Stochastic Process

Introduction and review of probability

Lecturer: Xiao Ma

Office: 631G, Tel: 3994 3112

maxiao@mail.sysu.edu.cn

School of Data and Computer Science

Sun Yat-sen University

GraduateCourse, Autumn2018

Chapter 1. Introduction and review of probability

Axioms of probability

- Given any sample space Ω and any class of events ε satisfying the axioms of events, a probability rule is a function $\Pr\{\cdot\}$ mapping each $A \in \varepsilon$ to a (finite) real number in such a way that the following three probability axioms hold

Axioms of probability

- Given any sample space Ω and any class of events ε satisfying the axioms of events, a probability rule is a function $\Pr\{\cdot\}$ mapping each $A \in \varepsilon$ to a (finite) real number in such a way that the following three probability axioms hold
  - ◇ $\Pr\{\Omega\} = 1$.

Axioms of probability

- Given any sample space Ω and any class of events ε satisfying the axioms of events, a probability rule is a function $\Pr\{\cdot\}$ mapping each $A \in \varepsilon$ to a (finite) real number in such a way that the following three probability axioms hold
  - ◇ $\Pr\{\Omega\} = 1$.
  - ◇ For every event A, $\Pr\{A\} \geq 0$.

Axioms of probability

- Given any sample space Ω and any class of events ε satisfying the axioms of events, a probability rule is a function $\Pr\{\cdot\}$ mapping each $A \in \varepsilon$ to a (finite) real number in such a way that the following three probability axioms hold

  ◇ $\Pr\{\Omega\} = 1$.

  ◇ For every event A, $\Pr\{A\} \geq 0$.

  ◇ The probability of the union of any sequence $A_1, A_2, \ldots$ of disjoint events is given by

$$\Pr\{\bigcup_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} \Pr\{A_n\}\}$$

where $\sum_{n=1}^{\infty} \Pr\{A_n\}$ is shorthand for $\lim_{n\to\infty} \sum_{n=1}^{\infty} \Pr\{A_n\}$.

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{\bigcup_{n=1}^{m} A_n\} = \sum_{n=1}^{m} \Pr\{A_n\}$, for $A_1, \ldots, A_m$ disjoint.

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{\bigcup_{n=1}^{m} A_n\} = \sum_{n=1}^{m} \Pr\{A_n\}$, for $A_1, \ldots, A_m$ disjoint.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{\bigcup_{n=1}^{m} A_n\} = \sum_{n=1}^{m} \Pr\{A_n\}$, for $A_1, \ldots, A_m$ disjoint.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.
  - ◇ $\Pr\{A\} \leq Pr\{B\} \leq 1$, for $A \subseteq B$.

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{\bigcup_{n=1}^{m} A_n\} = \sum_{n=1}^{m} \Pr\{A_n\}$, for $A_1, \ldots, A_m$ disjoint.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.
  - ◇ $\Pr\{A\} \leq Pr\{B\} \leq 1$, for $A \subseteq B$.
  - ◇ $\sum_n \Pr\{A_n\} \leq 1$, for $A_1, A_2, \ldots$disjoint.

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{\bigcup_{n=1}^{m} A_n\} = \sum_{n=1}^{m} \Pr\{A_n\}$, for $A_1, ..., A_m$ disjoint.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.
  - ◇ $\Pr\{A\} \leq Pr\{B\} \leq 1$, for $A \subseteq B$.
  - ◇ $\sum_n \Pr\{A_n\} \leq 1$, for $A_1, A_2, ...$disjoint.
  - ◇ $\Pr\{\bigcup_{n=1}^{\infty} A_n\} = \lim_{m\to\infty} \Pr\{\bigcup_{n=1}^{m} A_n\}$

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{\bigcup_{n=1}^{m} A_n\} = \sum_{n=1}^{m} \Pr\{A_n\}$, for $A_1, ..., A_m$ disjoint.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.
  - ◇ $\Pr\{A\} \leq Pr\{B\} \leq 1$, for $A \subseteq B$.
  - ◇ $\sum_n \Pr\{A_n\} \leq 1$, for $A_1, A_2, ...$disjoint.
  - ◇ $\Pr\{\bigcup_{n=1}^{\infty} A_n\} = \lim_{m\to\infty} \Pr\{\bigcup_{n=1}^{m} A_n\}$
  - ◇ $\Pr\{\bigcup_{n=1}^{\infty} A_n\} = \lim_{n\to\infty} \Pr\{A_n\}$, for $A_1 \subseteq A_2 \subseteq ...$

Axioms of probability

- It is surprising that this is all that is needed to avoid paradoxes. A few simple consequences are:
  - ◇ $\Pr\{\varnothing\} = 0$.
  - ◇ $\Pr\{\bigcup_{n=1}^{m} A_n\} = \sum_{n=1}^{m} \Pr\{A_n\}$, for $A_1, ..., A_m$ disjoint.
  - ◇ $\Pr\{A^c\} = 1 - \Pr\{A\}$.
  - ◇ $\Pr\{A\} \leq Pr\{B\} \leq 1$, for $A \subseteq B$.
  - ◇ $\sum_n \Pr\{A_n\} \leq 1$, for $A_1, A_2, ...$disjoint.
  - ◇ $\Pr\{\bigcup_{n=1}^{\infty} A_n\} = \lim_{m\to\infty} \Pr\{\bigcup_{n=1}^{m} A_n\}$
  - ◇ $\Pr\{\bigcup_{n=1}^{\infty} A_n\} = \lim_{n\to\infty} \Pr\{A_n\}$, for $A_1 \subseteq A_2 \subseteq ...$
  - ◇ $\Pr\{\bigcap_{n=1}^{\infty} A_n\} = \lim_{n\to\infty} \Pr\{A_n\}$, for $A_1 \supseteq A_2 \supseteq ...$

Expectations

Remark

The distribution function of a r.v. X often contains more detail than necessary. The expectation $\overline{X} = \mathrm{E}[X]$ is sometimes all that is needed.

Expectations

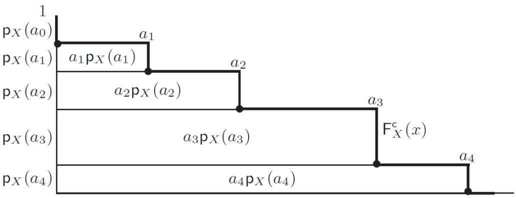
Remark

The distribution function of a r.v. X often contains more detail than necessary. The expectation $\overline{X} = \mathrm{E}[X]$ is sometimes all that is needed.

- for discrete X, $\mathrm{E}[X] = \sum_i x_i P_X(x_i)$;
- for continuous X, $\mathrm{E}[X] = \int y f_X(y) \mathrm{d}y$;
- for arbitrary nonnegative X, $\mathrm{E}[X] = \int F_X^c(y) \mathrm{d}y$;
- for arbitrary X, $\mathrm{E}[X] = \int_{-\infty}^{0} F_X(y) \mathrm{d}y + \int_0^{\infty} F_X^c(y) \mathrm{d}y$.

where $F_X^c(x) = 1 - F_X(x) = \Pr\{X > x\}$.

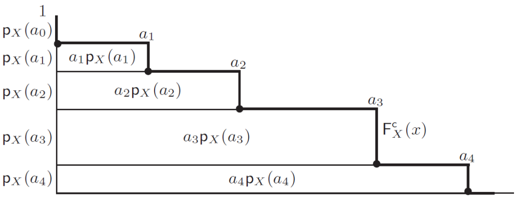
Expectations



Look at discrete case.

$$\int F_X^c(y)\mathrm{d}y = \sum_i a_i P_X(a_i).$$

Expectations



Look at discrete case.

$$\int F_X^c(y)\mathrm{d}y = \sum_i a_i P_X(a_i).$$

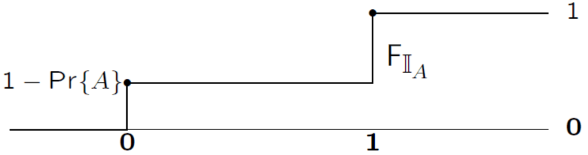
Almost as important is the standard deviation,

$$\sigma_X = \sqrt{\mathrm{E}[(X - \overline{X})^2]}$$

Indicator Random Variables

For every event A in a probability model, an indicator r.v. $\mathbb{1}_A$ is defined where $\mathbb{1}_A(\omega) = 1$ for $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ otherwise. Note that $\mathbb{1}_A$ is a binary r.v..

$$P_{\mathbb{1}_A}(0) = 1 - \Pr\{A\}; \qquad P_{\mathbb{1}_A}(1) = \Pr\{A\}.$$



$$\mathrm{E}[\mathbb{1}_A] = \Pr\{A\}; \qquad \sigma_{\mathbb{1}_A} = \sqrt{\Pr\{A\}(1 - \Pr\{A\})}.$$

Theorems about r.v.s can thus be applied to events.

Chapter 1. Introduction and review of probability

Basic inequalities

Theorem (the Markov inequality)

if a non-negative r.v. X has a mean $\mathrm{E}[X]$, then for every $y > 0$,

$$\Pr\{X \geqslant y\} \leqslant \frac{\mathrm{E}[X]}{y} \tag{1}$$

Basic inequalities

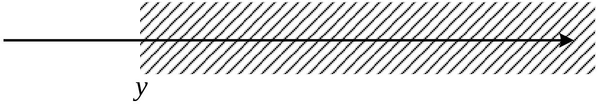
Theorem (the Markov inequality)

if a non-negative r.v. X has a mean $\mathrm{E}[X]$, then for every $y > 0$,

$$\Pr\{X \geqslant y\} \leqslant \frac{\mathrm{E}[X]}{y} \tag{1}$$

Proof.

$$\mathrm{E}[X] = \int_0^{+\infty} x dF_X \geqslant \int_y^{+\infty} x dF_X \geqslant y \int_y^{+\infty} dF_X = y\Pr\{X \geqslant y\}.$$



Basic inequalities

Theorem (the Chebyshev inequality)

if X has a mean $\mathrm{E}[X]$ and finite variance σ_X^2, then for every $\delta > 0$,

$$\Pr\{|X - \mathrm{E}[X]| \geqslant \delta\} \leqslant \frac{\sigma_X^2}{\delta^2} \tag{2}$$

Basic inequalities

Theorem (the Chebyshev inequality)

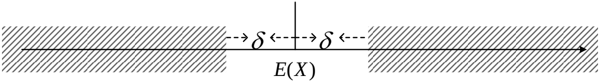
if X has a mean $\mathrm{E}[X]$ and finite variance σ_X^2, then for every $\delta > 0$,

$$\Pr\{|X - \mathrm{E}[X]| \geqslant \delta\} \leqslant \frac{\sigma_X^2}{\delta^2} \tag{2}$$

Proof.

apply the Markov inequality on $(X - \mathrm{E}[X])^2$,

$$\Pr\{|X - \mathrm{E}[X]| \geqslant \delta\} = \Pr\{|X - \mathrm{E}[X]|^2 \geqslant \delta^2\} \leqslant \frac{\mathrm{E}[(X - \mathrm{E}[X])^2]}{\delta^2} = \frac{\sigma_X^2}{\delta^2}.$$



□

Basic inequalities

Theorem (Chernoff bounds)

for all $s \geqslant 0$,

$$\Pr\{X \geqslant y\} \leqslant \mathrm{E}[e^{sX}]e^{-sy} \tag{3}$$

Basic inequalities

Theorem (Chernoff bounds)

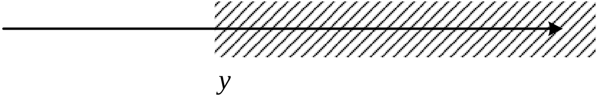
for all $s \geqslant 0$,

$$\Pr\{X \geqslant y\} \leqslant \mathrm{E}[e^{sX}]e^{-sy} \tag{3}$$

Proof.

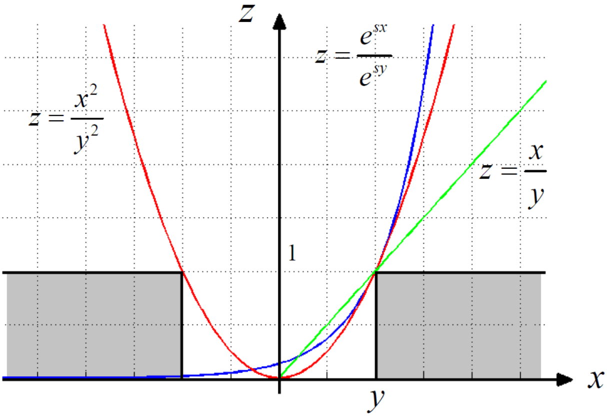
apply the Markov inequality on e^{sX},

$$\Pr\{X \geqslant y\} = \Pr\left(e^{sX} \geqslant e^{sy}\right) \leqslant \frac{\mathrm{E}[e^{sX}]}{e^{sy}}.$$



Basic inequalities

Consider the indicator r.v. $Z = \mathbb{1}_A$



Basic inequalities

Consider the indicator r.v. $Z = \mathbb{1}_A$

- Take $A = \{\omega : X(\omega) \geqslant y\}$, then

$$Z \leqslant \frac{X}{y} \Rightarrow \Pr\{X \geqslant y\} = E(Z) \leqslant \frac{E(X)}{y}.$$

- If X has 0 mean, take $A = \{\omega : |X(\omega)| \geqslant y\}$, then

$$Z \leqslant \frac{X^2}{y^2} \Rightarrow \Pr\{|X| \geqslant y\} = E(Z) \leqslant \frac{E(X^2)}{y^2} = \frac{\sigma_X^2}{y^2}.$$

- Take $A = \{\omega : X(\omega) \geqslant y\}$, then

$$Z \leqslant \frac{e^{sX}}{e^{sy}} \Rightarrow \Pr\{X \geqslant y\} = E(Z) \leqslant \frac{E(e^{sX})}{e^{sy}}.$$

Chapter 1. Introduction and review of probability

Weak law of large numbers

Theorem (Weak Law of large numbers (WLLN))

For each integer $n \geqslant 1$, *let* $S_n = X_1 + \cdots + X_n$, *where* $X_1, X_2, \ldots$ *are i.i.d. r.v.s satisfying* $\mathrm{E}[|X|] < \infty$. *Then for any* $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathrm{Pr}\left\{ |\frac{S_n}{n} - \mathrm{E}[X]| > \varepsilon \right\} = 0. \tag{4}$$

Weak law of large numbers

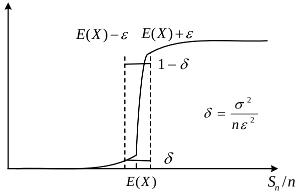
Theorem (Weak Law of large numbers (WLLN))

For each integer $n \geqslant 1$, let $S_n = X_1 + \cdots + X_n$, where $X_1, X_2, \ldots$ are i.i.d. r.v.s satisfying $\mathrm{E}[|X|] < \infty$. Then for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\left\{ \left|\frac{S_n}{n} - \mathrm{E}[X]\right| > \varepsilon \right\} = 0. \tag{4}$$

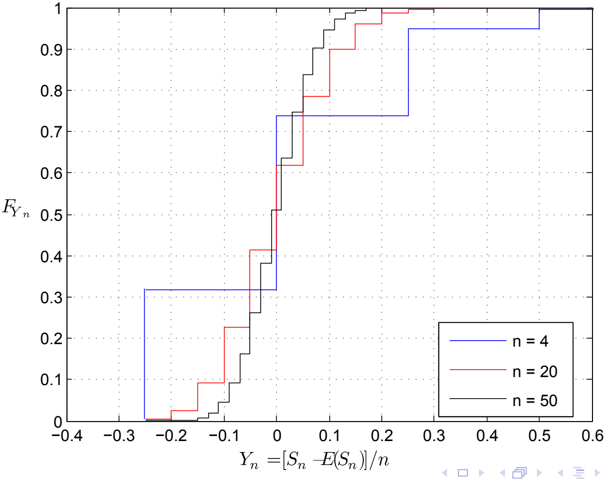
Proof.

apply the Chebyshev inequality on $\frac{S_n}{n}$.



Weak law of large numbers

Remark. The WLLN says that $\frac{S_n}{n}$ converge to $\overline{X}$ in probability.



The central limit theorem

Theorem (the central limit theorem (CLT))

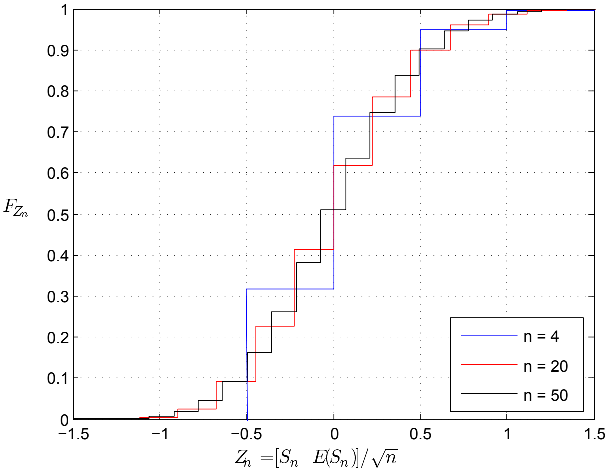
Let $X_1, X_2, \ldots$ be i.i.d. r.v.s with finite mean $\overline{X}$ and finite variance σ^2. Then for every $z \in \mathbb{R}$,

$$\lim_{n\to\infty} \Pr\left\{ \left| \frac{S_n - n\overline{X}}{\sigma\sqrt{n}} \right| \leqslant z \right\} = \Phi(z). \tag{5}$$

where $\Phi(z)$ is the CDF of Gaussian distribution with mean 0 and variance 1.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y$$

The central limit theorem



The central limit theorem

Remark

- *Paradox: The CLT says something very strong about how* $\frac{S_n}{n}$ *converges to* $\overline{X}$*, but convergence in distribution is a very weak form of convergence.*

The central limit theorem

Remark

- *Paradox: The CLT says something very strong about how* $\frac{S_n}{n}$ *converges to* $\overline{X}$*, but convergence in distribution is a very weak form of convergence.*
- *Resolution: The r.v.s that converge in distribution in the CLT are* $\frac{S_n - n\overline{X}}{\sqrt{n}\sigma_X}$*. Those that converge in probability to 0 are* $\frac{S_n - n\overline{X}}{n}$*.*

The central limit theorem

Remark

- *Paradox: The CLT says something very strong about how* $\frac{S_n}{n}$ *converges to* $\overline{X}$*, but convergence in distribution is a very weak form of convergence.*
- *Resolution: The r.v.s that converge in distribution in the CLT are* $\frac{S_n - n\overline{X}}{\sqrt{n}\sigma_X}$*. Those that converge in probability to 0 are* $\frac{S_n - n\overline{X}}{n}$*.*
- *Example: The CLT says that*

$$\lim_{n\to\infty} \Pr\left\{\frac{S_n - n\overline{X}}{n} \leqslant 0\right\} = \frac{1}{2}$$

This cannot be deduced from the WLLN.

Convergence

Definition

A sequence $X_1, X_2, \ldots$ of random variables is said to **converge in distribution** to a random variable X if

$$\lim_{n\to\infty} F_n(x) = F(x), \tag{6}$$

for each $x \in \mathbb{R}$ at which F is continuous. Here F_n and F are the cumulative distribution functions of random variables X_n and X correspondingly.

Convergence

Definition

A sequence $X_1, X_2, \ldots$ of random variables is said to **converge in distribution** to a random variable X if

$$\lim_{n\to\infty} F_n(x) = F(x), \tag{6}$$

for each $x \in \mathbb{R}$ at which F is continuous. Here F_n and F are the cumulative distribution functions of random variables X_n and X correspondingly.

Definition

A sequence $X_1, X_2, \ldots$ of random variables is said to **converge in mean square** to X if

$$\lim_{n\to\infty} \mathrm{E}[(X_n - X)^2] = 0, \tag{7}$$

Convergence of random variables

Definition

A sequence X_n of random variables **converges in probability** towards X if for all $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\{|X_n - X| > \varepsilon\} = 0. \tag{8}$$

Convergence of random variables

Definition

To say that the sequence X_n converges almost surely or almost everywhere or **with probability 1** or strongly towards X means that

$$\Pr\{\lim_{n\to\infty} X_n = X\} = 1. \tag{9}$$

Convergence of random variables

Definition

To say that the sequence X_n converges almost surely or almost everywhere or **with probability 1** or strongly towards X means that

$$\Pr\{\lim_{n\to\infty} X_n = X\} = 1. \tag{9}$$

Remark

convergence with probability 1 $\Rightarrow$ *convergence in probability; convergence in mean square* $\Rightarrow$ *convergence in probability; convergence in probability* $\Rightarrow$ *convergence in distribution;*

Thank you for your attention!