18.175: Lecture 11

Independent sums and large deviations

Scott Sheffield

MIT

Outline

Recollections

Large deviations

Outline

Recollections

Large deviations

Recall Borel-Cantelli lemmas

- **First Borel-Cantelli lemma:** If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n \ \text{i.o.}) = 0$.

Recall Borel-Cantelli lemmas

- **First Borel-Cantelli lemma:** If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n \text{ i.o.}) = 0$.
- **Second Borel-Cantelli lemma:** If A_n are independent, then $\sum_{n=1}^{\infty} P(A_n) = \infty$ implies $P(A_n \text{ i.o.}) = 1$.

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 maximal inequality

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

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

Kolmogorov maximal inequality

- **Thoerem:** 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 $Y_i = X_i 1_{(|X_i| \leq A)}$. Then $\sum X_i$ converges a.s. if and only if the following are all true:
  - $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$

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

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 $Y_i = X_i 1_{(|X_i| \le A)}$. 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).

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).
- To prove sufficiency, apply Borel-Cantelli to see that probability that $X_n \neq Y_n$ i.o. is zero. Subtract means from Y_n, reduce to case that each Y_n has mean zero. Apply Kolmogorov maximal inequality.

Outline

Recollections

Large deviations

Outline

Recollections

Large deviations

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 $X_1 \ldots X_n$ are i.i.d. copies of X and $Z = X_1 + \ldots + X_n$ 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 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 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.