[AI2613 Lecture 6]: Concentration Inequalities, Martingale

June 15, 2023

1 Chernoff Bounds

Recall the Markov inequality and Chebyshev's inequality we introduced before. They are used to prove that a random variable is concentrated around its expectation.

If we apply Markov inequality to

$$\mathbf{Pr}\left[f(X) \geq f(t)\right]$$

with $f(x) = e^{\alpha x}$ where $\alpha > 0$, then the bound amounts to bound $\mathbf{E}\left[e^{\alpha X}\right]$ which is the *moment generating function* of X.

When the random variable X can be written as the sum of independent Bernoulli variables, its moment generating function is easy to estimate and we obtain sharp concentration bounds.

Theorem 1 (Chernoff Bound) . *Let $X_1, \dots, X_n$ be independent random variables such that $X_i \sim \mathrm{Ber}(p_i)$ for each $i = 1, 2, \dots, n$. Let $X = \sum_{i=1}^n X_i$ and denote $\mu \triangleq \mathbf{E}[X] = \sum_{i=1}^n p_i$, we have*

$$\mathbf{Pr}\left[X \geq (1+\delta)\mu\right] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

If $0 < \delta < 1$, then we have

$$\mathbf{Pr}\left[X \leq (1-\delta)\mu\right] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$

Proof. We only prove the upper tail bound and the proof of lower tail bound is similar. For every $\alpha > 0$, we have

$$\mathbf{Pr}\left[X \geq (1+\delta)\mu\right] = \mathbf{Pr}\left[e^{\alpha X} \geq e^{\alpha(1+\delta)\mu}\right] \leq \frac{\mathbf{E}\left[e^{\alpha X}\right]}{e^{\alpha(1+\delta)\mu}}.$$

Therefore, we need to estimate the moment generating function $\mathbf{E}\left[e^{\alpha X}\right]$. Since $X = \sum_{i=1}^n X_i$ is the sum of independent Bernoulli variables, we have

$$\mathbf{E}\left[e^{\alpha X}\right] = \mathbf{E}\left[e^{\alpha \sum_{i=1}^n X_i}\right] = \underline{\mathbf{E}\left[\prod_{i=1}^n e^{\alpha X_i}\right] = \prod_{i=1}^n \mathbf{E}\left[e^{\alpha X_i}\right]}.$$

①

Since $X_i \sim \mathrm{Ber}(p_i)$, we can compute $\mathbf{E}\left[e^{\alpha X_i}\right]$ directly:

$$\underline{\mathbf{E}\left[e^{\alpha X_i}\right] = p_i e^{\alpha} + (1-p_i) = 1 + (e^{\alpha}-1)p_i \leq e^{((e^{\alpha}-1)p_i)}}.$$

②

Therefore,

$$\mathbf{E}\left[e^{\alpha X}\right] \leq \prod_{i=1}^n e^{((e^{\alpha}-1)p_i)} = e^{\left((e^{\alpha}-1)\sum_{i=1}^n p_i\right)} = e^{((e^{\alpha}-1)\mu)}.$$

本节讲述Concentration Ineq 以及详细总结Random Walk中一些定量计算问题（之前讨论的是收敛性，回归性等问题）

首先介绍Chernoff bound（Ineq），Chernoff bound 用于n个独立Bernoulli r.v.之和的r.v.的聚集性.

证明方法与Chebishev类似，

都运用 $Pr[X>a] \Rightarrow Pr[f(x)>f(a)]$，

其中cheb运用 $f=(X-EX)^2$.

而Chernoff运用类似 $f(x)=e^{\alpha x}$，

这种方法关键在于 $E[e^{\alpha x}]$ 可解，

其中 ① 运用了独立性

② 运用了Bernulli r.v. 条件

另外，注意 $f(x)=e^{\alpha x}$ 又称矩生成函数，因其泰勒展开为各阶矩的线性组合，权值由参数 α 控制.

因此，最后的bound含 α 后，可取 α 范围内的极值，取最tight的bound.

方法在下一页有展示.

Therefore,

$$\mathbf{Pr}\left[X \le (1+\delta)\mu\right] \le \frac{\mathbf{E}\left[e^{\alpha x}\right]}{e^{\alpha(1+\delta)\mu}} \le \left(\frac{e^{(e^{\alpha}-1)}}{e^{(\alpha(1+\delta))}}\right)^{\mu}$$

Note that above holds for any $\alpha > 0$. Therefore, we can choose α so as to minimize $\frac{e^{(e^{\alpha}-1)}}{e^{(\alpha(1+\delta))}}$. To this end, we let $\left(\frac{e^{(e^{\alpha}-1)}}{e^{(\alpha(1+\delta))}}\right)' = 0$. This gives $\alpha = \log(1+\delta)$. Therefore

$$\mathbf{Pr}\left[X \le (1+\delta)\mu\right] \le \left(\frac{e^{(e^{\alpha}-1)}}{e^{(\alpha(1+\delta))}}\right)^{\mu} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

□

The following form of Chernoff bound is more convenient to use (but weaker):

Chernoff bound 往往被简化成如左形式，便于使用.

Corollary 2 *For any* $0 < \delta < 1$,

$$\mathbf{Pr}\left[X \ge (1+\delta)\mu\right] \le \exp\left\{\left(-\frac{\delta^2}{3}\mu\right)\right\}$$

$$\mathbf{Pr}\left[X \le (1-\delta)\mu\right] \le \exp\left\{\left(-\frac{\delta^2}{2}\mu\right)\right\}$$

Proof. We only prove the upper tail. It suffices to verify that for $0 < \delta < 1$, we have

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \le \exp\left\{\left(-\frac{\delta^2}{3}\right)\right\}$$

Taking logarithm of both sides, this is equivalent to

$$\delta - (1+\delta)\ln(1+\delta) \le -\frac{\delta^2}{3}$$

Let $f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3}$ and note that

$$f'(\delta) = -\ln(1+\delta) + \frac{2}{3}\delta, \quad f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}.$$

Then for $0 < \delta < 1/2$, $f''(\delta) < 0$, and for $1/2 < \delta < 1$, $f''(\delta) > 0$. Therefore, $f'(\delta)$ first decreases and then increases in $[0,1]$. Also note that $f'(0) = 0$, $f'(1) < 0$ and $f'(\delta) \le 0$ when $0 \le \delta \le 1$. Therefore $f(\delta) \le f(0) = 0$. □

Seems not taught in class.

Example 1 (Tossing p-coins) . *Consider a p-coin that we get a head with probability p when tossing it. If we toss a p-coin n times, the average number of heads is pn.We want to determine the value δ such that with high probability (say 99%), the total number of heads is in the interval of $[(1-\delta)pn, (1+\delta)pn]$. We use Chernoff bound to determine δ.*

Let X denote the total number of heads, and $X_i \sim \mathrm{Ber}(p)$ be the indicator of whether the i-th toss gives a head. Then by Chernoff bound, we have

$$\mathbf{Pr}\left[|X - pn| \ge \delta \cdot pn\right] \le 2\exp\left\{\left(-\frac{\delta^2}{3}\cdot pn\right)\right\} \le 0.01$$

So if p is a constant, it suffices to choose

$$\delta = \Omega\left(\frac{1}{\sqrt{n}}\right).$$

2 *Hoeffding's Inequality*

One of annoying restrictions of Chernoff bound is that each X_i needs to be a Bernoulli random variable. We first relax this requirement by introducing Hoeffding's inequality which allows X_i to follow any distribution, provided its value is almost surely bounded.

我们可以发现 Chernoff bound 使用了蓝框(见前)的两个条件. 其中独立性很重要, 而 Bernoulli r.v. 条件可放宽.
由此诞生了 Hoeffiding's Ineq.
其将条件 Bernoulli r.v. 替换为①.
结论变为②.
证明见左.

Theorem 3 (Hoeffding's Inequality) *Let $X_1, \dots, X_n$ be independent random variables where each $X_i \in [a_i, b_i]$ for certain $a_i \le b_i$ with probability 1. Let $X = \sum_{i=1}^n X_i$ and $\mu \triangleq \mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i]$, then* ①

$$\mathbf{Pr}[|X - \mu| \ge t] \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \quad ②$$

for all $t \ge 0$.

It is instructive to compare Hoeffding and Chernoff when X_i's are independent Bernoulli variables. Formally, let $X_1, \dots, X_n$ be i.i.d. random variables where $X_i \sim \mathtt{Ber}(p)$ for all $i = 1, \dots, n$. Set $X = \sum_{i=1}^n X_i$ and denote $\mathbf{E}[X] = np$ by μ. By Hoeffding's inequality, we have

$$\mathbf{Pr}[|X - \mu| \ge t] \le 2\exp\left(-\frac{2t^2}{n}\right).$$

By Chernoff Bound, we have

$$\mathbf{Pr}[|X - \mu| \ge t] \le 2\exp\left(-\frac{t^2}{3pn}\right).$$

Comparing the exponent, it is easy to see that for $p > 1/6$, Hoeffding's inequality is tighter up to a certain constant factor. However, for smaller p, Chernoff bound is significantly better than Hoeffding's inequality.

We consider the balls-in-a-bag problem. There are g green balls and r red balls in a bag. These balls are the all same except for the color. We want to estimate the ratio $\frac{r}{r+g}$ by drawing balls. There are two scenarios.

- Draw balls with replacement. Let $X_i = \mathbf{1}[\text{the } i\text{-th ball is red}]$. Let $X = \sum_{i=1}^n X_i$. Then clearly each $X_i \sim \mathrm{Ber}\left(\frac{r}{r+g}\right)$ and $\mathbf{E}[X] = n \cdot \frac{r}{r+b}$.

  Since all X_i's are independent, we can directly apply Hoeffding's inequality and obtain

  $$\mathbf{Pr}[|X - \mathbf{E}[X]| \ge t] \le 2\exp\left(-\frac{2t^2}{n}\right).$$

- Draw balls without replacement. Again we let $Y_i = \mathbf{1}[\text{the } i\text{-th ball is red}]$, then unlike the case of drawing with replacement, variables in $\{Y_i\}$ are dependent. Let $Y = \sum_{i=1}^n Y_i$. We first calculate $\mathbf{E}[Y]$.

  For every $i \ge 1$, $\mathbf{E}[Y_i]$ is the probability that the i-th draw is a red ball. Note that drawing without replacement is equivalent to first drawing a

我们额外来讨论一下 1-d RW 的一些定量结果（定性结果见 lec4-lec5）
(w.p. $\frac{1}{2}$)

设 Z_i 为 i 时刻的位置，$\in \{\cdots,-2,-1,0,1,2,\cdots\}$，$Z_0=0$.

一个 trick 是取 $X_t = Z_t - Z_{t-1}$，$= \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$

1. $E[Z_i]=0$. 因为 $Z_i = Z_0 + \sum_{t=1}^{i} X_t$，
取 $\mathbb{E}$，由 $\mathbb{E}$ 线性性，
$\mathbb{E}[Z_i] = 0 + \sum \mathbb{E}[X_t]$，
而 $\mathbb{E}[X_t] = \frac{1}{2}\times 1 + \frac{1}{2}\times -1 = 0$.
故 $\mathbb{E}[Z_i]=0$.

2. $\Pr[\exists t \in [0, kn^2], \text{ s.t. } Z_t = n] \geq 1-\frac{1}{k}$ 即 kn^2 内达到 n 的概率不低于 $1-\frac{1}{k}$

证明见 $\boxed{\text{lec5}}$，运用 T_n 的分类讨论，列方程求 $E[T_n]$，并 bound 本结论的方法.

3. $E[|Z_i|]$!!? 和 $\Pr\Big[\big||Z_i| - E[|Z_i|]\big| \geq t\Big]$

证明：$|Z_i| = \underbrace{|Z_i|-|Z_{i-1}|}_{Y_{i-1}} + |Z_{i-1}|-|Z_{i-2}| \cdots\cdots + |Z_1|-|Z_0|$

$= \sum_{i=0}^{n-1} Y_i$

作左右期望：$\mathbb{E}[|Z_i|] = \sum_{i=0}^{n-1} \mathbb{E}[Y_i]$

☆ 注意 $\mathbb{E}[Y_i]$ 与 $E[X_i]$ 不同，其不会为 0，特别地，在 $Z_i=0$ 时 $\mathbb{E}[Y_i]=1$
其他时刻 $\mathbb{E}[Y_i]=0$

故 $\mathbb{E}[|Z_i|] = \sum \mathbb{E}[Y_i \mid Z_i=0] + \underbrace{\sum \mathbb{E}[Y_i \mid Z_i=1]}_{0}$

$= \sum_{i=0}^{n-1} \Pr[Z_i=0]$，这个在 Lec 2、3 中记 convergence 时讨论过.

$= \sum_{i=0}^{n-1} \Pr[Z_{2i}=0]$

$$= \sum_{i=0}^{[\frac{n-1}{2}]} \binom{2i}{i} / 2^{2i}$$

↓

在 2i 个点中找 i 个为 '+1' move 的概率，用排列公式算。

$$\approx \sum_{i=0}^{[\frac{n-1}{2}]} \sqrt{\frac{2}{\pi i}} \approx c \cdot \sqrt{n} \quad c \text{ 为 constant.}$$

最后用本节的 concentration Ineq 算概率.

由 Hoeffding Ineq, $\because Y_i \in \{-1, 1\}$.

$\therefore a_i = -1, b_i = 1$

$$\Pr[|X - \mu| \geq t] \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

又 $\because |Z_i| = \sum Y_i$, $\mu = E[|Z_i|] = \sum E[Y_i] = c \cdot \sqrt{n}$.

取 $t \sim O(\sqrt{n}) = k\sqrt{n}$

$\therefore$ 代入 Ineq, $\Pr\left[\big||Z_i| - c\sqrt{n}\big| \geq k\sqrt{n}\right] \leq (\cdots) \exp(-\frac{c^2}{2})$

$\therefore E[|Z_i|] = c \cdot \sqrt{n}$, $\Pr \leq (\cdots) \exp(-\frac{c^2}{2})$ 求毕.

uniform permutation of $r+g$ balls and drawing each ball one by one in that order. Therefore, the probabilty of $Y_i = 1$ is $\frac{r\cdot(r+g-1)!}{(r+g)!} = \frac{r}{r+g}$. So we have $\mathbf{E}[Y] = n \cdot \frac{r}{r+g}$.

However, since $\{Y_i\}$ are dependent, we cannot apply Hoeffding's inequality directly. This motivate us to generalize it by removing the requirement of independence.

3 *Martingale*

We develop the theory of martingale, which is a core concept in probability theory. We use martingale to get rid of the independence requirement in the concentration inequalities mentioned above.

⇒ 在上文的Hoeffding Ineq中，只有 X_i 独立才有 $X_i \in [a_i, b_i]$ ⇒ ΣX_i 范围的工具。

但如果引入鞅论，如 $\{X_t\}$ 是鞅，则无须独立条件也能bound住 ΣX_i

Consider a fair gambling game in which the expected gain in each round is zero. As a result, regardless of how much one bets in each round, the money in expectation remains the same. The balances after each round form a *martingale.*

Def 1

Definition 4 (Martingale) *Let $\{X_n\}_{n\ge0}$ and $\{Z_n\}_{n\ge0}$ be two sequences of random variables. Let $Z_n = \sum_{t=0}^{n} X_t$.*[1] *We say $\{Z_n\}_{n\ge0}$ is a martingale w.r.t. $\{X_n\}_{n\ge0}$ if*

$$\mathbf{E}[Z_{n+1} \mid X_0, X_1 \dots, X_n] = Z_n.$$

[1] Consider the problem of fair gambling where X_n is the gain of n-th round and $Z_n = \sum_{t=1}^{n} X_n$. $\{Z_n\}_{n\ge0}$ is not necessarily a Markov chain. The value X_n may depend on information before round $n-1$.

上展示了两种鞅定义，def 1 利用随机过程 $\{X_t\}$ 的变量定义，def 2 将 $X_0 \to \overline{X}_{0,1} \to \overline{X}_{0,2}$ 的过程抽象为 $\mathcal{F}_0 \to \mathcal{F}_1 \cdots$

Sometimes we say a single sequence $\{X_n\}_{n\ge0}$ is a martingale if it is a martingale w.r.t. itself. Formally, if for every $n \ge 0$, it holds that

abbr-def 1

$$\mathbf{E}[X_{n+1} \mid X_0, \dots, X_n] = X_n.$$

For convenience, from now on we use $\overline{X}_{i,j} = (X_i, X_{i+1} \dots, X_j)$ to simplify the notations. The conditional expectation $\mathbf{E}\left[Z_{n+1} \,\middle|\, \overline{X}_{0,n}\right]$ is equivalent to $\mathbf{E}\left[Z_{n+1} \,\middle|\, \sigma(\overline{X}_{0,n})\right]$ where $\sigma(\overline{X}_{0,n})$ is the σ-algebra generated by $X_0, \dots, X_n$. The motivates us to define martingale in a more general way.

对鞅的一种直观理解在于：将Def 1 左一取期望，则 $\mathbb{E}[Z_{n+1}] = \mathbb{E}[Z_n]$，故鞅可以理解为一种已知 $X_0 \sim X_n$ 信息时作决策，但决策结果的期望收益是0（$\boxed{\begin{matrix} \mathbb{E}Z_t = \mathbb{E}Z_{t+1} \\ Z_t = \Sigma X_i \end{matrix}}$）的一种随机过程.

Def 2

Definition 5 (Martingale (defined by filtration)) *Let $\{\mathcal{F}_n\}_{n\ge0}$ be a sequence of σ-algebras. We call such σ-algebra sequence a filtration if it satisfies*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \cdots.$$

Given a filtration $\{\mathcal{F}_n\}_{n\ge0}$, let $\{Z_n\}_{n\ge0}$ be a stochastic process that Z_n is $\mathcal{F}_n$-measurable for every $n \ge 0$. Then we say $\{Z_n\}_{n\ge0}$ is a martingale w.r.t. $\{\mathcal{F}_n\}_{n\ge0}$ if for every $n \ge 0$

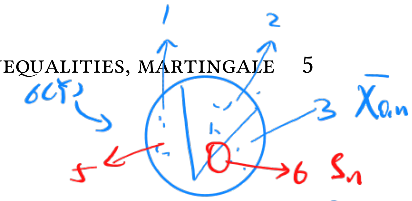
$$\mathbf{E}[Z_{n+1} | \mathcal{F}_n] = Z_n.$$

If $\mathbf{E}[Z_{n+1}|\mathcal{F}_n] \le Z_n$ in Definition 5, we call $\{Z_n\}_{n\ge0}$ a supermartingale w.r.t. $\{\mathcal{F}_n\}_{n\ge0}$. Similarly, if $\mathbf{E}[Z_{n+1}|\mathcal{F}_n] \ge Z_n$, we call it a submartingale.

Exp 1

Example 2 (1-D Random Walk) *Consider a random walk on $\mathbb{Z}$ starting from 0. The probability to the left and the probability to the right are both $\frac{1}{2}$ at each step. Denote the action at the n-th step by a uniform random variable*

接下来我们举若干例子，证明其为Martingale，以此加深理解.



$X_n \in \{-1,+1\}$. *Let* $S_n = \sum_{k=0}^{n} X_k$. *Then we can verify* $\{S_n\}_{n\geq 0}$ *is a martingale w.r.t.* $\{X_n\}_{n\geq 0}$ *(or w.r.t.* $\{S_n\}_{n\geq 0}$*) by noticing that*

因为S_n是$\overline{X}_{0,n}$可测，故在$\overline{X}_{0,n}$ condition下 $E[S_n \mid \overline{X}_{0,n}]$和$S_n$是同一个r.v.

$$\mathbf{E}\left[S_{n+1} \,\middle|\, \overline{X}_{0,n}\right] = \mathbf{E}\left[S_n + X_{n+1} \,\middle|\, \overline{X}_{0,n}\right] = S_n + \mathbf{E}\left[X_{n+1} \,\middle|\, \overline{X}_{0,n}\right] = S_n.$$

如上图所示，给定一个ω，$E[S_n \mid \overline{X}_{0,n}] = E[S_n \mid X^{-1}X(\omega)]$

More generally, if $\mathbf{E}\left[X_{n+1} \,\middle|\, \overline{X}_{0,n}\right] = \mu$, *we define* $Y_k = X_k - \mu$ *and* $S'_n \triangleq \sum_{k=0}^{n} Y_k = S_n - (n+1)\mu$. *Then* S'_n *is a martingale w.r.t.* $\{Y_n\}_{n\geq 0}$.

而由于S_n是$\sigma(\overline{X}_{0,n})$可测，故取一集点$X^{-1}X(\omega)$只对应唯一$X$的值，比如取$\omega$，使$\overline{X}_{0,n}(\omega)=2$

Example 3 *Consider a sequence of random variables* $\{X_n\}_{n\geq 0}$ *where* $\mathbf{E}\left[X_n \,\middle|\, \overline{X}_{0,n-1}\right] = 1$ *for all* $n \geq 1$. *Let* $P_n = \prod_{k=0}^{n} X_k$. *Then we can verify* $\{P_n\}_{n\geq 0}$ *is a martingale w.r.t.* $\{X_n\}_{n\geq 0}$ *by verifying that*

Exp 2

则$X^{-1}X(\omega)$为所有II区的ω，

$$\mathbf{E}\left[P_{n+1} \,\middle|\, \overline{X}_{0,n}\right] = \mathbf{E}\left[P_n \cdot X_{n+1} \,\middle|\, \bar{X}_{0,n}\right] = P_n \cdot \mathbf{E}\left[X_{n+1} \,\middle|\, \overline{X}_{0,n}\right] = P_n.$$

则$E[S_n \mid \overline{X}_{0,n}](\omega) = E[S_n \mid X^{-1}X(\omega)]$

可见X在II区中均为6，故期望也为6，

$E[S_n \mid \overline{X}_{0,n}](\omega) = 6$，而$S_n(\omega)$由$\omega$点本身在II区中，故$S_n(\omega)=6$，

可见二者是相同的。

关键在于S_n的$\sigma(\overline{X}_{0,n})$可测使$X^{-1}X(\omega)$不会使$S_n$有不同映射值，故期望可去掉，二者都是$\omega \mapsto \mathbb{R}$的r.v.

Example 4 (Galton-Watson Process) *Recall the Galton-Watson process we discussed in the last lecture. Suppose that all the individuals reproduce independently of each other and have the same offspring distribution. Let* G_t *be the number of individuals of the* t*-th generation. Each individual of generation* t *gives birth to a random number of children of generation* $t+1$. *Denote by* $X_{t,k}$ *the number of children of the* k*-th individual in the* t*-th generation. Assume that* $X_{t,k}$ *are i.i.d. and let* $\mu \triangleq \mathbf{E}\left[X_{t,k}\right]$. *Then we have* $G_{t+1} = \sum_{k=1}^{G_t} X_{t,k}$. *Thus,*

Exp 3

$$\mathbf{E}\left[G_{t+1} \mid G_t\right] = \mathbf{E}\left[\sum_{k=1}^{G_t} X_{t,k} \,\middle|\, G_t\right] = G_t \cdot \mathbf{E}\left[X_{t,1}\right] = \mu G_t.$$

Define $M_t = \mu^{-t} G_t$. *Then*

$$\mathbf{E}\left[M_{t+1} \mid G_t\right] = \mu^{-t-1}\mathbf{E}\left[G_{t+1} \mid G_t\right] = \mu^{-t} G_t = M_t.$$

That is, $\{M_t\}_{t\geq 0}$ *is a martingale w.r.t.* $\{G_t\}_{t\geq 0}$.

Exp 4

Example 5 (Pólya's urn) *Suppose there are some white balls and black balls in an urn. All of these balls are identical except the colors. Consider the following stochastic process: each round we pick a ball uniformly at random and observe its color; then we return the ball, and add an additional ball of the same color into the urn. We repeat the process, and our goal is to study the sequence of colors of the selected balls.*

W.l.o.g. assume that we start from only one white ball and one black ball in the urn, and the index of rounds starts from 2. Then after round n, *there are exactly* n *balls in the urn. Let* X_n *be the number of black balls after round* n, *and* $Z_n = \frac{X_n}{n}$ *is the ratio of black balls after round* n. *Clearly* $Z_2 = \frac{1}{2}$. *Then we have*

Example 5 shows that X_n does not have to be i.i.d..

$$\begin{aligned}\mathbf{E}\left[Z_{n+1} \,\middle|\, \bar{X}_{2,n}\right] &= \frac{1}{n+1}\mathbf{E}\left[X_{n+1} \,\middle|\, \overline{X}_{2,n}\right] \\ &= \frac{1}{n+1}\left(Z_n(X_n+1) + (1-Z_n)X_n\right) = \frac{Z_n + X_n}{n+1} = Z_n.\end{aligned}$$

分析随机过程，有Z_n概率为X_n+1个黑球，有$1-Z_n$概率为X_n个黑球（因为上轮摸黑我们就加一个黑球）然后化简即可。

That is, $\{Z_n\}_{n\geq 2}$ *is a martingale w.r.t.* $\{X_n\}_{n\geq 2}$.

References