IOE 516

Stochastic Processes II

Winter Term, 2024

Prof. Xiuli Chao

Email: xchao@umich.edu

Recap

- In the last lecture, we discussed two operations applications. Simple policies are developed and are shown, using concentration inequality, near-optimal when T is large.

- We also started to discuss an important class of stochastic process, martingales.

- Today, we will first present more martingale examples, and then discuss the main results of martingale theory.

Review of martingale

- **Definition.** A stochastic sequence $\{X_n; n \geq 0\}$ is called a martingale process if

  (i) $E[|X_n|] < \infty$, and

  (ii) the expected value of the next period is the same as the value of the current period

$$E[X_{n+1} \mid X_0, X_1, \ldots, X_n] = X_n, \qquad n = 1, \ldots.$$

Law of total expectation

- For the next example, we need to use Law of Total Expectation

$$E[X] = E[E[X|Y]]$$

- The result holds, of course, on a conditional space as well: If we the probability measure is the conditional probability space given some Z (maybe a vector), then

$$E[X|Z] = E[E[X|Y,Z]|Z].$$

- What's your intuitive reasoning?

Example 4

- Suppose $Y_1, Y_2, \ldots$ is a sequence of observations (maybe dependent of one another). Then for any random quantity of interest X, the process

$$X_n = E[X|Y_1, \ldots, Y_n], \quad n = 1, 2, \ldots,$$

  is a martingale.

- **Proof.**

$$\begin{aligned} E[X_{n+1}|Y_1, \ldots Y_n] &= E[E[X|Y_1, \ldots, Y_n, Y_{n+1}]|Y_1, \ldots, Y_n] \\ &= E[X|Y_1, \ldots, Y_n] \\ &= X_n. \end{aligned}$$

Remark

- This martingale

$$X_n = E[X|Y_1, \ldots, Y_n]$$

  is known as **Doob martingale**.

- It is well known from prediction theory that, given $Y_1, \ldots, Y_n$, the best prediction value of a random variable X (for minimizing mean square error) is $E[X|Y_1, \ldots, Y_n]$.

- This previous example states that the adaptive predictions of the value of a random variable is a martingale.

Example 5

- Suppose $X_1, X_2, \ldots$ is a sequence of random variables, neither independent or identically distributed. Then

$$Y_n = \sum_{i=1}^{n} \Big(X_i - E[X_i | X_1, \ldots, X_{i-1}] \Big)$$

and $Y_0 = 0$ is a martingale.

Martingale difference and martingale representation

- If martingale X_n is written as

$$X_n = X_0 + \sum_{i=1}^{n} Z_i,$$

then Z_i is called the martingale difference.

- **Claim:** Any martingale can be represented in the form of martingale difference with $Z_i = X_i - X_{i-1}, i = 1, 2, \ldots$.

- Martingale difference satisfies, for all $n \geq 0$, by $X_{n+1} = X_n + Z_{n+1}$ and, by $E[X_{n+1}|\mathcal{F}_n] = X_n$,

$$E[Z_{n+1}|\mathcal{F}_n] = 0.$$

- Thus $E[Z_n] = 0$ for all n. Furthermore, for any $i < j$,

$$E[Z_i Z_j] = 0.$$

- **Why?** Reason is

$$E[Z_i Z_j] = E[E[Z_i Z_j|\mathcal{F}_{j-1}]] = E[Z_i E[Z_j|\mathcal{F}_{j-1}]] = 0.$$

- Thus, martingale differences are uncorrelated, though they may not be independent.

- This shows that, martingale process is a generalization of sum of i.i.d. random variables.

Main results of martingale theory

- The main results in martingale theory:
  - Azuma-Hoeffding inequality (for martingale)
  - Martingale Stopping theorem
  - Kolmogorov maximal inequality (for sub-martingale)
  - Martingale Convergence theorem

- The first and third are concentration inequalities, the second is on L^1-convergence, while the last is a.s. convergence. We discuss the first three results in detail, and last one briefly.

Azuma-Hoeffding inequality

- If $X_0, X_1, X_2, \ldots$ is a martingale with $X_0 = \mu$, and there exist $a_i \le b_i$ such that

  $$a_i \le X_i - X_{i-1} \le b_i, \qquad i = 1, 2, \ldots.$$

  Then, for any $\epsilon > 0$,

  $$P(X_n - \mu > \epsilon) \le e^{-2\epsilon^2 / \sum_{i=1}^n (b_i - a_i)^2}$$

- Thus, Hoeffding inequality extends to martingale process.

Recall the following?

- If X has support on $[a, b]$ and has mean 0, then

$$E[e^{\theta X_1}] \leq e^{\theta^2 (b-a)^2/8}.$$

- Remember our proof? that uses

$$e^{\theta X} \leq \frac{b - X}{b - a} e^{\theta a} + \frac{X - a}{b - a} e^{\theta b}$$

- Then it follows

$$\begin{aligned} E[e^{\theta X}] &\leq \frac{b}{b-a} e^{\theta a} - \frac{a}{b-a} e^{\theta b} \\ &= e^{\theta a + \log\left(1 + \frac{a}{b-a} - \frac{a}{b-a} e^{\theta(b-a)}\right)} \\ &\leq \cdots \end{aligned}$$

Proof

- WLOG assume $\mu_i = 0$ for all i. Then, for any $\theta > 0$,

$$\begin{aligned}
P(X_n > \epsilon) &= P(e^{\theta X_n} > e^{\theta\epsilon}) \\
&\leq e^{-\theta\epsilon} E[e^{\theta X_n}] \\
&= e^{-\theta\epsilon} E[E[e^{\theta X_n} | X_1, \ldots, X_{n-1}]] \\
&\leq e^{-\theta\epsilon} E[e^{\theta X_{n-1}} E[e^{\theta Z_n} | X_1, \ldots, X_{n-1}]] \\
&\leq e^{-\theta\epsilon} E[e^{\theta Z_{n-1}}] e^{\theta^2 (b_n - a_n)^2/8} \\
&\leq e^{-\theta\epsilon + \theta^2 \sum_{i=1}^n (b_i - a_i)^2/8}
\end{aligned}$$

- Choosing $\theta = 4\epsilon / \sum_{i=1}^n (b_i - a_i)^2$ yields

$$P(X_n > \epsilon) \leq e^{-2\epsilon^2 / \sum_{i=1}^n (b_i - a_i)^2}.$$

Back to McDiarmid inequality

- With the machinery of martingale, we are now ready to prove McDiarmid inequality. Suppose $X_1, X_2, \ldots$ are independent random variables. If g satisfies,

$$\sup_{x_1,\ldots,x_n,x_i'} |g(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n) - g(x_1,\ldots,x_{i-1},x_i',x_{i+1},\ldots,x_n)| \le c_i.$$

  Then, we have

$$P(g(X_1,\ldots,X_n) - E[g(X_1,\ldots,X_n)] > \epsilon) \le e^{-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2}}.$$

Proof of McDiarmid inequality

- For $i = 1, \ldots, n$, consider Doob martingale

$$Y_i = E[g(X_1, X_2, \ldots, X_n)|X_1, \ldots, X_i].$$

- Clearly,

$$g(X_1, \ldots, X_n) = E[g(X_1, \ldots, X_n)] + \sum_{i=1}^{n} (Y_i - Y_{i-1}).$$

- Note that, for any $X_1 = x_1, \ldots, x_i = x_i$, by independence of X_i's,

$$\begin{aligned} Y_i - Y_{i-1} &= E[g(X_1, \ldots, X_{i-1}, X_i, \ldots, X_n)|X_1, \ldots, X_i] \\ &\quad -E[g(X_1, \ldots, X_{i-1}, X_i, \ldots, X_n)|X_1, \ldots, X_{i-1}] \\ &= E[g(x_1, \ldots, x_{i-1}, x_i, \ldots, X_n|) \\ &\quad -E[g(x_1, \ldots, x_{i-1}, X_i, \ldots, X_n) \\ &\le c_i \end{aligned}$$

- The above is true for any $X_1, \ldots, X_i$.

- It follows from Azuma-Hoeffding inequality that

$$P\Big(g(X_1, \ldots, X_n) - E[g(X_1, \ldots, X_n)] > \epsilon\Big) \le e^{-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2}}.$$

- **Remember:** McDiarmid inequality, or bounded difference inequality, can be very useful to you in your research.

Questions

- **Question 1:** Where is independence of $X_1, \ldots, X_n$ used in the proof?

- **Question 2:** Does the result hold true when these random variables are dependent?

Stopping time

- **Definition.** Let $X_1, X_2, \ldots$ be a stochastic sequence. An integer-valued random variable N is called a stopping time with respect to $X_1, X_2, \ldots$ if the event $\{N = n\}$ is completed determined by $X_1, X_2, \ldots, X_n$.

- The part of the sentence "with respect to $X_1, X_2, \ldots$ " is often omitted when no confusing may arise in the context.

Remark

- Typically, when we discuss random variables, we only consider r.v.'s that are a.s. finite, or $P(X < \infty) = 1$. Such random variables are called "regular".

- By convention, when discussing stopping times, we usually include the cases that can take "infinity", i.e., we allow $P(N < \infty) < 1$.

Example 1

- Let

$$P(X_n = 1) = 1/2 = 1 - P(X_n = -1), \qquad j = 1, 2, \ldots.$$

- Let $S_n = \sum_{i=1}^{n} X_i$ and $S_0 = 0$, and define

$$N = \inf\{n : S_n = 10\}.$$

- Is N a stopping time?

Example 2

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. non-negative random variables and

$$S_n = \sum_{i=1}^{n} X_i.$$

- Given time t, let

$$N = \max\{n : S_n \leq t\}.$$

- Is N a stopping time?

Question 1

- Is $N + 1$ a stopping time?

Question 2

- Suppose $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. We know that, for given number n, we have

$$E\Big[\sum_{i=1}^{n} X_i\Big] = nE[X_1].$$

- Let N be a stopping time. Do we have

$$E\Big[\sum_{i=1}^{N} X_i\Big] = E[N]E[X_1]?$$

Example

- Let
$$P(X_n = 1) = 1/2 = 1 - P(X_n = -1), \qquad j = 1, 2, \ldots.$$

- Define $S_n = \sum_{i=1}^{n} X_i$ and
$$N = \text{inf}\{n : S_n = 10\}.$$

- Do we have
$$E\Big[\sum_{i=1}^{N} X_i\Big] = E[N]E[X_1]?$$

Wald's equation

- Let $X_1, X_2, \ldots$ be i.i.d. with $E[|X_1|] < \infty$, and N is a stopping time with $E[N] < \infty$, then

$$E\Big[\sum_{i=1}^{N} X_i\Big] = E[N]E[X_1].$$

Recall the following result?

- If X_n is such that $\sum_{n=1}^{\infty} E[|X_n|] < \infty$, then

$$E\Big[\sum_{n=1}^{\infty} X_n\Big] = \sum_{n=1}^{\infty} E[X_n]$$

- This is called Fubini's theorem. You just need to know it to apply it.

Proof

$$
\begin{aligned}
E\Big[\sum_{i=1}^{N} X_i\Big] &= E\Big[\sum_{i=1}^{\infty} X_i 1[i \le N]\Big] \\
&= \sum_{i=1}^{\infty} E[X_i 1[i \le N]] \\
&= \sum_{i=1}^{\infty} E[E[X_i 1[i \le N] | F_{i-1}]] \\
&= \sum_{i=1}^{\infty} E[1[i \le N] E[X_i | F_{i-1}]] \\
&= \sum_{i=1}^{\infty} E[1[i \le N]] E[X_1] \\
&= E\Big[\sum_{i=1}^{\infty} 1[i \le N]\Big] E[X_1] \\
&= E[N] E[X_1]
\end{aligned}
$$

Remark

- The previous result is also known as Wald's first equation. There is a Wald's second equation: If X_i's are i.i.d. with mean 0 and finite variance, $E[X_1^2| < \infty$, and N a stopping time with $E[N] < \infty$. Then

$$Var\Big(\sum_{i=1}^{N} X_i\Big) = E[N] \cdot Var(X_1).$$

- This will be left to you as a homework problem.

Stopped time

- Let N be a stopping time with respect to $X_1, X_2, \ldots$, and n is a given integer. Then

  $$n \wedge N := \min\{n, N\}$$

  is called a **stopped time**.

- Why? The process $\{n \wedge N; n = 1, 2, \ldots\}$ stops at N!

Claim

- Given n. Stopped time is a stopping time.

- Need to show that, for $k = 1, 2, \ldots$, $n \wedge N = k$ is determined by the process $X_1, \ldots, X_k$.

- If $k < n$, then it is equivalent to $N = k$, so it is ...

- If $k \geq n$, then it is equivalent to $N \geq n$, hence it is ...