Computing probabilities by conditioning

Ex Two players take turns flipping a coin.
The first player to obtain a head wins.
What is the prob. that $\underbrace{\text{the player who starts wins}}_{E}$?

Condition on the first flip

player 1 wins   player 1 flips H   or T

$$P(E) = \underbrace{P(E|H)}_{1}\,\underbrace{P(H)}_{\frac{1}{2}} + \underbrace{P(E|T)}_{}\,\underbrace{P(T)}_{\frac{1}{2}}$$

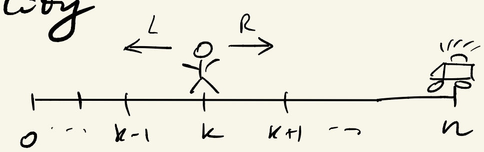
game resets, player 2 starts

$P(\text{the player who starts loses}) = 1 - P(E)$

$\Rightarrow P(E) = \frac{1}{2} + (1-P(E))\cdot\frac{1}{2}$. Solving gives $P(E) = 2/3$

Ex (Gambler's ruin) Consider a simple random walk starting at $k \in [0,n]$. What is the probability of $\underbrace{\text{reaching } n \text{ before reaching } 0}_{=: E_k}$?

payoff   bankrupcy



Condition on 1^{st} step, L or R:

$$P(E_k) = P(E_k|L)\,P(L) + P(E_k|R)\,P(R)$$
$$= P(E_{k-1})\cdot\frac{1}{2} + P(E_{k+1})\cdot\frac{1}{2}$$

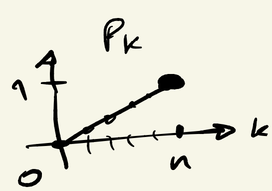
walk "resets" at $k-1$ walk "resets" at $k+1$

Denoting $p_k = P(E_k)$, we obtain

$$\begin{cases} p_k = \frac{1}{2}(p_{k-1} + p_{k+1}), & k = 1,\dots,n-1 \\ p_0 = 0; \quad p_n = 1 \end{cases}$$

$n+1$ linear equations in $n+1$ unknowns. Solve $\rightarrow$

$$\boxed{p_k = \frac{k}{n}}$$



Ex Secretary problem, a.k.a. Best prize problem

- We are presented with n prizes, in sequence.
- Upon seeing a prize, we must accept it (and end the game) or reject it (and move to the next prize). No going back.
- The only info we have at $\forall$ time is how the current prize compares to the prizes already seen.
- We want to $\underbrace{\boxed{\text{pick the best prize.}}}_{E}$ What shall we do?

- Strategy: reject the first k prizes; accept the first one that is better than those k.

Let's compute $P(E)$ and optimize k.

- Condition on the position of best prize:

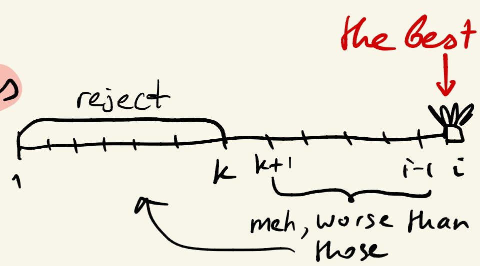
$$\boxed{B_i = \text{"}i\text{-th prize is the best"}.}$$

$$\text{L.T.P.} \Rightarrow P(E) = \sum_{i=1}^{n} \underbrace{P(E \mid B_i)}_{=?} \underbrace{P(B_i)}_{=1/n}$$

- $\forall\ i \leq k$ $\quad P(E \mid B_i) = 0$ (we reject the first k prizes)
- $\forall\ i > k$: assume B_i occurs, i.e. i-th prize is the best.

  We pick it iff all prizes $k+1, \dots, i-1$ are worse than the first k. (Otherwise we lose it)

$\Rightarrow E$ occurs $\Leftrightarrow$ the best prize among the first $i-1$ prizes is among the first k prizes.



This happens with prob. $\dfrac{k}{i-1}$

$$\Rightarrow \quad P(E \mid B_i) = \frac{k}{i-1}$$

$$\Rightarrow P(E) = \sum_{i=k+1}^{n} \frac{k}{i-1} \cdot \frac{1}{n} = \frac{k}{n} \sum_{i=k}^{n-1} \frac{1}{i} \simeq \frac{k}{n} \int_k^n \frac{dx}{x} = \frac{k}{n} \ln \frac{n}{k} = -\lambda \ln \lambda \quad \text{where } \lambda = \frac{k}{n}$$

Maximize $\Rightarrow \lambda = 1/e$, $P(E) = 1/e$ $\Rightarrow$

optional stopping

Ans Strategy = reject the first $\frac{n}{e}$ prizes, accept the first better than all rejected.
Probability to pick the best prize $= 1/e = 0.37$. Regardless of n!

Properties of Conditional Expectation

Heuristic: whatever properties expectation satisfies, conditional expectation satisfies a.s. Such as:

Prop (a) (linearity) $\mathbb{E}[aX+bY\,|\,\mathcal{F}] = a\,\mathbb{E}[X|\mathcal{F}] + b\,\mathbb{E}[Y|\mathcal{F}]$ a.s.

(b) (Monotonicity) $X \le Y$ a.s. $\Rightarrow \mathbb{E}[X|\mathcal{F}] \le \mathbb{E}[Y|\mathcal{F}]$ a.s.

(c) (MCT) If $X_n \ge 0$, $X_n \uparrow X$ a.s., $\mathbb{E}X < \infty$, then $\mathbb{E}[X_n|\mathcal{F}] \uparrow \mathbb{E}[X|\mathcal{F}]$ a.s.

(d) (Fatou) If $X_n \ge 0$, $\mathbb{E}\liminf_n X_n < \infty$ then $\mathbb{E}[\liminf_n X_n|\mathcal{F}] \le \liminf_n \mathbb{E}[X_n|\mathcal{F}]$ a.s.

(e) (DCT) If $|X_n| \le Z$, $\mathbb{E}Z < \infty$, $X_n \to X$ a.s. then $\mathbb{E}[X_n|\mathcal{F}] \to \mathbb{E}[X|\mathcal{F}]$ a.s. (and in L^1)

(f) (Jensen) If $\varphi: \mathbb{R} \to \mathbb{R}$ is convex, $\mathbb{E}|X| < \infty$, $\mathbb{E}|\varphi(X)| < \infty$, then $\varphi(\mathbb{E}[X|\mathcal{F}]) \le \mathbb{E}[\varphi(X)|\mathcal{F}]$ a.s.

Proof of (c) $Y_n := X - X_n \downarrow 0$ a.s. It suffices to show that

$$E_n := \mathbb{E}[Y_n|\mathcal{F}] \downarrow 0.$$

Monotonicity (b) $\Rightarrow E_n \downarrow$ a.s. and $E_n \ge 0$

$\Rightarrow E_n \downarrow Z$ for some r.v. $Z \ge 0$. WTS: $Z = 0$.

E_n is $\mathcal{F}$-mble $\Rightarrow Z$ is $\mathcal{F}$-mble.

$$\Rightarrow \forall F \in \mathcal{F}: \quad \mathbb{E}Z \underset{\text{usual MCT}}{=} \lim_n \mathbb{E}E_n \underset{\text{def of cond. exp. } (F=\Omega)}{=} \lim_n \mathbb{E}Y_n \overset{\text{usual MCT}}{=} 0.$$

$Z \ge 0 \Rightarrow Z = 0$. $\square$

Some properties specific to conditional expectation:

Prop (a) (law of total expectation) $\mathbb{E}\big[\mathbb{E}[X\mid\mathcal{F}]\big] = \mathbb{E}[X]$ a.s.

$\Uparrow$

(b) (towering) $\mathbb{E}\big[\mathbb{E}[X\mid\mathcal{F}_1]\mid\mathcal{F}_2\big] = \begin{cases}\mathbb{E}[X\mid\mathcal{F}_1] & \text{if } \mathcal{F}_1\subset\mathcal{F}_2\\ \mathbb{E}[X\mid\mathcal{F}_2] & \text{if } \mathcal{F}_2\subset\mathcal{F}_1\end{cases}$

(c) (independence) $X \perp \mathcal{F} \Rightarrow \mathbb{E}[X\mid\mathcal{F}] = \mathbb{E}[X]$ a.s.

(d) (conditional const) X is $\mathcal{F}$-measurable $\Rightarrow \mathbb{E}[XY\mid\mathcal{F}] = X\cdot\mathbb{E}[Y\mid\mathcal{F}]$ a.s.

Proof (a) follows from def of conditional exp. with $F=\Omega$.

(d) We need to check:

$$\mathbb{E}\big[X\cdot\mathbb{E}[Y\mid\mathcal{F}]\,\mathbb{1}_F\big] = \mathbb{E}\big[XY\,\mathbb{1}_F\big] \quad \forall F\in\mathcal{F}.$$

1. for indicators $X=\mathbb{1}_E$, $E\in\mathcal{F}$

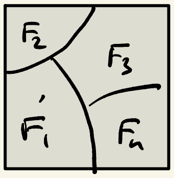
$$\mathbb{E}\big[\mathbb{E}[Y\mid\mathcal{F}]\,\mathbb{1}_{E\cap F}\big] \underset{\text{def of } \mathbb{E}[Y\mid\mathcal{F}],\ \text{since } E\cap F\in\mathcal{F}}{=\!=} \mathbb{E}\big[Y\,\mathbb{1}_{E\cap F}\big].$$

2. for simple r.v.'s X: use linearity
3. for $X, Y\geq 0$: take simple r.v.'s $X_n\uparrow X$, use CMCT (Prop (c) p.148).
4. for general integrable X, Y: decompose $X=X^+-X^-$, $Y=Y^+-Y^-$. $\square$

- Law of total expectation (p.149) for $\mathcal{F}=\sigma(F_1\sqcup\ldots\sqcup F_n)$:

$$\mathbb{E}[X] = \mathbb{E}\big[\underbrace{\mathbb{E}[X\mid\mathcal{F}]}_{\substack{\| \\ \mathbb{E}[X\mid F_i] \text{ if } F_i \text{ occurs (p.149)}}}\big] = \boxed{\sum_i \mathbb{E}[X\mid F_i]\,P(F_i)}$$

Ω

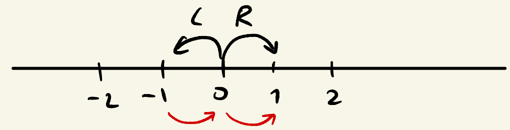


For $X=\mathbb{1}_E$, this is the same as the law of total probability (p.141)

Ex (Hitting time) For a simple random walk $S_n = X_1 + \dots + X_n$

consider T_{01} = #steps to get from 0 to 1, i.e. $T_{01} = \min\{n : S_n = 1\}$

$E[T_{01}] = ?$ Condition on 1^{st} step.



$$= \underbrace{E[T_{01} | R]}_{=1} \cdot \underbrace{P(R)}_{=1/2} + \underbrace{E[T_{01} | L]}_{=} \cdot \underbrace{P(L)}_{=1/2}$$

$E[T_{-1,1}] = E[T_{-1,0} + T_{0,1}] = E[T_{-1,0}] + E[T_{0,1}] = 2E[T_{0,1}]$ by symmetry

$\Rightarrow E[T_{0,1}] = \frac{1}{2} + E[T_{0,1}] \Rightarrow \boxed{E[T_{0,1}] = \infty}$

- In particular, expected time to return to 0 is also ∞, despite it being $< \infty$ a.s. (p.75).

SKIPPED

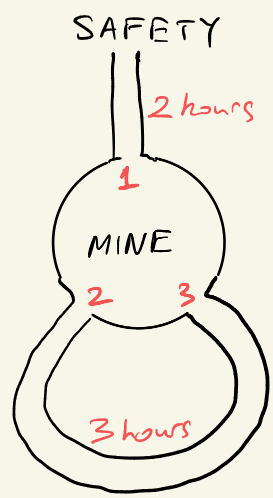
Ex2 A miner is trapped in a mine with 3 doors.

One door leads to safety in 2 hours of travel

The other two are connected by a loop, 3 hrs.

At each time, the miner chooses a door at random (and can't remember which doors were chosen)

Expected $\underbrace{\text{time to reach safety}}_{X} = ?$



Condition on the first choice of the door

$$E[X] = \underbrace{E[X|D_1]}_{=2} \underbrace{P(D_1)}_{=1/3} + \underbrace{E[X|D_2]}_{=3+E[X]} \underbrace{P(D_2)}_{=1/3} + \underbrace{E[X|D_3]}_{=3+E[X]} \underbrace{P(D_3)}_{=1/3}$$

$X|D_2$ has the same distribution as $3+X$ (after 3 hrs, situation resets)

Same for $X|D_3$.

$\Rightarrow E[X] = \frac{2}{3} + (3 + E[X]) \cdot \frac{2}{3}$. Solving $\Rightarrow$

$E[X] = 6$ hours.

Conditional expectation as a projection

- $L^2(\Omega, \Sigma, \mathbb{P}) = \{$all r.v's X such that $\mathbb{E}X^2 < \infty\}$ <u>Hilbert space</u>:

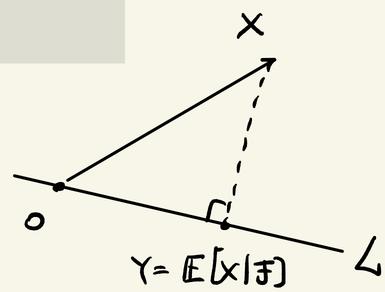
  Inner product: $\langle X, Y\rangle := \mathbb{E}XY$ Norm: $\|X\| = \sqrt{\langle X, Y\rangle}$

- Fix $\mathcal{F} \subset \Sigma$.

  $L := \{$all r.v's $X \in L^2(\Omega, \Sigma, \mathbb{P})$ that are $\mathcal{F}$-measurable$\}$

  ↑ closed linear subspace of $L^2(\Omega, \Sigma, \mathbb{P})$.

Prop (Projection) $\mathbb{E}[X|\mathcal{F}]$ is the orthogonal projection of $X \in L^2(\Omega, \Sigma, \mathbb{P})$ onto L



Proof WTS: $Y := \mathbb{E}(X|\mathcal{F}) \perp X - Y$

$$\Leftrightarrow \langle Y, X-Y\rangle = 0$$

$$\Leftrightarrow \mathbb{E}[Y(X-Y)] = 0$$

$$\Leftrightarrow \mathbb{E}[YX] = \mathbb{E}[Y^2]$$

L.T.E (Prop(a) p149) $\parallel$

$$\mathbb{E}\big[\mathbb{E}[YX|\mathcal{F}]\big] = \mathbb{E}\big[Y\,\underbrace{\mathbb{E}[X|\mathcal{F}]}_{=Y}\big] = \mathbb{E}[Y^2]. \quad ☺ \qquad \square$$

Y is $\mathcal{F}$-mble

conditional const (Prop(d) p.149)

- Orthogonal proj outputs the nearest point in L $\Rightarrow$

Cor (Best predictor) $\mathbb{E}[X|\mathcal{F}] = \underset{Y:\,\mathcal{F}\text{-mble}}{\operatorname{argmin}}\ \mathbb{E}(X-Y)^2$

HW: prove that $\mathbb{E}[\cdot|\mathcal{F}]$ is a contraction in L^p (p. 54 2019)