18.175: Lecture 10

Characteristic functions and central limit theorem

Scott Sheffield

MIT

Outline

Large deviations

Characteristic functions and central limit theorem

Outline

Large deviations

Characteristic functions and central limit theorem

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.

Large deviations

- Consider i.i.d. random variables X_i. Can we show that $P(S_n \geq na) \to 0$ exponentially fast when $a > E[X_i]$?

Large deviations

- Consider i.i.d. random variables X_i. Can we show that $P(S_n \geq na) \to 0$ 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).

General large deviation principle

- More general framework: a *large deviation principle* describes limiting behavior as $n \to \infty$ of family $\{\mu_n\}$ of measures on measure space $(\mathcal{X}, \mathcal{B})$ in terms of a *rate function* I.

General large deviation principle

- More general framework: a *large deviation principle* describes limiting behavior as $n \to \infty$ of family $\{\mu_n\}$ of measures on measure space $(\mathcal{X}, \mathcal{B})$ in terms of a *rate function* I.
- The **rate function** is a lower-semicontinuous map $I : \mathcal{X} \to [0, \infty]$. (The sets $\{x : I(x) \le a\}$ are closed — rate function called "good" if these sets are compact.)

General large deviation principle

- More general framework: a *large deviation principle* describes limiting behavior as $n \to \infty$ of family $\{\mu_n\}$ of measures on measure space $(\mathcal{X}, \mathcal{B})$ in terms of a *rate function* I.
- The **rate function** is a lower-semicontinuous map $I : \mathcal{X} \to [0, \infty]$. (The sets $\{x : I(x) \le a\}$ are closed — rate function called "good" if these sets are compact.)
- **DEFINITION:** $\{\mu_n\}$ satisfy LDP with rate function I and speed n if for all $\Gamma \in \mathcal{B}$,

$$-\inf_{x \in \Gamma^0} I(x) \le \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \le \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \le -\inf_{x \in \overline{\Gamma}} I(x).$$

General large deviation principle

- More general framework: a *large deviation principle* describes limiting behavior as $n \to \infty$ of family $\{\mu_n\}$ of measures on measure space $(\mathcal{X}, \mathcal{B})$ in terms of a *rate function* I.
- The **rate function** is a lower-semicontinuous map $I : \mathcal{X} \to [0, \infty]$. (The sets $\{x : I(x) \le a\}$ are closed — rate function called "good" if these sets are compact.)
- **DEFINITION:** $\{\mu_n\}$ satisfy LDP with rate function I and speed n if for all $\Gamma \in \mathcal{B}$,

$$-\inf_{x \in \Gamma^0} I(x) \le \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \le \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \le -\inf_{x \in \overline{\Gamma}} I(x).$$

- **INTUITION:** when "near x" the probability density function for μ_n is tending to zero like $e^{-I(x)n}$, as $n \to \infty$.

General large deviation principle

- More general framework: a *large deviation principle* describes limiting behavior as $n \to \infty$ of family $\{\mu_n\}$ of measures on measure space $(\mathcal{X}, \mathcal{B})$ in terms of a *rate function* I.
- The **rate function** is a lower-semicontinuous map $I : \mathcal{X} \to [0, \infty]$. (The sets $\{x : I(x) \le a\}$ are closed — rate function called "good" if these sets are compact.)
- **DEFINITION:** $\{\mu_n\}$ satisfy LDP with rate function I and speed n if for all $\Gamma \in \mathcal{B}$,

$$-\inf_{x \in \Gamma^0} I(x) \le \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \le \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \le -\inf_{x \in \overline{\Gamma}} I(x).$$

- **INTUITION:** when "near x" the probability density function for μ_n is tending to zero like $e^{-I(x)n}$, as $n \to \infty$.
- **Simple case:** I is continuous, Γ is closure of its interior.

General large deviation principle

- More general framework: a *large deviation principle* describes limiting behavior as $n \to \infty$ of family $\{\mu_n\}$ of measures on measure space $(\mathcal{X}, \mathcal{B})$ in terms of a *rate function* I.
- The **rate function** is a lower-semicontinuous map $I : \mathcal{X} \to [0, \infty]$. (The sets $\{x : I(x) \leq a\}$ are closed — rate function called "good" if these sets are compact.)
- **DEFINITION:** $\{\mu_n\}$ satisfy LDP with rate function I and speed n if for all $\Gamma \in \mathcal{B}$,

$$-\inf_{x \in \Gamma^0} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq -\inf_{x \in \overline{\Gamma}} I(x).$$

- **INTUITION:** when "near x" the probability density function for μ_n is tending to zero like $e^{-I(x)n}$, as $n \to \infty$.
- **Simple case:** I is continuous, Γ is closure of its interior.
- **Question:** How would I change if we replaced the measures μ_n by weighted measures $e^{(\lambda n, \cdot)}\mu_n$?

General large deviation principle

- More general framework: a *large deviation principle* describes limiting behavior as $n \to \infty$ of family $\{\mu_n\}$ of measures on measure space $(\mathcal{X}, \mathcal{B})$ in terms of a *rate function* I.
- The **rate function** is a lower-semicontinuous map $I : \mathcal{X} \to [0, \infty]$. (The sets $\{x : I(x) \le a\}$ are closed — rate function called "good" if these sets are compact.)
- **DEFINITION:** $\{\mu_n\}$ satisfy LDP with rate function I and speed n if for all $\Gamma \in \mathcal{B}$,

$$-\inf_{x \in \Gamma^0} I(x) \le \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \le \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \le -\inf_{x \in \overline{\Gamma}} I(x).$$

- **INTUITION:** when "near x" the probability density function for μ_n is tending to zero like $e^{-I(x)n}$, as $n \to \infty$.
- **Simple case:** I is continuous, Γ is closure of its interior.
- **Question:** How would I change if we replaced the measures μ_n by weighted measures $e^{(\lambda n, \cdot)} \mu_n$?
- Replace $I(x)$ by $I(x) - (\lambda, x)$? What is $\inf_x I(x) - (\lambda, x)$?

Cramer's theorem

- Let μ_n be law of empirical mean $A_n = \frac{1}{n}\sum_{j=1}^n X_j$ for i.i.d. vectors $X_1, X_2, \ldots, X_n$ in $\mathbb{R}^d$ with same law as X.

Cramer's theorem

- Let μ_n be law of empirical mean $A_n = \frac{1}{n}\sum_{j=1}^n X_j$ for i.i.d. vectors $X_1, X_2, \ldots, X_n$ in $\mathbb{R}^d$ with same law as X.
- Define **log moment generating function** of X by

$$\Lambda(\lambda) = \Lambda_X(\lambda) = \log M_X(\lambda) = \log \mathbb{E} e^{(\lambda, X)},$$

where $(\cdot,\cdot)$ is inner product on $\mathbb{R}^d$.

Cramer's theorem

- Let μ_n be law of empirical mean $A_n = \frac{1}{n}\sum_{j=1}^n X_j$ for i.i.d. vectors $X_1, X_2, \ldots, X_n$ in $\mathbb{R}^d$ with same law as X.
- Define **log moment generating function** of X by

$$\Lambda(\lambda) = \Lambda_X(\lambda) = \log M_X(\lambda) = \log \mathbb{E}e^{(\lambda, X)},$$

  where $(\cdot,\cdot)$ is inner product on $\mathbb{R}^d$.
- Define **Legendre transform** of Λ by

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\}.$$

Cramer's theorem

- Let μ_n be law of empirical mean $A_n = \frac{1}{n}\sum_{j=1}^n X_j$ for i.i.d. vectors $X_1, X_2, \ldots, X_n$ in $\mathbb{R}^d$ with same law as X.
- Define **log moment generating function** of X by

$$\Lambda(\lambda) = \Lambda_X(\lambda) = \log M_X(\lambda) = \log \mathbb{E} e^{(\lambda, X)},$$

  where $(\cdot,\cdot)$ is inner product on $\mathbb{R}^d$.
- Define **Legendre transform** of Λ by

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\}.$$

- **CRAMER'S THEOREM:** μ_n satisfy LDP with convex rate function Λ^*.

Thinking about Cramer's theorem

- Let μ_n be law of empirical mean $A_n = \frac{1}{n}\sum_{j=1}^n X_j$.

Thinking about Cramer's theorem

- Let μ_n be law of empirical mean $A_n = \frac{1}{n}\sum_{j=1}^n X_j$.
- **CRAMER'S THEOREM:** μ_n satisfy LDP with convex rate function

$$I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\},$$

where $\Lambda(\lambda) = \log M(\lambda) = \mathbb{E}e^{(\lambda, X_1)}$.

Thinking about Cramer's theorem

- Let μ_n be law of empirical mean $A_n = \frac{1}{n}\sum_{j=1}^n X_j$.
- **CRAMER'S THEOREM:** μ_n satisfy LDP with convex rate function
$$I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\},$$
where $\Lambda(\lambda) = \log M(\lambda) = \mathbb{E}e^{(\lambda, X_1)}$.
- This means that for all $\Gamma \in \mathcal{B}$ we have this **asymptotic lower bound** on probabilities $\mu_n(\Gamma)$
$$-\inf_{x \in \Gamma^0} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma),$$
so (up to sub-exponential error) $\mu_n(\Gamma) \geq e^{-n \inf_{x \in \Gamma^0} I(x)}$.

Thinking about Cramer's theorem

- Let μ_n be law of empirical mean $A_n = \frac{1}{n}\sum_{j=1}^n X_j$.
- **CRAMER'S THEOREM:** μ_n satisfy LDP with convex rate function

$$I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\},$$

  where $\Lambda(\lambda) = \log M(\lambda) = \mathbb{E}e^{(\lambda, X_1)}$.
- This means that for all $\Gamma \in \mathcal{B}$ we have this **asymptotic lower bound** on probabilities $\mu_n(\Gamma)$

$$-\inf_{x \in \Gamma^0} I(x) \le \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma),$$

  so (up to sub-exponential error) $\mu_n(\Gamma) \ge e^{-n \inf_{x \in \Gamma^0} I(x)}$.
- and this **asymptotic upper bound** on the probabilities $\mu_n(\Gamma)$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \le -\inf_{x \in \overline{\Gamma}} I(x),$$

  which says (up to subexponential error) $\mu_n(\Gamma) \le e^{-n \inf_{x \in \overline{\Gamma}} I(x)}$.

Proving Cramer upper bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d}\{(\lambda, x) - \Lambda(\lambda)\}$.

Proving Cramer upper bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d}\{(\lambda, x) - \Lambda(\lambda)\}$.
- For simplicity, assume that Λ is defined for all x (which implies that X has moments of all orders and Λ and Λ^* are strictly convex, and the derivatives of Λ and Λ' are inverses of each other). It is also enough to consider the case X has mean zero, which implies that $\Lambda(0) = 0$ is a minimum of Λ, and $\Lambda^*(0) = 0$ is a minimum of Λ^*.

Proving Cramer upper bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d}\{(\lambda, x) - \Lambda(\lambda)\}$.
- For simplicity, assume that Λ is defined for all x (which implies that X has moments of all orders and Λ and Λ^* are strictly convex, and the derivatives of Λ and Λ' are inverses of each other). It is also enough to consider the case X has mean zero, which implies that $\Lambda(0) = 0$ is a minimum of Λ, and $\Lambda^*(0) = 0$ is a minimum of Λ^*.
- We aim to show (up to subexponential error) that $\mu_n(\Gamma) \leq e^{-n \inf_{x \in \bar{\Gamma}} I(x)}$.

Proving Cramer upper bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d}\{(\lambda, x) - \Lambda(\lambda)\}$.
- For simplicity, assume that Λ is defined for all x (which implies that X has moments of all orders and Λ and Λ^* are strictly convex, and the derivatives of Λ and Λ' are inverses of each other). It is also enough to consider the case X has mean zero, which implies that $\Lambda(0) = 0$ is a minimum of Λ, and $\Lambda^*(0) = 0$ is a minimum of Λ^*.
- We aim to show (up to subexponential error) that $\mu_n(\Gamma) \leq e^{-n \inf_{x \in \bar{\Gamma}} I(x)}$.
- If Γ were singleton set $\{x\}$ we could find the λ corresponding to x, so $\Lambda^*(x) = (x, \lambda) - \Lambda(\lambda)$. Note then that
$$\mathbb{E}e^{(n\lambda, A_n)} = \mathbb{E}e^{(\lambda, S_n)} = M_X^n(\lambda) = e^{n\Lambda(\lambda)},$$
and also $\mathbb{E}e^{(n\lambda, A_n)} \geq e^{n(\lambda, x)}\mu_n\{x\}$. Taking logs and dividing by n gives $\Lambda(\lambda) \geq \frac{1}{n}\log \mu_n + (\lambda, x)$, so that $\frac{1}{n}\log \mu_n(\Gamma) \leq -\Lambda^*(x)$, as desired.

Proving Cramer upper bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d}\{(\lambda, x) - \Lambda(\lambda)\}$.
- For simplicity, assume that Λ is defined for all x (which implies that X has moments of all orders and Λ and Λ^* are strictly convex, and the derivatives of Λ and Λ' are inverses of each other). It is also enough to consider the case X has mean zero, which implies that $\Lambda(0) = 0$ is a minimum of Λ, and $\Lambda^*(0) = 0$ is a minimum of Λ^*.
- We aim to show (up to subexponential error) that $\mu_n(\Gamma) \leq e^{-n \inf_{x \in \bar{\Gamma}} I(x)}$.
- If Γ were singleton set $\{x\}$ we could find the λ corresponding to x, so $\Lambda^*(x) = (x, \lambda) - \Lambda(\lambda)$. Note then that
$$\mathbb{E}e^{(n\lambda, A_n)} = \mathbb{E}e^{(\lambda, S_n)} = M_X^n(\lambda) = e^{n\Lambda(\lambda)},$$
and also $\mathbb{E}e^{(n\lambda, A_n)} \geq e^{n(\lambda, x)}\mu_n\{x\}$. Taking logs and dividing by n gives $\Lambda(\lambda) \geq \frac{1}{n}\log \mu_n + (\lambda, x)$, so that $\frac{1}{n}\log \mu_n(\Gamma) \leq -\Lambda^*(x)$, as desired.
- General Γ: cut into finitely many pieces, bound each piece?

Proving Cramer lower bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d}\{(\lambda, x) - \Lambda(\lambda)\}$.

Proving Cramer lower bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\}$.
- We aim to show that asymptotically $\mu_n(\Gamma) \geq e^{-n \inf_{x \in \Gamma^0} I(x)}$.

Proving Cramer lower bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{(\lambda, x) - \Lambda(\lambda)\}$.
- We aim to show that asymptotically $\mu_n(\Gamma) \geq e^{-n \inf_{x \in \Gamma^0} I(x)}$.
- It's enough to show that for each given $x \in \Gamma^0$, we have that asymptotically $\mu_n(\Gamma) \geq e^{-nI(x)}$.

Proving Cramer lower bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d}\{(\lambda, x) - \Lambda(\lambda)\}$.
- We aim to show that asymptotically $\mu_n(\Gamma) \geq e^{-n \inf_{x\in\Gamma^0} I(x)}$.
- It's enough to show that for each given $x \in \Gamma^0$, we have that asymptotically $\mu_n(\Gamma) \geq e^{-nI(x)}$.
- Idea is to weight law of each X_i by $e^{(\lambda,x)}$ to get a new measure whose expectation is in the interior of x. In this new measure, A_n is "typically" in Γ for large Γ, so the probability is of order 1.

Proving Cramer lower bound

- Recall that $I(x) = \Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d}\{(\lambda, x) - \Lambda(\lambda)\}$.
- We aim to show that asymptotically $\mu_n(\Gamma) \geq e^{-n \inf_{x \in \Gamma^0} I(x)}$.
- It's enough to show that for each given $x \in \Gamma^0$, we have that asymptotically $\mu_n(\Gamma) \geq e^{-nI(x)}$.
- Idea is to weight law of each X_i by $e^{(\lambda,x)}$ to get a new measure whose expectation is in the interior of x. In this new measure, A_n is "typically" in Γ for large Γ, so the probability is of order 1.
- But by how much did we have to modify the measure to make this typical? Aren't we weighting the law of A_n by about $e^{-nI(x)}$ near x?

Outline

Large deviations

Characteristic functions and central limit theorem

Outline

Large deviations

Characteristic functions and central limit theorem

Characteristic functions

- Let X be a random variable.

Characteristic functions

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

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- 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 $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic function ϕ_X similar to moment generating function M_X.

Characteristic functions

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic function ϕ_X similar to moment generating function M_X.
- $\phi_{X+Y} = \phi_X\phi_Y$, 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 $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic function ϕ_X similar to moment generating function M_X.
- $\phi_{X+Y} = \phi_X\phi_Y$, 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 $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic function ϕ_X similar to moment generating function M_X.
- $\phi_{X+Y} = \phi_X\phi_Y$, 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 $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- Recall that by definition $e^{it} = \cos(t) + i\sin(t)$.
- Characteristic function ϕ_X similar to moment generating function M_X.
- $\phi_{X+Y} = \phi_X \phi_Y$, 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 are well defined at all t for all random variables X.

Characteristic function properties

- $\phi(0) = 1$

Characteristic function properties

- $\phi(0) = 1$
- $\phi(-t) = \overline{\phi(t)}$

Characteristic function properties

- $\phi(0) = 1$
- $\phi(-t) = \overline{\phi(t)}$
- $|\phi(t)| = |Ee^{itX}| \leq E|e^{itX}| = 1.$

Characteristic function properties

- $\phi(0) = 1$
- $\phi(-t) = \overline{\phi(t)}$
- $|\phi(t)| = |Ee^{itX}| \leq E|e^{itX}| = 1.$
- $|\phi(t+h) - \phi(t)| \leq E|e^{ihX} - 1|$, so $\phi(t)$ uniformly continuous on $(-\infty, \infty)$

Characteristic function properties

- $\phi(0) = 1$
- $\phi(-t) = \overline{\phi(t)}$
- $|\phi(t)| = |Ee^{itX}| \leq E|e^{itX}| = 1.$
- $|\phi(t+h) - \phi(t)| \leq E|e^{ihX} - 1|$, so $\phi(t)$ uniformly continuous on $(-\infty, \infty)$
- $Ee^{it(aX+b)} = e^{itb}\phi(at)$

Characteristic function examples

- **Coin:** If $P(X = 1) = P(X = -1) = 1/2$ then $\phi_X(t) = (e^{it} + e^{-it})/2 = \cos t$.

Characteristic function examples

- **Coin:** If $P(X=1)=P(X=-1)=1/2$ then $\phi_X(t)=(e^{it}+e^{-it})/2=\cos t$.
- That's periodic. Do we always have periodicity if X is a random integer?

Characteristic function examples

- **Coin:** If $P(X = 1) = P(X = -1) = 1/2$ then $\phi_X(t) = (e^{it} + e^{-it})/2 = \cos t$.
- That's periodic. Do we always have periodicity if X is a random integer?
- **Poisson:** If X is Poisson with parameter λ then $\phi_X(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{itk}}{k!} = \exp(\lambda(e^{it} - 1))$.

Characteristic function examples

- **Coin:** If $P(X = 1) = P(X = -1) = 1/2$ then $\phi_X(t) = (e^{it} + e^{-it})/2 = \cos t$.
- That's periodic. Do we always have periodicity if X is a random integer?
- **Poisson:** If X is Poisson with parameter λ then $\phi_X(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{itk}}{k!} = \exp(\lambda(e^{it} - 1))$.
- Why does doubling λ amount to squaring ϕ_X?

Characteristic function examples

- **Coin:** If $P(X=1) = P(X=-1) = 1/2$ then $\phi_X(t) = (e^{it} + e^{-it})/2 = \cos t$.
- That's periodic. Do we always have periodicity if X is a random integer?
- **Poisson:** If X is Poisson with parameter λ then $\phi_X(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{itk}}{k!} = \exp(\lambda(e^{it} - 1))$.
- Why does doubling λ amount to squaring ϕ_X?
- **Normal:** If X is standard normal, then $\phi_X(t) = e^{-t^2/2}$.

Characteristic function examples

- **Coin:** If $P(X = 1) = P(X = -1) = 1/2$ then $\phi_X(t) = (e^{it} + e^{-it})/2 = \cos t$.
- That's periodic. Do we always have periodicity if X is a random integer?
- **Poisson:** If X is Poisson with parameter λ then $\phi_X(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{itk}}{k!} = \exp(\lambda(e^{it} - 1))$.
- Why does doubling λ amount to squaring ϕ_X?
- **Normal:** If X is standard normal, then $\phi_X(t) = e^{-t^2/2}$.
- Is ϕ_X always real when the law of X is symmetric about zero?

Characteristic function examples

- **Coin:** If $P(X = 1) = P(X = -1) = 1/2$ then $\phi_X(t) = (e^{it} + e^{-it})/2 = \cos t$.
- That's periodic. Do we always have periodicity if X is a random integer?
- **Poisson:** If X is Poisson with parameter λ then $\phi_X(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{itk}}{k!} = \exp(\lambda(e^{it} - 1))$.
- Why does doubling λ amount to squaring ϕ_X?
- **Normal:** If X is standard normal, then $\phi_X(t) = e^{-t^2/2}$.
- Is ϕ_X always real when the law of X is symmetric about zero?
- **Exponential:** If X is standard exponential (density e^{-x} on $(0, \infty)$) then $\phi_X(t) = 1/(1 - it)$.

Characteristic function examples

- **Coin:** If $P(X = 1) = P(X = -1) = 1/2$ then $\phi_X(t) = (e^{it} + e^{-it})/2 = \cos t$.
- That's periodic. Do we always have periodicity if X is a random integer?
- **Poisson:** If X is Poisson with parameter λ then $\phi_X(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{itk}}{k!} = \exp(\lambda(e^{it} - 1))$.
- Why does doubling λ amount to squaring ϕ_X?
- **Normal:** If X is standard normal, then $\phi_X(t) = e^{-t^2/2}$.
- Is ϕ_X always real when the law of X is symmetric about zero?
- **Exponential:** If X is standard exponential (density e^{-x} on $(0, \infty)$) then $\phi_X(t) = 1/(1 - it)$.
- **Bilateral exponential:** if $f_X(t) = e^{-|x|}/2$ on $\mathbb{R}$ then $\phi_X(t) = 1/(1 + t^2)$. Use linearity of $f_X \to \phi_X$.

Fourier inversion formula

- If $f:\mathbb{R}\to\mathbb{C}$ is in L^1, write $\hat{f}(t) := \int_{-\infty}^{\infty} f(x)e^{-itx}dx$.

Fourier inversion formula

- If $f : \mathbb{R} \to \mathbb{C}$ is in L^1, write $\hat{f}(t) := \int_{-\infty}^{\infty} f(x) e^{-itx} dx$.
- **Fourier inversion:** If f is nice: $f(x) = \frac{1}{2\pi} \int \hat{f}(t) e^{itx} dt$.

Fourier inversion formula

- If $f : \mathbb{R} \to \mathbb{C}$ is in L^1, write $\hat{f}(t) := \int_{-\infty}^{\infty} f(x) e^{-itx} dx$.
- **Fourier inversion:** If f is nice: $f(x) = \frac{1}{2\pi} \int \hat{f}(t) e^{itx} dt$.
- Easy to check this when f is density function of a Gaussian. Use linearity of $f \to \hat{f}$ to extend to linear combinations of Gaussians, or to convolutions with Gaussians.

Fourier inversion formula

- If $f : \mathbb{R} \to \mathbb{C}$ is in L^1, write $\hat{f}(t) := \int_{-\infty}^{\infty} f(x)e^{-itx}dx$.
- **Fourier inversion:** If f is nice: $f(x) = \frac{1}{2\pi}\int \hat{f}(t)e^{itx}dt$.
- Easy to check this when f is density function of a Gaussian. Use linearity of $f \to \hat{f}$ to extend to linear combinations of Gaussians, or to convolutions with Gaussians.
- Show $f \to \hat{f}$ is an isometry of Schwartz space (endowed with L^2 norm). Extend definition to L^2 completion.

Fourier inversion formula

- If $f : \mathbb{R} \to \mathbb{C}$ is in L^1, write $\hat{f}(t) := \int_{-\infty}^{\infty} f(x)e^{-itx}dx$.
- **Fourier inversion:** If f is nice: $f(x) = \frac{1}{2\pi}\int \hat{f}(t)e^{itx}dt$.
- Easy to check this when f is density function of a Gaussian. Use linearity of $f \to \hat{f}$ to extend to linear combinations of Gaussians, or to convolutions with Gaussians.
- Show $f \to \hat{f}$ is an isometry of Schwartz space (endowed with L^2 norm). Extend definition to L^2 completion.
- **Convolution theorem:** If

$$h(x) = (f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)dy,$$

then

$$\hat{h}(t) = \hat{f}(t)\hat{g}(t).$$

Fourier inversion formula

- If $f:\mathbb{R}\to\mathbb{C}$ is in L^1, write $\hat{f}(t) := \int_{-\infty}^{\infty} f(x)e^{-itx}dx$.
- **Fourier inversion:** If f is nice: $f(x) = \frac{1}{2\pi}\int \hat{f}(t)e^{itx}dt$.
- Easy to check this when f is density function of a Gaussian. Use linearity of $f \to \hat{f}$ to extend to linear combinations of Gaussians, or to convolutions with Gaussians.
- Show $f \to \hat{f}$ is an isometry of Schwartz space (endowed with L^2 norm). Extend definition to L^2 completion.
- **Convolution theorem:** If

$$h(x) = (f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)dy,$$

then

$$\hat{h}(t) = \hat{f}(t)\hat{g}(t).$$

- **Possible application?**

$$\int 1_{[a,b]}(x)f(x)dx = (\widehat{1_{[a,b]}f})(0) = (\hat{f} * \widehat{1_{[a,b]}})(0) = \int \hat{f}(t)\widehat{1_{[a,b]}}(-t)dx.$$

Characteristic function inversion formula

- If the map $\mu_X \to \phi_X$ is linear, is the map $\phi \to \mu[a, b]$ (for some fixed $[a, b]$) a linear map? How do we recover $\mu[a, b]$ from ϕ?

Characteristic function inversion formula

- If the map $\mu_X \to \phi_X$ is linear, is the map $\phi \to \mu[a,b]$ (for some fixed $[a,b]$) a linear map? How do we recover $\mu[a,b]$ from ϕ?
- Say $\phi(t) = \int e^{itx}\mu(x)$.

Characteristic function inversion formula

- If the map $\mu_X \to \phi_X$ is linear, is the map $\phi \to \mu[a,b]$ (for some fixed $[a,b]$) a linear map? How do we recover $\mu[a,b]$ from ϕ?
- Say $\phi(t) = \int e^{itx}\mu(x)$.
- **Inversion theorem:**

$$\lim_{T\to\infty}(2\pi)^{-1}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt = \mu(a,b) + \frac{1}{2}\mu(\{a,b\})$$

Characteristic function inversion formula

- If the map $\mu_X \to \phi_X$ is linear, is the map $\phi \to \mu[a,b]$ (for some fixed $[a,b]$) a linear map? How do we recover $\mu[a,b]$ from ϕ?
- Say $\phi(t) = \int e^{itx}\mu(x)$.
- **Inversion theorem:**

$$\lim_{T\to\infty}(2\pi)^{-1}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt = \mu(a,b)+\frac{1}{2}\mu(\{a,b\})$$

- **Main ideas of proof:** Write

$$I_T = \int \frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt = \int_{-T}^{T}\int \frac{e^{-ita}-e^{-itb}}{it}e^{itx}\mu(x)dt.$$

Characteristic function inversion formula

- If the map $\mu_X \to \phi_X$ is linear, is the map $\phi \to \mu[a,b]$ (for some fixed $[a,b]$) a linear map? How do we recover $\mu[a,b]$ from ϕ?
- Say $\phi(t) = \int e^{itx}\mu(x)$.
- **Inversion theorem:**

$$\lim_{T\to\infty}(2\pi)^{-1}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt = \mu(a,b)+\frac{1}{2}\mu(\{a,b\})$$

- **Main ideas of proof:** Write

$$I_T = \int \frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt = \int_{-T}^{T}\int \frac{e^{-ita}-e^{-itb}}{it}e^{itx}\mu(x)dt.$$

- Observe that $\frac{e^{-ita}-e^{-itb}}{it} = \int_a^b e^{-ity}dy$ has modulus bounded by $b-a$.

Characteristic function inversion formula

- If the map $\mu_X \to \phi_X$ is linear, is the map $\phi \to \mu[a,b]$ (for some fixed $[a,b]$) a linear map? How do we recover $\mu[a,b]$ from ϕ?
- Say $\phi(t) = \int e^{itx}\mu(x)$.
- **Inversion theorem:**

$$\lim_{T\to\infty}(2\pi)^{-1}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt = \mu(a,b) + \frac{1}{2}\mu(\{a,b\})$$

- **Main ideas of proof:** Write

$$I_T = \int \frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt = \int_{-T}^{T}\int \frac{e^{-ita}-e^{-itb}}{it}e^{itx}\mu(x)dt.$$

- Observe that $\frac{e^{-ita}-e^{-itb}}{it} = \int_a^b e^{-ity}dy$ has modulus bounded by $b-a$.
- That means we can use Fubini to compute I_T.

Bochner's theorem

- Given any function ϕ and any points $t_1, \ldots, t_n$, we can consider the matrix with i, j entry given by $\phi(t_i - t_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.

Bochner's theorem

- Given any function ϕ and any points $t_1, \ldots, t_n$, we can consider the matrix with i, j entry given by $\phi(t_i - t_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{C}$ with $\phi(0) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.

Bochner's theorem

- Given any function ϕ and any points $t_1, \ldots, t_n$, we can consider the matrix with i, j entry given by $\phi(t_i - t_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{C}$ with $\phi(0) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.
- Why positive definite?

Bochner's theorem

- Given any function ϕ and any points $t_1, \ldots, t_n$, we can consider the matrix with i, j entry given by $\phi(t_i - t_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{C}$ with $\phi(0) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.
- Why positive definite?
- Write $Y = \sum_{j=1}^{n} a_j e^{t_j X}$. This is a complex-valued random variable. What is $\mathbb{E}|Y|^2$?

Bochner's theorem

- Given any function ϕ and any points $t_1, \ldots, t_n$, we can consider the matrix with i, j entry given by $\phi(t_i - t_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{C}$ with $\phi(0) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.
- Why positive definite?
- Write $Y = \sum_{j=1}^n a_j e^{t_j X}$. This is a complex-valued random variable. What is $\mathbb{E}|Y|^2$?
- $Y\overline{Y} = \sum_{j=1}^n \sum_{k=1}^n a_j \overline{a_k} e^{(t_i - t_j)X}$ and $\mathbb{E}Y\overline{Y} = \sum_{j=1}^n \sum_{k=1}^n a_j \overline{a_k} \phi(t_i - t_j)$.

Bochner's theorem

- Given any function ϕ and any points $t_1, \ldots, t_n$, we can consider the matrix with i, j entry given by $\phi(t_i - t_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{C}$ with $\phi(0) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.
- Why positive definite?
- Write $Y = \sum_{j=1}^n a_j e^{t_j X}$. This is a complex-valued random variable. What is $\mathbb{E}|Y|^2$?
- $Y\overline{Y} = \sum_{j=1}^n \sum_{k=1}^n a_j \overline{a_k} e^{(t_i - t_j)X}$ and $\mathbb{E}Y\overline{Y} = \sum_{j=1}^n \sum_{k=1}^n a_j \overline{a_k} \phi(t_i - t_j)$.
- Set of possible characteristic functions is a pretty nice set.

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.

- **Slightly stronger theorem:** If $\mu_n \implies \mu_\infty$ then $\phi_n(t) \to \phi_\infty(t)$ for all t. Conversely, if $\phi_n(t)$ converges to a limit that is continuous at 0, then the associated sequence of distributions μ_n is tight and converges weakly to measure μ with characteristic function ϕ.

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.

- **Slightly stronger theorem:** If $\mu_n \implies \mu_\infty$ then $\phi_n(t) \to \phi_\infty(t)$ for all t. Conversely, if $\phi_n(t)$ converges to a limit that is continuous at 0, then the associated sequence of distributions μ_n is tight and converges weakly to measure μ with characteristic function ϕ.
- **Proof ideas:** First statement easy (since $X_n \implies X$ implies $Eg(X_n) \to Eg(X)$ for any bounded continuous g). For second statement, try to use fact that $u^{-1}\int_{-u}^{u}(1-\phi(t))dt \to 0$ to get tightness of the μ_n. Then note that any subsequential limit of the μ_n must be equal to μ. Use this to argue that $\int fd\mu_n$ converges to $\int fd\mu$ for every bounded continuous f.

Moments, derivatives, CLT

- If $\int |x|^n \mu(x) < \infty$ then the characteristic function ϕ of μ has a continuous derivative of order n given by $\phi^{(n)}(t) = \int (ix)^n e^{itx} \mu(dx)$.

Moments, derivatives, CLT

- If $\int |x|^n \mu(x) < \infty$ then the characteristic function ϕ of μ has a continuous derivative of order n given by $\phi^{(n)}(t) = \int (ix)^n e^{itx} \mu(dx)$.
- Indeed, if $E|X|^2 < \infty$ and $EX = 0$ then $\phi(t) = 1 - t^2 E(X^2)/2o(t^2)$.

Moments, derivatives, CLT

- If $\int |x|^n \mu(x) < \infty$ then the characteristic function ϕ of μ has a continuous derivative of order n given by $\phi^{(n)}(t) = \int (ix)^n e^{itx} \mu(dx)$.
- Indeed, if $E|X|^2 < \infty$ and $EX = 0$ then $\phi(t) = 1 - t^2 E(X^2)/2o(t^2)$.
- This and the continuity theorem together imply the central limit theorem.

Moments, derivatives, CLT

- If $\int |x|^n \mu(x) < \infty$ then the characteristic function ϕ of μ has a continuous derivative of order n given by $\phi^{(n)}(t) = \int (ix)^n e^{itx} \mu(dx)$.
- Indeed, if $E|X|^2 < \infty$ and $EX = 0$ then $\phi(t) = 1 - t^2 E(X^2)/2o(t^2)$.
- This and the continuity theorem together imply the central limit theorem.
- **Theorem:** Let $X_1, X_2, \ldots$ by i.i.d. with $EX_i = \mu$, $\mathrm{Var}(X_i) = \sigma^2 \in (0, \infty)$. If $S_n = X_1 + \ldots + X_n$ then $(S_n - n\mu)/(\sigma n^{1/2})$ converges in law to a standard normal.