Advanced Probability Theory (Math541)

Instructor: Kani Chen

(Classic)/Modern Probability Theory (1900-1960)

Primitive/Classic Probability:

(16th-19th century)

- Gerolamo Cardano (1501-1576)
  “Liber de Ludo Aleae” (games of chance)

Primitive/Classic Probability:

(16th-19th century)

- Gerolamo Cardano (1501-1576)
  "Liber de Ludo Aleae" (games of chance)
- Galilei Galileo (1564-1642)

Primitive/Classic Probability:

(16th-19th century)

- Gerolamo Cardano (1501-1576)
  “Liber de Ludo Aleae” (games of chance)
- Galilei Galileo (1564-1642)
- Pierre de Fermat (1601-1665) and Blaise Pascal (1623-1662)

Primitive/Classic Probability:

(16th-19th century)

- Gerolamo Cardano (1501-1576)
  "Liber de Ludo Aleae" (games of chance)
- Galilei Galileo (1564-1642)
- Pierre de Fermat (1601-1665) and Blaise Pascal (1623-1662)
- Jakob Bernoulli (1654 -1705) *Bernoulli trial/distribution/r.v/numbers*
  Daniel (1700-1782) (*utility function*)
  Johann (1667-1748) (*L'Hopital rule*)

- Abraham de Moivre (1667-1754)
  "Doctrine of Chances"

- Abraham de Moivre (1667-1754)
  "Doctrine of Chances"
- Pierre-Simon Laplace (1749-1827)
  Laplace transform
  De Moivre-Laplace Theorem (the central limit theorem)

- Abraham de Moivre (1667-1754)
  "Doctrine of Chances"
- Pierre-Simon Laplace (1749-1827)
  Laplace transform
  De Moivre-Laplace Theorem (the central limit theorem)
- Simeon Denis Poisson (1781-1840).
  Poisson distribution/process, the law of rare events

- Abraham de Moivre (1667-1754)
  “Doctrine of Chances”
- Pierre-Simon Laplace (1749-1827)
  Laplace transform
  De Moivre-Laplace Theorem (the central limit theorem)
- Simeon Denis Poisson (1781-1840).
  Poisson distribution/process, the law of rare events
- Emile Borel (1871-1956).
  Borel sets/measurable, Borel-Cantelli lemma, Borel strong law.

Foundation of modern probability:

- Keynes, J. M. (1921): *A Treatise on Probability.*

Foundation of modern probability:

- Keynes, J. M. (1921): *A Treatise on Probability.*
- von Mises, R. (1928): *Probability, Statistics and Truth.*
  (1931): *Mathematical Theory of Probability and Statistics.*

Foundation of modern probability:

- Keynes, J. M. (1921): *A Treatise on Probability.*
- von Mises, R. (1928): *Probability, Statistics and Truth.* (1931): *Mathematical Theory of Probability and Statistics.*
- Kolmogorov, A. (1933): *Foundations of the Theory of Probability.* Kolmogorov's axioms: Probability space trio: $(\Omega, \mathcal{F}, P)$.

Measure-theoretic Probabilities

(Chapter 1 begins)

- Sets, set operations ($\cap$, $\cup$ and complement), set of sets/subsets, algebra, σ-algebra, measurable space $(\Omega, \mathcal{F})$, product space ...

Measure-theoretic Probabilities

(Chapter 1 begins)

- Sets, set operations ($\cap$, $\cup$ and complement), set of sets/subsets, algebra, σ-algebra, measurable space $(\Omega, \mathcal{F})$, product space ...
- measures, probabilities, product measure, independence of events, conditional probabilities, conditional expectation with respect to σ-algebra.

Measure-theoretic Probabilities

(Chapter 1 begins)

- Sets, set operations ($\cap$, $\cup$ and complement), set of sets/subsets, algebra, σ-algebra, measurable space $(\Omega, \mathcal{F})$, product space ...
- measures, probabilities, product measure, independence of events, conditional probabilities, conditional expectation with respect to σ-algebra.
- Transformation/map, measurable transformation/map, random variables/elements defined through mapping $X = X(\omega)$.

Measure-theoretic Probabilities

(Chapter 1 begins)

- Sets, set operations ($\cap$, $\cup$ and complement), set of sets/subsets, algebra, σ-algebra, measurable space $(\Omega, \mathcal{F})$, product space ...
- measures, probabilities, product measure, independence of events, conditional probabilities, conditional expectation with respect to σ-algebra.
- Transformation/map, measurable transformation/map, random variables/elements defined through mapping $X = X(\omega)$.
- Expectation/integral.

Measure-theoretic Probabilities

(Chapter 1 begins)

- Sets, set operations ($\cap$, $\cup$ and complement), set of sets/subsets, algebra, σ-algebra, measurable space $(\Omega, \mathcal{F})$, product space ...
- measures, probabilities, product measure, independence of events, conditional probabilities, conditional expectation with respect to σ-algebra.
- Transformation/map, measurable transformation/map, random variables/elements defined through mapping $X = X(\omega)$.
- Expectation/integral.
- Caratheodory's extension theorem, Kolmogorov's extension theorem, Dynkin's $\pi - \lambda$ theorem, The Radon-Nikodym Theorem.

Convergence.

- Convergence modes:
  convergence almost sure (strong),
  in probability,
  in L^p (most commonly, L^1 or L^2),
  in distribution/law (weak).
  The relations of the convergence modes.

Convergence.

- Convergence modes:
  convergence almost sure (strong),
  in probability,
  in L^p (most commonly, L^1 or L^2),
  in distribution/law (weak).
  The relations of the convergence modes.
- Dominated convergence theorem,
  (extension to uniformly integrable r.v.s.)
  monotone convergence theorem.
  Fatou's lemma.

Law of Large numbers.

$X_1, ..., X_n, ...$ are iid random variables with mean μ. Let $S_n = \sum_{i=1}^n X_i$.

- Weak law of Large numbers:

$$S_n/n \to \mu, \qquad \text{in probability.}$$

i.e., for any $\epsilon > 0$,

$$P(|S_n/n - \mu| > \epsilon) \to 0.$$

Law of Large numbers.

$X_1, ..., X_n, ...$ are iid random variables with mean μ. Let $S_n = \sum_{i=1}^n X_i$.

- Weak law of Large numbers:

$$S_n/n \to \mu, \qquad \text{in probability.}$$

  i.e., for any $\epsilon > 0$,

$$P(|S_n/n - \mu| > \epsilon) \to 0.$$

- Kolmogorov's Strong law of Large numbers:

$$S_n/n \to \mu, \qquad \text{almost surely.}$$

Law of Large numbers.

$X_1, ..., X_n, ...$ are iid random variables with mean μ. Let $S_n = \sum_{i=1}^n X_i$.

- Weak law of Large numbers:

$$S_n/n \to \mu, \qquad \text{in probability.}$$

  i.e., for any $\epsilon > 0$,

$$P(|S_n/n - \mu| > \epsilon) \to 0.$$

- Kolmogorov's Strong law of Large numbers:

$$S_n/n \to \mu, \qquad \text{almost surely.}$$

- Consistency in statistical estimation.

Large deviation (strengthening weak law)

For fixed $t > 0$, how fast is $P(S_n/n - \mu > t) \to 0$?

- Under regularity conditions,

$$\frac{1}{n} \log P(S_n/n - \mu > t) \approx \gamma(t),$$

where γ is determined by the commmon distribution of X_i.

Large deviation (strengthening weak law)

For fixed $t > 0$, how fast is $P(S_n/n - \mu > t) \to 0$?

- Under regularity conditions,

$$\frac{1}{n} \log P(S_n/n - \mu > t) \approx \gamma(t),$$

  where γ is determined by the commmon distribution of X_i.
- Special case, X_i are iid $N(0,1)$, then $\gamma(t) = -t^2/2$.
  Note that $1 - \Phi(x) \approx \phi(x)/x$.

Law of iterated logarithm (strengthening strong law)

$S_n/n - \mu \to 0$ a.s.
Is there a proper a_n such that the "limit" of S_n/a_n is nonzero finite?

- Kolmogorov's law of iterated logarithm.

$$\limsup_{n\to\infty} \frac{S_n/n - \mu}{\sqrt{2\sigma^2 n \log\log(n)}} \to 1, \quad \text{a.s.}$$

where $\sigma^2 = var(X_i)$.

Law of iterated logarithm (strengthening strong law)

$S_n/n - \mu \to 0$ a.s.
Is there a proper a_n such that the "limit" of S_n/a_n is nonzero finite?

- Kolmogorov's law of iterated logarithm.

$$\limsup_{n\to\infty} \frac{S_n/n - \mu}{\sqrt{2\sigma^2 n \log\log(n)}} \to 1, \quad \text{a.s.}$$

  where $\sigma^2 = var(X_i)$.

- Marcinkiewicz-Zygmund strong law: for $0 < p < 2$,

$$\frac{S_n - nc}{n^{1/p}} \to 0, \quad \text{a.s.},$$

  if and only if $E(|X_i|^p) < \infty$, where $c = \mu$ for $1 \le p < 2$.

Convergence of Series.

The convergence of S_n for independent $X_1, ..., X_n$.

- Khintchine's convergence theorem: if $E(X_i) = 0$ and $\sum_n var(X_n) < \infty$, then S_n converges a.s. as well as in L^2.

Convergence of Series.

The convergence of S_n for independent $X_1, ..., X_n$.

- Khintchine's convergence theorem: if $E(X_i) = 0$ and $\sum_n var(X_n) < \infty$, then S_n converges a.s. as well as in L^2.
- Kolmogorov's three series theorem:
  S_n converges a.s. if and only if $\sum_n P(|X_n| > 1) < \infty$,
  $\sum_n E(X_n 1_{\{|X_n| \leq 1\}}) < \infty$, and $\sum_n var(X_n 1_{\{|X_n| \leq 1\}}) < \infty$.

Convergence of Series.

The convergence of S_n for independent $X_1, ..., X_n$.

- Khintchine's convergence theorem: if $E(X_i) = 0$ and $\sum_n var(X_n) < \infty$, then S_n converges a.s. as well as in L^2.
- Kolmogorov's three series theorem:
  S_n converges a.s. if and only if $\sum_n P(|X_n| > 1) < \infty$,
  $\sum_n E(X_n 1_{\{|X_n| \le 1\}}) < \infty$, and $\sum_n var(X_n 1_{\{|X_n| \le 1\}}) < \infty$.
- Kronecker Lemma:
  If $0 < b_n \uparrow \infty$ and $\sum_n (a_n/b_n) < \infty$ then $\sum_{j=1}^n a_j/b_n \to 0$.
  A technique to show law of large numbers via convergence of series.

The central limit theorem (Chapter 2).

De Moivre-Lapalace Theorem

For $X_1, X_2 ...$ independent, under proper conditions,

$$\frac{S_n - E(S_n)}{\sqrt{\text{var}(S_n)}} \to N(0,1) \quad \text{in distribution.}$$

$$i.e., \qquad P\Big(\frac{S_n - E(S_n)}{\sqrt{\text{var}(S_n)}} < x\Big) \to P(N(0,1) < x)$$

for all x.

The conditions and extensions

- Lyapunov.

The conditions and extensions

- Lyapunov.
- Lindeberg (1922) condition: X_j is mean 0 with variance σ_j^2, such that

$$\sum_{j=1}^{n} E(X_j^2 1_{\{|X_j|>\epsilon s_n\}}) = o(s_n^2)$$

for all $\epsilon > 0$, where $s_n^2 = \sum_{j=1}^{n} \sigma_j^2$.

The conditions and extensions

- Lyapunov.
- Lindeberg (1922) condition: X_j is mean 0 with variance σ_j^2, such that
$$\sum_{j=1}^{n} E(X_j^2 1_{\{|X_j|>\epsilon s_n\}}) = o(s_n^2)$$
for all $\epsilon > 0$, where $s_n^2 = \sum_{j=1}^{n} \sigma_j^2$.
- Feller (1935): If CLT holds, $\sigma_n/s_n \to 0$ and $s_n \to \infty$, then the above Lindeberg condition holds.

The conditions and extensions

- Lyapunov.
- Lindeberg (1922) condition: X_j is mean 0 with variance σ_j^2, such that

$$\sum_{j=1}^{n} E(X_j^2 1_{\{|X_j|>\epsilon s_n\}}) = o(s_n^2)$$

  for all $\epsilon > 0$, where $s_n^2 = \sum_{j=1}^n \sigma_j^2$.
- Feller (1935): If CLT holds, $\sigma_n/s_n \to 0$ and $s_n \to \infty$, then the above Lindeberg condition holds.
- Extensions to martingales, Markov process ...

The conditions and extensions

- Lyapunov.
- Lindeberg (1922) condition: X_j is mean 0 with variance σ_j^2, such that
$$\sum_{j=1}^{n} E(X_j^2 1_{\{|X_j|>\epsilon s_n\}}) = o(s_n^2)$$
for all $\epsilon > 0$, where $s_n^2 = \sum_{j=1}^{n} \sigma_j^2$.
- Feller (1935): If CLT holds, $\sigma_n/s_n \to 0$ and $s_n \to \infty$, then the above Lindeberg condition holds.
- Extensions to martingales, Markov process ...
- Inference in statistics (accuracy justification of estimation: confidence interval, test of hypothesis.)

Rate of convergence to normality

Suppose $X_1, ..., X_n$ are iid.

- Berry-Esseen Bound:

$$\left| P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} < x \right) - P(N(0,1) < x) \right| \leq \frac{c\gamma/\sigma^3}{n^{1/2}}$$

for all x, where $\gamma = E(|X_i|^3)$ and c is a *universal* constant.

Rate of convergence to normality

Suppose $X_1, ..., X_n$ are iid.

- Berry-Esseen Bound:

$$\left| P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} < x \right) - P(N(0,1) < x) \right| \le \frac{c\gamma/\sigma^3}{n^{1/2}}$$

  for all x, where $\gamma = E(|X_i|^3)$ and c is a *universal* constant.
- Extensions, e.g., to non iid cases, etc..

Rate of convergence to normality

Suppose $X_1, ..., X_n$ are iid.

- Berry-Esseen Bound:

$$\left| P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} < x \right) - P(N(0,1) < x) \right| \leq \frac{c\gamma/\sigma^3}{n^{1/2}}$$

  for all x, where $\gamma = E(|X_i|^3)$ and c is a *universal* constant.
- Extensions, e.g., to non iid cases, etc..
- Edgeworth expansion.

Random Walk (Chapter 3)

Given $X_1, X_2, ...$ iid, we study the behavior $\{S_1, S_2, ...\}$ as a sequence of random variables.

- Stopping times.
  T is an integer-valued r.v. such that $T = n$ only depends on the values of $X_1, ..., X_n$, on, in other words, the values $S_1, ..., S_n$.

Random Walk (Chapter 3)

Given $X_1, X_2, ...$ iid, we study the behavior $\{S_1, S_2, ...\}$ as a sequence of random variables.

- Stopping times.
  T is an integer-valued r.v. such that $T = n$ only depends on the values of $X_1, ..., X_n$, on, in other words, the values $S_1, ..., S_n$.
- Wald's identity/equation:
  Under proper conditions, $E(S_T) = \mu E(T)$.

Random Walk (Chapter 3)

Given $X_1, X_2, \ldots$ iid, we study the behavior $\{S_1, S_2, \ldots\}$ as a sequence of random variables.

- Stopping times.
  T is an integer-valued r.v. such that $T = n$ only depends on the values of $X_1, \ldots, X_n$, on, in other words, the values $S_1, \ldots, S_n$.
- Wald's identity/equation:
  Under proper conditions, $E(S_T) = \mu E(T)$.
- Interpretation: if X_i is the gain/loss of the i-th game, which is fair in the sense that $E(X_i) = 0$. Then S_n is the cumulative gain/loss in the first n games. Under proper conditions, any exit strategy (stopping time) shall still break even.

Tool box:

- Indicator functions.
  (e.g.,Borel-Cantelli Lemma $\longrightarrow$ Borel's strong law of large numbers $\longrightarrow$ Kolmogorov's 0-1 law.)

Tool box:

- Indicator functions.
  (e.g.,Borel-Cantelli Lemma $\longrightarrow$ Borel's strong law of large numbers $\longrightarrow$ Kolmogorov's 0-1 law.)
- Truncation.

Tool box:

- Indicator functions.
  (e.g.,Borel-Cantelli Lemma $\longrightarrow$ Borel's strong law of large numbers $\longrightarrow$ Kolmogorov's 0-1 law.)
- Truncation.
- Characteristic functions/moment generating functions. (in proving the CLT and its convergence rates)

Inequalities

- Jensen, Holder, Cauchy-Schwartz, Lyapunov, Minkowski.

Inequalities

- Jensen, Holder, Cauchy-Schwartz, Lyapunov, Minkowski.
- Markov, Chebyshev.

Inequalities

- Jensen, Holder, Cauchy-Schwartz, Lyapunov, Minkowski.
- Markov, Chebyshev.
- Kolmogorov.

Inequalities

- Jensen, Holder, Cauchy-Schwartz, Lyapunov, Minkowski.
- Markov, Chebyshev.
- Kolmogorov.
- Exponential: Berstein, Bennett, Hoeffding,

Inequalities

- Jensen, Holder, Cauchy-Schwartz, Lyapunov, Minkowski.
- Markov, Chebyshev.
- Kolmogorov.
- Exponential: Berstein, Bennett, Hoeffding,
- Doob (for martingale).

Inequalities

- Jensen, Holder, Cauchy-Schwartz, Lyapunov, Minkowski.
- Markov, Chebyshev.
- Kolmogorov.
- Exponential: Berstein, Bennett, Hoeffding,
- Doob (for martingale).
- Khintchine, Marcinkiewicz-Zygmund, Burkholder-Gundy,

Martingales (Chapter 4).

1. Conditional expectation with respect to σ-algebra.
2. Definition of martingales,
3. Inequalities.
4. Optional sampling theorem.
5. Martingale convergence theorem.