Note:

Say $Y = f(X_1, X_2, \dots, X_n)$, then $Z_0 = \mathbb{E}[f(X_1, X_2, \dots, X_n)] = \mathbb{E}[Y]$

no dependence on X_i

$$Z_1 = \mathbb{E}[f(X_1, X_2, \dots, X_n) \mid X_1]$$

$$\vdots$$

$$Z_n = \mathbb{E}[f(X_1, X_2, \dots, X_n) \mid X_1, \dots, X_n] = Y$$

So the $Z_0, Z_1, \dots, Z_n$ give increasingly refined estimates of the value of the R.V. $Y = f(X_1, \dots, X_n)$

$\boxed{\text{Doob Martingale, in action.}}$

ex. Throw n balls into m bins.

Let X_i = bin into which i^{th} ball lands, i.e. $X_i \in [m]$

Let Y = # of empty bins after all balls thrown, i.e. $Y = f(X_1, X_2, \dots, X_n)$

$Z_0 = \mathbb{E}[Y] = n\left(1 - \frac{1}{m}\right)^n$ ← by sum of indicators + linearity of $\mathbb{E}$

$\vdots$

$Z_i = \mathbb{E}[Y \mid X_1, \dots, X_i]$ ← = expected number of empty bins GIVEN "knowledge" of the position of the first i balls.

$\vdots$

$Z_n = Y$ ← since you know exactly the number of empty bins once all balls have bin thrown.

---

ex. Let $G \sim G_{n,p}$ be an Erdös-Reyni random graph (n vertices, edge $(v_i, v_j) \in E$ w.p. p)

Label the $m = \binom{n}{2}$ possible edges as $1, 2, \dots, m$ (in any manner) and let

$X_i = 1 \iff$ the i^{th} possible edge appears in G.

↳ $Pr(X_i = 1) = p$

Let f be any finite-valued function over graphs with n vertices. { e.g. Chromatic number, # of 3-cliques, # of independent sets, ...

Let $Z_0 = \mathbb{E}[f(G)]$, and $Z_i = \mathbb{E}[f(G) \mid X_1, \dots, X_i]$ for $0 < i \leq m$.

This is a Doob Martingale called "edge-exposure" martingale.

AZUMA-HOEFFDING

Let $Z_0, Z_1, \ldots$ be martingale w.r.t. $X_0, X_1, \ldots$ and suppose that $\exists (a_1, b_1), (a_2, b_2), \ldots$ s.t. $\forall i > 0$, $a_i \leq Z_i - Z_{i-1} \leq b_i$. Then $\forall t \geq 0$ and $\lambda > 0$

$$\Pr(|Z_t - Z_0| \geq \lambda) \leq 2\exp\left(-2\lambda^2 / \sum_{i=1}^{t} (b_i - a_i)^2\right)$$

As an example:

$$\Pr\Big(|Z_t - Z_0| \geq \varepsilon \underbrace{\sqrt{\tfrac{1}{t}\sum_{i=1}^{t}(b_i - a_i)^2}}_{\text{" root mean squared gap "}}\Big) \leq 2\exp(-2\varepsilon^2/t)$$

Consider this applied to a Doob martingale for $Y = f(X_1, \ldots, X_n)$, i.e.

$Z_0 = \mathbb{E}[f(X_1, \ldots, X_n)]$

$Z_i = \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i]$

So that $Z_n = f(X_1, \ldots, X_n)$.

Then $\Pr\big(|f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]| \geq \lambda\big) \leq 2\exp\left(\dfrac{-2\lambda^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$

where $a_i \leq Z_i - Z_{i-1} \leq b_i \Rightarrow a_i \leq \big| \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, \underline{X_i}] - \mathbb{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, \underline{X_{i-1}}] \big| \leq b_i$

that is, the expected value of f given "knowledge" of $X_1, \ldots, X_{i-1}$ doesn't change much (assuming $b_i - a_i$ is "small") when you learn X_i.

A more "relaxed" version, when the upper and lower bounds on $Z_i - Z_{i-1}$ are symmetric:

Given $b_1, \ldots, b_n$ s.t. $|Z_i - Z_{i-1}| \leq b_i \quad \forall i \in [n]$

$\forall t \geq 1$ and $\lambda > 0$, $\Pr(|Z_t - Z_0| \quad \lambda) \leq 2\exp\left(-\lambda^2 \big/ 2\sum_{i=1}^{n} b_i^2\right)$

(take the general version and set $a_i = -b_i$, and this falls out)

Not quite as tight as the general bound, but requires less information about the bounds

Stopping Times

Consider the "gambling" martingale where $X_1, \dots, X_i, \dots$ are the winnings of fair games and $Z_1, \dots, Z_i, \dots$ are the total winnings.

What if you decide, before you start gambling, to stop after n games?

- Observe that if $Z_0, Z_1, \dots, Z_n$ is a martingale w.r.t. $X_0, X_1, \dots, X_n$

$$Z_i = \mathbb{E}[Z_{i+1} \mid X_0, \dots, X_i] \quad \text{by def.}$$

repeating

$$\Rightarrow \mathbb{E}[Z_i] = \mathbb{E}\left[\mathbb{E}[Z_{i+1} \mid X_0, \dots, X_i]\right] = \mathbb{E}[Z_{i+1}]$$

$$\Rightarrow \mathbb{E}[Z_n] = \mathbb{E}[Z_0]$$

$\Rightarrow$ If the number of games is fixed in advance, $\mathbb{E}[Z_n] = \mathbb{E}[Z_0] = 0$

What if the stopping strategy is dependent on the $\{X_i\}$, i.e. not fixed in advance?
(Or includes some "on the side" randomness?)
What can we say about the expected winnings when play stops?

- Let T be a non-negative integer-valued R.V.
  We say T is a STOPPING TIME for $Z_0, Z_1, \dots$ if the event $T=n$ is independent of all the R.V.s $\{Z_{n+j} \mid Z_1, \dots, Z_n, j>0\}$ ☹??

(Think of a "stopping time" as a randomized strategy for determining when to stop. It can only use info. about what has happened so far.)

$\Rightarrow$ Whether or not $T=n$ depends only on $Z_0, \dots, Z_n$
(OR: you don't have to "look into the future" to decide if it is time to stop.)

Martingale stopping theorem

Let $Z_0, Z_1, \dots$ be a martingale w.r.t. $X_1, X_2, \dots$ If T is a stopping time for $X_1, X_2, \dots$ then $\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$ if at least one of the following holds:

1. $\exists c$ s.t. $|Z_i| \le c \;\; \forall i \ge 0$
2. $\exists c$ s.t. $\Pr(T \le c) = 1$ (i.e. T is bounded by a constant)
3. $\exists c$ s.t. $\mathbb{E}\left[\,|Z_{i+1} - Z_i| \mid X_1, X_2, \dots, X_i\right] < c$ and $T < \infty$

Why do we need this if the games are fair? Shouldn't it be the case that

$\mathbb{E}[X_i]=0 \Rightarrow \mathbb{E}[Z_T]=\mathbb{E}[Z_0]=0$?! Nope.

Consider the strategy: stop the first time T that $Z_T > B$ for some constant $B>0$

IF this happens, then $\mathbb{E}[Z_T] > B > 0$.

BUT it may never happen! $\Rightarrow$ you may lose an ∞ amount if you play this way

The condition $\mathbb{E}[Z_T]=\mathbb{E}[Z_0]=0$, informally, says that you can determine at time T when, on average, you can walk away with the same amount of money that you came in with.

---

ex: Sequence of fair games, win \$1 or lose \$1 w.p. $1/2$ (each); $X_1, X_2, \dots$ are the winnings of each game. $Z_0, Z_1, \dots$ are the total winnings after each game with $Z_0=0$.

STRATEGY: Stop as soon as you lose ℓ_1 dollars or win ℓ_2 dollars.

What is the probability that you win ℓ_2 before you lose ℓ_1?

↳ $\{Z_i\}$ is martingale w.r.t. $\{X_i\}$ ✓

↳ the strategy yields a proper stopping time for $X_1, X_2, \dots$ ✓

$\Rightarrow$ By the stopping theorem

$$\mathbb{E}[Z_T]=\mathbb{E}[Z_0]=0$$

Let $q=\Pr[\text{"stop with winnings } \ell_2\text{"}]$. Then $\mathbb{E}[Z_T]=\ell_2(q)-\ell_1(1-q)$

↑ Z_T has two possible values: ℓ_2 w.p. q, $-\ell_1$ w.p. $(1-q)$

Hence $0=\ell_2 q-\ell_1(1-q)$ or

$$\boxed{q=\frac{\ell_1}{\ell_1+\ell_2}}$$