \geq b\} e^{tb}$.

Recall: moment generating functions

- Let X be a random variable.
- The **moment generating function** of X is defined by $M(t) = M_X(t) := E[e^{tX}]$.
- When X is discrete, can write $M(t) = \sum_x e^{tx} p_X(x)$. So $M(t)$ is a weighted average of countably many exponential functions.
- When X is continuous, can write $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$. So $M(t)$ is a weighted average of a continuum of exponential functions.
- We always have $M(0) = 1$.
- If $b > 0$ and $t > 0$ then $E[e^{tX}] \geq E[e^{t \min\{X,b\}}] \geq P\{X \geq b\} e^{tb}$.
- If X takes both positive and negative values with positive probability then $M(t)$ grows at least exponentially fast in $|t|$ as $|t| \to \infty$.

Recall: moment generating functions for i.i.d. sums

- We showed that if $Z = X + Y$ and X and Y are independent, then $M_Z(t) = M_X(t)M_Y(t)$

Recall: moment generating functions for i.i.d. sums

- We showed that if $Z = X + Y$ and X and Y are independent, then $M_Z(t) = M_X(t)M_Y(t)$
- If $X_1 \dots X_n$ are i.i.d. copies of X and $Z = X_1 + \dots + X_n$ then what is M_Z?

Recall: moment generating functions for i.i.d. sums

- We showed that if $Z = X + Y$ and X and Y are independent, then $M_Z(t) = M_X(t)M_Y(t)$
- If $X_1 \ldots X_n$ are i.i.d. copies of X and $Z = X_1 + \ldots + X_n$ then what is M_Z?
- Answer: M_X^n. Follows by repeatedly applying formula above.

Recall: moment generating functions for i.i.d. sums

- We showed that if $Z = X + Y$ and X and Y are independent, then $M_Z(t) = M_X(t)M_Y(t)$
- If $X_1 \ldots X_n$ are i.i.d. copies of X and $Z = X_1 + \ldots + X_n$ then what is M_Z?
- Answer: M_X^n. Follows by repeatedly applying formula above.
- This a big reason for studying moment generating functions. It helps us understand what happens when we sum up a lot of independent copies of the same random variable.

Large deviations

- Consider i.i.d. random variables X_i. Want to show that if $\phi(\theta) := M_{X_i}(\theta) = E\exp(\theta X_i)$ is less than infinity for some $\theta > 0$, then $P(S_n \geq na) \to 0$ exponentially fast when $a > E[X_i]$.

Large deviations

- Consider i.i.d. random variables X_i. Want to show that if $\phi(\theta) := M_{X_i}(\theta) = E\exp(\theta X_i)$ is less than infinity for some $\theta > 0$, then $P(S_n \geq na) \to 0$ exponentially fast when $a > E[X_i]$.
- Kind of a quantitative form of the weak law of large numbers. The empirical average A_n is *very* unlikely to be ϵ away from its expected value (where "very" means with probability less than some exponentially decaying function of n).

Large deviations

- Consider i.i.d. random variables X_i. Want to show that if $\phi(\theta) := M_{X_i}(\theta) = E\exp(\theta X_i)$ is less than infinity for some $\theta > 0$, then $P(S_n \geq na) \to 0$ exponentially fast when $a > E[X_i]$.
- Kind of a quantitative form of the weak law of large numbers. The empirical average A_n is *very* unlikely to be ϵ away from its expected value (where "very" means with probability less than some exponentially decaying function of n).
- Write $\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na)$. It gives the "rate" of exponential decay as a function of a.

Outline

Outline

Kolmogorov zero-one law and three-series theorem

Large deviations

DeMoivre-Laplace limit theorem

Weak convergence

Characteristic functions

DeMoivre-Laplace limit theorem

- Let X_i be i.i.d. random variables. Write $S_n = \sum_{i=1}^{n} X_n$.

DeMoivre-Laplace limit theorem

- Let X_i be i.i.d. random variables. Write $S_n = \sum_{i=1}^{n} X_n$.
- Suppose each X_i is 1 with probability p and 0 with probability $q = 1 - p$.

DeMoivre-Laplace limit theorem

- Let X_i be i.i.d. random variables. Write $S_n = \sum_{i=1}^n X_n$.
- Suppose each X_i is 1 with probability p and 0 with probability $q = 1 - p$.
- **DeMoivre-Laplace limit theorem:**

$$\lim_{n\to\infty} P\{a \le \frac{S_n - np}{\sqrt{npq}} \le b\} \to \Phi(b) - \Phi(a).$$

DeMoivre-Laplace limit theorem

- Let X_i be i.i.d. random variables. Write $S_n = \sum_{i=1}^n X_n$.
- Suppose each X_i is 1 with probability p and 0 with probability $q = 1 - p$.
- **DeMoivre-Laplace limit theorem:**

$$\lim_{n\to\infty} P\{a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\} \to \Phi(b) - \Phi(a).$$

- Here $\Phi(b) - \Phi(a) = P\{a \leq Z \leq b\}$ when Z is a standard normal random variable.

DeMoivre-Laplace limit theorem

- Let X_i be i.i.d. random variables. Write $S_n = \sum_{i=1}^n X_n$.
- Suppose each X_i is 1 with probability p and 0 with probability $q = 1 - p$.
- **DeMoivre-Laplace limit theorem:**

$$\lim_{n\to\infty} P\{a \le \frac{S_n - np}{\sqrt{npq}} \le b\} \to \Phi(b) - \Phi(a).$$

- Here $\Phi(b) - \Phi(a) = P\{a \le Z \le b\}$ when Z is a standard normal random variable.
- $\frac{S_n - np}{\sqrt{npq}}$ describes "number of standard deviations that S_n is above or below its mean".

DeMoivre-Laplace limit theorem

- Let X_i be i.i.d. random variables. Write $S_n = \sum_{i=1}^{n} X_n$.
- Suppose each X_i is 1 with probability p and 0 with probability $q = 1 - p$.
- **DeMoivre-Laplace limit theorem:**

$$\lim_{n\to\infty} P\{a \le \frac{S_n - np}{\sqrt{npq}} \le b\} \to \Phi(b) - \Phi(a).$$

- Here $\Phi(b) - \Phi(a) = P\{a \le Z \le b\}$ when Z is a standard normal random variable.
- $\frac{S_n - np}{\sqrt{npq}}$ describes "number of standard deviations that S_n is above or below its mean".
- **Proof idea:** use binomial coefficients and Stirling's formula.

DeMoivre-Laplace limit theorem

- Let X_i be i.i.d. random variables. Write $S_n = \sum_{i=1}^n X_n$.
- Suppose each X_i is 1 with probability p and 0 with probability $q = 1 - p$.
- **DeMoivre-Laplace limit theorem:**

$$\lim_{n\to\infty} P\{a \le \frac{S_n - np}{\sqrt{npq}} \le b\} \to \Phi(b) - \Phi(a).$$

- Here $\Phi(b) - \Phi(a) = P\{a \le Z \le b\}$ when Z is a standard normal random variable.
- $\frac{S_n - np}{\sqrt{npq}}$ describes "number of standard deviations that S_n is above or below its mean".
- **Proof idea:** use binomial coefficients and Stirling's formula.
- Question: Does similar statement hold if X_i are i.i.d. from some other law?

DeMoivre-Laplace limit theorem

- Let X_i be i.i.d. random variables. Write $S_n = \sum_{i=1}^n X_n$.
- Suppose each X_i is 1 with probability p and 0 with probability $q = 1 - p$.
- **DeMoivre-Laplace limit theorem:**

$$\lim_{n\to\infty} P\{a \le \frac{S_n - np}{\sqrt{npq}} \le b\} \to \Phi(b) - \Phi(a).$$

- Here $\Phi(b) - \Phi(a) = P\{a \le Z \le b\}$ when Z is a standard normal random variable.
- $\frac{S_n - np}{\sqrt{npq}}$ describes "number of standard deviations that S_n is above or below its mean".
- **Proof idea:** use binomial coefficients and Stirling's formula.
- Question: Does similar statement hold if X_i are i.i.d. from some other law?
- **Central limit theorem:** Yes, if they have finite variance.

Local $p = 1/2$ DeMoivre-Laplace limit theorem

- **Stirling:** $n! \sim n^n e^{-n}\sqrt{2\pi n}$ where $\sim$ means ratio tends to one.

Local $p = 1/2$ DeMoivre-Laplace limit theorem

- **Stirling:** $n! \sim n^n e^{-n}\sqrt{2\pi n}$ where $\sim$ means ratio tends to one.
- **Theorem:** If $2k/\sqrt{2n} \to x$ then $P(S_{2n} = 2k) \sim (\pi n)^{-1/2} e^{-x^2/2}$.

Local $p = 1/2$ DeMoivre-Laplace limit theorem

- **Stirling:** $n! \sim n^n e^{-n}\sqrt{2\pi n}$ where $\sim$ means ratio tends to one.
- **Theorem:** If $2k/\sqrt{2n} \to x$ then $P(S_{2n} = 2k) \sim (\pi n)^{-1/2} e^{-x^2/2}$.
- Recall $P(S_{2n} = 2k) = \binom{2n}{n+k} 2^{-2n} = 2^{-2n} \frac{(2n)!}{(n+k)!(n-k)!}$.

Outline

Kolmogorov zero-one law and three-series theorem

Large deviations

DeMoivre-Laplace limit theorem

Weak convergence

Characteristic functions

Outline

Kolmogorov zero-one law and three-series theorem

Large deviations

DeMoivre-Laplace limit theorem

Weak convergence

Characteristic functions

Weak convergence

- Let X be random variable, X_n a sequence of random variables.

Weak convergence

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.

Weak convergence

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- Also say that the $F_n = F_{X_n}$ converge weakly to $F = F_X$.

Weak convergence

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- Also say that the $F_n = F_{X_n}$ converge weakly to $F = F_X$.
- **Example:** X_i chosen from $\{-1, 1\}$ with i.i.d. fair coin tosses: then $n^{-1/2}\sum_{i=1}^n X_i$ converges in law to a normal random variable (mean zero, variance one) by DeMoivre-Laplace.

Weak convergence

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- Also say that the $F_n = F_{X_n}$ converge weakly to $F = F_X$.
- **Example:** X_i chosen from $\{-1, 1\}$ with i.i.d. fair coin tosses: then $n^{-1/2} \sum_{i=1}^n X_i$ converges in law to a normal random variable (mean zero, variance one) by DeMoivre-Laplace.
- **Example:** If X_n is equal to $1/n$ a.s. then X_n converge weakly to an X equal to 0 a.s. Note that $\lim_{n\to\infty} F_n(0) \neq F(0)$ in this case.

Weak convergence

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- Also say that the $F_n = F_{X_n}$ converge weakly to $F = F_X$.
- **Example:** X_i chosen from $\{-1, 1\}$ with i.i.d. fair coin tosses: then $n^{-1/2} \sum_{i=1}^n X_i$ converges in law to a normal random variable (mean zero, variance one) by DeMoivre-Laplace.
- **Example:** If X_n is equal to $1/n$ a.s. then X_n converge weakly to an X equal to 0 a.s. Note that $\lim_{n\to\infty} F_n(0) \neq F(0)$ in this case.
- **Example:** If X_i are i.i.d. then the empirical distributions converge a.s. to law of X_1 (Glivenko-Cantelli).

Weak convergence

- Let X be random variable, X_n a sequence of random variables.
- Say X_n **converge in distribution** or **converge in law** to X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all $x \in \mathbb{R}$ at which F_X is continuous.
- Also say that the $F_n = F_{X_n}$ converge weakly to $F = F_X$.
- **Example:** X_i chosen from $\{-1, 1\}$ with i.i.d. fair coin tosses: then $n^{-1/2}\sum_{i=1}^n X_i$ converges in law to a normal random variable (mean zero, variance one) by DeMoivre-Laplace.
- **Example:** If X_n is equal to $1/n$ a.s. then X_n converge weakly to an X equal to 0 a.s. Note that $\lim_{n\to\infty} F_n(0) \neq F(0)$ in this case.
- **Example:** If X_i are i.i.d. then the empirical distributions converge a.s. to law of X_1 (Glivenko-Cantelli).
- **Example:** Let X_n be the nth largest of $2n+1$ points chosen i.i.d. from fixed law.

Convergence results

- **Theorem:** If $F_n \to F_\infty$, then we can find corresponding random variables Y_n on a common measure space so that $Y_n \to Y_\infty$ almost surely.

Convergence results

- **Theorem:** If $F_n \to F_\infty$, then we can find corresponding random variables Y_n on a common measure space so that $Y_n \to Y_\infty$ almost surely.
- **Proof idea:** Take $\Omega = (0,1)$ and $Y_n = \sup\{y : F_n(y) < x\}$.

Convergence results

- **Theorem:** If $F_n \to F_\infty$, then we can find corresponding random variables Y_n on a common measure space so that $Y_n \to Y_\infty$ almost surely.
- **Proof idea:** Take $\Omega = (0,1)$ and $Y_n = \sup\{y : F_n(y) < x\}$.
- **Theorem:** $X_n \implies X_\infty$ if and only if for every bounded continuous g we have $Eg(X_n) \to Eg(X_\infty)$.

Convergence results

- **Theorem:** If $F_n \to F_\infty$, then we can find corresponding random variables Y_n on a common measure space so that $Y_n \to Y_\infty$ almost surely.
- **Proof idea:** Take $\Omega = (0,1)$ and $Y_n = \sup\{y : F_n(y) < x\}$.
- **Theorem:** $X_n \implies X_\infty$ if and only if for every bounded continuous g we have $Eg(X_n) \to Eg(X_\infty)$.
- **Proof idea:** Define X_n on common sample space so converge a.s., use bounded convergence theorem.

Convergence results

- **Theorem:** If $F_n \to F_\infty$, then we can find corresponding random variables Y_n on a common measure space so that $Y_n \to Y_\infty$ almost surely.
- **Proof idea:** Take $\Omega = (0,1)$ and $Y_n = \sup\{y : F_n(y) < x\}$.
- **Theorem:** $X_n \implies X_\infty$ if and only if for every bounded continuous g we have $Eg(X_n) \to Eg(X_\infty)$.
- **Proof idea:** Define X_n on common sample space so converge a.s., use bounded convergence theorem.
- **Theorem:** Suppose g is measurable and its set of discontinuity points has μ_X measure zero. Then $X_n \implies X_\infty$ implies $g(X_n) \implies g(X)$.

Convergence results

- **Theorem:** If $F_n \to F_\infty$, then we can find corresponding random variables Y_n on a common measure space so that $Y_n \to Y_\infty$ almost surely.
- **Proof idea:** Take $\Omega = (0,1)$ and $Y_n = \sup\{y : F_n(y) < x\}$.
- **Theorem:** $X_n \implies X_\infty$ if and only if for every bounded continuous g we have $Eg(X_n) \to Eg(X_\infty)$.
- **Proof idea:** Define X_n on common sample space so converge a.s., use bounded convergence theorem.
- **Theorem:** Suppose g is measurable and its set of discontinuity points has μ_X measure zero. Then $X_n \implies X_\infty$ implies $g(X_n) \implies g(X)$.
- **Proof idea:** Define X_n on common sample space so converge a.s., use bounded convergence theorem.

Compactness

- **Theorem:** Every sequence F_n of distribution has subsequence converging to right continuous nondecreasing F so that $\lim F_{n(k)}(y) = F(y)$ at all continuity points of F.

Compactness

- **Theorem:** Every sequence F_n of distribution has subsequence converging to right continuous nondecreasing F so that $\lim F_{n(k)}(y) = F(y)$ at all continuity points of F.
- Limit may not be a distribution function.

Compactness

- **Theorem:** Every sequence F_n of distribution has subsequence converging to right continuous nondecreasing F so that $\lim F_{n(k)}(y) = F(y)$ at all continuity points of F.
- Limit may not be a distribution function.
- Need a "tightness" assumption to make that the case. Say μ_n are **tight** if for every ϵ we can find an M so that $\mu_n[-M, M] < \epsilon$ for all n. Define tightness analogously for corresponding real random variables or distributions functions.

Compactness

- **Theorem:** Every sequence F_n of distribution has subsequence converging to right continuous nondecreasing F so that $\lim F_{n(k)}(y) = F(y)$ at all continuity points of F.
- Limit may not be a distribution function.
- Need a "tightness" assumption to make that the case. Say μ_n are **tight** if for every ϵ we can find an M so that $\mu_n[-M, M] < \epsilon$ for all n. Define tightness analogously for corresponding real random variables or distributions functions.
- **Theorem:** Every subsequential limit of the F_n above is the distribution function of a probability measure if and only if the F_n are tight.

Total variation norm

- If we have two probability measures μ and ν we define the **total variation distance** between them is $||\mu - \nu|| := \sup_B |\mu(B) - \nu(B)|$.

Total variation norm

- If we have two probability measures μ and ν we define the **total variation distance** between them is $||\mu - \nu|| := \sup_B |\mu(B) - \nu(B)|$.
- Intuitively, it two measures are close in the total variation sense, then (most of the time) a sample from one measure looks like a sample from the other.

Total variation norm

- If we have two probability measures μ and ν we define the **total variation distance** between them is $||\mu - \nu|| := \sup_B |\mu(B) - \nu(B)|$.
- Intuitively, it two measures are close in the total variation sense, then (most of the time) a sample from one measure looks like a sample from the other.
- Convergence in total variation norm is much stronger than weak convergence.

Outline

Kolmogorov zero-one law and three-series theorem

Large deviations

DeMoivre-Laplace limit theorem

Weak convergence

Characteristic functions

Outline

Kolmogorov zero-one law and three-series theorem

Large deviations

DeMoivre-Laplace limit theorem

Weak convergence

Characteristic functions

Characteristic functions

- Let X be a random variable.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X\phi_Y$, just as $M_{X+Y} = M_X M_Y$, if X and Y are independent.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X\phi_Y$, just as $M_{X+Y} = M_X M_Y$, if X and Y are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X\phi_Y$, just as $M_{X+Y} = M_X M_Y$, if X and Y are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$.
- And if X has an mth moment then $E[X^m] = i^m\phi_X^{(m)}(0)$.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$. Like $M(t)$ except with i thrown in.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic functions are similar to moment generating functions in some ways.
- For example, $\phi_{X+Y} = \phi_X \phi_Y$, just as $M_{X+Y} = M_X M_Y$, if X and Y are independent.
- And $\phi_{aX}(t) = \phi_X(at)$ just as $M_{aX}(t) = M_X(at)$.
- And if X has an mth moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.
- But characteristic functions have an advantage: they are well defined at all t for all random variables X.

Continuity theorems

- **Lévy's continuity theorem:** if

$$\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$$

for all t, then X_n converge in law to X.

Continuity theorems

- **Lévy's continuity theorem:** if

$$\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$$

for all t, then X_n converge in law to X.

- By this theorem, we can prove the weak law of large numbers by showing $\lim_{n\to\infty} \phi_{A_n}(t) = \phi_\mu(t) = e^{it\mu}$ for all t. In the special case that $\mu = 0$, this amounts to showing $\lim_{n\to\infty} \phi_{A_n}(t) = 1$ for all t.

Continuity theorems

- **Lévy's continuity theorem:** if

$$\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$$

for all t, then X_n converge in law to X.

- By this theorem, we can prove the weak law of large numbers by showing $\lim_{n\to\infty} \phi_{A_n}(t) = \phi_\mu(t) = e^{it\mu}$ for all t. In the special case that $\mu = 0$, this amounts to showing $\lim_{n\to\infty} \phi_{A_n}(t) = 1$ for all t.
- **Moment generating analog:** if moment generating functions $M_{X_n}(t)$ are defined for all t and n and $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$ for all t, then X_n converge in law to X.