Probability and Stochastic Processes II - Lecture 5d

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/stac62/staC6320242.html

2024

- what seems like a slightly more general definition of a martingale: the s.p. $\{Y_n : n \geq 0\}$ is a *martingale* with respect to the s.p. $\{X_n : n \geq 0\}$ whenever $Y_n : (\Omega, \mathcal{A}_{X_0,\ldots,X_n}) \to (R^1, \mathcal{B}^1), E(|Y_n|) < \infty$ and

$$E(Y_{n+1} \mid X_0, \ldots, X_n) = Y_n$$

for every $n \in \mathbb{N}_0$.

- certainly the previous definition of a martingale satisfies this where $\{Y_n : n \geq 0\}$ is just the same process as $\{X_n : n \geq 0\}$, i.e., $\{(n, Y_n) : n \geq 0\} = \{(n, X_n) : n \geq 0\}$

- note the requirement $Y_n : (\Omega, \mathcal{A}_{X_0,\ldots,X_n}) \to (R^1, \mathcal{B}^1)$ is just saying that Y_n can be thought of as a function of $(X_0, \ldots, X_n)$ and recall $\mathcal{A}_{X_0,\ldots,X_n} \subset \mathcal{A}$ so Y_n is a valid r.v.

- if T is a stopping time for $\{X_n : n \geq 0\}$ then it is not necessarily true that T is a stopping time for $\{Y_n : n \geq 0\}$ since $\mathcal{A}_{Y_0,\ldots,Y_n} \subset \mathcal{A}_{X_0,\ldots,X_n}$ but we can still consider Y_T as, via the same argument used in Prop. V.7 (just replace the X's by Y's), Y_T is a r.v.

- in fact we can define $\{Y_n : n \geq 0\}$ as a martingale according to Definition V.3 but for the stopping time results we only need that T be a stopping time for the underlying stochastic process $\{X_n : n \geq 0\}$

- this follows because

$$\begin{aligned} & E(Y_{n+1} \mid Y_0, \ldots, Y_n) \overset{*}{=} E\left(E(Y_{n+1} \mid X_0, \ldots, X_n) \mid Y_0, \ldots, Y_n\right) \\ = \ & E(Y_n \mid Y_0, \ldots, Y_n) = Y_n \end{aligned}$$

where for * we use the general definition of conditional expectation provided previously

Definition V.2 For random variable Y with $E(|Y|) < \infty$ and sub σ-algebra $\mathcal{C} \subset \mathcal{A}$, then $E(Y \mid \mathcal{C})$ is defined as the unique function satisfying

$$\begin{aligned} &\text{(i) } E(Y \mid \mathcal{C}) : \quad (\Omega, \mathcal{C}) \to (R^1, \mathcal{B}^1), \\ &\text{(ii) } E(HY) = \quad E(HE(Y \mid \mathcal{C})) \text{ for every } H : (\Omega, \mathcal{C}) \to (R^1, \mathcal{B}^1) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{s.t. } E(|HY|) < \infty. \end{aligned}$$

and the following result

Proposition V.14 Suppose $\mathcal{C}, \mathcal{D}$ are sub σ-algebras of $\mathcal{A}$ (i) when $\mathcal{C} \subset \mathcal{D}$ then $E(E(Y \mid \mathcal{C}) \mid \mathcal{D}) = E(Y \mid \mathcal{C})$ and (ii) $E(E(Y \mid \mathcal{D}) \mid \mathcal{C}) = E(Y \mid \mathcal{C})$.

- **notes**

1. All equalities hold *wp*1.

2. This generalizes the property of conditional probability

$$P_{(\cdot \mid B)}(A \mid C) = P(A \mid B \cap C)$$

where $P_{(\cdot \mid B)} = P(\cdot \mid B)$

3. This gives TTE for conditional expectations, namely, if $\mathcal{C} \subset \mathcal{D}$, then $E(Y \mid \mathcal{C}) = E(E(Y \mid \mathcal{D}) \mid \mathcal{C})$ since $\mathcal{C} = \mathcal{C} \cap \mathcal{D}$ (recall $E(Y \mid \{\phi, \Omega\}) = E(Y)$).

Proof: (i) Suppose that $\mathcal{C} \subset \mathcal{D}$. Then $E(E(Y \mid \mathcal{C}) \mid \mathcal{D}) = E(Y \mid \mathcal{C})$ since, if $K : (\Omega, \mathcal{D}) \to (R^1, \mathcal{B}^1)$, then $E(K \mid \mathcal{D}) = K$ and

$$E(Y \mid \mathcal{C}) : (\Omega, \mathcal{C}) \to (R^1, \mathcal{B}^1)$$

implies $E(Y \mid \mathcal{C}) : (\Omega, \mathcal{D}) \to (R^1, \mathcal{B}^1)$ as $\mathcal{C} \subset \mathcal{D}$.

(ii) We require $E(E(Y \mid \mathcal{D}) \mid \mathcal{C}) : (\Omega, \mathcal{C}) \to (R^1, \mathcal{B}^1)$ and $E(HE(Y \mid \mathcal{D})) = E(HE(E(Y \mid \mathcal{D}) \mid \mathcal{C}))$ for every $H : (\Omega, \mathcal{C}) \to (R^1, \mathcal{B}^1)$. But now $E(HE(Y \mid \mathcal{D})) = E(HY)$ because $H : (\Omega, \mathcal{D}) \to (R^1, \mathcal{B}^1)$ as $\mathcal{C} \subset \mathcal{D}$. Also $E(HY) = E(HE(Y \mid \mathcal{C}))$ by definition and so we have

$$E(HE(E(Y \mid \mathcal{D}) \mid \mathcal{C})) = E(HE(Y \mid \mathcal{C}))$$

for every $H : (\Omega, \mathcal{C}) \to (R^1, \mathcal{B}^1)$ and we can conclude that $E(Y \mid \mathcal{C}) = E(E(Y \mid \mathcal{D}) \mid \mathcal{C})$ since both are r.v.'s wrt $\mathcal{C}$. ■

- **note** - when $\mathcal{C} \subset \mathcal{D}$ we can write

$$E(E(Y \mid \mathcal{C}) \mid \mathcal{D}) = E(E(Y \mid \mathcal{D}) \mid \mathcal{C}) = E(Y \mid \mathcal{C} \cap \mathcal{D}) = E(Y \mid \mathcal{C})$$

but this result does not hold in general without the nesting (see counterexample at the end of this lecture)

- so * follows from

$$E(Y_{n+1} \mid Y_0, \ldots, Y_n) = E(Y_{n+1} \mid \mathcal{A}_{Y_0,\ldots,Y_n})$$
$$E(Y_{n+1} \mid X_0, \ldots, X_n) = E(Y_{n+1} \mid \mathcal{A}_{X_0,\ldots,X_n})$$

since $\mathcal{C} = A_{Y_0,\ldots,Y_n} \subset \mathcal{D} = A_{X_0,\ldots,X_n}$

Martingale Convergence

Proposition V.15 *(Martingale Convergence Theorem)* If a martingale $\{X_n : n \in \mathbb{N}_0\}$ is bounded below (there exists consctant c s.t. $P(X_n > c$ for all $n) = 1)$ or is bounded above (there exists consctant c s.t. $P(X_n < c$ for all $n) = 1)$, then there exists a r.v. X s.t. $X_n \overset{wp1}{\to} X$.

Proof: Accept.

Example V.10 (*Simple symmetric random walk***)**

- if $\{X_n : n \in \mathbb{N}_0\}$ is a ssrw then it is a martingale but we know $f_{ij} = 1$ for all $i, j \in \mathbb{Z}$ so X_n doesn't converge and note that this process is not bounded below or above

- but if we consider the gambler's ruin problem we see that $X_n \overset{wp1}{\to} X$ where $P(X = 0) = a/c$ and $P(X = c) = (c - a)/c$ and note this martingale is bounded above and below ■

Example V.11

- $\{X_n : n \in \mathbb{N}_0\}$ with $X_0 \ = 1$ is a MC with statespace $\mathcal{S} = \{2^m : m \in \mathbb{Z}\}$ with $p_{i,2i} = 1/3, p_{i,i/2} = 2/3$ for $i \in \mathcal{S}$

- $E(X_{n+1} \,|\, X_0, \ldots, X_n) = \frac{1}{3}2X_n + \frac{2}{3}\frac{1}{2}X_n = \frac{2}{3}X_n + \frac{1}{3}X_n = X_n$ so this is a martingale and it is bounded below by 0

- so what is X s.t. $X_n \stackrel{wp1}{\to} X$?

- note that $Y_n = \log_2 X_n$ is a MC on $\mathbb{Z}$ with transition probabilities $q_{i,i+1} = 1/3, p_{i,i-1} = 2/3$ so in fact $Y_n = \sum_{i=0}^n Z_i$ with $Z_0 = 0$ and $Z_1, Z_2, \ldots \sim 2\text{Bernoulli}(1/3) - 1$ is a srw where$E(Z_i) = 2/3 - 1 = -1/3$

- then by the SLLN

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n} Z_i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Z_i = -\frac{1}{3}\right) = 1$$

which implies

$$P\left(\lim_{n\to\infty} Y_n = \lim_{n\to\infty}\sum_{i=1}^{n} Z_i = -\infty\right) = 1$$

and so

$$X_n = 2^{Y_n} \overset{wp1}{\longrightarrow} 0$$

and X is degenerate at 0 ■

Exercise V.13 Text 3.5.5

Exercise V.14 Text 3.5.7

Example V.12 *The Branching Process (Galton-Watson Process)*

- $\{X_n : n \in \mathbb{N}_0\}$ with $X_0 = a \in \mathbb{N}_0$ is a branching process when $X_n \in \mathbb{N}_0$ and when $X_n = m$ then

$$X_{n+1} = Z_{1,n} + \cdots + Z_{m,n}$$

where $Z_{1,n}, \ldots, Z_{m,n} \overset{i.i.d.}{\sim} P_Z$, for all m and n, where P_Z is a probablity distribution on $\mathbb{N}_0$ called the *offspring distribution*

- if there m sons of a family alive at generation n the i-th individual gives rise to $Z_{i,n}$ new individuals then disappears and concern is with whether or not the family name dies out $(X_n = 0)$

- e.g. a mass of a fissile substance where $X_n = \#$ of free neutrons capable of splitting an atom to create more free neutrons and we are interested in whether the chain reaction dies out $(X_n = 0)$, stays reasonably stable or grows in size

- clearly $\{X_n : n \in \mathbb{N}_0\}$ is a MC and it is time homgeneous with $p_{00} = 1, p_{ij} = P_Z(Z_1 + \ldots + Z_i = j)$ where $Z_1, \ldots, Z_j \overset{i.i.d.}{\sim} P_Z$

- also suppose μ is the mean of P_Z and that it is finite, then

$$E(X_{n+1} \mid X_0, \ldots, X_n) = E(Z_1 + \ldots + Z_{X_n} \mid X_0, \ldots, X_n) = \mu X_n$$
$$E(X_n) = \mu E(X_{n-1}) = \mu^2 E(X_{n-2}) = \cdots = \mu^n E(X_0) = \mu^n a$$

and so a martingale iff $\mu = 1$

- therefore, the expected size of the population satisfies

$$E(X_n) \to \begin{cases} 0 & \text{if } \mu < 1 \\ a & \text{if } \mu = 1 \\ \infty & \text{if } \mu > 1 \end{cases}$$

Case 1: $\mu < 1$

- now $E(X_n) = \sum_{i=1}^{\infty} P(X_n \geq i) \geq P(X_n \geq 1)$ and so $P(X_n \geq 1) \to 0$ when $\mu < 1$ so chain reaction stops and $P(\lim_{n \to \infty} X_n = 0) = 1$

- in general

$$P_a(X_1 = 0) = P(Z_1 = 0, \ldots, Z_a = 0) = (P(Z_1 = 0))^a = P_Z^a(\{0\}) > 0$$

when $P_Z(\{0\}) > 0$ and so $P(X_n = 0) > 0$ for every n

Case 2: $\mu > 1$

- fact: $P(\lim_{n\to\infty} X_n = \infty) > 0$ but this probability is not necessarily 1 since $P(\lim_{n\to\infty} X_n = 0) > 0$ whenever $P_Z(\{0\}) > 0$

- intuition: $W_n = X_n/\mu^n$ is a martingale bounded from below, so by Martingale Convergence Theorem, $W_n \overset{wp1}{\to} W$ for some r.v. W and if $P(W > 0) > 0$, then for any $\omega \in \{W > 0\}$ we have

$$\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} \mu^n \frac{X_n(\omega)}{\mu^n} = \infty \text{ with probability 1}$$

Case 3: $\mu = 1$

- then $\{X_n : n \in \mathbb{N}_0\}$ is a martingale bounded from below, so by the Martingale Convergence Theorem there is a r.v. X s.t. $X_n \overset{wp1}{\to} X$

- the branching process is degenerate when $P_Z(\{1\}) = 1$ (then $\mu = 1$) and $P(X = a) = 1$

- consider now the nongenerate case, namely, $P_Z(\{1\}) < 1$

- since $\mu = 1$, this implies that $0 < P_Z(\{0, 1\}) < 1$ as well

- since X_n and X are integer-valued this means $X_n(\omega) = j$ for some j for all $n > N_\omega$ for some N_ω (since X is integer-valued)

Lemma V.16 If $j > 0$ then $P(X = j) = 0$ whenever $P_Z(\{0, 1\}) < 1$.

- therefore $P(X = 0) = 1$ and so $P(X_n = 0 \text{ for some } n) = 1$ and extinction is guaranteed

Proof: Using the continuity of probability measure and definition of the limit infimum of a sequence of sets

$$
\begin{aligned}
& P(X=j) = P(X_{n+1} = X_n = j \text{ for all } n \text{ large enough}) \\
= \;& P(Z_{1,n} + \cdots + Z_{j,n} = j \text{ for all } n \text{ large enough}) \\
= \;& P(\liminf_{n\to\infty} \{Z_{1,n} + \cdots + Z_{j,n} = j\}) \\
= \;& P(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \{Z_{1,k} + \cdots + Z_{j,k} = j\}) \\
= \;& \lim_{n\to\infty} P(\cap_{j=n}^{\infty} \{Z_{1,k} + \cdots + Z_{j,k} = j\}) \\
= \;& \lim_{n\to\infty} P(\lim_{m\to\infty} \cap_{k=n}^{n+m} \{Z_{1,k} + \cdots + Z_{j,k} = j\}) \\
= \;& \lim_{n\to\infty} \lim_{m\to\infty} P(\cap_{k=n}^{n+m} \{Z_{1,k} + \cdots + Z_{j,k} = j\}) \\
= \;& \lim_{n\to\infty} \lim_{m\to\infty} \prod_{k=n}^{m+n} P(\{Z_{1,k} + \cdots + Z_{j,k} = j\}) \text{ by independence} \\
= \;& 0 \text{ whenever } P(\{Z_{1,k} + \cdots + Z_{j,k} = j\}) < 1
\end{aligned}
$$

which follows from $0 < P_Z(\{0,1\}) < 1$. ■

Example V.13 *Stock Options - Discrete*

- let X_n denote the price of a stock (say BCE) at the end of trading day n

- a *European call* on a particular stock is an option to buy a stock at *strike price* K (say dollars) at a fixed future *strike time* S (some future trading day)

- we ignore commissions

- you can buy or sell such an option (called *covered* if you sell when you own the stock)

- consider the case of a buyer who pays C for the option

- if $X_S \leq K$ then the option won't be exercised and you lose C

- if $X_S > K$ then the option will be exercised and you gain $X_S - K - C$

- what price C should you buy (sell) the option?

- this is determined by the *no-arbitrage probabilities*

- an arbitrage is a situation in the financial markets where you can invest with no possibility of a loss but a possibility of a gain

- it is a basic principle of finance that such opportunities will exist only momentarily and quickly disappear (free money with no risk) so a market with no arbitrage opportunities is an idealization with practical relevance

- consider a portfolio consisting of x shares of the stock and y calls (x or y can be negative called *shorting*) and suppose the stock can take on only the values $X_S \in \{u, d\}$ at fixed time S where $d < a, K < u$ and $X_0 = a$

- for an arbitrage (with the same profit if $X_S = u$ or $X_S = d$) we must have

$$x(u-a) + y(u-K-C) = x(d-a) - yC \text{ so}$$
$$x(u-d) + y(u-K) = 0 \text{ and } y = -x\frac{u-d}{u-K}$$

which implies the net profit is

$$x(d-a) + x\frac{u-d}{u-K}C$$

- no-arbitrage requres this to be 0 so

$$0 = (d - a) + \frac{u - d}{u - K} C \text{ or}$$

$$C = (a - d)\frac{u - K}{u - d} = (u - K)\left(\frac{a - d}{u - d}\right) + 0\left(1 - \frac{a - d}{u - d}\right)$$

- the *no-arbitrge probabilities* are given by

$$\frac{a - d}{u - d}, 1 - \frac{a - d}{u - d} = \frac{u - a}{u - d}$$

where $\frac{a-d}{u-d}$ is the no-arbitrage probability that the price of the stock rises and so under these probabilities

$$E(X_S) = u\frac{a - d}{u - d} + d\frac{u - a}{u - d} = a$$

so the stock price is a martingale and the expected value of the option at time S is

$$(u - K)\frac{a - d}{u - d} + 0\frac{u - a}{u - d} = C$$

and so is also a martingale

- this analysis is for a single period (only prices X_0, X_S) and two prices for the stock but it can be generalized and always the concept of a martingale arising from the principle of no-arbitrge plays a role

■

Exercise V.15 Text 3.6.3

Exercise V.16 Text 3.7.9

Example V.14 Counterexample, when $\mathcal{C}, \mathcal{D}$ aren't nested, to

$$E(E(Y \mid \mathcal{C}) \mid \mathcal{D}) = E(E(Y \mid \mathcal{D}) \mid \mathcal{C}) = E(Y \mid \mathcal{C} \cap \mathcal{D})$$

- let $\Omega = \{a, b, c\}, \mathcal{A} = 2^{\Omega}, P =$ uniform, $Y(a) = 1, Y(b) = 2, Y(c) = 3$

$$\mathcal{C} = \{\phi, \Omega, \{a\}, \{b, c\}\}, \mathcal{D} = \{\phi, \Omega, \{b\}, \{a, c\}\}, \mathcal{C} \cap \mathcal{D} = \{\phi, \Omega\}$$

so $E(Y \mid \mathcal{C} \cap \mathcal{D}) = E(Y) = 2$, and

ω	$E(Y \mid \mathcal{C})(\omega)$	$E(Y \mid \mathcal{D})(\omega)$	$E(E(Y \mid \mathcal{C}) \mid \mathcal{D})(\omega)$	$E(E(Y \mid \mathcal{D}) \mid \mathcal{C})(\omega)$
a	1	$\frac{1}{2}1 + \frac{1}{2}3 = 2$	$\frac{1}{2}1 + \frac{1}{2}\frac{5}{2} = \frac{9}{4}$	2
b	$\frac{1}{2}2 + \frac{1}{2}3 = \frac{5}{2}$	2	$\frac{5}{2}$	$\frac{1}{2}2 + \frac{1}{2}2 = 2$
c	$\frac{5}{2}$	2	$\frac{9}{4}$	2

Understanding conditioning on a sigma algebra and conditioning in general. (optional)

- suppose we have a probability model $(\Omega, \mathcal{A}, P)$

- also, we have an *information processor Info* s.t. when ω occurs then $Info(\omega)$ prescribes the truth value of every event in the σ-algebra $\mathcal{C} \subset \mathcal{A}$

- as such all expectations (and thus probabilities) should be based on the original probability model and the information $Info(\omega)$

- so, for example, our belief that $A \in \mathcal{A}$ is true changes to $P(A \,|\, \mathcal{C})(\omega) = E(I_A \,|\, \mathcal{C})(\omega)$

Example - if $\mathcal{C} = \{\phi, C, C^c, \Omega\}$ then

$$P(A \,|\, \mathcal{C})(\omega) = \begin{cases} P(A \,|\, C) & \text{if} \quad \omega \in C \\ P(A \,|\, C^c) & \text{if} \quad \omega \in C^c \end{cases}$$

■

- suppose there are two information processors $Info_{\mathcal{C}}$ and $Info_{\mathcal{D}}$ labelled by their corresponding σ-algebras

- if we are told $Info_{\mathcal{C}}(\omega)$ and $Info_{\mathcal{D}}(\omega)$, then what information does this correspond to?

- note - if we know the truth value of every event in a class of sets $\mathcal{C}^*$, then we also know the truth value of every element of the σ-algebra $\mathcal{C} = \sigma(\mathcal{C}^*)$ $=$ the smallest σ-algebra containing $\mathcal{C}^*$ (the σ-algebra generated by $\mathcal{C}^*$)

- so when there are two information processors $Info_{\mathcal{C}}$ and $Info_{\mathcal{D}}$ this corresponds to the information processor $Info_{\sigma(\mathcal{C}\cup\mathcal{D})}$ where $\sigma(\mathcal{C}\cup\mathcal{D})$ is the smallest σ-algebra containing $\mathcal{C}$ and $\mathcal{D}$

Example (continued) - suppose in addition to $Info_{\mathcal{C}}(\omega)$ we are told the value of $Info_{\mathcal{D}}(\omega)$ where $\mathcal{D} = \{\phi, D, D^c, \Omega\}$

- then $\sigma(\mathcal{C}\cup\mathcal{D}) = \sigma(\{C \cap D, C \cap D^c, C^c \cap D, C^c \cap D^c\})$ where

$$\{C \cap D, C \cap D^c, C^c \cap D, C^c \cap D^c\}$$

is the partition of Ω generated by C and D

- then

$$P(A\,|\,\sigma(\mathcal{C}\cup\mathcal{D}))(\omega)=\begin{cases} P(A\,|\,C\cap D) & \text{if} \quad \omega\in C\cap D \\ P(A\,|\,C\cap D^c) & \text{if} \quad \omega\in C\cap D^c \\ P(A\,|\,C^c\cap D) & \text{if} \quad \omega\in C^c\cap D \\ P(A\,|\,C^c\cap D^c) & \text{if} \quad \omega\in C^c\cap D^c \end{cases}$$

■

- the fundamental *principle of conditional probability*: when $\mathcal{C}$ contains a finest partition of Ω, then observing $Info_{\mathcal{C}}(\omega)$ means we must condition on the true element of this partition

- note the Borel sets $\mathcal{B}^1$ contain the partition $\{\{x\}:x\in\mathbb{R}^1\}$ but $\sigma(\{\{x\}:x\in\mathbb{R}^1\})\neq\mathcal{B}^1$

- but for r.v. $X:(\Omega,\mathcal{A})\rightarrow(\mathbb{R}^1,\mathcal{B}^1)$ then $\mathcal{A}_X$ contains the finest partition $\{X^{-1}\{x\}:x\in\mathbb{R}^1\}$

note - the proposition about iterated conditional expectations when $\mathcal{C}$ and $\mathcal{D}$ are nested is a result about *averages* (which is what an expectation is) and is not about a principle of inference (which is what conditional probability is)
- so, for example, when $\mathcal{C} \subset \mathcal{D}$, then the information in $Info_{\mathcal{C}}(\omega)$ is contained in the information $Info_{\mathcal{D}}(\omega)$ and the principle of conditional probability says that the correct conditional expectations are given by $E(Y \mid \mathcal{D}) \neq E(Y \mid \mathcal{C})$ and Prop. V.14 only says that

$$E(E(Y \mid \mathcal{C}) \mid \mathcal{D}) = E(E(Y \mid \mathcal{D}) \mid \mathcal{C}) = E(Y \mid \mathcal{C})$$