18.175: Lecture 5

Moment generating functions

Scott Sheffield

MIT

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Recall definitions

- **Probability space** is triple $(\Omega, \mathcal{F}, P)$ where Ω is sample space, $\mathcal{F}$ is set of events (the σ-algebra) and $P : \mathcal{F} \to [0, 1]$ is the probability function.

Recall definitions

- **Probability space** is triple $(\Omega, \mathcal{F}, P)$ where Ω is sample space, $\mathcal{F}$ is set of events (the σ-algebra) and $P : \mathcal{F} \to [0, 1]$ is the probability function.
- σ-**algebra** is collection of subsets closed under complementation and countable unions. Call $(\Omega, \mathcal{F})$ a measure space.

Recall definitions

- **Probability space** is triple $(\Omega, \mathcal{F}, P)$ where Ω is sample space, $\mathcal{F}$ is set of events (the σ-algebra) and $P : \mathcal{F} \to [0, 1]$ is the probability function.
- σ-**algebra** is collection of subsets closed under complementation and countable unions. Call $(\Omega, \mathcal{F})$ a measure space.
- **Measure** is function $\mu : \mathcal{F} \to \mathbb{R}$ satisfying $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$ and countable additivity: $\mu(\cup_i A_i) = \sum_i \mu(A_i)$ for disjoint A_i.

Recall definitions

- **Probability space** is triple $(\Omega, \mathcal{F}, P)$ where Ω is sample space, $\mathcal{F}$ is set of events (the σ-algebra) and $P : \mathcal{F} \to [0,1]$ is the probability function.
- σ-**algebra** is collection of subsets closed under complementation and countable unions. Call $(\Omega, \mathcal{F})$ a measure space.
- **Measure** is function $\mu : \mathcal{F} \to \mathbb{R}$ satisfying $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$ and countable additivity: $\mu(\cup_i A_i) = \sum_i \mu(A_i)$ for disjoint A_i.
- Measure μ is **probability measure** if $\mu(\Omega) = 1$.

Recall definitions

- **Probability space** is triple $(\Omega, \mathcal{F}, P)$ where Ω is sample space, $\mathcal{F}$ is set of events (the σ-algebra) and $P : \mathcal{F} \to [0, 1]$ is the probability function.
- σ-**algebra** is collection of subsets closed under complementation and countable unions. Call $(\Omega, \mathcal{F})$ a measure space.
- **Measure** is function $\mu : \mathcal{F} \to \mathbb{R}$ satisfying $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$ and countable additivity: $\mu(\cup_i A_i) = \sum_i \mu(A_i)$ for disjoint A_i.
- Measure μ is **probability measure** if $\mu(\Omega) = 1$.
- The **Borel** σ-**algebra** $\mathcal{B}$ on a topological space is the smallest σ-algebra containing all open sets.

Recall definitions

- Real random variable is function $X : \Omega \to \mathbb{R}$ such that the preimage of every Borel set is in $\mathcal{F}$.

Recall definitions

- Real random variable is function $X : \Omega \to \mathbb{R}$ such that the preimage of every Borel set is in $\mathcal{F}$.
- Note: to prove X is measurable, it is enough to show that the pre-image of every open set is in $\mathcal{F}$.

Recall definitions

- Real random variable is function $X : \Omega \to \mathbb{R}$ such that the preimage of every Borel set is in $\mathcal{F}$.
- Note: to prove X is measurable, it is enough to show that the pre-image of every open set is in $\mathcal{F}$.
- Can talk about σ-algebra generated by random variable(s): smallest σ-algebra that makes a random variable (or a collection of random variables) measurable.

Lebesgue integration

- Lebesgue: If you can measure, you can integrate.

Lebesgue integration

- Lebesgue: If you can measure, you can integrate.
- In more words: if $(\Omega, \mathcal{F})$ is a measure space with a measure μ and $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable, then we can define $\int f d\mu$ (for non-negative f, also if both $f \vee 0$ and $-f \wedge 0$ and have finite integrals...)

Lebesgue integration

- Lebesgue: If you can measure, you can integrate.
- In more words: if $(\Omega, \mathcal{F})$ is a measure space with a measure μ and $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable, then we can define $\int f d\mu$ (for non-negative f, also if both $f \vee 0$ and $-f \wedge 0$ and have finite integrals...)
- Idea: start with $\mu(\Omega) < \infty$ case, then define integral, verify linearity and positivity (a.e. non-negative functions have non-negative integrals) in 4 cases:

Lebesgue integration

- Lebesgue: If you can measure, you can integrate.
- In more words: if $(\Omega, \mathcal{F})$ is a measure space with a measure μ and $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable, then we can define $\int f d\mu$ (for non-negative f, also if both $f \vee 0$ and $-f \wedge 0$ and have finite integrals...)
- Idea: start with $\mu(\Omega) < \infty$ case, then define integral, verify linearity and positivity (a.e. non-negative functions have non-negative integrals) in 4 cases:
  - f takes only finitely many values.

Lebesgue integration

- Lebesgue: If you can measure, you can integrate.
- In more words: if $(\Omega, \mathcal{F})$ is a measure space with a measure μ and $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable, then we can define $\int f d\mu$ (for non-negative f, also if both $f \vee 0$ and $-f \wedge 0$ and have finite integrals...)
- Idea: start with $\mu(\Omega) < \infty$ case, then define integral, verify linearity and positivity (a.e. non-negative functions have non-negative integrals) in 4 cases:
  - f takes only finitely many values.
  - f is bounded (hint: reduce to previous case by rounding down or up to nearest multiple of ϵ for $\epsilon \to 0$).

Lebesgue integration

- Lebesgue: If you can measure, you can integrate.
- In more words: if $(\Omega, \mathcal{F})$ is a measure space with a measure μ and $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable, then we can define $\int f d\mu$ (for non-negative f, also if both $f \vee 0$ and $-f \wedge 0$ and have finite integrals...)
- Idea: start with $\mu(\Omega) < \infty$ case, then define integral, verify linearity and positivity (a.e. non-negative functions have non-negative integrals) in 4 cases:
  - f takes only finitely many values.
  - f is bounded (hint: reduce to previous case by rounding down or up to nearest multiple of ϵ for $\epsilon \to 0$).
  - f is non-negative (hint: reduce to previous case by taking $f \wedge N$ for $N \to \infty$).

Lebesgue integration

- Lebesgue: If you can measure, you can integrate.
- In more words: if $(\Omega, \mathcal{F})$ is a measure space with a measure μ and $f : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable, then we can define $\int f d\mu$ (for non-negative f, also if both $f \vee 0$ and $-f \wedge 0$ and have finite integrals...)
- Idea: start with $\mu(\Omega) < \infty$ case, then define integral, verify linearity and positivity (a.e. non-negative functions have non-negative integrals) in 4 cases:
  - f takes only finitely many values.
  - f is bounded (hint: reduce to previous case by rounding down or up to nearest multiple of ϵ for $\epsilon \to 0$).
  - f is non-negative (hint: reduce to previous case by taking $f \wedge N$ for $N \to \infty$).
  - f is any measurable function (hint: treat positive/negative parts separately, difference makes sense if both integrals finite).

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Expectation

- Given probability space $(\Omega, \mathcal{F}, P)$ and random variable X, we write $EX = \int XdP$. Always defined if $X \geq 0$, or if integrals of $\max\{X, 0\}$ and $\min\{X, 0\}$ are separately finite.

Expectation

- Given probability space $(\Omega, \mathcal{F}, P)$ and random variable X, we write $EX = \int X dP$. Always defined if $X \geq 0$, or if integrals of $\max\{X, 0\}$ and $\min\{X, 0\}$ are separately finite.
- Since expectation is an integral, we can interpret our basic properties of integrals (as well as results to come: Jensen's inequality, Hölder's inequality, Fatou's lemma, monotone convergence, dominated convergence, etc.) as properties of expectation.

Expectation

- Given probability space $(\Omega, \mathcal{F}, P)$ and random variable X, we write $EX = \int XdP$. Always defined if $X \geq 0$, or if integrals of $\max\{X, 0\}$ and $\min\{X, 0\}$ are separately finite.
- Since expectation is an integral, we can interpret our basic properties of integrals (as well as results to come: Jensen's inequality, Hölder's inequality, Fatou's lemma, monotone convergence, dominated convergence, etc.) as properties of expectation.
- EX^k is called k**th moment of** X. Also, if $m = EX$ then $E(X - m)^2$ is called the **variance** of X.

Properties of expectation/integration

- **Jensen's inequality:** If μ is probability measure and $\phi : \mathbb{R} \to \mathbb{R}$ is convex then $\phi(\int f d\mu) \leq \int \phi(f) d\mu$. If X is random variable then $E\phi(X) \geq \phi(EX)$.

Properties of expectation/integration

- **Jensen's inequality:** If μ is probability measure and $\phi : \mathbb{R} \to \mathbb{R}$ is convex then $\phi(\int f d\mu) \leq \int \phi(f) d\mu$. If X is random variable then $E\phi(X) \geq \phi(EX)$.
- **Main idea of proof:** Approximate ϕ below by linear function L that agrees with ϕ at EX.

Properties of expectation/integration

- **Jensen's inequality:** If μ is probability measure and $\phi : \mathbb{R} \to \mathbb{R}$ is convex then $\phi(\int f d\mu) \leq \int \phi(f) d\mu$. If X is random variable then $E\phi(X) \geq \phi(EX)$.
- **Main idea of proof:** Approximate ϕ below by linear function L that agrees with ϕ at EX.
- **Applications:** Utility, hedge fund payout functions.

Properties of expectation/integration

- **Jensen's inequality:** If μ is probability measure and $\phi : \mathbb{R} \to \mathbb{R}$ is convex then $\phi(\int f d\mu) \leq \int \phi(f) d\mu$. If X is random variable then $E\phi(X) \geq \phi(EX)$.
- **Main idea of proof:** Approximate ϕ below by linear function L that agrees with ϕ at EX.
- **Applications:** Utility, hedge fund payout functions.
- **Hölder's inequality:** Write $\|f\|_p = (\int |f|^p d\mu)^{1/p}$ for $1 \leq p < \infty$. If $1/p + 1/q = 1$, then $\int |fg| d\mu \leq \|f\|_p \|g\|_q$.

Properties of expectation/integration

- **Jensen's inequality:** If μ is probability measure and $\phi : \mathbb{R} \to \mathbb{R}$ is convex then $\phi(\int f d\mu) \le \int \phi(f) d\mu$. If X is random variable then $E\phi(X) \ge \phi(EX)$.
- **Main idea of proof:** Approximate ϕ below by linear function L that agrees with ϕ at EX.
- **Applications:** Utility, hedge fund payout functions.
- **Hölder's inequality:** Write $\|f\|_p = (\int |f|^p d\mu)^{1/p}$ for $1 \le p < \infty$. If $1/p + 1/q = 1$, then $\int |fg| d\mu \le \|f\|_p \|g\|_q$.
- **Main idea of proof:** Rescale so that $\|f\|_p \|g\|_q = 1$. Use some basic calculus to check that for any positive x and y we have $xy \le x^p/p + y^q/p$. Write $x = |f|$, $y = |g|$ and integrate to get $\int |fg| d\mu \le \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p \|g\|_q$.

Properties of expectation/integration

- **Jensen's inequality:** If μ is probability measure and $\phi : \mathbb{R} \to \mathbb{R}$ is convex then $\phi(\int f d\mu) \leq \int \phi(f) d\mu$. If X is random variable then $E\phi(X) \geq \phi(EX)$.
- **Main idea of proof:** Approximate ϕ below by linear function L that agrees with ϕ at EX.
- **Applications:** Utility, hedge fund payout functions.
- **Hölder's inequality:** Write $\|f\|_p = (\int |f|^p d\mu)^{1/p}$ for $1 \leq p < \infty$. If $1/p + 1/q = 1$, then $\int |fg| d\mu \leq \|f\|_p \|g\|_q$.
- **Main idea of proof:** Rescale so that $\|f\|_p \|g\|_q = 1$. Use some basic calculus to check that for any positive x and y we have $xy \leq x^p/p + y^q/p$. Write $x = |f|$, $y = |g|$ and integrate to get $\int |fg| d\mu \leq \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p \|g\|_q$.
- **Cauchy-Schwarz inequality:** Special case $p = q = 2$. Gives $\int |fg| d\mu \leq \|f\|_2 \|g\|_2$. Says that dot product of two vectors is at most product of vector lengths.

Bounded convergence theorem

- **Bounded convergence theorem:** Consider *probability* measure μ and suppose $|f_n| \leq M$ a.s. for all n and some fixed $M > 0$, and that $f_n \to f$ in probability (i.e., $\lim_{n\to\infty} \mu\{x : |f_n(x) - f(x)| > \epsilon\} = 0$ for all $\epsilon > 0$). Then

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

(Build counterexample for infinite measure space using wide and short rectangles?...)

Bounded convergence theorem

- **Bounded convergence theorem:** Consider *probability* measure μ and suppose $|f_n| \leq M$ a.s. for all n and some fixed $M > 0$, and that $f_n \to f$ in probability (i.e., $\lim_{n\to\infty} \mu\{x : |f_n(x) - f(x)| > \epsilon\} = 0$ for all $\epsilon > 0$). Then

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

  (Build counterexample for infinite measure space using wide and short rectangles?...)
- **Main idea of proof:** for any ϵ, δ can take n large enough so $\int |f_n - f| d\mu < M\delta + \epsilon$.

Fatou's lemma

- **Fatou's lemma:** If $f_n \geq 0$ then

$$\liminf_{n\to\infty} \int f_n d\mu \geq \int (\liminf_{n\to\infty} f_n) d\mu.$$

(Counterexample for opposite-direction inequality using thin and tall rectangles?)

Fatou's lemma

- **Fatou's lemma:** If $f_n \geq 0$ then
$$\liminf_{n\to\infty} \int f_n d\mu \geq \int (\liminf_{n\to\infty} f_n) d\mu.$$
(Counterexample for opposite-direction inequality using thin and tall rectangles?)
- **Main idea of proof:** first reduce to case that the f_n are increasing by writing $g_n(x) = \inf_{m\geq n} f_m(x)$ and observing that $g_n(x) \uparrow g(x) = \liminf_{n\to\infty} f_n(x)$. Then truncate, used bounded convergence, take limits.

More integral properties

- **Monotone convergence:** If $f_n \geq 0$ and $f_n \uparrow f$ then
$$\int f_n d\mu \uparrow \int f d\mu.$$

More integral properties

- **Monotone convergence:** If $f_n \geq 0$ and $f_n \uparrow f$ then
$$\int f_n d\mu \uparrow \int f d\mu.$$
- **Main idea of proof:** one direction obvious, Fatou gives other.

More integral properties

- **Monotone convergence:** If $f_n \geq 0$ and $f_n \uparrow f$ then

$$\int f_n d\mu \uparrow \int f d\mu.$$

- **Main idea of proof:** one direction obvious, Fatou gives other.
- **Dominated convergence:** If $f_n \to f$ a.e. and $|f_n| \leq g$ for all n and g is integrable, then $\int f_n d\mu \to \int f d\mu$.

More integral properties

- **Monotone convergence:** If $f_n \geq 0$ and $f_n \uparrow f$ then
$$\int f_n d\mu \uparrow \int f d\mu.$$
- **Main idea of proof:** one direction obvious, Fatou gives other.
- **Dominated convergence:** If $f_n \to f$ a.e. and $|f_n| \leq g$ for all n and g is integrable, then $\int f_n d\mu \to \int f d\mu$.
- **Main idea of proof:** Fatou for functions $g + f_n \geq 0$ gives one side. Fatou for $g - f_n \geq 0$ gives other.

Computing expectations

- Change of variables. Measure space $(\Omega, \mathcal{F}, P)$. Let X be random variable in $(S, \mathcal{S})$ with distribution μ. Then if $f(S, \mathcal{S}) \to (R, \mathcal{R})$ is measurable we have $Ef(X) = \int_S f(y)\mu(dy)$.

Computing expectations

- Change of variables. Measure space $(\Omega, \mathcal{F}, P)$. Let X be random variable in $(S, \mathcal{S})$ with distribution μ. Then if $f(S, \mathcal{S}) \to (R, \mathcal{R})$ is measurable we have $Ef(X) = \int_S f(y)\mu(dy)$.
- Prove by checking for indicators, simple functions, non-negative functions, integrable functions.

Computing expectations

- Change of variables. Measure space $(\Omega, \mathcal{F}, P)$. Let X be random variable in $(S, \mathcal{S})$ with distribution μ. Then if $f(S, \mathcal{S}) \to (R, \mathcal{R})$ is measurable we have $Ef(X) = \int_S f(y)\mu(dy)$.
- Prove by checking for indicators, simple functions, non-negative functions, integrable functions.
- Examples: normal, exponential, Bernoulli, Poisson, geometric...

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Moment generating functions

- Let X be a random variable.

Moment generating functions

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

Moment generating functions

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

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.

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.

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 $M(0) = 1$.

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 $E[e^{tX}] \geq E[e^{t \min\{X,b\}}] \geq P\{X \geq b\} e^{tb}$.

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 $|t| \to \infty$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.
- Then $M'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}]$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.
- Then $M'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}]$.
- in particular, $M'(0) = E[X]$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.
- Then $M'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}]$.
- in particular, $M'(0) = E[X]$.
- Also $M''(t) = \frac{d}{dt}M'(t) = \frac{d}{dt}E[Xe^{tX}] = E[X^2e^{tX}]$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.
- Then $M'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}]$.
- in particular, $M'(0) = E[X]$.
- Also $M''(t) = \frac{d}{dt}M'(t) = \frac{d}{dt}E[Xe^{tX}] = E[X^2e^{tX}]$.
- So $M''(0) = E[X^2]$. Same argument gives that nth derivative of M at zero is $E[X^n]$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.
- Then $M'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}]$.
- in particular, $M'(0) = E[X]$.
- Also $M''(t) = \frac{d}{dt}M'(t) = \frac{d}{dt}E[Xe^{tX}] = E[X^2e^{tX}]$.
- So $M''(0) = E[X^2]$. Same argument gives that nth derivative of M at zero is $E[X^n]$.
- Interesting: knowing all of the derivatives of M at a single point tells you the moments $E[X^k]$ for all integer $k \geq 0$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.
- Then $M'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}]$.
- in particular, $M'(0) = E[X]$.
- Also $M''(t) = \frac{d}{dt}M'(t) = \frac{d}{dt}E[Xe^{tX}] = E[X^2e^{tX}]$.
- So $M''(0) = E[X^2]$. Same argument gives that nth derivative of M at zero is $E[X^n]$.
- Interesting: knowing all of the derivatives of M at a single point tells you the moments $E[X^k]$ for all integer $k \geq 0$.
- Another way to think of this: write $e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \dots$.

Moment generating functions actually generate moments

- Let X be a random variable and $M(t) = E[e^{tX}]$.
- Then $M'(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}]$.
- in particular, $M'(0) = E[X]$.
- Also $M''(t) = \frac{d}{dt}M'(t) = \frac{d}{dt}E[Xe^{tX}] = E[X^2e^{tX}]$.
- So $M''(0) = E[X^2]$. Same argument gives that nth derivative of M at zero is $E[X^n]$.
- Interesting: knowing all of the derivatives of M at a single point tells you the moments $E[X^k]$ for all integer $k \geq 0$.
- Another way to think of this: write $e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \ldots$.
- Taking expectations gives $E[e^{tX}] = 1 + tm_1 + \frac{t^2m_2}{2!} + \frac{t^3m_3}{3!} + \ldots$, where m_k is the kth moment. The kth derivative at zero is m_k.

Moment generating functions for independent sums

- Let X and Y be independent random variables and $Z = X + Y$.

Moment generating functions for independent sums

- Let X and Y be independent random variables and $Z = X + Y$.
- Write the moment generating functions as $M_X(t) = E[e^{tX}]$ and $M_Y(t) = E[e^{tY}]$ and $M_Z(t) = E[e^{tZ}]$.

Moment generating functions for independent sums

- Let X and Y be independent random variables and $Z = X + Y$.
- Write the moment generating functions as $M_X(t) = E[e^{tX}]$ and $M_Y(t) = E[e^{tY}]$ and $M_Z(t) = E[e^{tZ}]$.
- If you knew M_X and M_Y, could you compute M_Z?

Moment generating functions for independent sums

- Let X and Y be independent random variables and $Z = X + Y$.
- Write the moment generating functions as $M_X(t) = E[e^{tX}]$ and $M_Y(t) = E[e^{tY}]$ and $M_Z(t) = E[e^{tZ}]$.
- If you knew M_X and M_Y, could you compute M_Z?
- By independence, $M_Z(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$ for all t.

Moment generating functions for independent sums

- Let X and Y be independent random variables and $Z = X + Y$.
- Write the moment generating functions as $M_X(t) = E[e^{tX}]$ and $M_Y(t) = E[e^{tY}]$ and $M_Z(t) = E[e^{tZ}]$.
- If you knew M_X and M_Y, could you compute M_Z?
- By independence, $M_Z(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$ for all t.
- In other words, adding independent random variables corresponds to multiplying moment generating functions.

Moment generating functions for sums of i.i.d. random variables

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

Moment generating functions for sums of i.i.d. random variables

- 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?

Moment generating functions for sums of i.i.d. random variables

- 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.

Moment generating functions for sums of i.i.d. random variables

- 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.

Other observations

- If $Z = aX$ then can I use M_X to determine M_Z?

Other observations

- If $Z = aX$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{taX}] = M_X(at)$.

Other observations

- If $Z = aX$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{taX}] = M_X(at)$.
- If $Z = X + b$ then can I use M_X to determine M_Z?

Other observations

- If $Z = aX$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{taX}] = M_X(at)$.
- If $Z = X + b$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{tX+bt}] = e^{bt}M_X(t)$.

Other observations

- If $Z = aX$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{taX}] = M_X(at)$.
- If $Z = X + b$ then can I use M_X to determine M_Z?
- Answer: Yes. $M_Z(t) = E[e^{tZ}] = E[e^{tX+bt}] = e^{bt}M_X(t)$.
- Latter answer is the special case of $M_Z(t) = M_X(t)M_Y(t)$ where Y is the constant random variable b.

Existence issues

- Seems that unless $f_X(x)$ decays superexponentially as x tends to infinity, we won't have $M_X(t)$ defined for all t.

Existence issues

- Seems that unless $f_X(x)$ decays superexponentially as x tends to infinity, we won't have $M_X(t)$ defined for all t.
- What is M_X if X is standard Cauchy, so that $f_X(x) = \frac{1}{\pi(1+x^2)}$.

Existence issues

- Seems that unless $f_X(x)$ decays superexponentially as x tends to infinity, we won't have $M_X(t)$ defined for all t.
- What is M_X if X is standard Cauchy, so that $f_X(x) = \frac{1}{\pi(1+x^2)}$.
- Answer: $M_X(0) = 1$ (as is true for any X) but otherwise $M_X(t)$ is infinite for all $t \neq 0$.

Existence issues

- Seems that unless $f_X(x)$ decays superexponentially as x tends to infinity, we won't have $M_X(t)$ defined for all t.
- What is M_X if X is standard Cauchy, so that $f_X(x) = \frac{1}{\pi(1+x^2)}$.
- Answer: $M_X(0) = 1$ (as is true for any X) but otherwise $M_X(t)$ is infinite for all $t \neq 0$.
- Informal statement: moment generating functions are not defined for distributions with fat tails.

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Markov's and Chebyshev's inequalities

- **Markov's inequality:** Let X be non-negative random variable. Fix $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.

Markov's and Chebyshev's inequalities

- **Markov's inequality:** Let X be non-negative random variable. Fix $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Proof:** Consider a random variable Y defined by $Y = \begin{cases} a & X \geq a \\ 0 & X < a \end{cases}$. Since $X \geq Y$ with probability one, it follows that $E[X] \geq E[Y] = aP\{X \geq a\}$. Divide both sides by a to get Markov's inequality.

Markov's and Chebyshev's inequalities

- **Markov's inequality:** Let X be non-negative random variable. Fix $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Proof:** Consider a random variable Y defined by $Y = \begin{cases} a & X \geq a \\ 0 & X < a \end{cases}$. Since $X \geq Y$ with probability one, it follows that $E[X] \geq E[Y] = aP\{X \geq a\}$. Divide both sides by a to get Markov's inequality.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

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

Markov's and Chebyshev's inequalities

- **Markov's inequality:** Let X be non-negative random variable. Fix $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Proof:** Consider a random variable Y defined by $Y = \begin{cases} a & X \geq a \\ 0 & X < a \end{cases}$. Since $X \geq Y$ with probability one, it follows that $E[X] \geq E[Y] = aP\{X \geq a\}$. Divide both sides by a to get Markov's inequality.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

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

- **Proof:** Note that $(X - \mu)^2$ is a non-negative random variable and $P\{|X - \mu| \geq k\} = P\{(X - \mu)^2 \geq k^2\}$. Now apply Markov's inequality with $a = k^2$.

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be non-negative random variable with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be non-negative random variable with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

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

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be non-negative random variable with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

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

- Inequalities allow us to deduce limited information about a distribution when we know only the mean (Markov) or the mean and variance (Chebyshev).

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be non-negative random variable with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

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

- Inequalities allow us to deduce limited information about a distribution when we know only the mean (Markov) or the mean and variance (Chebyshev).
- **Markov:** if $E[X]$ is small, then it is not too likely that X is large.

Markov and Chebyshev: rough idea

- **Markov's inequality:** Let X be non-negative random variable with finite mean. Fix a constant $a > 0$. Then $P\{X \geq a\} \leq \frac{E[X]}{a}$.
- **Chebyshev's inequality:** If X has finite mean μ, variance σ^2, and $k > 0$ then

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

- Inequalities allow us to deduce limited information about a distribution when we know only the mean (Markov) or the mean and variance (Chebyshev).
- **Markov:** if $E[X]$ is small, then it is not too likely that X is large.
- **Chebyshev:** if $\sigma^2 = \mathrm{Var}[X]$ is small, then it is not too likely that X is far from its mean.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.
- Then the value $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ is called the *empirical average* of the first n trials.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.
- Then the value $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ is called the *empirical average* of the first n trials.
- We'd guess that when n is large, A_n is typically close to μ.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.
- Then the value $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ is called the *empirical average* of the first n trials.
- We'd guess that when n is large, A_n is typically close to μ.
- Indeed, **weak law of large numbers** states that for all $\epsilon > 0$ we have $\lim_{n\to\infty} P\{|A_n - \mu| > \epsilon\} = 0$.

Statement of weak law of large numbers

- Suppose X_i are i.i.d. random variables with mean μ.
- Then the value $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ is called the *empirical average* of the first n trials.
- We'd guess that when n is large, A_n is typically close to μ.
- Indeed, **weak law of large numbers** states that for all $\epsilon > 0$ we have $\lim_{n\to\infty} P\{|A_n - \mu| > \epsilon\} = 0$.
- Example: as n tends to infinity, the probability of seeing more than $.50001n$ heads in n fair coin tosses tends to zero.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- By additivity of expectation, $\mathbb{E}[A_n] = \mu$.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- By additivity of expectation, $\mathbb{E}[A_n] = \mu$.
- Similarly, $\mathrm{Var}[A_n] = \frac{n\sigma^2}{n^2} = \sigma^2/n$.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- By additivity of expectation, $\mathbb{E}[A_n] = \mu$.
- Similarly, $\mathrm{Var}[A_n] = \frac{n\sigma^2}{n^2} = \sigma^2/n$.
- By Chebyshev $P\{|A_n - \mu| \geq \epsilon\} \leq \frac{\mathrm{Var}[A_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$.

Proof of weak law of large numbers in finite variance case

- As above, let X_i be i.i.d. random variables with mean μ and write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$.
- By additivity of expectation, $\mathbb{E}[A_n] = \mu$.
- Similarly, $\mathrm{Var}[A_n] = \frac{n\sigma^2}{n^2} = \sigma^2/n$.
- By Chebyshev $P\{|A_n - \mu| \geq \epsilon\} \leq \frac{\mathrm{Var}[A_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$.
- No matter how small ϵ is, RHS will tend to zero as n gets large.

L^2 weak law of large numbers

- Say X_i and X_j are uncorrelated if $E(X_i X_j) = EX_i EX_j$.

L^2 weak law of large numbers

- Say X_i and X_j are uncorrelated if $E(X_i X_j) = EX_i EX_j$.
- Chebyshev/Markov argument works whenever variables are uncorrelated (does not actually require independence).

What else can you do with just variance bounds?

- Having "almost uncorrelated" X_i is sometimes enough: just need variance of A_n to go to zero.

What else can you do with just variance bounds?

- Having "almost uncorrelated" X_i is sometimes enough: just need variance of A_n to go to zero.
- Toss αn bins into n balls. How many bins are filled?

What else can you do with just variance bounds?

- Having "almost uncorrelated" X_i is sometimes enough: just need variance of A_n to go to zero.
- Toss αn bins into n balls. How many bins are filled?
- When n is large, the number of balls in the first bin is approximately a Poisson random variable with expectation α.

What else can you do with just variance bounds?

- Having "almost uncorrelated" X_i is sometimes enough: just need variance of A_n to go to zero.
- Toss αn bins into n balls. How many bins are filled?
- When n is large, the number of balls in the first bin is approximately a Poisson random variable with expectation α.
- Probability first bin contains no ball is $(1 - 1/n)^{\alpha n} \approx e^{-\alpha}$.

What else can you do with just variance bounds?

- Having "almost uncorrelated" X_i is sometimes enough: just need variance of A_n to go to zero.
- Toss αn bins into n balls. How many bins are filled?
- When n is large, the number of balls in the first bin is approximately a Poisson random variable with expectation α.
- Probability first bin contains no ball is $(1 - 1/n)^{\alpha n} \approx e^{-\alpha}$.
- We can explicitly compute variance of the number of bins with no balls. Allows us to show that fraction of bins with no balls concentrates about its expectation, which is $e^{-\alpha}$.

How do you extend to random variables without variance?

- Assume X_n are i.i.d. non-negative instances of random variable X with finite mean. Can one prove law of large numbers for these?

How do you extend to random variables without variance?

- Assume X_n are i.i.d. non-negative instances of random variable X with finite mean. Can one prove law of large numbers for these?
- Try truncating. Fix large N and write $A = X1_{X>N}$ and $B = X1_{X\le N}$ so that $X = A + B$. Choose N so that EB is very small. Law of large numbers holds for A.

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Outline

Integration

Expectation

Moment generating functions

Weak law of large numbers: Markov/Chebyshev approach

Weak law of large numbers: characteristic function approach

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?
- In this strange and delightful case A_n actually has the same probability distribution as X.

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?
- In this strange and delightful case A_n actually has the same probability distribution as X.
- In particular, the A_n are not tightly concentrated around any particular value even when n is very large.

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?
- In this strange and delightful case A_n actually has the same probability distribution as X.
- In particular, the A_n are not tightly concentrated around any particular value even when n is very large.
- But weak law holds as long as $E[|X|]$ is finite, so that μ is well defined.

Extent of weak law

- Question: does the weak law of large numbers apply no matter what the probability distribution for X is?
- Is it always the case that if we define $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$ then A_n is typically close to some fixed value when n is large?
- What if X is Cauchy?
- In this strange and delightful case A_n actually has the same probability distribution as X.
- In particular, the A_n are not tightly concentrated around any particular value even when n is very large.
- But weak law holds as long as $E[|X|]$ is finite, so that μ is well defined.
- One standard proof uses 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

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

Continuity theorems

- 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.

Continuity theorems

- 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.
- The weak law of large numbers can be rephrased as the statement that A_n converges in law to μ (i.e., to the random variable that is equal to μ with probability one).

Continuity theorems

- 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.
- The weak law of large numbers can be rephrased as the statement that A_n converges in law to μ (i.e., to the random variable that is equal to μ with probability one).
- **Lévy's continuity theorem (coming later):** 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

- 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.
- The weak law of large numbers can be rephrased as the statement that A_n converges in law to μ (i.e., to the random variable that is equal to μ with probability one).
- **Lévy's continuity theorem (coming later):** 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 weak law of large numbers by showing $\lim_{n\to\infty} \phi_{A_n}(t) = \phi_\mu(t) = e^{it\mu}$ for all t. When $\mu = 0$, 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, for all t, $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$, then X_n converge in law to X.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X - \mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X]=0$, we have $\phi'_X(0) = E[\frac{\partial}{\partial t}e^{itX}]_{t=0} = iE[X] = 0$.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X]=0$, we have $\phi_X'(0) = E[\frac{\partial}{\partial t} e^{itX}]_{t=0} = iE[X] = 0$.
- Write $g(t) = \log \phi_X(t)$ so $\phi_X(t) = e^{g(t)}$. Then $g(0)=0$ and (by chain rule) $g'(0) = \lim_{\epsilon\to 0} \frac{g(\epsilon)-g(0)}{\epsilon} = \lim_{\epsilon\to 0} \frac{g(\epsilon)}{\epsilon} = 0$.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X]=0$, we have $\phi'_X(0) = E[\frac{\partial}{\partial t}e^{itX}]_{t=0} = iE[X] = 0$.
- Write $g(t) = \log \phi_X(t)$ so $\phi_X(t) = e^{g(t)}$. Then $g(0)=0$ and (by chain rule) $g'(0) = \lim_{\epsilon\to 0}\frac{g(\epsilon)-g(0)}{\epsilon} = \lim_{\epsilon\to 0}\frac{g(\epsilon)}{\epsilon} = 0$.
- Now $\phi_{A_n}(t) = \phi_X(t/n)^n = e^{ng(t/n)}$. Since $g(0)=g'(0)=0$ we have $\lim_{n\to\infty} ng(t/n) = \lim_{n\to\infty} t\frac{g(\frac{t}{n})}{\frac{t}{n}} = 0$ if t is fixed. Thus $\lim_{n\to\infty} e^{ng(t/n)} = 1$ for all t.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X]=0$, we have $\phi_X'(0) = E[\frac{\partial}{\partial t}e^{itX}]_{t=0} = iE[X] = 0$.
- Write $g(t) = \log \phi_X(t)$ so $\phi_X(t) = e^{g(t)}$. Then $g(0)=0$ and (by chain rule) $g'(0) = \lim_{\epsilon\to 0} \frac{g(\epsilon)-g(0)}{\epsilon} = \lim_{\epsilon \to 0}\frac{g(\epsilon)}{\epsilon} = 0$.
- Now $\phi_{A_n}(t) = \phi_X(t/n)^n = e^{ng(t/n)}$. Since $g(0) = g'(0) = 0$ we have $\lim_{n\to\infty} ng(t/n) = \lim_{n\to\infty} t\frac{g(\frac{t}{n})}{\frac{t}{n}} = 0$ if t is fixed. Thus $\lim_{n\to\infty} e^{ng(t/n)} = 1$ for all t.

Proof sketch for weak law of large numbers, finite mean case

- As above, let X_i be i.i.d. instances of random variable X with mean zero. Write $A_n := \frac{X_1+X_2+\ldots+X_n}{n}$. Weak law of large numbers holds for i.i.d. instances of X if and only if it holds for i.i.d. instances of $X-\mu$. Thus it suffices to prove the weak law in the mean zero case.
- Consider the characteristic function $\phi_X(t) = E[e^{itX}]$.
- Since $E[X]=0$, we have $\phi_X'(0) = E[\frac{\partial}{\partial t} e^{itX}]_{t=0} = iE[X] = 0$.
- Write $g(t) = \log \phi_X(t)$ so $\phi_X(t) = e^{g(t)}$. Then $g(0)=0$ and (by chain rule) $g'(0) = \lim_{\epsilon\to 0} \frac{g(\epsilon)-g(0)}{\epsilon} = \lim_{\epsilon\to 0}\frac{g(\epsilon)}{\epsilon} = 0$.
- Now $\phi_{A_n}(t) = \phi_X(t/n)^n = e^{ng(t/n)}$. Since $g(0)=g'(0)=0$ we have $\lim_{n\to\infty} ng(t/n) = \lim_{n\to\infty} t\frac{g(\frac{t}{n})}{\frac{t}{n}} = 0$ if t is fixed. Thus $\lim_{n\to\infty} e^{ng(t/n)} = 1$ for all t.
- By Lévy's continuity theorem, the A_n converge in law to 0 (i.e., to the random variable that is 0 with probability one).