

Martingales I: basic definitions and concepts

Elvis Cui

MS in Biostatistics, UCLA

February 7, 2020

Overview

- 1 Conditional Expectation
- 2 Martingales: definitions
- 3 Stopping times and stopping processes
- 4 Predictable processes and Doob's decomposition
- 5 Reversed time martingales
- 6 Symmetric tail σ -algebra and exchangeability
- 7 Hewitt-Savage theorem
- 8 U-statistics
 - U-statistics are reversed martingales
- 9 Conditional independence and de Finetti theorem
- 10 Polya's urn model and U-statistic

Conditional Expectation I

Definition 1 (Kolmogorov's conditional expectation)

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a sub- σ -field $\mathcal{G} \in \mathcal{F}$, a random variable $X \in \mathcal{F}$ s.t. $\mathbb{E}|X| < \infty$, we define the **conditional expectation of X given \mathcal{G}** to be any random variable Y s.t.

- $Y \in \mathcal{G}$
- $\forall A \in \mathcal{G}, \int_A X dP = \int_A Y dP$

The existence of conditional can be derived from **Radon-Nikodym theorem**.

Conditional Expectation II

Definition 2 (Conditional expectation as projection)

Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a sub- σ -field $\mathcal{G} \in \mathcal{F}$, let $L_2(\mathbb{P})$ be the space of all square integrable random variables with probability measure \mathbb{P} . Let $\|X\|_2 = (\mathbb{E}(X^2))^{1/2}$ and

$$\mathbb{H} = \{Z \in L_2(\mathbb{P}) : Z \in \mathcal{G}\}$$

Then by **projection theorem in Hilbert spaces**, for any $X \in L_2(\mathbb{P})$, there exists a essentially unique $\hat{X} \in \mathbb{H}$ s.t. for any $Y \in \mathbb{H}$, $\|\hat{X} - X\|_2 \leq \|Y - X\|_2$. Thus, for any $Y \in \mathbb{H}$, we have

$$\mathbb{E}XY = \mathbb{E}\hat{X}Y$$

Let $Y(\omega) = 1\{\omega \in A\}$, $A \in \mathcal{G}$, we find \hat{X} is the **conditional expectation** of X .

Definition 3 (Filtration, adaptation and martingales)

Let \mathcal{F}_n be a **filtration**, i.e., an increasing sequence of σ -fields. A sequence X_n is said to be **adapted** to \mathcal{F}_n iff $X_n \in \mathcal{F}_n$ for any n . The adapted sequence X_n is a martingale if

- $\mathbb{E}|X_n| < \infty$
- $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$

If " $=$ " is replaced with " \leq " (\geq), then the sequence is called **super-martingale** (sub-martingale), respectively.

For examples of martingales, see (i) Durrett, Probability, Section 4.2, 2019. (ii) Billingsley, Probability and measure, chapter 35, 2010.

Martingale difference sequence

Definition 4 (Martingale difference)

A sequence of RV $\{X_n : n \geq 0\}$ adapted to a filtration \mathcal{F}_n is referred to as a **Martingale difference sequence** iff $\mathbb{E}(X_n | \mathcal{F}_{n-1}) = 0$ for all $n \geq 1$. Similarly, the sequence forms a sub-(super)-martingale difference iff $\mathbb{E}(X_n | \mathcal{F}_{n-1}) \geq 0$ (≤ 0).

Stopping times and stopping processes

Definition 5 (Stopping times)

A **stopping time** T w.r.t. the filtration \mathcal{F}_n is an extended RV taking values in $\mathbb{N} \cup \{+\infty\}$ s.t. for any finite $n \geq 0$,

$$\{T = n\} \in \mathcal{F}_n$$

A constant random variable $T = N$ is always a stopping time.

Definition 6 (Stopping processes)

Given $\{Z_n : n = 1, \dots\}$ a process adapted to \mathcal{F}_n and a stopping time T , we define the **stopped process** Z^T as the collection of random variables:

$$Z_{T \wedge n} = \sum_{k=0}^{n-1} 1(T = k)Z_k + Z_n 1(T \geq n)$$

Predictable processes and Doob's decomposition

Definition 7 (Predictable process)

$\{X_n : n \geq 0\}$ is called predictable w.r.t. $\{\mathcal{F}_n\}$ iff X_n is measurable w.r.t. \mathcal{F}_{n-1} for any $n \geq 1$.

Theorem 8 (Doob's decomposition)

Let $X = \{X_n : n \geq 0\}$ be a process adapted to the filtration $\{\mathcal{F}_n\}$ s.t. $\mathbb{E}|X_n| < \infty$. Then the process can be decomposed as

$$X_n = X_0 + M_n + \Lambda_n \text{ } \mathbb{P}\text{-a.s.}$$

Where $\{M_n\}$ is a mean zero martingale and $\{\Lambda_n\}$ is a predictable process s.t. $\mathbb{P}(\Lambda_0 = 0) = 1$.

Reversed time martingales

Definition 9 (Reversed filtration)

Given $(\Omega, \mathcal{F}, \mathbb{P})$, a reversed filtration $\{\mathcal{G}_n : n \geq 0\}$ is a decreasing sequence of σ -algebras contained in \mathcal{F} , i.e., $\mathcal{G}_0 \supseteq \mathcal{G}_1 \supseteq \dots$.

Definition 10 (Reversed time martingales)

Let $\{X_n : n \geq 0\}$ be adapted to $\{\mathcal{G}_n\}$ s.t. $\mathbb{E}|X_n| < \infty$. Then Z_n is referred to as a reversed martingale iff $\mathbb{E}(Z_n | \mathcal{G}_{n+1}) = Z_{n+1}$ for all n . Similarly, sub or super martingale iff \geq or \leq for all n .

Definition 11 (Tail σ -algebra)

Let X_1, X_2, \dots be a sequence of RVs defined on $(\Omega, \mathcal{F}, \mathbb{P})$. **The tail σ -algebra** associated with this sequence is given by

$$\mathcal{T}_\infty = \bigcap_{n \geq 1} \mathcal{T}_n$$

where

$$\mathcal{T}_n = \sigma(\cup_{j=1}^{\infty} \sigma(X_n, \dots, X_{n+j}))$$

Exchangeable distribution and symmetric sets

Definition 12 (Exchangeability)

$\{X_k : k \geq 1\}$ has **exchangeable distribution** iff all finite dimensional distribution of X are exchangeable, i.e., for every n and any permutation $\pi \in \text{Perm}(n)$,

$$(X_1, \dots, X_n) =^d (X_{\pi(1)}, \dots, X_{\pi(n)})$$

Definition 13 (Symmetric set)

A set $A \in \mathcal{B}(\mathbb{R}^n)$ is called **symmetric** iff it is invariant w.r.t. permutations of coordinates.

$$(x_1, \dots, x_n) \in A \text{ iff } (x_{\pi(1)}, \dots, x_{\pi(n)}) \in A$$

The family $\mathcal{B}_{\text{sym}}(\mathbb{R}^n)$ of all symmetric Borel sets is a σ -algebra.

Definition 14 (Symmetric tail σ -algebra)

Let $\mathcal{S}_n^0 = \{C \in \mathcal{F} : 1_C = 1_A(X_1, \dots, X_n), A \in \mathcal{B}_{\text{sym}}(\mathbb{R}^n)\}$ and $\mathcal{T}_{n+1} = \sigma(X_{n+j} : j \geq 1)$. We define

$$\mathcal{S}_n = \sigma(\mathcal{S}_n^0, \mathcal{T}_{n+1})$$

That is, the smallest σ -algebra in \mathcal{F} generated by products of symmetric functions and measurable functions. **The symmetric tail σ -algebra** is defined to be

$$\mathcal{S}_\infty = \bigcap_{n \geq 1} \mathcal{S}_n$$

Obviously, we have $\mathcal{T}_\infty \subseteq \mathcal{S}_\infty$.

Theorem 15 (Hewitt-Savage)

Let \mathcal{S}_∞ be the symmetric tail σ -algebra generated by a sequence of **exchangeable random variables** $\{X_n\}$. Let \mathcal{T}_∞ be the usual tail σ -algebra. Then \mathcal{S}_∞ and \mathcal{T}_∞ are essentially equal, i.e., for any $A \in \mathcal{S}_\infty$, there exists a $B \in \mathcal{T}_\infty$ s.t. $\mathbb{P}(A \Delta B) = 0$.

Definition 16 (U-statistics)

Let $h(x_1, \dots, x_m)$ be an m argument function. In the language of statistics, it is called **the kernel of degree m** . The **U-statistic** corresponding to this kernel is defined as

$$\mathbb{U}_{n,m}(h) = \frac{(n-m)!}{n!} \sum_{i \in I_n^m} h(X_{i_1}, \dots, X_{i_m})$$

Where

$$I_n^m = \{i = (i_1, \dots, i_m) : 1 \leq i_k \leq n, i_k \neq i_l \text{ for } k \neq l\}$$

U-statistic as reversed martingale

Lemma 17

Let X have exchangeable distribution and let $\mathbb{E}(|h(X_1, \dots, X_m)|) < \infty$.
Then

$$\mathbb{U}_{n,m}(h) = \mathbb{E}(h(X_1, \dots, X_m) | \mathcal{S}_n)$$

and $\{\mathbb{U}_{n,m}(h)\}$ is a reversed time martingale w.r.t. $\{\mathcal{S}_n\}$.

Proof.

- Step 1: $\mathbb{U}_{n,m}(h) \in \mathcal{S}_n$.
- Step 2: Suppose g symmetric indicators, f measurable indicators, then show that: $\mathbb{E}(hgf) = \mathbb{E}(h_{\pi}gf)$.
- Step 3: Tower property.



Reversed time martingale convergence theorem

Theorem 18

Let X have exchangeable distribution and $\mathbb{E}(|h(X_1, \dots, X_m)|) < \infty$. Then we have

$$\mathbb{U}_{n,m}(h) \rightarrow \mathbb{E}[h(X_1, \dots, X_m) | \mathcal{S}_\infty] \text{ a.s. and in } L_1$$

This is a direct consequence of reversed time martingale convergence theorem. If h is a bounded kernel then the convergence also holds for the V statistics

$$\mathbb{V}_{n,m}(h) = \frac{n_{(m)}}{n^m} \mathbb{U}_{n,m}(h) + \sum_{r=1}^{m-1} \frac{1}{n^{m-r}} \left[\frac{n_{(r)}}{n^m} \mathbb{U}_{n,m}(h^{(r)}) \right]$$

where $h^{(r)}$ is the kernel h evaluated at sequences (i_1, \dots, i_m) having exactly r distinct indices.

Conditional independence

Definition 19 (Conditional independence)

Given $(\Omega, \mathcal{F}, \mathbb{P})$ and three sub- σ -algebras $\mathcal{F}_1, \mathcal{F}_2, \mathcal{G}$. We say that \mathcal{F}_1 and \mathcal{F}_2 are **conditionally independent given \mathcal{G}** if

$$\mathbb{E}(1_{B_1 \cap B_2} | \mathcal{G}) = \mathbb{E}(1_{B_1} | \mathcal{G}) \mathbb{E}(1_{B_2} | \mathcal{G}) \quad \mathbb{P} - a.s.$$

for all $B_1 \in \mathcal{F}_1$ and $B_2 \in \mathcal{F}_2$.

Lemma 20

The σ -algebras \mathcal{F}_1 and \mathcal{F}_2 are **conditionally independent** iff

$$\mathbb{E}(1_{B_1} | \sigma(\mathcal{G}, \mathcal{F}_2)) = \mathbb{E}(1_{B_1} | \mathcal{G}) \quad \mathbb{P} - a.s.$$

for any $B_1 \in \mathcal{F}_1$.

de Finetti theorem

Theorem 21 (de Finetti)

Let $\{X_n\}$ be an exchangeable sequence. Then there exists a σ -algebra $\mathcal{S}_0 \subseteq \mathcal{S}_\infty$ s.t. X_j 's are iid conditionally on \mathcal{S}_0 .

Example 22 (Kernels having product form)

For any bounded measurable kernels of the form

$$h(x_1, \dots, x_m) = \prod_{i=1}^m h_i(x_i)$$

Then for exchangeable sequence $\{X_n : n \geq 1\}$, we have :

$$\lim_n \mathbb{U}_{n,m}(h) = \lim_n \mathbb{E}(h(X_1, \dots, X_m) | \mathcal{S}_n) \rightarrow \prod_{i=1}^m \mathbb{E}(h_i(X_i) | \mathcal{S}_\infty)$$

Polya's Urn Model

An urn contains initially b black balls and r red balls. Balls are drawn at random and after each draw, the selected ball is returned to the urn with c additional balls of the same color.

Example 23 (Polya's urn)

Let A_i be the event that i^{th} draw results in a black ball. Then the random variables $1_{A_i}(\omega)$'s are exchangeable. Define

$$\mathbb{U}_{n,m}(h) = \frac{(n-m)!}{n!} \sum_{i_1 \neq \dots \neq i_m}^n 1_{A_{i_1}} \cdots 1_{A_{i_m}}$$

By reversed time martingale convergence theorem, we have

$$\mathbb{U}_{n,m} \rightarrow_{a.s., L_1} Z^m = [\mathbb{E}(1_{A_1} | \mathcal{S}_\infty)]^m$$

Therefore, we have $Z \sim \text{Beta}(\frac{b}{c}, \frac{r}{c})$ since the dist. of a bounded RV is completely determined by its moments.

Thank you !

Next time:

Martingale theory with applications to laws of large numbers.

Weak convergence.

Concentration inequalities.