18.175: Lecture 16

Central limit theorem variants

Scott Sheffield

MIT

Outline

CLT idea

CLT variants

Outline

CLT idea

CLT variants

Recall Fourier inversion formula

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

Recall 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: $f(x) = \frac{1}{2\pi} \int \hat{f}(t) e^{itx} dt$.

Recall 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.

Recall 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.

Recall 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

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

Recall 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

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

- **Observation:** can define Fourier transforms of generalized functions. Can interpret finite measure as generalized function.

Recall Bochner's theorem

- Given any function ϕ and any points $x_1, \ldots, x_n$, we can consider the matrix with i, j entry given by $\phi(x_i - x_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.

Recall Bochner's theorem

- Given any function ϕ and any points $x_1, \ldots, x_n$, we can consider the matrix with i, j entry given by $\phi(x_i - x_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{R}$ with $\phi(1) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.

Recall Bochner's theorem

- Given any function ϕ and any points $x_1, \ldots, x_n$, we can consider the matrix with i, j entry given by $\phi(x_i - x_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{R}$ with $\phi(1) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.
- Positive definiteness kind of comes from fact that variances of random variables are non-negative.

Recall Bochner's theorem

- Given any function ϕ and any points $x_1, \ldots, x_n$, we can consider the matrix with i, j entry given by $\phi(x_i - x_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{R}$ with $\phi(1) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.
- Positive definiteness kind of comes from fact that variances of random variables are non-negative.
- The set of all possible characteristic functions is a pretty nice set.

Recall Bochner's theorem

- Given any function ϕ and any points $x_1, \ldots, x_n$, we can consider the matrix with i, j entry given by $\phi(x_i - x_j)$. Call ϕ **positive definite** if this matrix is always positive semidefinite Hermitian.
- Bochner's theorem: a continuous function from $\mathbb{R}$ to $\mathbb{R}$ with $\phi(1) = 1$ is a characteristic function of a some probability measure on $\mathbb{R}$ if and only if it is positive definite.
- Positive definiteness kind of comes from fact that variances of random variables are non-negative.
- The set of all possible characteristic functions is a pretty nice set.
- The Fourier transform is a natural map from set of all probability measures on $\mathbb{R}$ (which can be described by their distribution functions F) to the set of possible characteristic functions.

Recall continuity theorem

- **Strong continuity 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 a measure μ with characteristic function ϕ.

Recall CLT idea

- Let X be a random variable.

Recall CLT idea

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

Recall CLT idea

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- And if X has an mth moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.

Recall CLT idea

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- And if X has an mth moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.
- In particular, if $E[X] = 0$ and $E[X^2] = 1$ then $\phi_X(0) = 1$ and $\phi'_X(0) = 0$ and $\phi''_X(0) = -1$.

Recall CLT idea

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- And if X has an mth moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.
- In particular, if $E[X] = 0$ and $E[X^2] = 1$ then $\phi_X(0) = 1$ and $\phi'_X(0) = 0$ and $\phi''_X(0) = -1$.
- Write $L_X := -\log \phi_X$. Then $L_X(0) = 0$ and $L'_X(0) = -\phi'_X(0)/\phi_X(0) = 0$ and $L''_X = -(\phi''_X(0)\phi_X(0) - \phi'_X(0)^2)/ \phi_X(0)^2 = 1$.

Recall CLT idea

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- And if X has an mth moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.
- In particular, if $E[X] = 0$ and $E[X^2] = 1$ then $\phi_X(0) = 1$ and $\phi_X'(0) = 0$ and $\phi_X''(0) = -1$.
- Write $L_X := -\log \phi_X$. Then $L_X(0) = 0$ and $L_X'(0) = -\phi_X'(0)/\phi_X(0) = 0$ and $L_X'' = -(\phi_X''(0)\phi_X(0) - \phi_X'(0)^2)/\ \phi_X(0)^2 = 1$.
- If $V_n = n^{-1/2} \sum_{i=1}^n X_i$ where X_i are i.i.d. with law of X, then $L_{V_n}(t) = nL_X(n^{-1/2}t)$.

Recall CLT idea

- Let X be a random variable.
- The **characteristic function** of X is defined by $\phi(t) = \phi_X(t) := E[e^{itX}]$.
- And if X has an mth moment then $E[X^m] = i^m \phi_X^{(m)}(0)$.
- In particular, if $E[X] = 0$ and $E[X^2] = 1$ then $\phi_X(0) = 1$ and $\phi_X'(0) = 0$ and $\phi_X''(0) = -1$.
- Write $L_X := -\log \phi_X$. Then $L_X(0) = 0$ and $L_X'(0) = -\phi_X'(0)/\phi_X(0) = 0$ and $L_X'' = -(\phi_X''(0)\phi_X(0) - \phi_X'(0)^2)/\ \phi_X(0)^2 = 1$.
- If $V_n = n^{-1/2} \sum_{i=1}^n X_i$ where X_i are i.i.d. with law of X, then $L_{V_n}(t) = nL_X(n^{-1/2}t)$.
- When we zoom in on a twice differentiable function near zero (scaling vertically by n and horizontally by $\sqrt{n}$) the picture looks increasingly like a parabola.

Outline

CLT idea

CLT variants

Outline

CLT idea

CLT variants

Lindeberg-Feller theorem

- CLT is pretty special. What other kinds of sums are approximately Gaussian?

Lindeberg-Feller theorem

- CLT is pretty special. What other kinds of sums are approximately Gaussian?
- **Triangular arrays:** Suppose $X_{n,m}$ are independent expectation-zero random variables when $1 \leq m \leq n$.

Lindeberg-Feller theorem

- CLT is pretty special. What other kinds of sums are approximately Gaussian?
- **Triangular arrays:** Suppose $X_{n,m}$ are independent expectation-zero random variables when $1 \leq m \leq n$.
- Suppose $\sum_{m=1}^{n} EX_{n,m}^2 \to \sigma^2 > 0$ and for all ϵ, $\lim_{n\to\infty} E(|X_{n,m}|^2; |X_{n,m}| > \epsilon) = 0$.

Lindeberg-Feller theorem

- CLT is pretty special. What other kinds of sums are approximately Gaussian?
- **Triangular arrays:** Suppose $X_{n,m}$ are independent expectation-zero random variables when $1 \leq m \leq n$.
- Suppose $\sum_{m=1}^{n} EX_{n,m}^2 \to \sigma^2 > 0$ and for all ϵ, $\lim_{n\to\infty} E(|X_{n,m}|^2; |X_{n,m}| > \epsilon) = 0$.
- Then $S_n = X_{n,1} + X_{n,2} + \ldots + X_{n,n} \implies \sigma\chi$ (where χ is standard normal) as $n \to \infty$.

Lindeberg-Feller theorem

- CLT is pretty special. What other kinds of sums are approximately Gaussian?
- **Triangular arrays:** Suppose $X_{n,m}$ are independent expectation-zero random variables when $1 \leq m \leq n$.
- Suppose $\sum_{m=1}^n EX_{n,m}^2 \to \sigma^2 > 0$ and for all ϵ, $\lim_{n\to\infty} E(|X_{n,m}|^2; |X_{n,m}| > \epsilon) = 0$.
- Then $S_n = X_{n,1} + X_{n,2} + \ldots + X_{n,n} \implies \sigma\chi$ (where χ is standard normal) as $n \to \infty$.
- **Proof idea:** Use characteristic functions $\phi_{n,m} = \phi_{X_{n,m}}$. Try to get some uniform handle on how close they are to their quadratic approximations.

Berry-Esseen theorem

- If X_i are i.i.d. with mean zero, variance σ^2, and $E|X_i|^3 = \rho < \infty$, and $F_n(x)$ is distribution of $(X_1 + \ldots + X_n)/(\sigma\sqrt{n})$ and $\Phi(x)$ is standard normal distribution, then $|F_n(x) - \Phi(x)| \leq 3\rho/(\sigma^3\sqrt{n})$.

Berry-Esseen theorem

- If X_i are i.i.d. with mean zero, variance σ^2, and $E|X_i|^3 = \rho < \infty$, and $F_n(x)$ is distribution of $(X_1 + \ldots + X_n)/(\sigma\sqrt{n})$ and $\Phi(x)$ is standard normal distribution, then $|F_n(x) - \Phi(x)| \leq 3\rho/(\sigma^3\sqrt{n})$.
- Provided one has a third moment, CLT convergence is very quick.

Berry-Esseen theorem

- If X_i are i.i.d. with mean zero, variance σ^2, and $E|X_i|^3 = \rho < \infty$, and $F_n(x)$ is distribution of $(X_1 + \ldots + X_n)/(\sigma\sqrt{n})$ and $\Phi(x)$ is standard normal distribution, then $|F_n(x) - \Phi(x)| \leq 3\rho/(\sigma^3\sqrt{n})$.
- Provided one has a third moment, CLT convergence is very quick.
- **Proof idea:** You can convolve with something that has a characteristic function with compact support. Play around with Fubini, error estimates.

Local limit theorems for walks on $\mathbb{Z}$

- Suppose $X \in b + h\mathbb{Z}$ a.s. for some fixed constants b and h.

Local limit theorems for walks on $\mathbb{Z}$

- Suppose $X \in b + h\mathbb{Z}$ a.s. for some fixed constants b and h.
- Observe that if $\phi_X(\lambda) = 1$ for some $\lambda \neq 0$ then X is supported on (some translation of) $(2\pi/\lambda)\mathbb{Z}$. If this holds for all λ, then X is a.s. some constant. When the former holds but not the latter (i.e., ϕ_X is periodic but not identically 1) we call X a **lattice random variable**.

Local limit theorems for walks on $\mathbb{Z}$

- Suppose $X \in b + h\mathbb{Z}$ a.s. for some fixed constants b and h.
- Observe that if $\phi_X(\lambda) = 1$ for some $\lambda \neq 0$ then X is supported on (some translation of) $(2\pi/\lambda)\mathbb{Z}$. If this holds for all λ, then X is a.s. some constant. When the former holds but not the latter (i.e., ϕ_X is periodic but not identically 1) we call X a **lattice random variable**.
- Write $p_n(x) = P(S_n/\sqrt{n} = x)$ for $x \in \mathcal{L}_n := (nb + h\mathbb{Z})/\sqrt{n}$ and $n(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)$.

Local limit theorems for walks on $\mathbb{Z}$

- Suppose $X \in b + h\mathbb{Z}$ a.s. for some fixed constants b and h.
- Observe that if $\phi_X(\lambda) = 1$ for some $\lambda \neq 0$ then X is supported on (some translation of) $(2\pi/\lambda)\mathbb{Z}$. If this holds for all λ, then X is a.s. some constant. When the former holds but not the latter (i.e., ϕ_X is periodic but not identically 1) we call X a **lattice random variable**.
- Write $p_n(x) = P(S_n/\sqrt{n} = x)$ for $x \in \mathcal{L}_n := (nb + h\mathbb{Z})/\sqrt{n}$ and $n(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)$.
- Assume X_i are i.i.d. lattice with $EX_i = 0$ and $EX_i^2 = \sigma^2 \in (0, \infty)$. **Theorem:** As $n \to \infty$,

$$\Big| \sup_{x \in \mathcal{L}^n} \left| n^{1/2}/hp_n(x) - n(x) \right| \to 0.$$

Local limit theorems for walks on $\mathbb{Z}$

- Suppose $X \in b + h\mathbb{Z}$ a.s. for some fixed constants b and h.
- Observe that if $\phi_X(\lambda) = 1$ for some $\lambda \neq 0$ then X is supported on (some translation of) $(2\pi/\lambda)\mathbb{Z}$. If this holds for all λ, then X is a.s. some constant. When the former holds but not the latter (i.e., ϕ_X is periodic but not identically 1) we call X a **lattice random variable**.
- Write $p_n(x) = P(S_n/\sqrt{n} = x)$ for $x \in \mathcal{L}_n := (nb + h\mathbb{Z})/\sqrt{n}$ and $n(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)$.
- Assume X_i are i.i.d. lattice with $EX_i = 0$ and $EX_i^2 = \sigma^2 \in (0, \infty)$. **Theorem:** As $n \to \infty$,

$$\Big| \sup_{x \in \mathcal{L}^n} \left| n^{1/2}/hp_n(x) - n(x) \right| \to 0.$$

- **Proof idea:** Use characteristic functions, reduce to periodic integral problem. Note that for Y supported on $a + \theta\mathbb{Z}$, we have $P(Y = x) = \frac{1}{2\pi/\theta} \int_{-\pi/\theta}^{\pi/\theta} e^{-itx} \phi_Y(t) dt$.