18.175: Lecture 7

Sums of random variables

Scott Sheffield

MIT

Outline

Definitions

Sums of random variables

Outline

Definitions

Sums of random variables

Recall expectation definition

- Given probability space $(\Omega, \mathcal{F}, P)$ and random variable X (i.e., measurable function X from Ω to $\mathbb{R}$), we write $EX = \int XdP$.

Recall expectation definition

- Given probability space $(\Omega, \mathcal{F}, P)$ and random variable X (i.e., measurable function X from Ω to $\mathbb{R}$), we write $EX = \int XdP$.
- Expectation is always defined if $X \geq 0$ a.s., or if integrals of $\max\{X, 0\}$ and $\min\{X, 0\}$ are separately finite.

Strong law of large numbers

- **Theorem (strong law):** If $X_1, X_2, \ldots$ are i.i.d. real-valued random variables with expectation m and $A_n := n^{-1} \sum_{i=1}^n X_i$ are the *empirical means* then $\lim_{n\to\infty} A_n = m$ almost surely.

Strong law of large numbers

- **Theorem (strong law):** If $X_1, X_2, \dots$ are i.i.d. real-valued random variables with expectation m and $A_n := n^{-1} \sum_{i=1}^n X_i$ are the *empirical means* then $\lim_{n \to \infty} A_n = m$ almost surely.
- Last time we defined independent. We showed how to use Kolmogorov to construct infinite i.i.d. random variables on a measure space with a natural σ-algebra (in which the existence of a limit of the X_i is a measurable event). So we've come far enough to say that the statement makes sense.

Recall some definitions

- Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.

Recall some definitions

- Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.
- Random variables X and Y are independent if for all $C, D \in \mathcal{R}$, we have $P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$, i.e., the events $\{X \in C\}$ and $\{Y \in D\}$ are independent.

Recall some definitions

- Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.
- Random variables X and Y are independent if for all $C, D \in \mathcal{R}$, we have $P(X \in C, Y \in D) = P(X \in C)P(Y \in D)$, i.e., the events $\{X \in C\}$ and $\{Y \in D\}$ are independent.
- Two σ-fields $\mathcal{F}$ and $\mathcal{G}$ are independent if A and B are independent whenever $A \in \mathcal{F}$ and $B \in \mathcal{G}$. (This definition also makes sense if $\mathcal{F}$ and $\mathcal{G}$ are arbitrary algebras, semi-algebras, or other collections of measurable sets.)

Recall some definitions

- Say events $A_1, A_2, \ldots, A_n$ are independent if for each $I \subset \{1, 2, \ldots, n\}$ we have $P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.

Recall some definitions

- Say events $A_1, A_2, \ldots, A_n$ are independent if for each $I \subset \{1, 2, \ldots, n\}$ we have $P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.
- Say random variables $X_1, X_2, \ldots, X_n$ are independent if for any measurable sets $B_1, B_2, \ldots, B_n$, the events that $X_i \in B_i$ are independent.

Recall some definitions

- Say events $A_1, A_2, \ldots, A_n$ are independent if for each $I \subset \{1, 2, \ldots, n\}$ we have $P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.
- Say random variables $X_1, X_2, \ldots, X_n$ are independent if for any measurable sets $B_1, B_2, \ldots, B_n$, the events that $X_i \in B_i$ are independent.
- Say σ-algebras $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ if any collection of events (one from each σ-algebra) are independent. (This definition also makes sense if the $\mathcal{F}_i$ are algebras, semi-algebras, or other collections of measurable sets.)

Recall Kolmogorov

- **Kolmogorov extension theorem:** If we have consistent probability measures on $(\mathbb{R}^n, \mathcal{R}^n)$, then we can extend them uniquely to a probability measure on $\mathcal{R}^{\mathbb{N}}$.

Recall Kolmogorov

- **Kolmogorov extension theorem:** If we have consistent probability measures on $(\mathbb{R}^n, \mathcal{R}^n)$, then we can extend them uniquely to a probability measure on $\mathcal{R}^{\mathbb{N}}$.
- Proved using semi-algebra variant of Carathéeodory's extension theorem.

Extend Kolmogorov

- Kolmogorov extension theorem not generally true if replace $(\mathbb{R}, \mathcal{R})$ with any measure space.

Extend Kolmogorov

- Kolmogorov extension theorem not generally true if replace $(\mathbb{R}, \mathcal{R})$ with any measure space.
- But okay if we use **standard Borel spaces**. Durrett calls such spaces nice: a set $(S, \mathcal{S})$ is **nice** if have 1-1 map from S to $\mathbb{R}$ so that ϕ and ϕ^{-1} are both measurable.

Extend Kolmogorov

- Kolmogorov extension theorem not generally true if replace $(\mathbb{R}, \mathcal{R})$ with any measure space.
- But okay if we use **standard Borel spaces**. Durrett calls such spaces nice: a set $(S, \mathcal{S})$ is **nice** if have 1-1 map from S to $\mathbb{R}$ so that ϕ and ϕ^{-1} are both measurable.
- Are there any interesting nice measure spaces?

Extend Kolmogorov

- Kolmogorov extension theorem not generally true if replace $(\mathbb{R}, \mathcal{R})$ with any measure space.
- But okay if we use **standard Borel spaces**. Durrett calls such spaces nice: a set $(S, \mathcal{S})$ is **nice** if have 1-1 map from S to $\mathbb{R}$ so that ϕ and ϕ^{-1} are both measurable.
- Are there any interesting nice measure spaces?
- **Theorem:** Yes, lots. In fact, if S is a complete separable metric space M (or a Borel subset of such a space) and $\mathcal{S}$ is the set of Borel subsets of S, then $(S, \mathcal{S})$ is nice.

Extend Kolmogorov

- Kolmogorov extension theorem not generally true if replace $(\mathbb{R}, \mathcal{R})$ with any measure space.
- But okay if we use **standard Borel spaces**. Durrett calls such spaces nice: a set $(S, \mathcal{S})$ is **nice** if have 1-1 map from S to $\mathbb{R}$ so that ϕ and ϕ^{-1} are both measurable.
- Are there any interesting nice measure spaces?
- **Theorem:** Yes, lots. In fact, if S is a complete separable metric space M (or a Borel subset of such a space) and $\mathcal{S}$ is the set of Borel subsets of S, then $(S, \mathcal{S})$ is nice.
- **separable** means containing a countable dense set.

Standard Borel spaces

- **Main idea of proof:** Reduce to case that diameter less than one (e.g., by replacing $d(x,y)$ with $d(x,y)/(1+d(x,y))$). Then map M continuously into $[0,1]^{\mathbb{N}}$ by considering countable dense set $q_1, q_2, \ldots$ and mapping x to $\big(d(q_1,x), d(q_2,x), \ldots\big)$. Then give measurable one-to-one map from $[0,1]^{\mathbb{N}}$ to $[0,1]$ via binary expansion (to send $\mathbb{N} \times \mathbb{N}$-indexed matrix of 0's and 1's to an $\mathbb{N}$-indexed sequence of 0's and 1's).

Standard Borel spaces

- **Main idea of proof:** Reduce to case that diameter less than one (e.g., by replacing $d(x,y)$ with $d(x,y)/(1+d(x,y))$). Then map M continuously into $[0,1]^{\mathbb{N}}$ by considering countable dense set $q_1, q_2, \ldots$ and mapping x to $\big(d(q_1,x), d(q_2,x), \ldots\big)$. Then give measurable one-to-one map from $[0,1]^{\mathbb{N}}$ to $[0,1]$ via binary expansion (to send $\mathbb{N}\times\mathbb{N}$-indexed matrix of 0's and 1's to an $\mathbb{N}$-indexed sequence of 0's and 1's).
- In practice: say I want to let Ω be set of closed subsets of a disc, or planar curves, or functions from one set to another, etc. If I want to construct natural σ-algebra $\mathcal{F}$, I just need to produce metric that makes Ω complete and separable (and if I have to enlarge Ω to make it complete, that might be okay). Then I check that the events I care about belong to this σ-algebra.

Fubini's theorem

- Consider σ-finite measure spaces $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$.

Fubini's theorem

- Consider σ-finite measure spaces $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$.
- Let $\Omega = X \times Y$ and $\mathcal{F}$ be product σ-algebra.

Fubini's theorem

- Consider σ-finite measure spaces $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$.
- Let $\Omega = X \times Y$ and $\mathcal{F}$ be product σ-algebra.
- Check: unique measure μ on $\mathcal{F}$ with $\mu(A \times B) = \mu_1(A)\mu_2(B)$.

Fubini's theorem

- Consider σ-finite measure spaces $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$.
- Let $\Omega = X \times Y$ and $\mathcal{F}$ be product σ-algebra.
- Check: unique measure μ on $\mathcal{F}$ with $\mu(A \times B) = \mu_1(A)\mu_2(B)$.
- **Fubini's theorem:** If $f \geq 0$ or $\int |f| d\mu < \infty$ then

$$\int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = \int_{X\times Y} f d\mu =$$

$$\int_Y \int_X f(x,y)\mu_1(dx)\mu_2(dy).$$

Fubini's theorem

- Consider σ-finite measure spaces $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$.
- Let $\Omega = X \times Y$ and $\mathcal{F}$ be product σ-algebra.
- Check: unique measure μ on $\mathcal{F}$ with $\mu(A \times B) = \mu_1(A)\mu_2(B)$.
- **Fubini's theorem:** If $f \geq 0$ or $\int |f| d\mu < \infty$ then

$$\int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = \int_{X \times Y} f d\mu =$$

$$\int_Y \int_X f(x,y)\mu_1(dx)\mu_2(dy).$$

- **Main idea of proof:** Check definition makes sense: if f measurable, show that restriction of f to slice $\{(x,y) : x = x_0\}$ is measurable as function of y, and the integral over slice is measurable as function of x_0. Check Fubini for indicators of rectangular sets, use $\pi - \lambda$ to extend to measurable indicators. Extend to simple, bounded, L^1 (or non-negative) functions.

Non-measurable Fubini counterexample

- What if we take total ordering $\prec$ or reals in $[0, 1]$ (such that for each y the set $\{x : x \prec y\}$ is countable) and consider indicator function of $\{(x, y) : x \prec y\}$?

More observations

- If X_i are independent with distributions μ_i, then $(X_1, \ldots, X_n)$ has distribution $\mu_1 \times \ldots \mu_n$.

More observations

- If X_i are independent with distributions μ_i, then $(X_1, \ldots, X_n)$ has distribution $\mu_1 \times \ldots \mu_n$.
- If X_i are independent and satisfy either $X_i \geq 0$ for all i or $E|X_i| < \infty$ for all i then

$$E \prod_{i=1}^{n} X_i = \prod_{i=1}^{n} X_i.$$

Outline

Definitions

Sums of random variables

Outline

Definitions

Sums of random variables

Summing two random variables

- Say we have independent random variables X and Y with density functions f_X and f_Y.

Summing two random variables

- Say we have independent random variables X and Y with density functions f_X and f_Y.
- Now let's try to find $F_{X+Y}(a) = P\{X + Y \leq a\}$.

Summing two random variables

- Say we have independent random variables X and Y with density functions f_X and f_Y.
- Now let's try to find $F_{X+Y}(a) = P\{X + Y \leq a\}$.
- This is the integral over $\{(x, y) : x + y \leq a\}$ of $f(x, y) = f_X(x)f_Y(y)$. Thus,

Summing two random variables

- Say we have independent random variables X and Y with density functions f_X and f_Y.
- Now let's try to find $F_{X+Y}(a) = P\{X + Y \leq a\}$.
- This is the integral over $\{(x, y) : x + y \leq a\}$ of $f(x, y) = f_X(x)f_Y(y)$. Thus,
- $$P\{X + Y \leq a\} = \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x)f_Y(y)dxdy$$
$$= \int_{-\infty}^{\infty} F_X(a - y)f_Y(y)dy.$$

Summing two random variables

- Say we have independent random variables X and Y with density functions f_X and f_Y.
- Now let's try to find $F_{X+Y}(a) = P\{X + Y \le a\}$.
- This is the integral over $\{(x, y) : x + y \le a\}$ of $f(x, y) = f_X(x)f_Y(y)$. Thus,
- $$P\{X + Y \le a\} = \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)dxdy$$
  $$= \int_{-\infty}^{\infty} F_X(a - y)f_Y(y)dy.$$
- Differentiating both sides gives $f_{X+Y}(a) = \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy$.

Summing two random variables

- Say we have independent random variables X and Y with density functions f_X and f_Y.
- Now let's try to find $F_{X+Y}(a) = P\{X + Y \le a\}$.
- This is the integral over $\{(x, y) : x + y \le a\}$ of $f(x, y) = f_X(x)f_Y(y)$. Thus,
- $$P\{X + Y \le a\} = \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x)f_Y(y)dxdy$$
  $$= \int_{-\infty}^{\infty} F_X(a - y)f_Y(y)dy.$$
- Differentiating both sides gives $f_{X+Y}(a) = \frac{d}{da} \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy$.
- Latter formula makes some intuitive sense. We're integrating over the set of x, y pairs that add up to a.

Summing two random variables

- Say we have independent random variables X and Y with density functions f_X and f_Y.
- Now let's try to find $F_{X+Y}(a) = P\{X + Y \le a\}$.
- This is the integral over $\{(x,y) : x + y \le a\}$ of $f(x,y) = f_X(x)f_Y(y)$. Thus,
- $$P\{X + Y \le a\} = \int_{-\infty}^{\infty}\int_{-\infty}^{a-y} f_X(x)f_Y(y)dxdy$$
  $$= \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy.$$
- Differentiating both sides gives $f_{X+Y}(a) = \frac{d}{da}\int_{-\infty}^{\infty} F_X(a-y)f_Y(y)dy = \int_{-\infty}^{\infty} f_X(a-y)f_Y(y)dy$.
- Latter formula makes some intuitive sense. We're integrating over the set of x, y pairs that add up to a.
- Can also write $P(X + Y \le z) = \int F(z-y)dG(y)$.

Summing i.i.d. uniform random variables

- Suppose that X and Y are i.i.d. and uniform on $[0, 1]$. So $f_X = f_Y = 1$ on $[0, 1]$.

Summing i.i.d. uniform random variables

- Suppose that X and Y are i.i.d. and uniform on $[0, 1]$. So $f_X = f_Y = 1$ on $[0, 1]$.
- What is the probability density function of $X + Y$?

Summing i.i.d. uniform random variables

- Suppose that X and Y are i.i.d. and uniform on $[0,1]$. So $f_X = f_Y = 1$ on $[0,1]$.
- What is the probability density function of $X+Y$?
- $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy = \int_0^1 f_X(a-y)$ which is the length of $[0,1] \cap [a-1,a]$.

Summing i.i.d. uniform random variables

- Suppose that X and Y are i.i.d. and uniform on $[0,1]$. So $f_X = f_Y = 1$ on $[0,1]$.
- What is the probability density function of $X+Y$?
- $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy = \int_0^1 f_X(a-y)$ which is the length of $[0,1] \cap [a-1,a]$.
- That's a when $a \in [0,1]$ and $2-a$ when $a \in [0,2]$ and 0 otherwise.

Summing two normal variables

- X is normal with mean zero, variance σ_1^2, Y is normal with mean zero, variance σ_2^2.

Summing two normal variables

- X is normal with mean zero, variance σ_1^2, Y is normal with mean zero, variance σ_2^2.
- $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-x^2}{2\sigma_1^2}}$ and $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-y^2}{2\sigma_2^2}}$.

Summing two normal variables

- X is normal with mean zero, variance σ_1^2, Y is normal with mean zero, variance σ_2^2.
- $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-x^2}{2\sigma_1^2}}$ and $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-y^2}{2\sigma_2^2}}$.
- We just need to compute $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$.

Summing two normal variables

- X is normal with mean zero, variance σ_1^2, Y is normal with mean zero, variance σ_2^2.
- $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-x^2}{2\sigma_1^2}}$ and $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-y^2}{2\sigma_2^2}}$.
- We just need to compute $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$.
- We could compute this directly.

Summing two normal variables

- X is normal with mean zero, variance σ_1^2, Y is normal with mean zero, variance σ_2^2.
- $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-x^2}{2\sigma_1^2}}$ and $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-y^2}{2\sigma_2^2}}$.
- We just need to compute $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$.
- We could compute this directly.
- Or we could argue with a multi-dimensional bell curve picture that if X and Y have variance 1 then $f_{\sigma_1 X + \sigma_2 Y}$ is the density of a normal random variable (and note that variances and expectations are additive).

Summing two normal variables

- X is normal with mean zero, variance σ_1^2, Y is normal with mean zero, variance σ_2^2.
- $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-x^2}{2\sigma_1^2}}$ and $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-y^2}{2\sigma_2^2}}$.
- We just need to compute $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$.
- We could compute this directly.
- Or we could argue with a multi-dimensional bell curve picture that if X and Y have variance 1 then $f_{\sigma_1 X + \sigma_2 Y}$ is the density of a normal random variable (and note that variances and expectations are additive).
- Or use fact that if $A_i \in \{-1, 1\}$ are i.i.d. coin tosses then $\frac{1}{\sqrt{N}} \sum_{i=1}^{\sigma^2 N} A_i$ is approximately normal with variance σ^2 when N is large.

Summing two normal variables

- X is normal with mean zero, variance σ_1^2, Y is normal with mean zero, variance σ_2^2.
- $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-x^2}{2\sigma_1^2}}$ and $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-y^2}{2\sigma_2^2}}$.
- We just need to compute $f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) dy$.
- We could compute this directly.
- Or we could argue with a multi-dimensional bell curve picture that if X and Y have variance 1 then $f_{\sigma_1 X + \sigma_2 Y}$ is the density of a normal random variable (and note that variances and expectations are additive).
- Or use fact that if $A_i \in \{-1, 1\}$ are i.i.d. coin tosses then $\frac{1}{\sqrt{N}} \sum_{i=1}^{\sigma^2 N} A_i$ is approximately normal with variance σ^2 when N is large.
- Generally: if independent random variables X_j are normal (μ_j, σ_j^2) then $\sum_{j=1}^n X_j$ is normal $(\sum_{j=1}^n \mu_j, \sum_{j=1}^n \sigma_j^2)$.

L^2 weak laws of large numbers

- Say X_i and X_j are uncorrelated if $E(X_iX_j) = EX_iEX_j$.

L^2 weak laws of large numbers

- Say X_i and X_j are uncorrelated if $E(X_i X_j) = EX_i EX_j$.
- When random variables are uncorrelated, their variances add.

L^2 weak laws of large numbers

- Say X_i and X_j are uncorrelated if $E(X_i X_j) = EX_i EX_j$.
- When random variables are uncorrelated, their variances add.
- If we have a sequence $X_1, X_2, \ldots$ of uncorrelated random variables with common mean μ and uniformly bounded variance, and we write $S_n = X_1 + \ldots + X_n$, then $S_n/n \to \mu$ in probability.

L^2 weak laws of large numbers

- Say X_i and X_j are uncorrelated if $E(X_i X_j) = EX_i EX_j$.
- When random variables are uncorrelated, their variances add.
- If we have a sequence $X_1, X_2, \dots$ of uncorrelated random variables with common mean μ and uniformly bounded variance, and we write $S_n = X_1 + \dots + X_n$, then $S_n/n \to \mu$ in probability.
- Weak versus strong. Convergence in probability versus a.s. convergence.