Probability and Stochastic Processes II - Lecture 5a

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/stac62/staC632024.html

2024

Martingales

V.1 Review of Conditional Expectation and Probability

- recall for probability model $(\Omega, \mathcal{A}, P)$ and $A, C \in \mathcal{A}$ with $P(C) > 0$

$$P(A \,|\, C) = \frac{P(A \cap C)}{P(C)}$$

- for random vector $\mathbf{X}$ we want to calculate things like

$$P(\mathbf{X} \in B \,|\, \mathbf{Y} = \mathbf{y}) \text{ and } E(\mathbf{X} \,|\, \mathbf{Y} = \mathbf{y})$$

where $\mathbf{Y} = T(\mathbf{X})$ for some "smooth" $T : R^k \to R^l$

- this requires defining conditional probability and expectation when $P(\mathbf{Y} = \mathbf{y}) = 0$, at least in the absolutely continuous case

- also see Lectures 12 and 20 from STAC62

- this can typically be done via limits as in defining the conditional pdf of $\mathbf{X}$ at $\mathbf{x}$ given $\mathbf{Y} = \mathbf{y}$ by, when $\mathbf{x} \in T^{-1}\{\mathbf{y}\}$,

$$
\begin{aligned}
f_{\mathbf{X}\,|\,\mathbf{Y}}(\mathbf{x}\,|\,\mathbf{y}) &= \lim_{\delta_1 \downarrow 0, \delta_2 \downarrow 0} \left\{ \frac{P_{\mathbf{X}}(B_{\delta_1}(\mathbf{x}) \cap T^{-1}B_{\delta_2}(\mathbf{y}))}{Vol(B_{\delta_1}(\mathbf{x}) \cap T^{-1}B_{\delta_2}(\mathbf{y}))} \Big/ \frac{P_{\mathbf{Y}}(B_{\delta_2}(\mathbf{y}))}{Vol(B_{\delta_2}(\mathbf{y}))} \right\} \\
&\overset{\text{fact}}{=} \frac{f_{\mathbf{X}}(\mathbf{x}) J_T(x)}{f_{\mathbf{Y}}(\mathbf{y})}
\end{aligned}
$$

for $\mathbf{x} \in T^{-1}\{\mathbf{y}\}$ where (now allowing T to be many to one)

$$
J_T(\mathbf{x}) = \left| \det \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial T_l(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_l(\mathbf{x})}{\partial x_k} \end{pmatrix} \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial T_l(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_l(\mathbf{x})}{\partial x_k} \end{pmatrix}' \right|^{-1/2}
$$

- in the discrete case $J_T(\mathbf{x}) \equiv 1$, $f_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$ and $f_{\mathbf{Y}}(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y}) = P(T^{-1}\{\mathbf{y}\})$ so $f_{\mathbf{X}\,|\,\mathbf{Y}}(\mathbf{x}\,|\,\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})/f_{\mathbf{Y}}(\mathbf{y})$

Example V.1 *Projections*

- if $T(x_1, \ldots, x_k) = (x_1, x_2)$ then $l = 2$

$$J_T(\mathbf{x}) = \left| \det \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \frac{\partial T_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T2(\mathbf{x})}{\partial x_k} \end{pmatrix} \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \frac{\partial T_2(\mathbf{x})}{\partial x_1} \\ \vdots & \vdots \\ \frac{\partial T_1(\mathbf{x})}{\partial x_k} & \frac{\partial T_2(\mathbf{x})}{\partial x_k} \end{pmatrix} \right|^{-1/2}$$

$$= \left| \det \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \right|^{-1/2} = \left| \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right|^{-1/2} = 1$$

- also $f_{(X_1,X_2)}(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(X_1,\ldots,X_k)}(x_1, \ldots, x_k)\, dx_3 \cdots dx_k$ so

$$f_{(X_3,\ldots,X_k)\,|\,(X_1,X_2)}(x_3, \ldots, x_k \,|\, x_1, x_2) = \frac{f_{(X_1,\ldots,X_k)}(x_1, \ldots, x_k)}{f_{(X_1,X_2)}(x_1, x_2)}$$

■

Example V.2 *Projection conditionals of the* $N_k(\boldsymbol{\mu}, \Sigma)$

- suppose $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{X}_1 = T(\mathbf{X}) = (X_1, \ldots, X_l)^t$ for $l \leq k$

- partition $\boldsymbol{\mu}$ and Σ as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \text{ where } \boldsymbol{\mu}_1 \in R^l, \boldsymbol{\mu}_2 \in R^{k-l}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^t & \Sigma_{22} \end{pmatrix} \text{ where } \begin{matrix} \Sigma_{11} \in R^{l \times l}, & \Sigma_{12} \in R^{l \times (k-l)}, \\ & \Sigma_{22} \in R^{(k-l) \times (k-l)} \end{matrix}$$

$$\mathbf{X}_2 \,|\, \mathbf{X}_1 = \mathbf{x}_1 \sim N_{k-l}(\boldsymbol{\mu}_2 + \Sigma_{12}' \Sigma_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{12}' \Sigma_{11}^{-1} \Sigma_{12})$$

$$E(\mathbf{X}_2 \,|\, \mathbf{X}_1)(\mathbf{x}_1) = \boldsymbol{\mu}_2 + \Sigma_{12}' \Sigma_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1)$$

■

- in general, for r.v. Y where $E(|Y|) < \infty$ and random vector $\mathbf{X}$

$$E_{p_{Y\,|\,\mathbf{X}}}(Y\,|\,\mathbf{X})(\mathbf{x}) = \sum_y y p_{Y\,|\,\mathbf{X}}(y\,|\,\mathbf{x}) \text{ discrete case}$$

$$E_{f_{Y\,|\,\mathbf{X}}}(Y\,|\,\mathbf{X})(\mathbf{x}) = \int_{-\infty}^{\infty} y f_{Y\,|\,\mathbf{X}}(y\,|\,\mathbf{x})\, d\mathbf{x} \text{ absolutely continuous case}$$

- note

$$\begin{aligned}
\sum_y |y| p_{Y\,|\,\mathbf{X}}(y\,|\,\mathbf{x}) &= \sum_y |y| \frac{p_{(\mathbf{X},Y)}(\mathbf{x},y)}{p_{\mathbf{X}}(\mathbf{x})} \\
&= \frac{1}{p_{\mathbf{X}}(\mathbf{x})} \sum_{y:p_{(\mathbf{X},Y)}(\mathbf{x},y)>0} |y| p_{(\mathbf{X},Y)}(\mathbf{x},y) \\
&\le \frac{1}{p_{\mathbf{X}}(\mathbf{x})} \sum_{(\mathbf{z},y)} |y| p_{(\mathbf{X},Y)}(\mathbf{z},y) = \frac{1}{p_{\mathbf{X}}(\mathbf{x})} E(|Y|) < \infty
\end{aligned}$$

and similarly for a.c. case so conditional expectation is defined and finite

- recall that expectations for random vectors are defined coordinate-wise and probabilities can be obtained via expectations of indicator functions $P(A) = E(I_A)$

- what about conditional probability and expectations in general (not confined to being purely discrete or purely absolutely continuous or for infinite dimensional objects like some stochastic processes)?

- it is customary to give the general definition of conditional expectation based on providing its characterizing properties and we do this for random variables while conditional expectations of random vectors are obtained coordinate-wise

- think of $E_{p_{Y\,|\,\mathbf{X}}}(Y\,|\,\mathbf{X}) : (R^k, \mathcal{B}^k) \to (R^1, \mathcal{B}^1)$ (fact) and then define $E(Y\,|\,\mathbf{X}) : \Omega \to R^1$ by

$$E(Y\,|\,\mathbf{X})(\omega) = E_{P_{Y\,|\,\mathbf{X}}}(Y\,|\,\mathbf{X})(\mathbf{X}(\omega))$$

- now recall $\mathcal{A}_{\mathbf{X}} = \{A \in \mathcal{A} : A = \mathbf{X}^{-1}B \text{ for some } B \in \mathcal{B}^k\}$ which is a σ-algebra called the *σ-algebra on Ω generated by* $\mathbf{X}$

Proposition V.1 $E(Y\,|\,\mathbf{X}) : (\Omega, \mathcal{A}_{\mathbf{X}}) \to (R^1, \mathcal{B}^1)$.

Proof: For $B \in \mathcal{B}^1$ we have $(E(Y\,|\,\mathbf{X}))^{-1} B = \left(E_{P_{Y\,|\,\mathbf{X}}}(Y\,|\,\mathbf{X}) \circ \mathbf{X}\right)^{-1} B = \mathbf{X}^{-1}\left(E_{P_{Y\,|\,\mathbf{X}}}(Y\,|\,\mathbf{X})\right)^{-1} B$ and $\left(E_{P_{Y\,|\,\mathbf{X}}}(Y\,|\,\mathbf{X})\right)^{-1} B \in \mathcal{B}^k$ and since $\mathbf{X} : (\Omega, \mathcal{A}) \to (R^k, \mathcal{B}^k)$ this gives the result. ■

Definition V.1 For random variable Y with $E(|Y|) < \infty$ and random vector $\mathbf{X}$, then $E(Y \,|\, \mathbf{X})$ is defined as the unique function satisfying

$$\text{(i)} \qquad E(Y \,|\, \mathbf{X}) : (\Omega, \mathcal{A}_{\mathbf{X}}) \to (R^1, \mathcal{B}^1),$$

$$\text{(ii)} \qquad E(Yh(\mathbf{X})) = E(h(\mathbf{X})E(Y \,|\, \mathbf{X})) \text{ for every } h : (R^k, \mathcal{B}^k) \to (R^1, \mathcal{B}^1)$$

$$\text{s.t. } E(|Yh(\mathbf{X})|) < \infty.$$

- there is the need to prove that in general there exists a r.v. satisfying Definition V.1 but we will assume this is the case and then prove it is unique *wp*1 (both follow from the Radon-Nikodym Theorem which requires more measure theory)

- note we proved for the discrete case (**Proposition III.8.1** in STAC62)

If $h : (R^k, \mathcal{B}^k) \to (R^1, \mathcal{B}^1)$ is s.t. $E(|Yh(\mathbf{X})|) < \infty$, then $E(Yh(\mathbf{X})) = E(h(\mathbf{X})E(Y \mid \mathbf{X}))$.

Proof:

$$E(Yh(\mathbf{X})) = \sum_{(\mathbf{x},y)} yh(\mathbf{x})p_{(\mathbf{X},Y)}(\mathbf{x},y) = \sum_{(\mathbf{x},y)} yh(\mathbf{x})p_{\mathbf{X}}(\mathbf{x})\frac{p_{(\mathbf{X},Y)}(\mathbf{x},y)}{p_{\mathbf{X}}(\mathbf{x})}$$

$$= \sum_{(\mathbf{x},y)} yh(\mathbf{x})p_{\mathbf{X}}(\mathbf{x})p_{Y \mid \mathbf{X}}(y \mid \mathbf{x}) = \sum_{\mathbf{x}} h(\mathbf{x})\left(\sum_{y} yp_{Y \mid \mathbf{X}}(y \mid \mathbf{x})\right)p_{\mathbf{X}}(\mathbf{x})$$

$$= \sum_{\mathbf{x}} h(\mathbf{x})E_{p_{Y \mid \mathbf{X}}}(Y \mid \mathbf{X})(\mathbf{x})p_{\mathbf{X}}(\mathbf{x}) = E(h(\mathbf{X})E(Y \mid \mathbf{X})). \blacksquare$$

Exercise V.1 Establish the above result for the absolutely continuous case.

Proposition V.2 If f and g satisfy Definition V.1 then $P(\{\omega : f(\omega) \neq g(\omega)\}) = 0$ or $f \overset{wp1}{=} g$.

- in other words $f : (\Omega, \mathcal{A}_{\mathbf{X}}) \to (R^1, \mathcal{B}^1)$ and $E(Yh(\mathbf{X})) = E(h(\mathbf{X})f)$ for every $h : (R^k, \mathcal{B}^k) \to (R^1, \mathcal{B}^1)$ s.t. $E(|Yh(\mathbf{X})|) < \infty$ implies $f \overset{wp1}{=} E(Y \,|\, \mathbf{X})$

Proof: Since $f, g : (\Omega, \mathcal{A}_{\mathbf{X}}) \to (R^1, \mathcal{B}^1)$ then

$$\begin{aligned} A_+ &= \{\omega : f(\omega) - g(\omega) > 0\} \in \mathcal{A}_{\mathbf{X}} \\ A_- &= \{\omega : f(\omega) - g(\omega) < 0\} \in \mathcal{A}_{\mathbf{X}}. \end{aligned}$$

Since $A_+ \in \mathcal{A}_{\mathbf{X}}$ there exists $B_+ \in \mathcal{B}^k$ s.t. $A_+ = \mathbf{X}^{-1} B_+$.Therefore, if $h(\mathbf{X}) = I_{B_+}(\mathbf{X})$, then

$$E(Yh(\mathbf{X})) = E(h(\mathbf{X})f) = E(h(\mathbf{X})g)$$

which implies so $0 = E(I_{B_+}(\mathbf{X})(f - g))$ which implies $P(A_+) = 0$ since $I_{B_+}(\mathbf{X}(\omega))(f(\omega) - g(\omega)) \geq 0$ and this is strict when $\omega \in A_+$. Similarly, $P(A_-) = 0$ which gives the result. ■

Corollary V.3 $E(Yh(\mathbf{X})\,|\,\mathbf{X}) \overset{wp1}{=} h(\mathbf{X})E(Y\,|\,\mathbf{X})$ for every $h : (R^k, \mathcal{B}^k) \to (R^1, \mathcal{B}^1)$ s.t. $E(|Yh(\mathbf{X})|) < \infty$.

Proof: Immediate since $h(\mathbf{X})E(Y\,|\,\mathbf{X})$ satisfies Definition V.1.

note - $E(Y\,|\,\mathbf{X})$ has all the properties of E as it is an expectation such as being linear and the TTE holds, namely

Corollary V.4 (*Theorem of Total Expectation*) For random vector $(\mathbf{X}, Y)$ such that $E(|Y|) < \infty, E(Y) = E(E(Y\,|\,\mathbf{X}))$.

Proof: Put $h(\mathbf{x}) \equiv 1$. ■

- define conditional probability by $P(A\,|\,\mathbf{X}) = E(I_A\,|\,\mathbf{X})$ for $A \in \mathcal{A}$

Corollary V.5 (*Theorem of Total Probability*) If $A \in \mathcal{A}$, then $P(A) = E(P(A\,|\,\mathbf{X}))$.

- somerimes $E(Y \mid \mathbf{X})$ is denoted $E(Y \mid \mathcal{A}_{\mathbf{X}})$ and this leads to a general definition of the conditional expecation of a random variable given a sub σ-algebra $\mathcal{C} \subset \mathcal{A}$

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

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

- in Definition V.1 put $H(\omega) = h(X(\omega))$ and $\mathcal{C} = \mathcal{A}_{\mathbf{X}}$

- this can be applied to define $E(Y \mid \{X_t : t \in T\})$ the conditional expectation of Y given the process $\{X_t : t \in T\}$ by putting

$$\mathcal{C} = \mathcal{A}_{\{X_t : t \in T\}}$$

the σ-algebra generated by the stochastic process

V.2 Martingales

Definition V.3 A stochastic process $\{X_n : n \in \mathbb{N}_0\}$, where $E|X_n| < \infty$ and $E(X_{n+1} \,|\, X_0, \ldots, X_n) = X_n$ for every $n \in \mathbb{N}_0$, is a *martingale*.

Proposition V.6 If $\{X_n : n \in \mathbb{N}_0\}$ is a martingale and $m \leq n$, then $E(X_{n+1} \,|\, X_0, \ldots, X_m) = X_m$ and so $E(X_n) = E(X_0)$ for all n.

Proof: We have for every measureable h

$$\begin{aligned} & E(h(X_0, \ldots, X_{n-1})X_{n+1}) \\ = \;& E(h(X_0, \ldots, X_{n-1})E(X_{n+1} \,|\, X_0, \ldots, X_{n-1})) \text{ and} \\ & E(h(X_0, \ldots, X_{n-1})X_{n+1}) = E(h(X_0, \ldots, X_{n-1})E(X_{n+1} \,|\, X_0, \ldots, X_n)) \\ = \;& E(h(X_0, \ldots, X_{n-1})X_n) = E(h(X_0, \ldots, X_{n-1})E(X_n \,|\, X_0, \ldots, X_{n-1})) \end{aligned}$$

so $E(X_{n+1} \,|\, X_0, \ldots, X_{n-1}) = E(X_n \,|\, X_0, \ldots, X_{n-1}) = X_{n-1}$ which establishes the result. ■

- note - the definition can can also be written as follows: a stochastic process $\{X_n : n \in \mathbb{N}_0\}$, where $E|X_n| < \infty$ and $E(X_{n+1} \,|\, \mathcal{A}_{X_0,\ldots,X_n}) = X_n$ for every $n \in \mathbb{N}_0$, is a martingale

- we have that $\mathcal{A}_{X_0,\ldots,X_n} \subset \mathcal{A}_{X_0,\ldots,X_{n+1}} \subset \mathcal{A}$, so each $\mathcal{A}_{X_0,\ldots,X_n}$ is a sub σ-algebra of $\mathcal{A}$

- generally a set $\{\mathcal{A}_t; t \in T\}$ of sub σ-algebras of $\mathcal{A}$, with T totally ordered, is a *filtration* when $\mathcal{A}_s \subset \mathcal{A}_t$ when $s \le t$

- then a stochastic process $\{X_t; t \in T\}$ where $X_t : (\Omega, \mathcal{A}_t) \to (\mathbb{R}, \mathcal{B}^1)$, $E|X_t| < \infty$ and $E(X_t \,|\, \mathcal{A}_s) = X_s$ for every $s \le t$, is a *martingale with respect to the filtration* $\{\mathcal{A}_t; t \in T\}$

Example V.3 *When is a Markov chain a martingale?*

- if $\{X_n : n \in \mathbb{N}_0\}$ is also a MC with $S \subset \mathbb{Z}$ and the distribution of the state at time n has finite expectation for every n, then to be a martingale we must have

$$\begin{aligned} & E(X_{n+1} \,|\, X_0, \ldots, X_n)(i_0, \ldots, i_n) = E(X_{n+1} \,|\, X_n)(i_n) \\ = \; & \sum_j jP(X_{n+1} = j \,|\, X_n)(i_n) = i_n \end{aligned}$$

with time homogeneity this becomes

$$E(X_1 \,|\, X_0)(i) = \sum_i iP(X_1 \,|\, X_0)(j) = \sum_j jp_{ij} = i$$

- for example, a srw has $X_n = \sum_{i=0}^n Z_i$ with $Z_0 = 0$ and $Z_1, Z_2, \ldots \overset{i.i.d.}{\sim} 2\text{Bernoulli}(p) - 1$

$$E(X_1 \mid X_0)(i) = (i-1)(1-p) + (i+1)p = i + 2p - 1$$

and so is a martingale only when $p = 1/2$, namely, it is a ssrw ■

Exercise V.2 Text 3.1.6 shows converse to Proposition V.6 is false.

Example *The martingale gambling strategy*

- consider a game of coin tossing where a gambler bets on H which occurs with probability $1/2$, and if the gambler bets $\$x$ the payoff is $\$2x$ so the expected gain on a toss is $0.5(2x - x) - 0.5x = 0$

- the gambler adopts the following strategy: they bet $\$1$ on the first toss, if they lose this bet they bet $\$2$ on the next toss, if they lose this bet they bet $\$4$ on the next toss and generally if they lose the first n bets they bet $\$2^n$ on the next bet and they stop as soon as they win which happens with probability 1

- if the first H occurs at time n then gain is $2^n - (1 + 2 + \cdots + 2^{n-1}) = 2^n - 2^n + 1 = 1$ so this guarantees a profit

- but note that expected loss just before win is

$$\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n (2^n - 1) = \infty$$

so you need a big bank account if you want to use this strategy

- let X_n denote the gambler's gain (loss) at toss n

- so

$$X_{n+1} = \begin{cases} X_n & \text{if stopped by toss } n \\ X_n + 2^n & \text{if } H \text{ at toss } n \\ X_n - 2^n & \text{if } T \text{ at toss } n \end{cases}$$

- then

$$\begin{aligned} E(X_{n+1} \,|\, X_1, \ldots, X_n)(x_1, \ldots, x_n) &= x_n \text{ and so} \\ E(X_{n+1} \,|\, X_1, \ldots, X_n) &= X_n \end{aligned}$$

and $\{X_n : n = 1, 2, ...\}$ is a martingale

- **stopping times (rules)**

- an important aspect of probability theory is the concept of a stopping time

- for example, a stopping time is an important aspect of many clinical trials which proceed sequentially until some desirable (e.g. cures) or undesirable (e.g. deaths) outcomes have occurred

Definition V.4 An extended r.v. T concentrated on $\mathbb{N}_0 \cup \{\infty\}$ is a *stopping time* for stochastic process $\{X_n : n \in \mathbb{N}_0\}$ if $\{\omega : T(\omega) = n\} \in \mathcal{A}_{X_0,\ldots,X_n}$ for every n.

- the idea here is that we are observing a stochastic process $X_0, X_1, \ldots$ developing in time and at each time n a decision is made whether or not to stop observing the process based on the observed values of $X_0, X_1, \ldots, X_n$

- the requirement $\{\omega : T(\omega) = n\} \in \mathcal{A}_{X_0,\ldots,X_n}$ is simply saying that the decision is made based upon the present and past and does not involve the future (which of course hasn't been observed)

- so $\{\omega : T(\omega) = n\} = (X_0, X_1, \ldots, X_n)^{-1}B$ for some $B \in \mathcal{B}^{n+1}$ for some condition, as represented by B, on the observed values of $X_0, X_1, \ldots, X_n$

- also $\mathcal{A}_{X_0,\dots,X_n} \subset \mathcal{A}_{X_0,\dots,X_{n+1}}$ implies

$$\{\omega : T(\omega) \le n\} = \cup_{k=0}^{n} \{\omega : T(\omega) = k\} \in \mathcal{A}_{X_0,\dots,X_n}$$

Example V.4

- $T \equiv k$ for some fixed $k \in \mathbb{N}_0$ so

$$\begin{aligned} \{\omega : \quad T(\omega) = n\} &= \left\{ \begin{array}{ll} \phi & n \ne k \\ \Omega & n = k \end{array} \right. \\ &= \quad \left\{ \begin{array}{ll} (X_0, X_1, \dots, X_n)^{-1}\phi & n \ne k \\ (X_0, X_1, \dots, X_n)^{-1}\mathbb{R}^{n+1} & n = k \end{array} \right. \in \mathcal{A}_{X_0,\dots,X_n} \end{aligned}$$

since $\phi, \mathbb{R}^{n+1} \in \mathcal{B}^{n+1}$ so T is a stopping rule

- $T = \inf\{n : X_n \le x\}$ then

$$\{\omega : T(\omega) = n\} = \{\cap_{i=0}^{n-1} X_i^{-1}(x, \infty)\} \cap X_n^{-1}(-\infty, x] \in \mathcal{A}_{X_0,\dots,X_n}$$

since $X_i^{-1}(x, \infty), X_i^{-1}(-\infty, x] \in \mathcal{A}_{X_i}$ for all i so T is a stopping rule

- $T = \inf\{n : X_{n+1} \leq x\}$, then

$$\{\omega : T(\omega) = n\} = \left\{\cap_{i=0}^{n} X_i^{-1}(x, \infty)\right\} \cap X_{n+1}^{-1}(-\infty, x]$$

and this may not be in $\mathcal{A}_{X_0,\ldots,X_n}$ since we cannot guarantee that $X_{n+1}^{-1}(-\infty, x] \in \mathcal{A}_{X_0,\ldots,X_n}$ for every n unless the process was degenerate and so T is **no**t a valid stopping time because for $T = n$ to occur we have to look into the future ■

Example V.5 *A valid stopping time* T *can have* $T = \infty$ *with positive probability*

- $T = \infty$ means we never stop the process

- we have that $T = \inf\{n : X_n = x\}$ is a valid stopping time but if $P(X_n = x) = 0$ for all n then $P(T = \infty) = 1$ ■

- we restrict attention hereafter to stopping times that are finite, namely, $P(T = \infty) = 0$ because we want to consider the *stopped time variable* X_T which isn't defined when $T = \infty$

- this means we remove from Ω all those ω such that $T = \infty$ and note $\Omega \backslash \{T = \infty\} = \cup_{n=0}^{\infty}\{T = n\} \in \mathcal{A}$

Proposition V.7 When $P(T = \infty) = 0$ then X_T is a random variable.

Proof: After deleting all ω such that $T = \infty$,

$$X_T^{-1}(-\infty, c) = \cup_{n=0}^{\infty}\{T = n, X_n < c\} = \cup_{n=0}^{\infty}\{T = n\} \cap \{X_n < c\} \in \mathcal{A}.$$

■

- so concern is with the distribution of the stopped value X_T of the stochastic process

Proposition V.8 (*Optional stopping lemma*) If $\{X_n : n \in \mathbb{N}_0\}$ is a martingale and T is a bounded stopping time then $E(X_T) = E(X_0)$.

Proof: So there is a constant $m \in \mathbb{N}_0$ s.t. $P(T \leq m) = 1$. We have

$$
\begin{aligned}
& E(X_T - X_0) = E\left(\sum_{k=1}^{T}(X_k - X_{k-1})\right) \\
= \; & E\left(\sum_{k=1}^{m}(X_k - X_{k-1})I_{\{k\leq T\}}\right) = \sum_{k=1}^{m} E((X_k - X_{k-1})\, I_{\{k\leq T\}}) \\
= \; & \sum_{k=1}^{m} E((X_k - X_{k-1})\, I_{\{T\leq k-1\}^c}) = \sum_{k=1}^{m} E((X_k - X_{k-1})\,(1 - I_{\{T\leq k-1\}})) \\
= \; & \sum_{k=1}^{m} E(E((X_k - X_{k-1})\,(1 - I_{\{T\leq k-1\}}) \,|\, X_0, \ldots, X_{k-1})) \\
= \; & \sum_{k=1}^{m} E((1 - I_{\{T\leq k-1\}})E((X_k - X_{k-1}) \,|\, X_0, \ldots, X_{k-1}))
\end{aligned}
$$

since $(1 - I_{\{T\leq k-1\}})$ depends on $X_0, \ldots, X_{k-1}$ and using the definition of conditional expectation and since $E(X_k \,|\, X_0, \ldots, X_{k-1}) = X_{k-1}$ then

$$\begin{aligned} & E(X_T - X_0) \\ = & \sum_{k=1}^{m} E((1 - I_{\{T \le k-1\}})(E(X_k \,|\, X_0, \ldots, X_{k-1}) - X_{k-1})) \\ = & \sum_{k=1}^{m} E((1 - I_{\{T \le k-1\}})(X_{k-1} - X_{k-1})) = 0. \end{aligned}$$

Now $E(X_0) < \infty$ and $E(X_T - X_0) < \infty$ so
$E(X_T) = E(X_0) + E(X_T - X_0) < \infty$ which implies
$0 = E(X_T - X_0) = E(X_T) - E(X_0)$ which gives the result. ■

Example V.6

- a ssrw is a martingale and define stopping time

$$T = \min(10^{12}, \inf\{n \geq 0 : X_n = -5\})$$

Exercise V.3 Show T is a valid stopping time.

- T is bounded so Proposition V.8 applies which says $E(X_T) = E(X_0) = 0$

- but the chain is recurrent so we will hit -5 with probability 1 and almost certainly within the first 10^{12} steps starting from state 0 and so we are virtually certain (very high probability $q \approx 1$) $X_T = -5$

- so how can we have $E(X_T) = 0$?

$$\begin{aligned} & 0 = E(X_T) = qE(X_T \mid T \neq 10^{12}) + (1-q)E(X_T \mid T = 10^{12}) \\ = \quad & q(-5) + (1-q)E(X_T \mid T = 10^{12}) \text{ so} \\ & E(X_T \mid T = 10^{12}) = \frac{5q}{1-q} \end{aligned}$$

- this means that with a very small probability X_T is huge

■