Probability Theory & Mathematical Statistics

Concentration of Measure

尹一通 Nanjing University, 2024 Spring

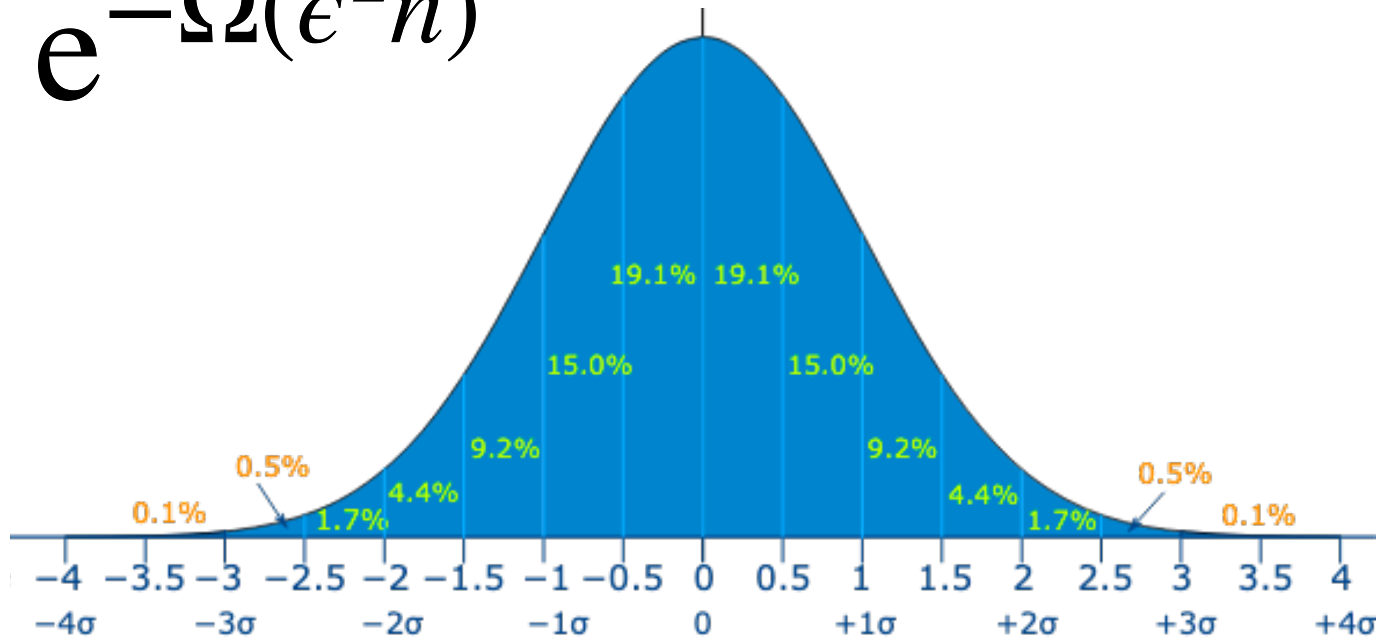
Bernoulli's Law of Large Number

In *Ars Conjectandi* (1713)

- If $X_1, X_2, \ldots \sim$ Bernoulli(p) be *i.i.d.*, then $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} p$, i.e. $\forall \epsilon > 0$,

$$\Pr\left(\left| \overline{X}_n - p \right| > \epsilon \right) \to 0 \quad \text{as } n \to \infty$$

- How fast is the convergence rate:
  - **Chebyshev**'s inequality gives: $< \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$
  - **CLT** (and **de Moivre—Laplace**): should be Gaussian-like $\mathrm{e}^{-\Omega(\epsilon^2 n)}$



Sum of Independent Trials

(Poisson binomial distribution)

- Let $X_1, \ldots, X_n \in \{0,1\}$ be independent trials (also called **Poisson trials**), which are not necessarily identically distributed, and let

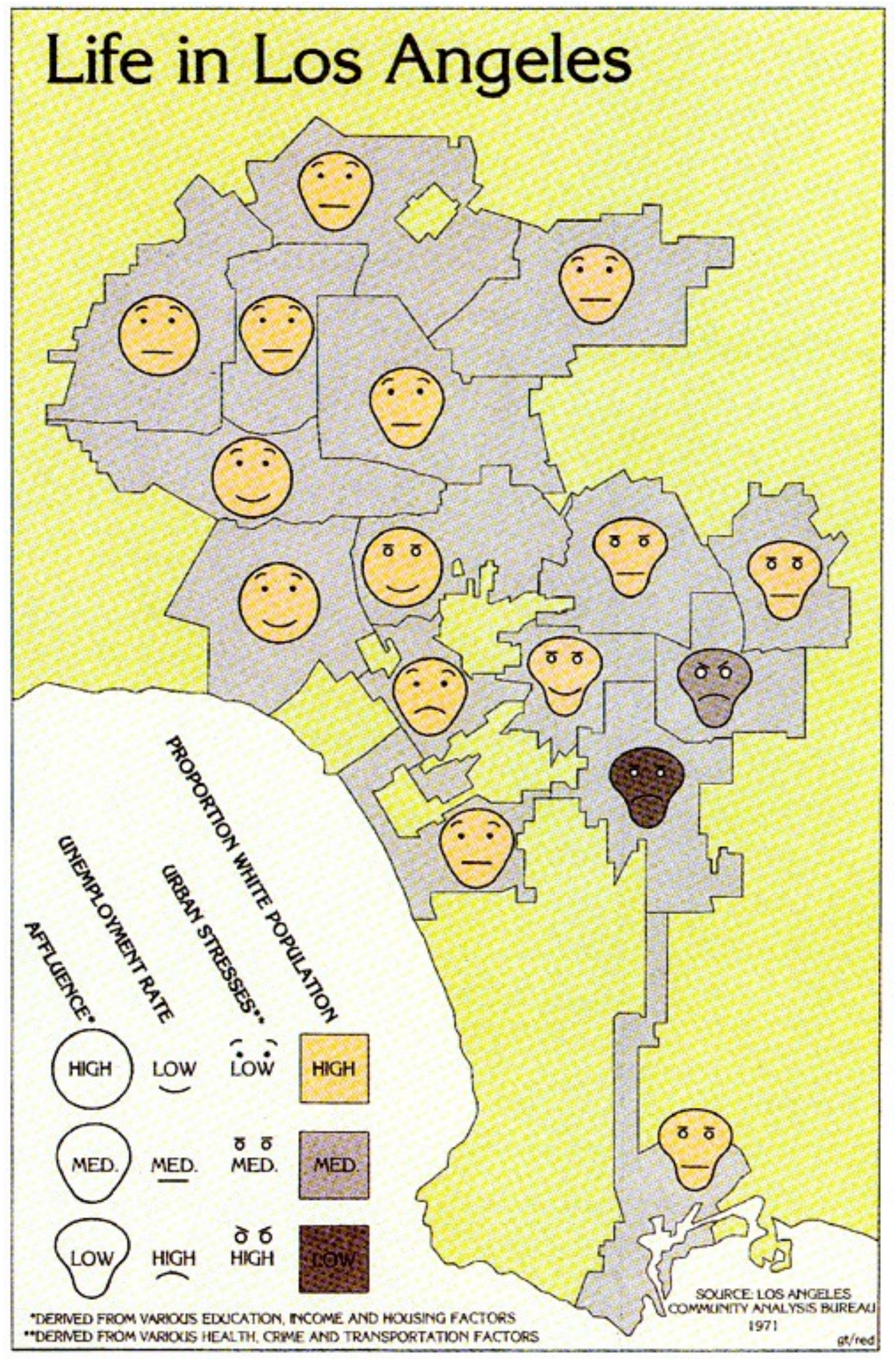
$$S_n = \sum_{i=1}^{n} X_i$$

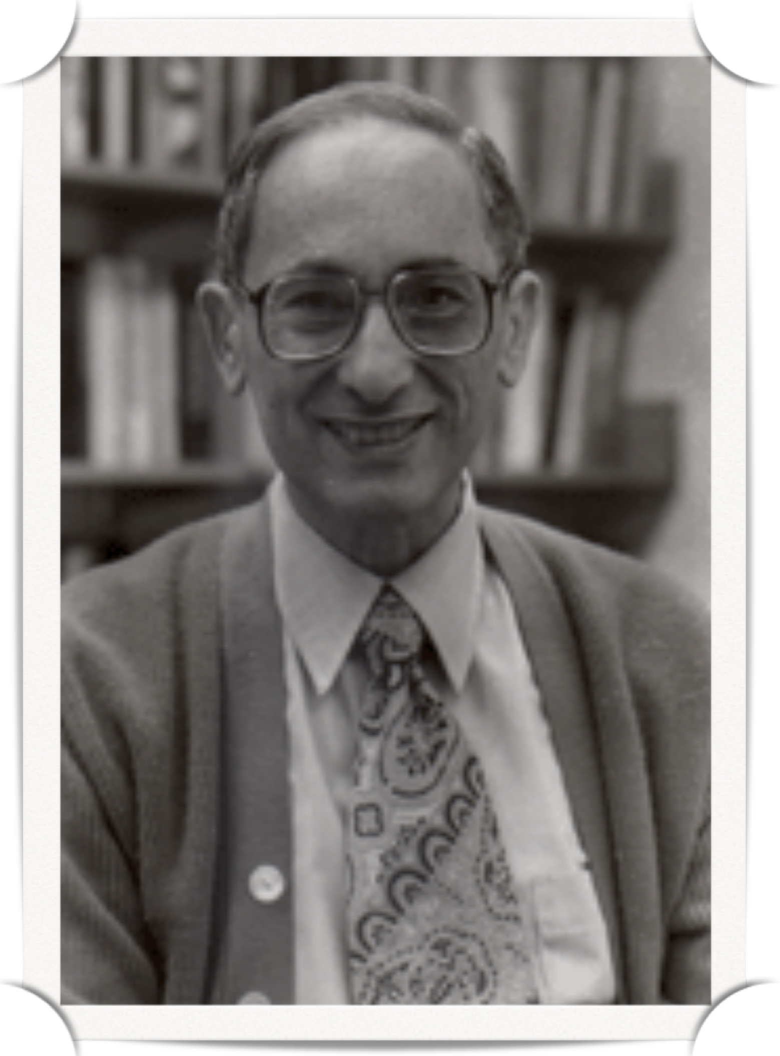
(called **Poisson binomial** random variable)

- Deviation / concentration / tail bounds:

$$\Pr\left(\,|S_n - \mathbb{E}[S_n]| \geq ?\right) \leq ?$$

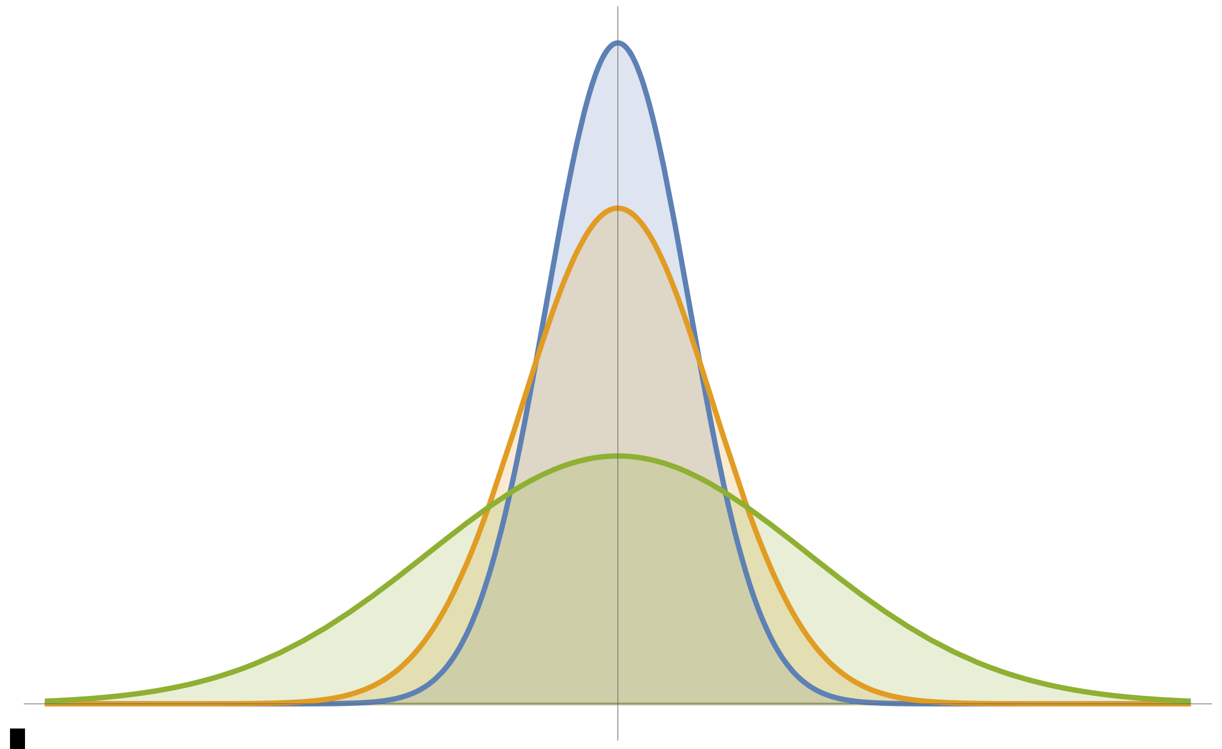
Chernoff-Hoeffding Bounds





Herman Chernoff

Chernoff Bound



- <u>Chernoff bound</u>: Let $X_1, \ldots, X_n \in \{0,1\}$ be **independent trials**

$$X = \sum_{i=1}^{n} X_i \quad \text{and} \quad \mu = \mathbb{E}[X]$$

(Poisson binomial RV with mean μ)

- For any $\delta > 0$,

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

- For any $0 < \delta < 1$,

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

Chernoff Bound (Upper Tail)

- Let $X_1, \ldots, X_n \in \{0,1\}$ be independent trials with $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$

For any $\delta > 0$: $\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$

Proof: $\Pr(X \geq (1+\delta)\mu) \leq \Pr\left(e^{tX} \geq e^{t(1+\delta)\mu}\right)$ (for any $t > 0$)

(Markov's inequality) $\leq e^{-t(1+\delta)\mu} \cdot \mathbb{E}\left[e^{tX}\right]$

Moment Generating Function

- The moment generating function (MGF) of a random variable X is

$$M_X(t) = \mathbb{E}[\mathrm{e}^{tX}] = \sum_{k\geq 0} \frac{t^k \mathbb{E}[X^k]}{k!}$$

- If $X_1, \ldots, X_n \in \{0,1\}$ are independent with $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$, then

$$M_X(t) = \mathbb{E}[\mathrm{e}^{tX}] \leq \mathrm{e}^{(\mathrm{e}^t - 1)\mu}$$

Proof: $\mathbb{E}\left[\mathrm{e}^{tX}\right] = \mathbb{E}\left[\prod_{i=1}^{n} \mathrm{e}^{tX_i}\right] = \prod_{i=1}^{n} \mathbb{E}\left[\mathrm{e}^{tX_i}\right] = \prod_{i=1}^{n} \left(\mathrm{e}^t p_i + (1 - p_i)\right) \leq \prod_{i=1}^{n} \mathrm{e}^{(\mathrm{e}^t - 1)p_i}$

(independence) (where $p_i = \mathbb{E}[X_i]$) (because $1 + x \leq \mathrm{e}^x$)

$\leq \mathrm{e}^{\sum_{i=1}^{n} (\mathrm{e}^t - 1)p_i} = \mathrm{e}^{(\mathrm{e}^t - 1)\mu}$ (where $\mu = \sum_{i=1}^{n} p_i$)

Chernoff Bound (Upper Tail)

- Let $X_1, \ldots, X_n \in \{0,1\}$ be independent trials with $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$

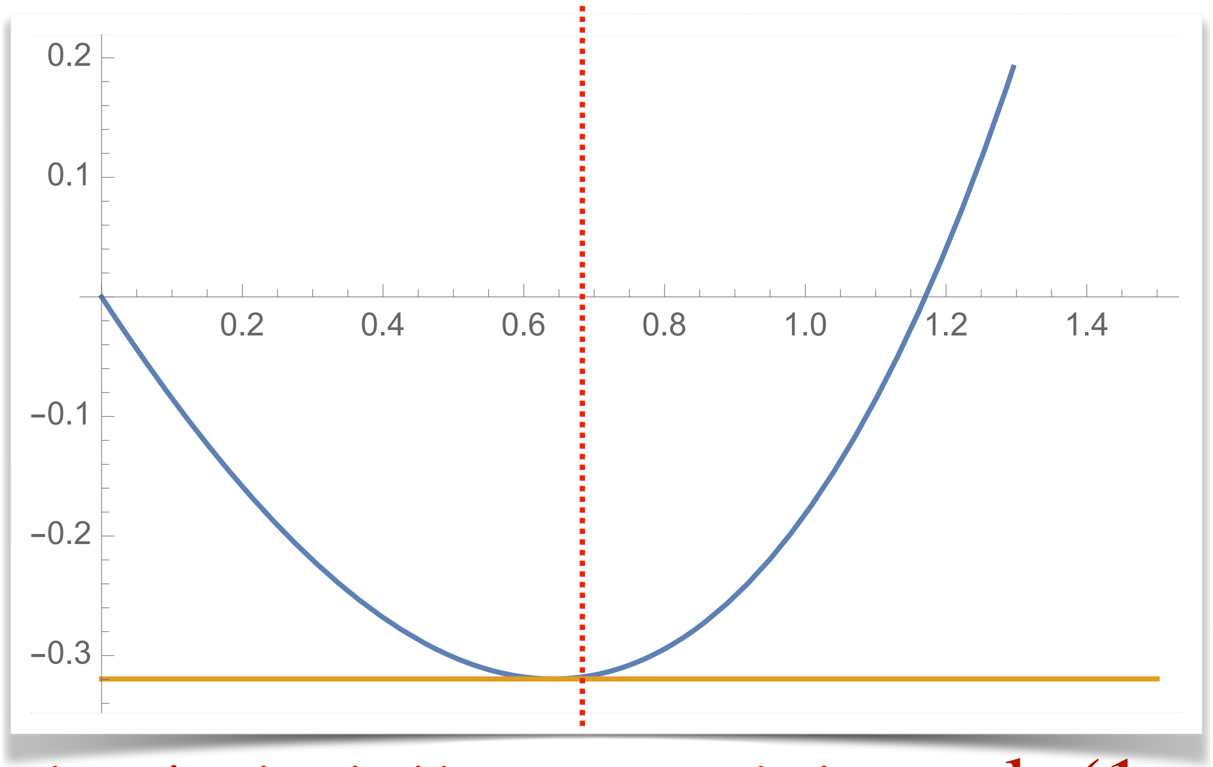
For any $\delta > 0$: $\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$

Proof: $\Pr(X \geq (1+\delta)\mu) \leq \Pr\left(e^{tX} \geq e^{t(1+\delta)\mu}\right)$ (for any $t > 0$)

(Markov's inequality) $\leq e^{-t(1+\delta)\mu} \cdot \mathbb{E}\left[e^{tX}\right] \leq e^{-t(1+\delta)\mu} \cdot e^{(e^t-1)\mu}$

$= e^{(e^t-1-t(1+\delta))\mu} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$

(choose $t = \ln(1+\delta)$)



(minimized at stationary point $t = \ln(1+\delta)$)

Chernoff Bound (Lower Tail)

- Let $X_1, \ldots, X_n \in \{0,1\}$ be independent trials with $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$

For any $0 < \delta < 1$: $\quad \Pr(X \le (1-\delta)\mu) \le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$

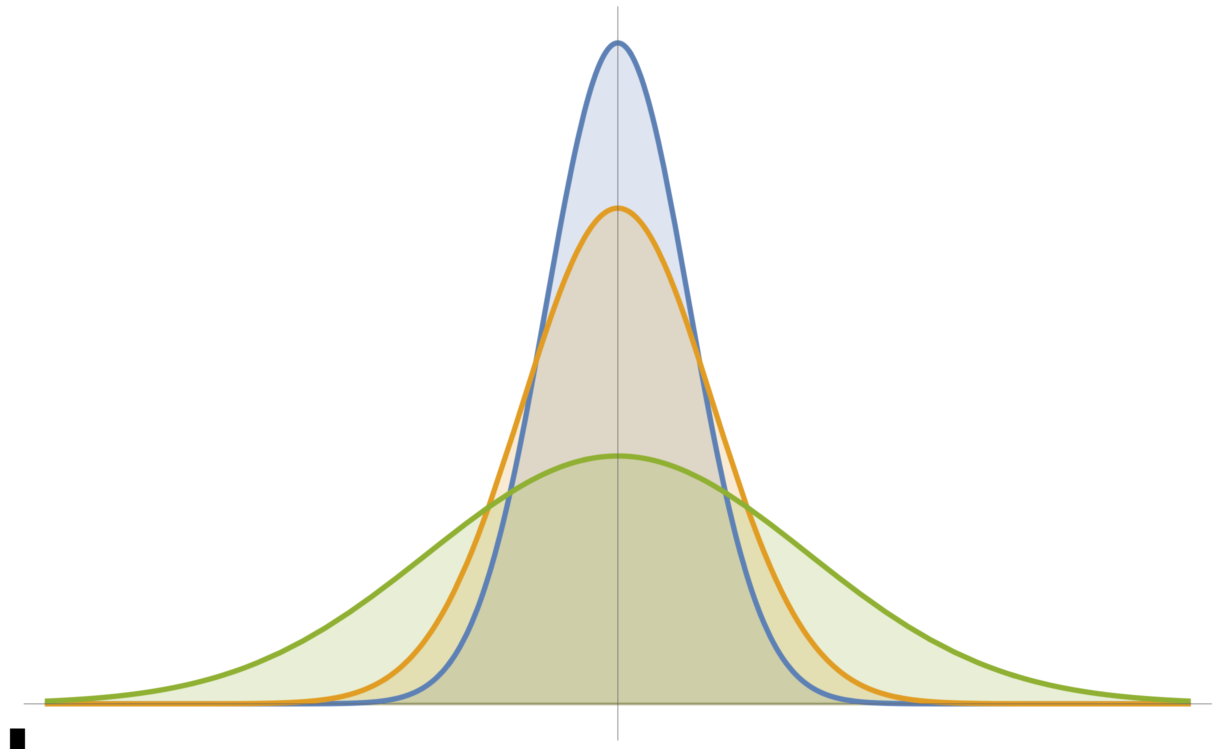
Proof: $\quad \Pr(X \le (1-\delta)\mu) \le \Pr\left(e^{tX} \ge e^{t(1-\delta)\mu}\right)$ (for any $t < 0$)

(Markov's inequality) $\quad \le e^{-t(1-\delta)\mu} \cdot \mathbb{E}\left[e^{tX}\right] \le e^{-t(1-\delta)\mu} \cdot e^{(e^t-1)\mu}$

$$= e^{(e^t - 1 - t(1-\delta))\mu} = \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

(choose $t = \ln(1-\delta)$)

Chernoff Bound



- <u>Chernoff bound</u>: Let $X_1, \ldots, X_n \in \{0,1\}$ be independent trials

$$X = \sum_{i=1}^{n} X_i \quad \text{and} \quad \mu = \mathbb{E}[X]$$

(Poisson binomial RV with mean μ)

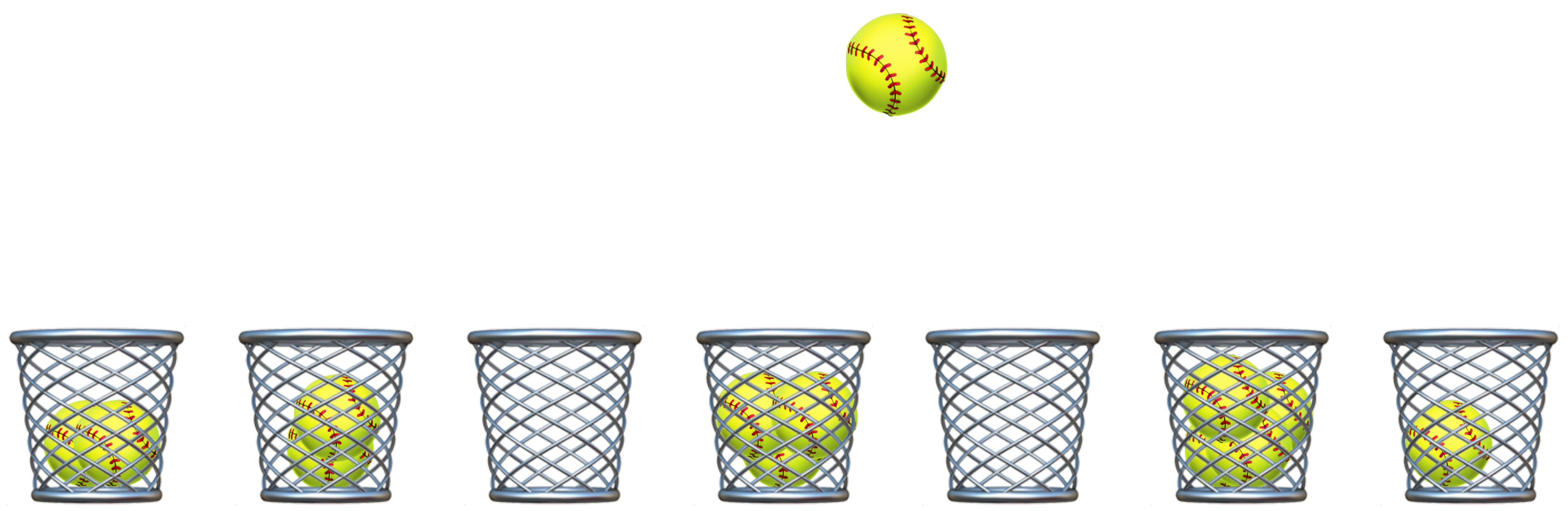
- For any $\delta > 0$,

$$\Pr(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \leq \begin{cases} e^{-\frac{\mu\delta^2}{3}} & \text{if } 0 < \delta < 1 \\ 2^{-(1+\delta)\mu} & \text{if } (1+\delta) \geq 2e \end{cases}$$

- For any $0 < \delta < 1$,

$$\Pr(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \leq e^{-\frac{\mu\delta^2}{2}}$$

Balls into Bins
(Multinomial distribution)



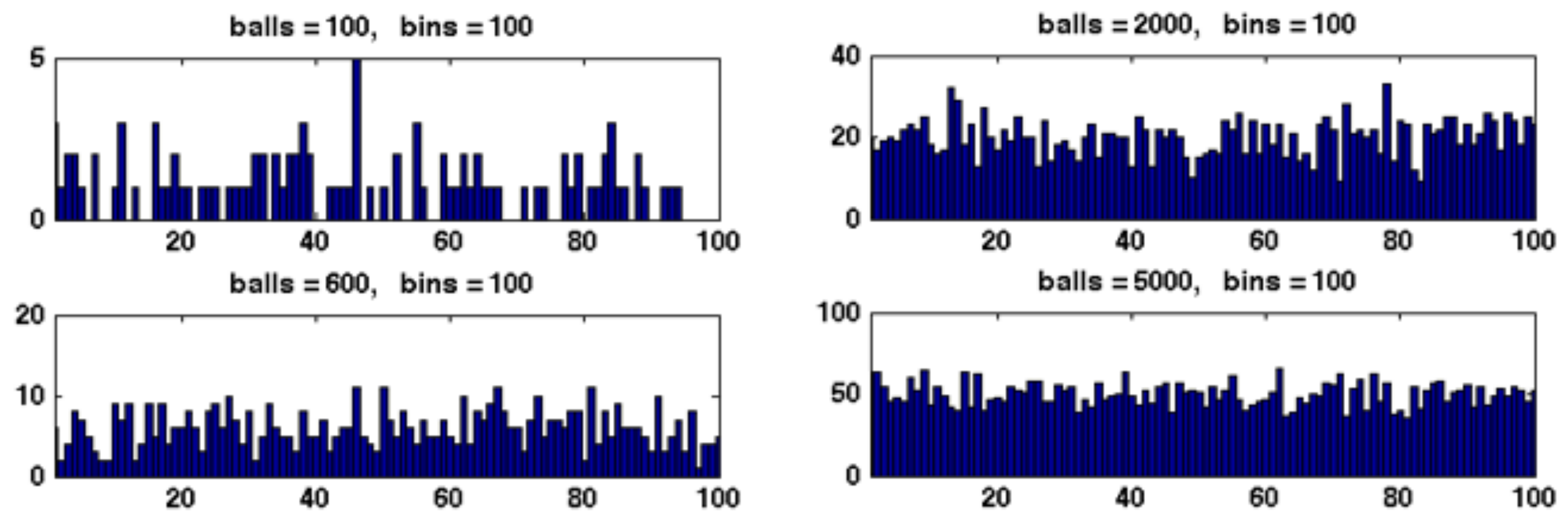
- Throw m balls into n bins ***u.a.r.*** Numbers of balls received in each bins:

$$(X_1, X_2, \ldots, X_n) \sim \textbf{multinomial distribution} \text{ of parameters } m, \underbrace{(1/n, \ldots, 1/n)}_{n}$$

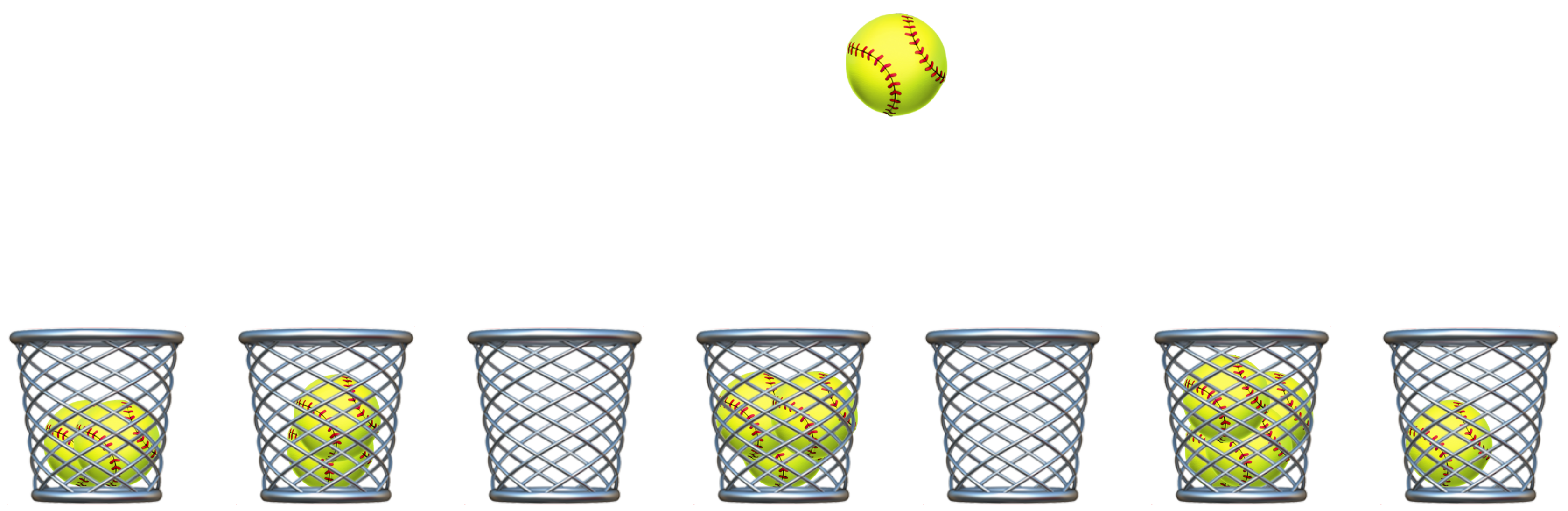
- <u>Occupancy problem</u>: the maximum load $\max_{1 \le i \le n} X_i$ ***w.h.p.*** (*with high probability*)

$$\max_{1 \le i \le n} X_i = \begin{cases} O\left(\frac{\log n}{\log \log n}\right) & \text{if } m = n \\ O\left(\frac{m}{n}\right) & \text{if } m \ge n \ln n \end{cases}$$

w.h.p. (with prob. $1 - o\left(\frac{1}{n}\right)$)



Occupancy Problem



- Throw m balls into n bins ***u.a.r.*** The i-th bin receives X_i balls:

$$\text{Marginally,} \quad X_i \sim \text{Bin}(m, 1/n) \quad \text{and} \quad \mu = \mathbb{E}\left[X_i\right] = \frac{m}{n}$$

- When $m = n$: $\mu = 1$

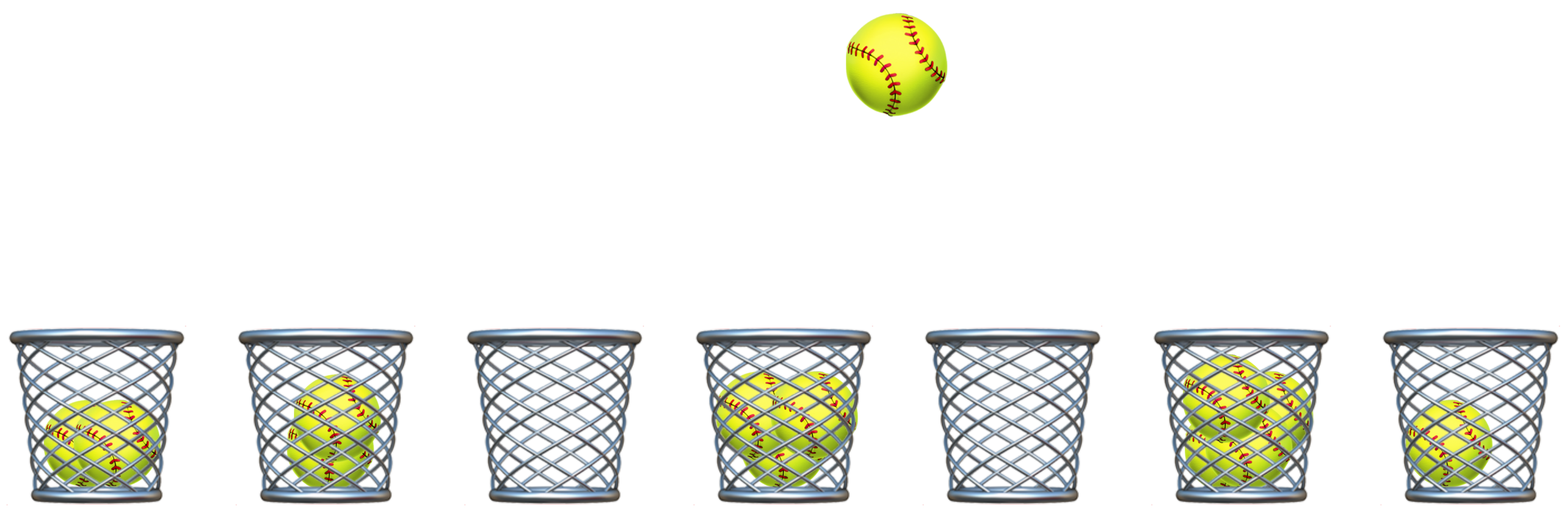
$$\Pr\left(X_i \geq L\right) = \Pr\left(X_i \geq L\mu\right) \leq \frac{\mathrm{e}^L}{\mathrm{e}L^L} \leq \frac{1}{n^2}$$

for $L = \dfrac{\mathrm{e}\ln n}{\ln\ln n}$

Chernoff: $\Pr(X_i \geq (1+\delta)\mu) \leq \left(\dfrac{\mathrm{e}^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$

- **Union bound:** $\Pr\left(\max_{1\leq i\leq n} X_i \geq L\right) \leq \sum_{i=1}^{n} \Pr\left(X_i \geq L\right) \leq \dfrac{1}{n}$ $\Longrightarrow$ $\max_{1\leq i\leq n} X_i = O\left(\dfrac{\log n}{\log\log n}\right)$ **w.h.p.**

Occupancy Problem



- Throw m balls into n bins ***u.a.r.*** The i-th bin receives X_i balls:

  Marginally, $X_i \sim \text{Bin}(m, 1/n)$ and $\mu = \mathbb{E}\left[X_i\right] = \dfrac{m}{n}$

- When $m \geq n \ln n$: $\mu \geq \ln n$

$$\Pr\left(X_i \geq \frac{2\mathrm{e}m}{n}\right) = \Pr\left(X_i \geq 2\mathrm{e}\mu\right) \leq 2^{-2\mathrm{e}\mu} \leq 2^{-2\mathrm{e}\ln n} \leq \frac{1}{n^2}$$

Chernoff: $\Pr\left(X_i \geq L\right) \leq 2^{-L}$ if $L \geq 2\mathrm{e}\mu$

- **Union bound:** $\Pr\left(\max_{1\leq i\leq n} X_i \geq \dfrac{2\mathrm{e}m}{n}\right) \leq \displaystyle\sum_{i=1}^{n} \Pr\left(X_i \geq \frac{2\mathrm{e}m}{n}\right) \leq \frac{1}{n}$ $\Longrightarrow$ $\max_{1\leq i\leq n} X_i = O\left(\dfrac{m}{n}\right)$ **w.h.p.**

Chernoff-Hoeffding Bound

- Chernoff-Hoeffding bound:

If $X_1, \ldots, X_n \in \{0,1\}$ are independent and $S_n = \sum_{i=1}^{n} X_i$, then for any $t > 0$,

$$\Pr\left(\left| S_n - \mathbb{E}[S_n] \right| \geq t \right) \leq 2 \exp\left(-\frac{2t^2}{n} \right)$$

In general, if $X_1, \ldots, X_n$ are independent and $X_i \in [a_i, b_i]$, $1 \leq i \leq n$, then for any $t > 0$,

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

Hoeffding's Lemma*

- Hoeffding's lemma: If $Y \in [a, b]$ **a.s.** and $\mathbb{E}[Y] = 0$, then its **MGF**

$$M_Y(\lambda) = \mathbb{E}\left[e^{\lambda Y}\right] \le e^{\lambda^2(b-a)^2/8}$$

Proof: Define $\Psi_Y(\lambda) = \ln \mathbb{E}\left[e^{\lambda Y}\right]$. It suffices to prove $\Psi_Y(\lambda) \le \lambda^2(b-a)^2/8$.

Taylor's expansion: $\exists \xi \in [0,\lambda]$ s.t. $\Psi_Y(\lambda) = \Psi_Y(0) + \lambda \Psi'_Y(0) + \frac{\lambda^2}{2}\Psi''_Y(\xi) \le \frac{\lambda^2(b-a)^2}{8}$

$M_Y(0) = 1$ and $M'_Y(0) = \mathbb{E}[Y] = 0 \implies \Psi_Y(0) = 0$ and $\Psi'_Y(0) = M'_Y(0)/M_Y(0) = 0$

$$\Psi''_Y(\xi) = \frac{M''_Y(\xi)}{M_Y(\xi)} - \frac{M'_Y(\xi)^2}{M_Y(\xi)^2} = \mathbb{E}\left[Y^2 e^{\xi Y}/M_Y(\xi)\right] - \mathbb{E}\left[Y e^{\xi Y}/M_Y(\xi)\right]^2 = \mathbf{Var}[Z]$$

for a new random variable Z with CDF $F_Z(z) = \int_{-\infty}^{z} \frac{e^{\xi y}}{\mathbb{E}[e^{\xi Y}]} \mathrm{d}F_Y(y)$ (Lebesgue-Stieltjes integral)

Notice also $Z \in [a, b]$ **a.s.** $\implies \mathbf{Var}[Z] = \mathbf{Var}\left[Z - \frac{a+b}{2}\right] \le \mathbb{E}\left[\left(Z - \frac{a+b}{2}\right)^2\right] \le \frac{(b-a)^2}{4}$

Chernoff-Hoeffding Bound

- Chernoff-Hoeffding bound: If $S_n = \sum_{i=1}^{n} X_i$ and $X_i \in [a_i, b_i]$ are independent, then

$$\forall t > 0, \quad \Pr\left(\left|S_n - \mathbb{E}[S_n]\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Proof: Let $Y_i = X_i - \mathbb{E}[X_i]$ and $Y = S_n - \mathbb{E}[S_n] = \sum_{i=1}^{n} Y_i \implies \mathbb{E}[Y] = \mathbb{E}[Y_i] = 0$

$$\Pr\left(S_n - \mathbb{E}[S_n] \geq t\right) = \Pr(Y \geq t) \leq \Pr\left(e^{\lambda Y} \geq e^{\lambda t}\right) \quad \text{(for any } \lambda > 0\text{)}$$

(Markov's inequality) $\leq e^{-\lambda t} \cdot \mathbb{E}\left[e^{\lambda Y}\right] = e^{-\lambda t} \cdot \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda Y_i}\right] = e^{-\lambda t} \cdot \prod_{i=1}^{n} \mathbb{E}\left[e^{\lambda Y_i}\right]$ (independence)

(Hoeffding's lemma) $\leq \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right) = \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$

by choosing: $\lambda = \frac{4t}{\sum_{i=1}^{n}(b_i - a_i)^2}$

Chernoff-Hoeffding Bound

- Chernoff-Hoeffding bound: If $S_n = \sum_{i=1}^{n} X_i$ and $X_i \in [a_i, b_i]$ are independent, then

$$\forall t > 0, \quad \Pr\left(\left|S_n - \mathbb{E}[S_n]\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Proof: Let $Y_i = X_i - \mathbb{E}[X_i]$ and $Y = S_n - \mathbb{E}[S_n] = \sum_{i=1}^{n} Y_i \implies \mathbb{E}[Y] = \mathbb{E}[Y_i] = 0$

$$\Pr\left(S_n - \mathbb{E}[S_n] \leq -t\right) = \Pr(Y \leq -t) \leq \Pr\left(e^{\lambda Y} \geq e^{-\lambda t}\right) \quad \text{(for any } \lambda < 0\text{)}$$

(Markov's inequality) $\leq e^{\lambda t} \cdot \mathbb{E}\left[e^{\lambda Y}\right] = e^{\lambda t} \cdot \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda Y_i}\right] = e^{\lambda t} \cdot \prod_{i=1}^{n} \mathbb{E}\left[e^{\lambda Y_i}\right]$ (independence)

(Hoeffding's lemma) $\leq \exp\left(\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right) = \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$

by choosing:

$$\lambda = \frac{-4t}{\sum_{i=1}^{n}(b_i - a_i)^2}$$

Chernoff-Hoeffding Bound

- Chernoff-Hoeffding bound:

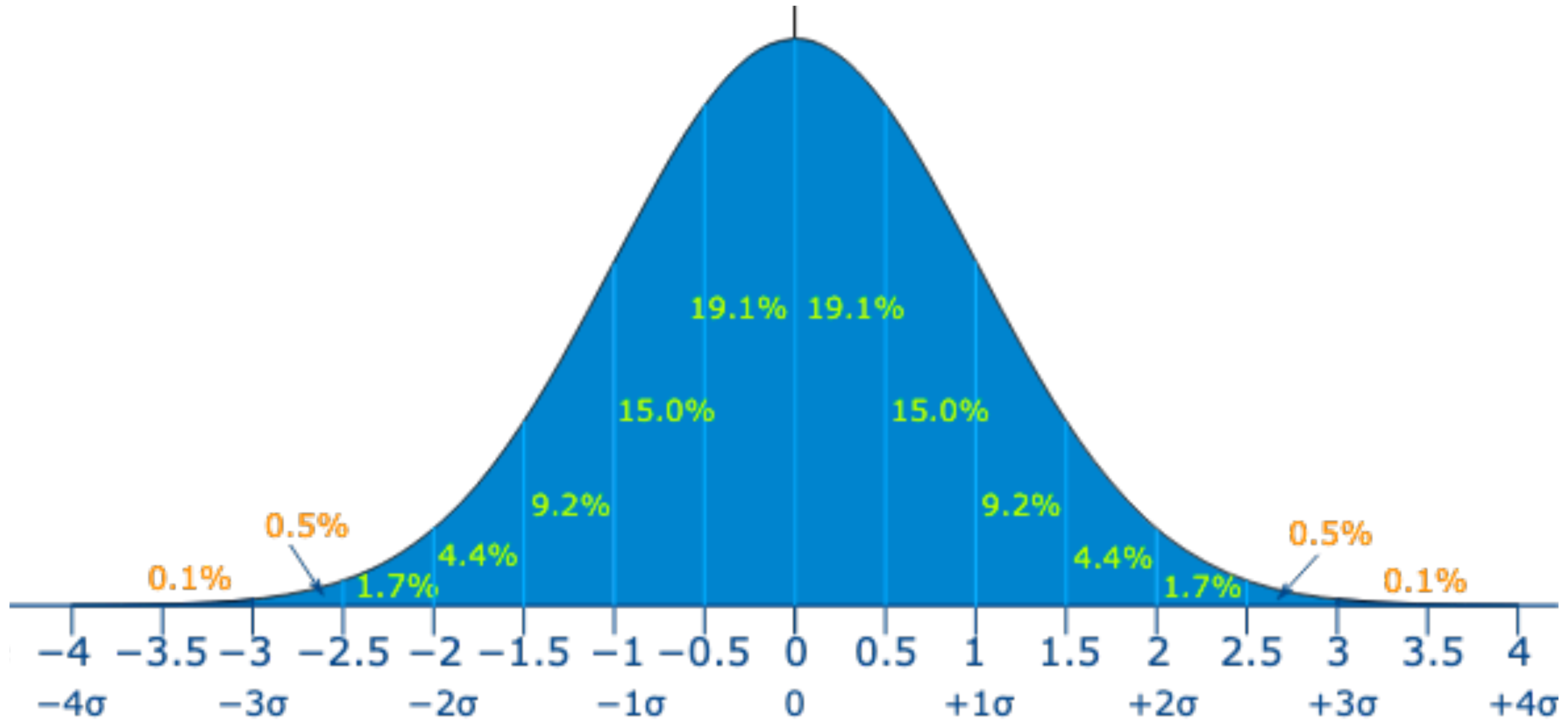
If $X_1, \ldots, X_n \in \{0,1\}$ are independent and $S_n = \sum_{i=1}^{n} X_i$, then for any $t > 0$,

$$\Pr\left(\left|S_n - \mathbb{E}[S_n]\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{n}\right)$$

In general, if $X_1, \ldots, X_n$ are independent and $X_i \in [a_i, b_i]$, $1 \leq i \leq n$, then for any $t > 0$,

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

Sub-Gaussian Tail



- <u>Chernoff-Hoeffding bound</u>:

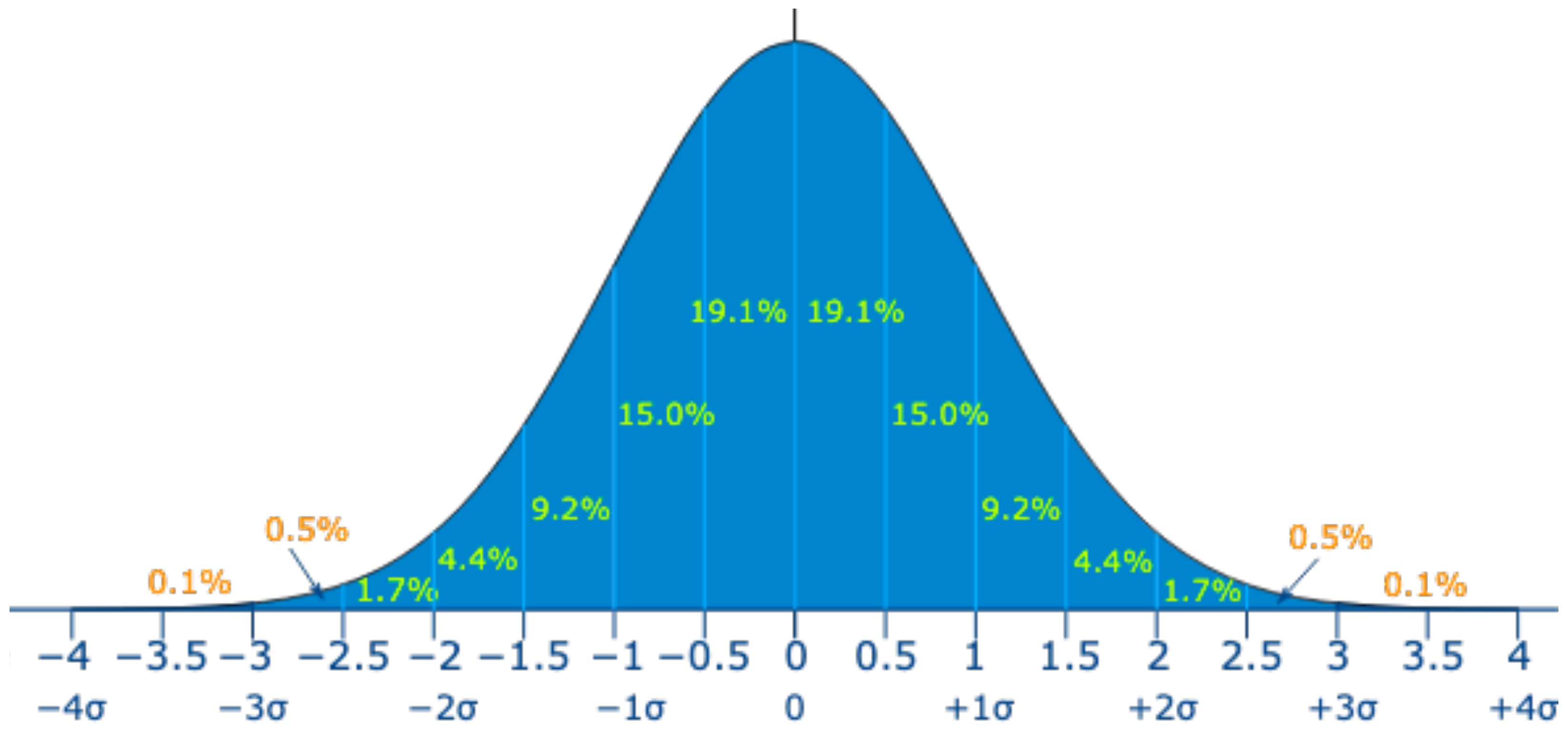
  If $X_1, \ldots, X_n \in \{0,1\}$ are independent **Poisson trials** and $S_n = \sum_{i=1}^{n} X_i$, then for any $t > 0$,

$$\Pr\left(\left|\frac{S_n - \mathbb{E}[S_n]}{\sqrt{n}/2}\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2}\right)$$

- Note that $\sigma(S_n) = \sqrt{\mathbf{Var}[S_n]} = \sqrt{\sum_{i=1}^{n} \mathbf{Var}[X_i]} \leq \sqrt{n}/2$

- The "worst-case" standardized S_n has a <u>sub-Gaussian</u> tail $e^{-\Omega(t^2)}$

Sub-Gaussian Tail



- <u>Chernoff-Hoeffding bound</u>:

  If $S_n = \sum_{i=1}^{n} X_i$, where $X_i \in [a_i, b_i]$, $1 \le i \le n$, are independent, then for any $t > 0$,

$$\Pr\left(\left|\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\sum_{i=1}^{n}(b_i - a_i)^2/4}}\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{2}\right)$$

- $Z \in [a,b] \implies \left|Z - \frac{a+b}{2}\right| \le \frac{b-a}{2} \implies \mathbf{Var}[Z] = \mathbf{Var}\left[Z - \frac{a+b}{2}\right] \le \mathbb{E}\left[\left(Z - \frac{a+b}{2}\right)^2\right] \le \frac{(b-a)^2}{4}$

- The "worst-case" standardized S_n has a <u>sub-Gaussian</u> tail $e^{-\Omega(t^2)}$

Controlling a Fair Voting

- In a society of n isolated (independent) and neutral (uniform) peoples, how many peoples are there enough to manipulate the result of a majority vote with $1-\delta$ certainty?
- Let $S_n = X_1 + \cdots + X_n$ for i.i.d. Bernoulli random variables $X_1, \ldots, X_n$ with parameter 1/2.

$$\Pr\left(\left|S_n - (n - S_n)\right| \geq t\right) = \Pr\left(\left|S_n - \mathbb{E}[S_n]\right| \geq \frac{t}{2}\right)$$

$$\leq 2\exp\left(-\frac{t^2}{2n}\right) \leq \delta$$

- A clique of $t \geq \sqrt{2n\ln(2/\delta)}$ peoples is enough

Error Reduction (two-sided case)

- Decision problem $f : \{0,1\}^* \to \{0,1\}$.
- Monte Carlo randomized algorithm $\mathscr{A}$ with ***two-sided*** error:
  - $\forall x \in \{0,1\}^*: \quad \Pr(\mathscr{A}(x) = f(x)) \geq \frac{1}{2} + p$
- $\mathscr{A}^n$: independently run $\mathscr{A}$ for n times, return majority of the n outputs

$$\Pr(\mathscr{A}^n(x) \neq f(x)) = \Pr\left(S_n \leq \frac{n}{2}\right) = \Pr\left(S_n \leq \mathbb{E}[S_n] - pn\right) \leq \exp(-2p^2 n) \leq \delta$$

where $S_n = X_1 + \cdots X_n$, and $X_i = I[\mathscr{A}(x) = f(x)$ in ith run]

when $n \geq \frac{1}{2p^2} \ln \frac{1}{\delta}$

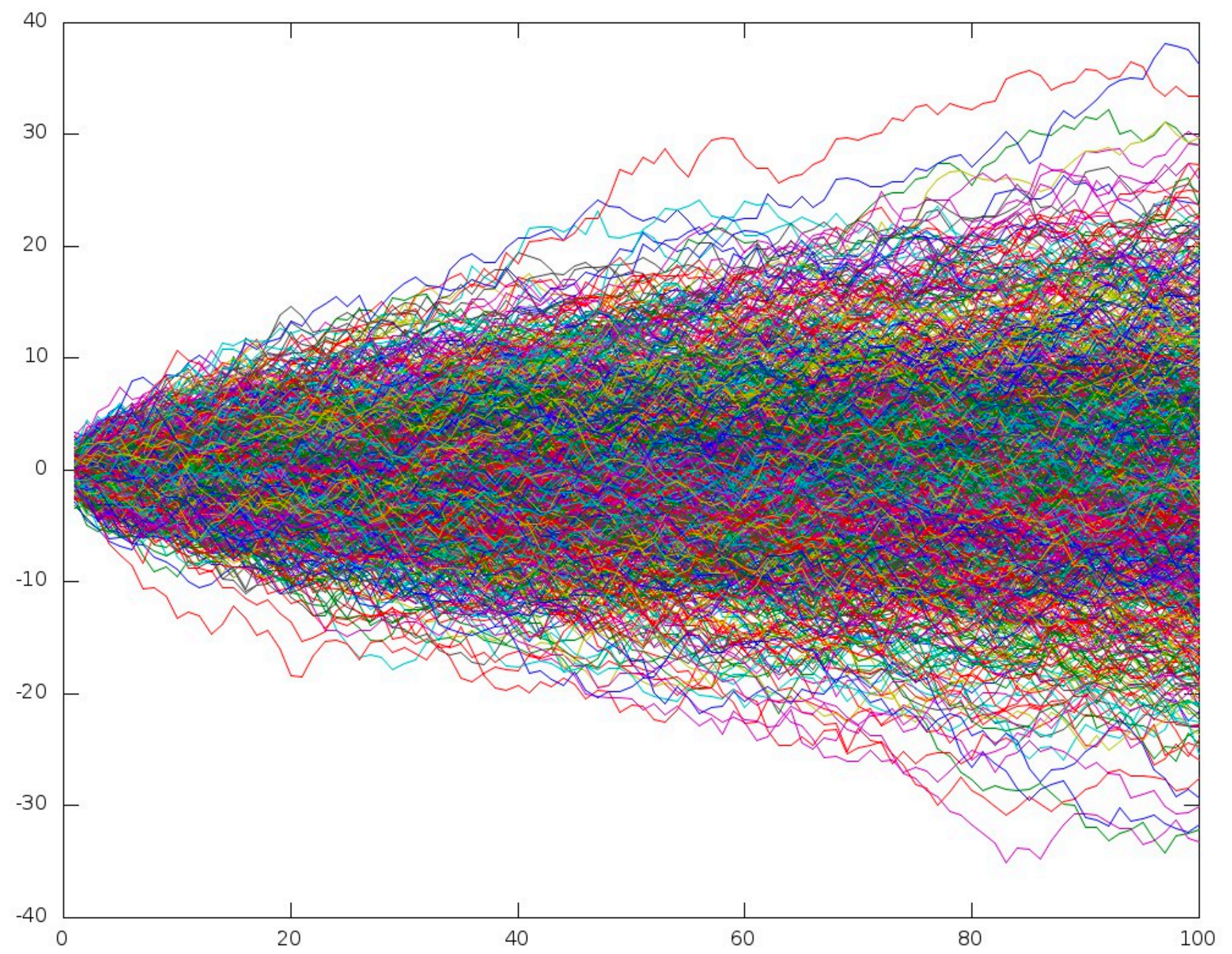
The Median Trick

- Computation problem $f : \{0,1\}^* \to \mathbb{R}$.
- *Randomized approximation* algorithm $\mathscr{A}$: $\forall x \in \{0,1\}^*$,
  - $\Pr\left(\mathscr{A}(x) \in (1 \pm \epsilon)f(x)\right) = \Pr\left((1-\epsilon)f(x) \le \mathscr{A}(x) \le (1+\epsilon)f(x)\right) \ge \frac{1}{2} + p$
- $\mathscr{A}^n$: independently run $\mathscr{A}$ for n times, return **median** of the n outputs
  - Let $X_i = I[\mathscr{A}(x) \in (1 \pm \epsilon)f(x)$ in the ith run of $\mathscr{A}(x)] \implies \mathbb{E}[X_i] \ge 1/2 + p$
  - **Observation:** $\mathscr{A}^n(x) \in (1 \pm \epsilon)f(x)$ if $S_n = X_1 + \cdots + X_n > \frac{n}{2}$

$$\Pr\left(\mathscr{A}(x) \notin (1 \pm \epsilon)f(x)\right) \le \Pr\left(S_n \le \frac{n}{2}\right) \le \Pr\left(S_n \le \mathbb{E}[S_n] - np\right) \le e^{-2p^2n} \le \delta$$

when $n \ge \frac{1}{2p^2} \ln \frac{1}{\delta}$

The Method of Bounded Differences



The Method of Bounded Differences

- <u>McDiarmid's Inequality</u>:

  Let $X_1, \ldots, X_n$ be **independent** random variables, where $X_i \in \mathcal{X}_i$ for all i.

  If $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfies the **<u>bounded differences property</u>**:

$$\forall i\text{:} \quad \sup_{x_1 \in \mathcal{X}_1, \ldots, x_n \in \mathcal{X}_n, x_i' \in \mathcal{X}_i} \left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right| \le c_i$$

  then for any $t > 0$,

$$\Pr\left(\left| f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \right| \ge t \right) \le 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2} \right)$$

- **Chernoff-Hoeffding:** f is sum of $[a_i, b_i]$-bounded variables

Every Lipschitz function is approximately a constant function in product measures.

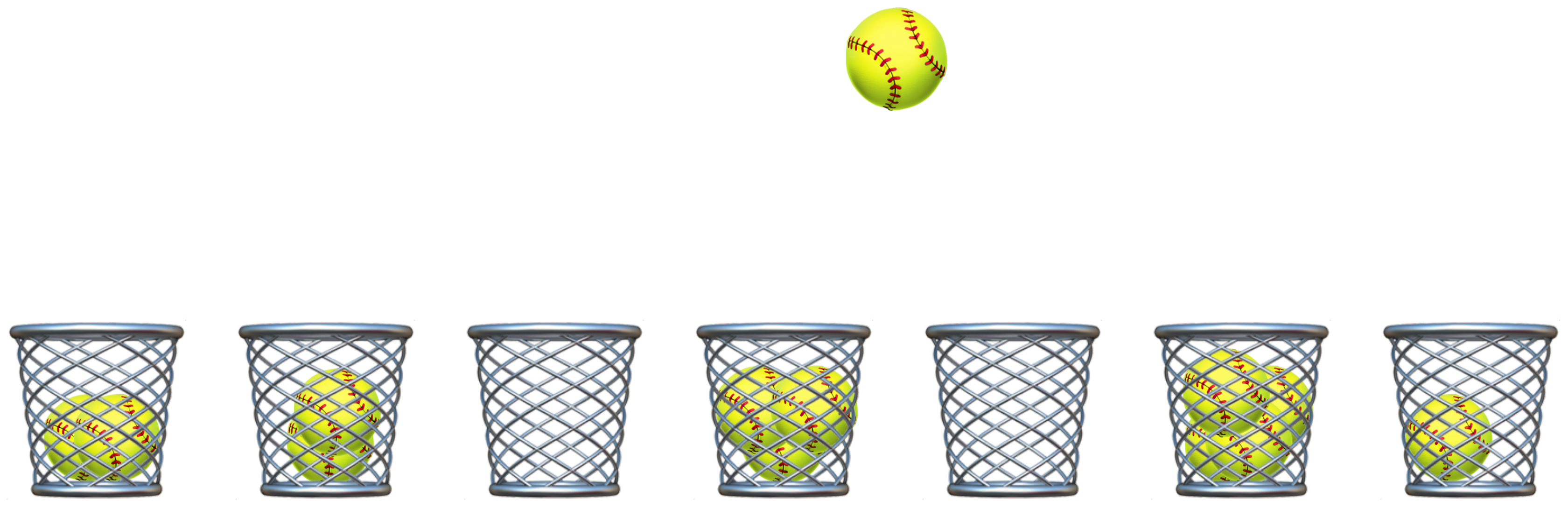
Pattern Matching

- $s = (s_1, \ldots, s_n) \in Q^n$: uniform random string of n letters from alphabet Q with $|Q| = q$
- For **pattern** $\pi \in Q^k$, let X be the number of appearances of π in s as substring

$$X = \sum_{i=1}^{n-k+1} I[\pi = s_{i,i+1,\ldots,i+k-1}] \quad = f(s_1, \ldots, s_n) \text{ has } k\text{-bounded differences}$$

- **Linearity of expectation:** $\mathbb{E}[X] = \sum_{i=1}^{n-k+1} \mathbb{E}[I_i] = (n-k+1)q^{-k}$
- **McDiarmid's Inequality:** $\Pr\left(\left|X - \mathbb{E}[X]\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{nk^2}\right)$

Empty Bins



- m balls are thrown into n bins. Let Y be the number of empty bins.

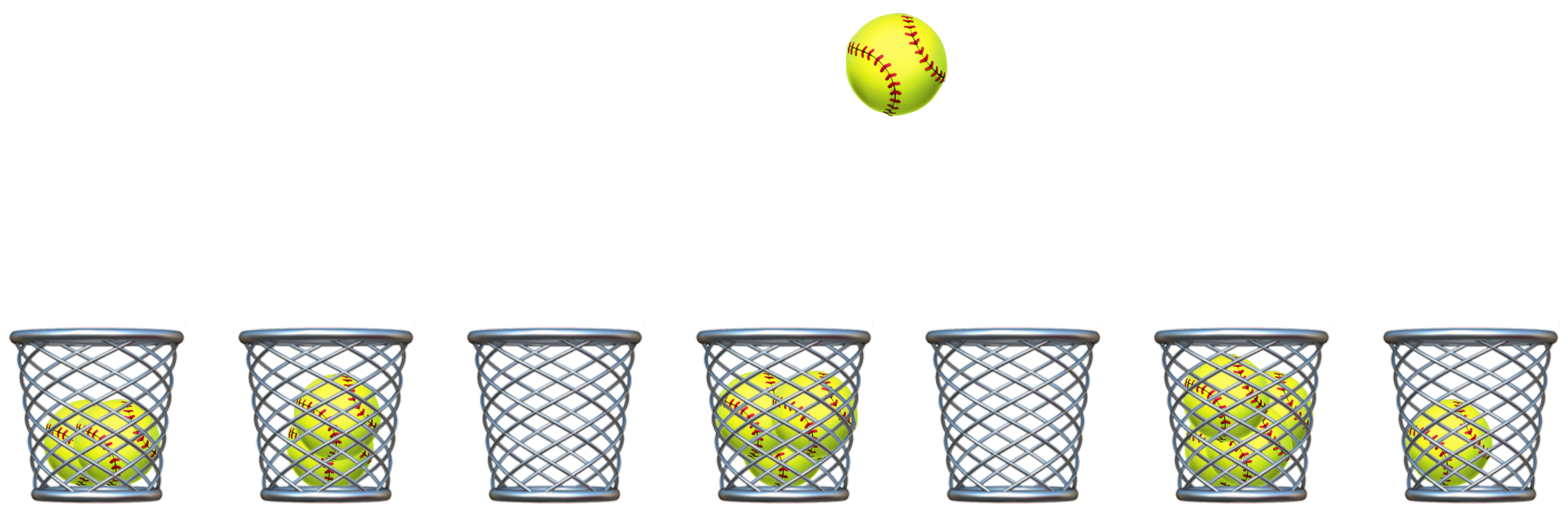
$$Y = \sum_{i=1}^{n} I[i\text{th bin is empty}]$$

- **Linearity of expectation:**

$$\mathbb{E}[Y] = \sum_{i=1}^{n} \Pr(i\text{th bin is empty}) = n\left(1 - \frac{1}{n}\right)^m$$

- Deviation: $\Pr\left(\left|Y - \mathbb{E}[Y]\right| \geq t\right) < ?$

Empty Bins



- m balls are thrown into n bins. Let X_j be the bin that receives the jth ball.

$X_1, \ldots, X_m \in [n]$ are uniform and independent.

- Let Y be the number of empty bins: (Applies to any $f(X_1, \ldots, X_m)$ with bounded differences)

$$Y = n - |\{X_1, X_2, \ldots, X_m\}|$$

has 1-bounded differences

- **McDiarmid's Inequality:**

$$\Pr\left(\left|Y - \mathbb{E}[Y]\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{m}\right)$$

The Method of Bounded Differences

- <u>McDiarmid's Inequality</u>:

  Let $X_1, \ldots, X_n$ be **independent** random variables, where $X_i \in \mathcal{X}_i$ for all i.

  If $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfies the <u>bounded differences property</u>:

$$\forall i: \quad \sup_{x_1 \in \mathcal{X}_1, \ldots, x_n \in \mathcal{X}_n, x_i' \in \mathcal{X}_i} \left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right| \le c_i$$

  then for any $t > 0$,

$$\Pr\left(\left| f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \right| \ge t \right) \le 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2} \right)$$

Every Lipschitz function is approximately a constant function in product measures.

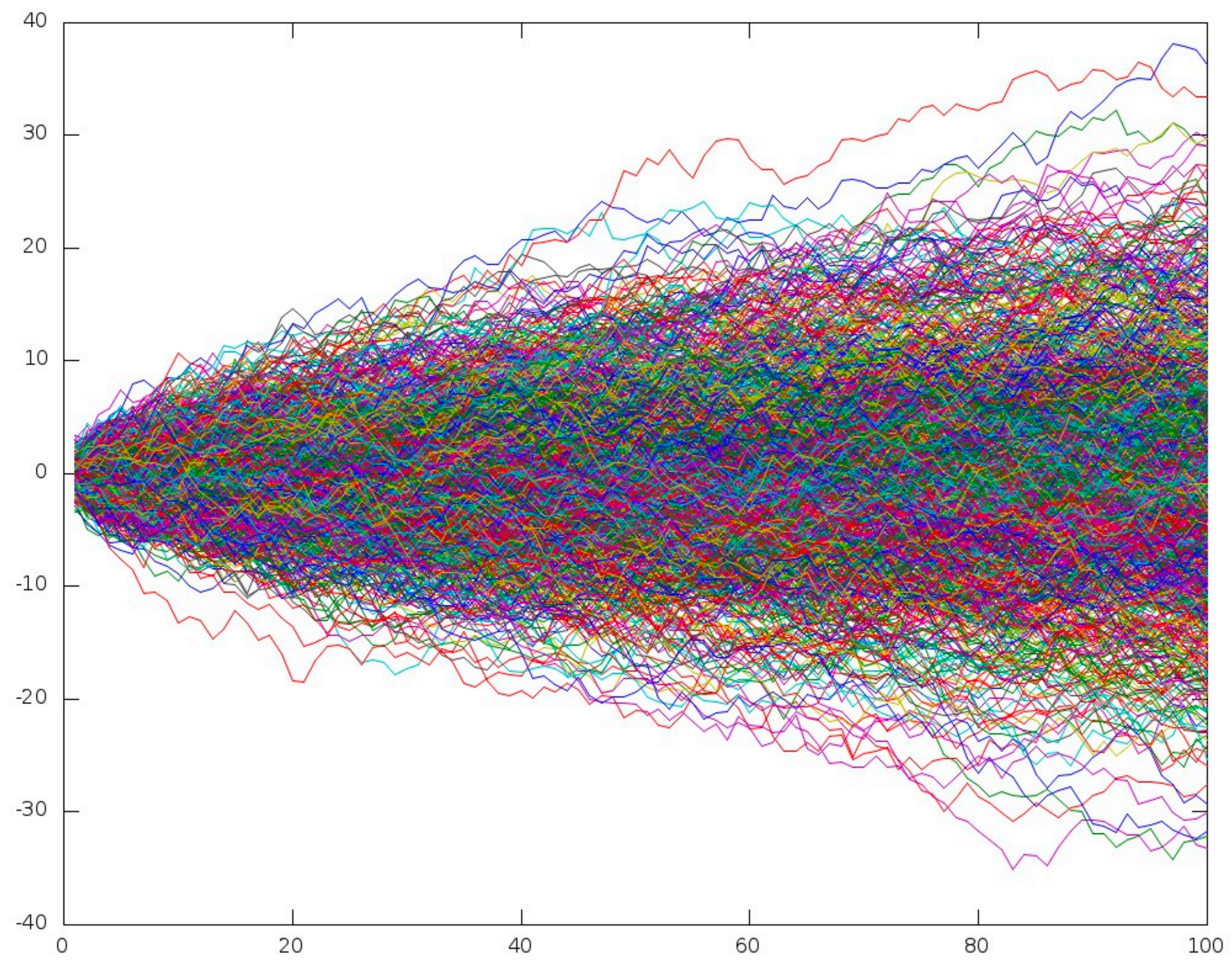
Doob Sequence

- The **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$

$$Y_0 = \mathbb{E}\left[f(X_1, \ldots, X_n)\right] \dashrightarrow f(X_1, \ldots, X_n) = Y_n$$

no information   full information

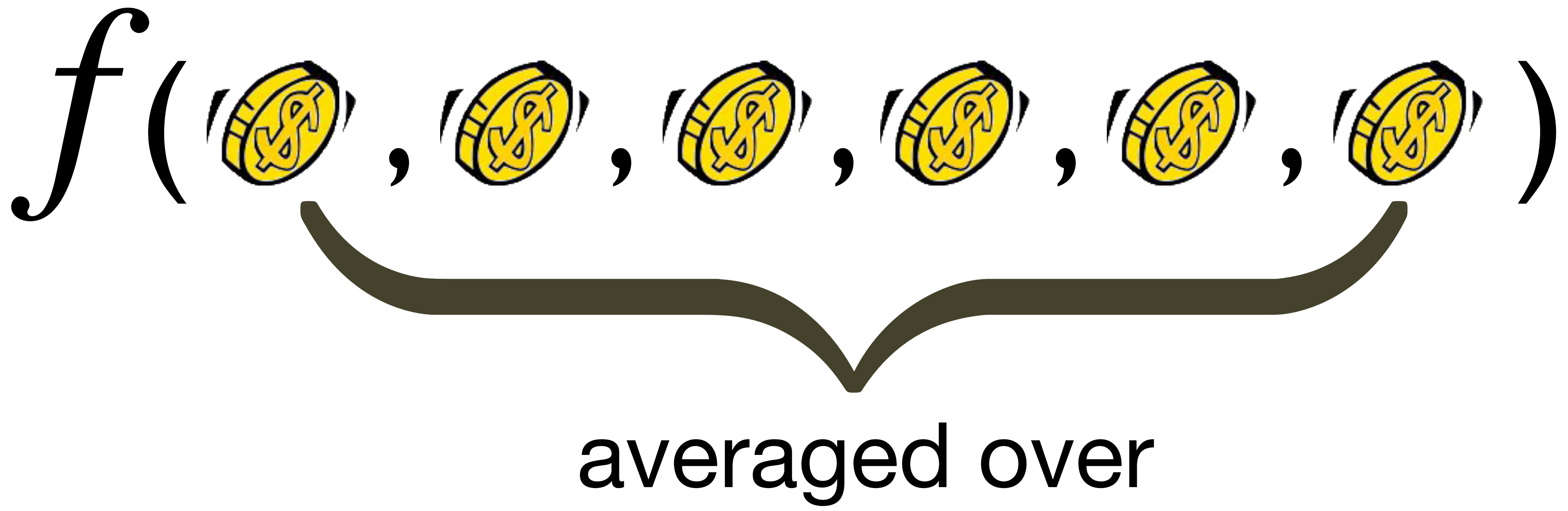


$$\Pr\left(\left|f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]\right| < t\right)$$

Doob Sequence

- The **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$

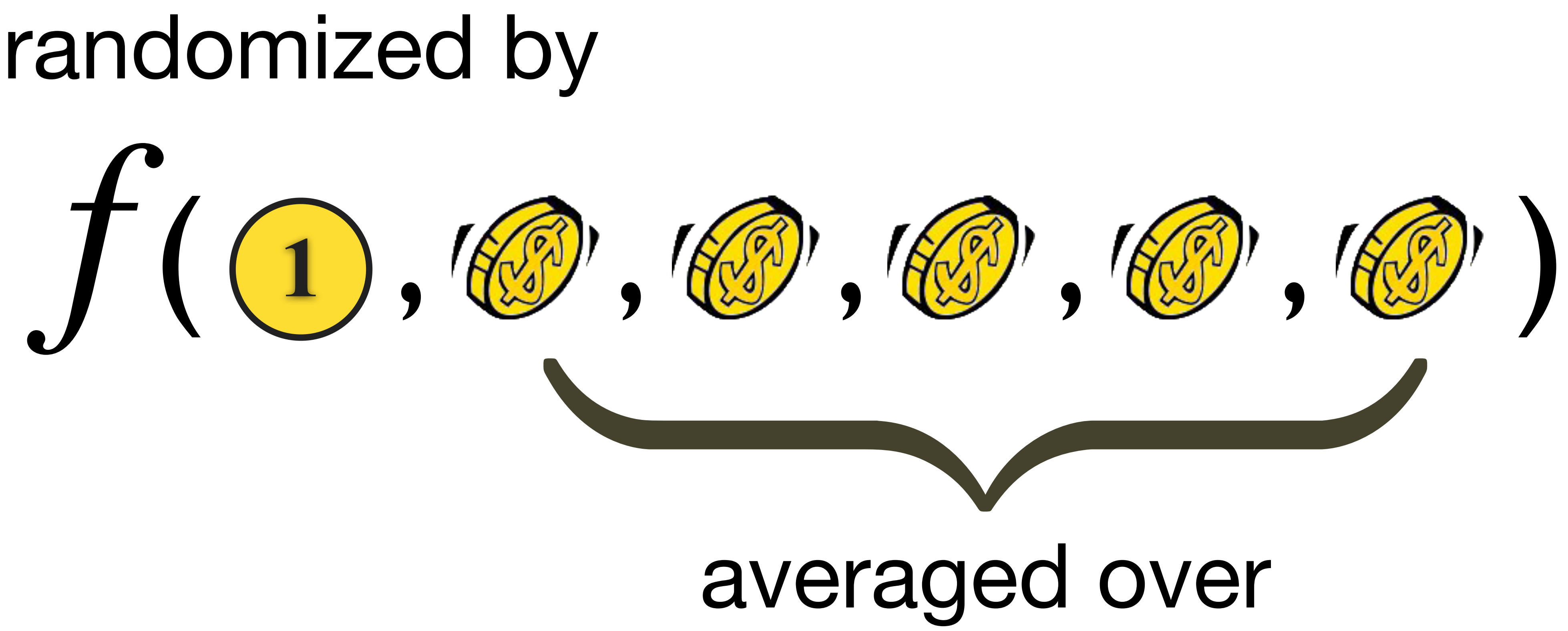


$$\mathbb{E}[f] = Y_0$$

Doob Sequence

- The **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \leq i \leq n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$

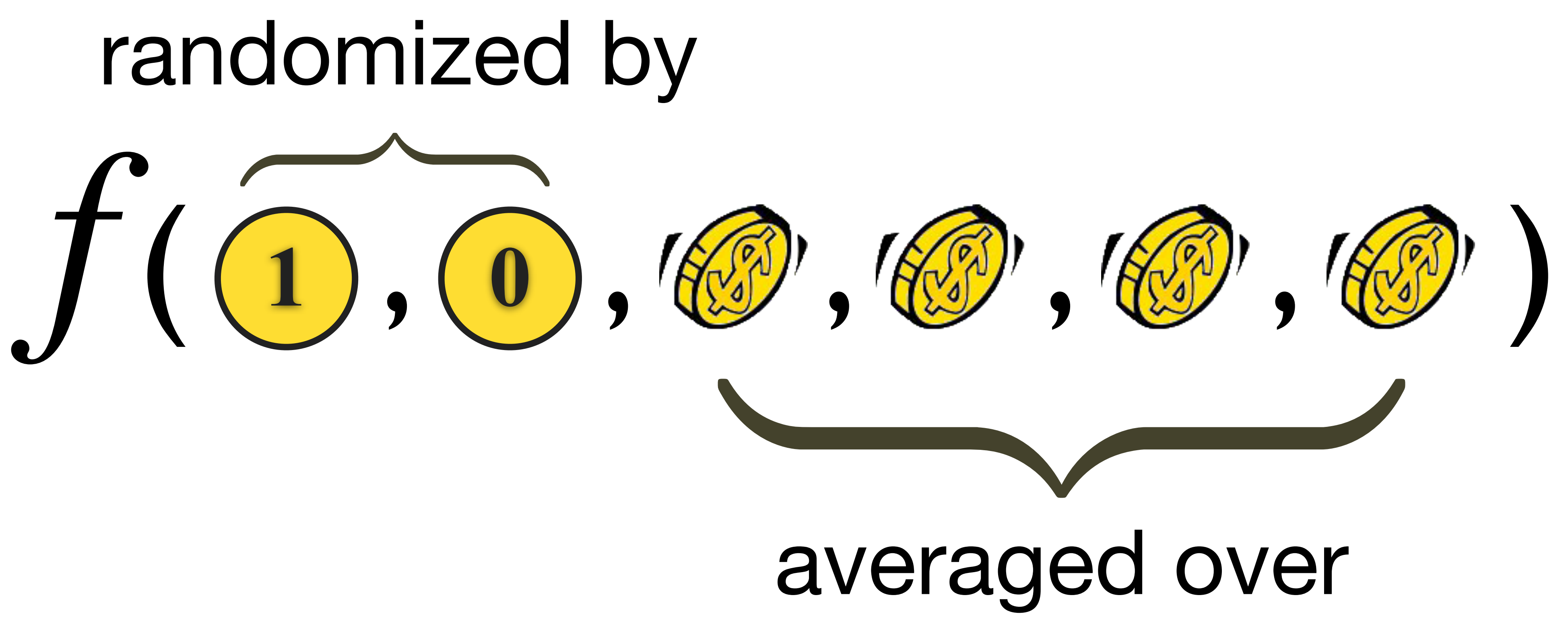


$$\mathbb{E}[f] = Y_0 \to Y_1$$

Doob Sequence

- The **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$

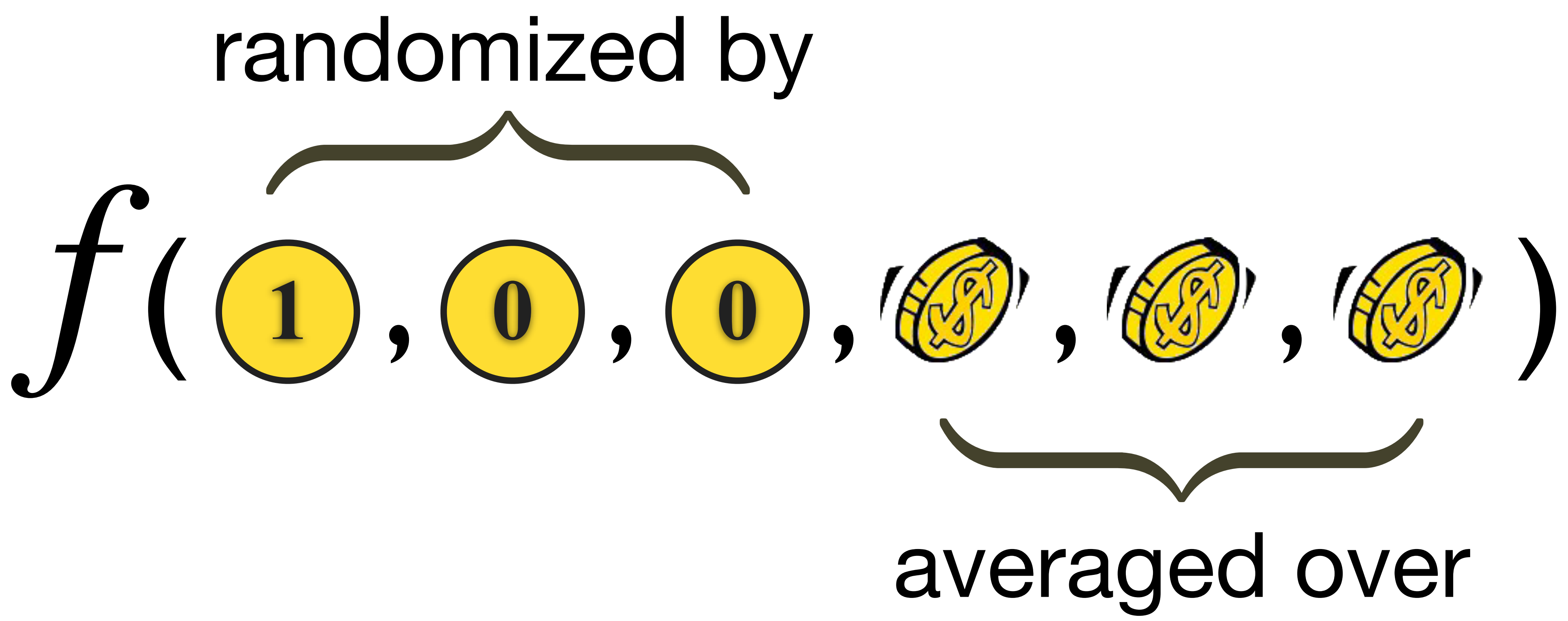


$$\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2$$

Doob Sequence

- The <u>**Doob sequence**</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$

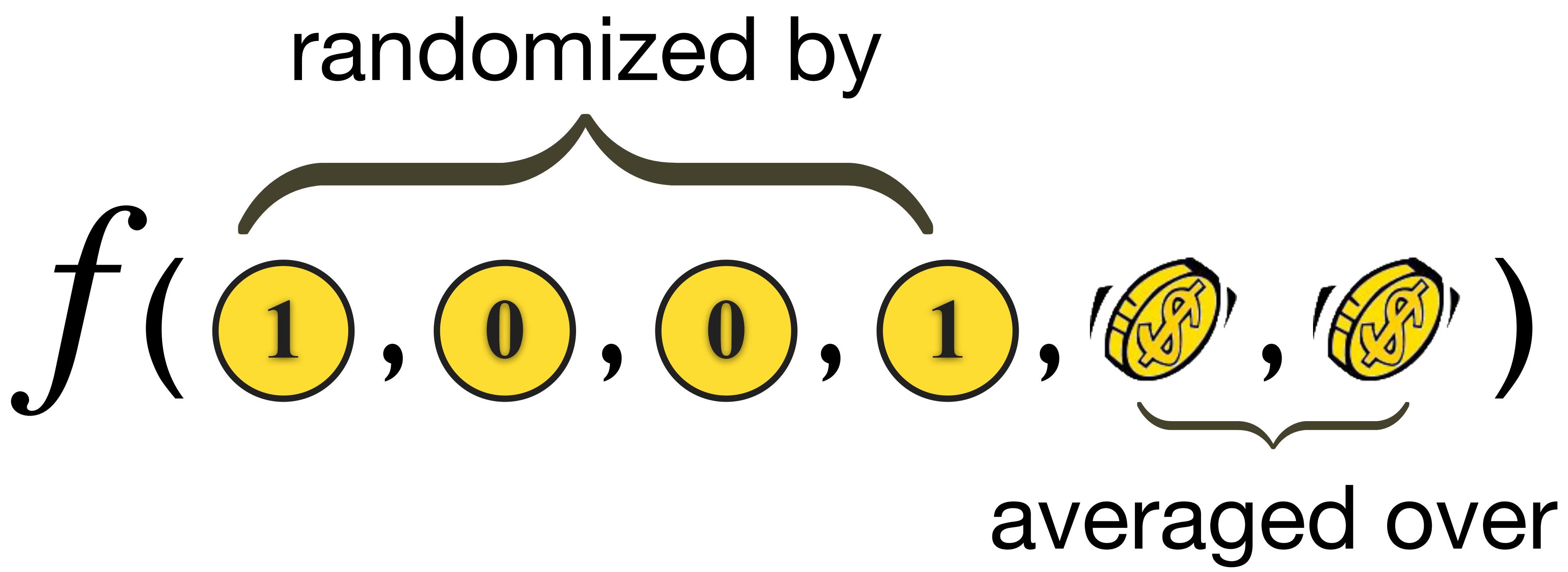


$$\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2 \to Y_3$$

Doob Sequence

- The <u>**Doob sequence**</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \leq i \leq n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$

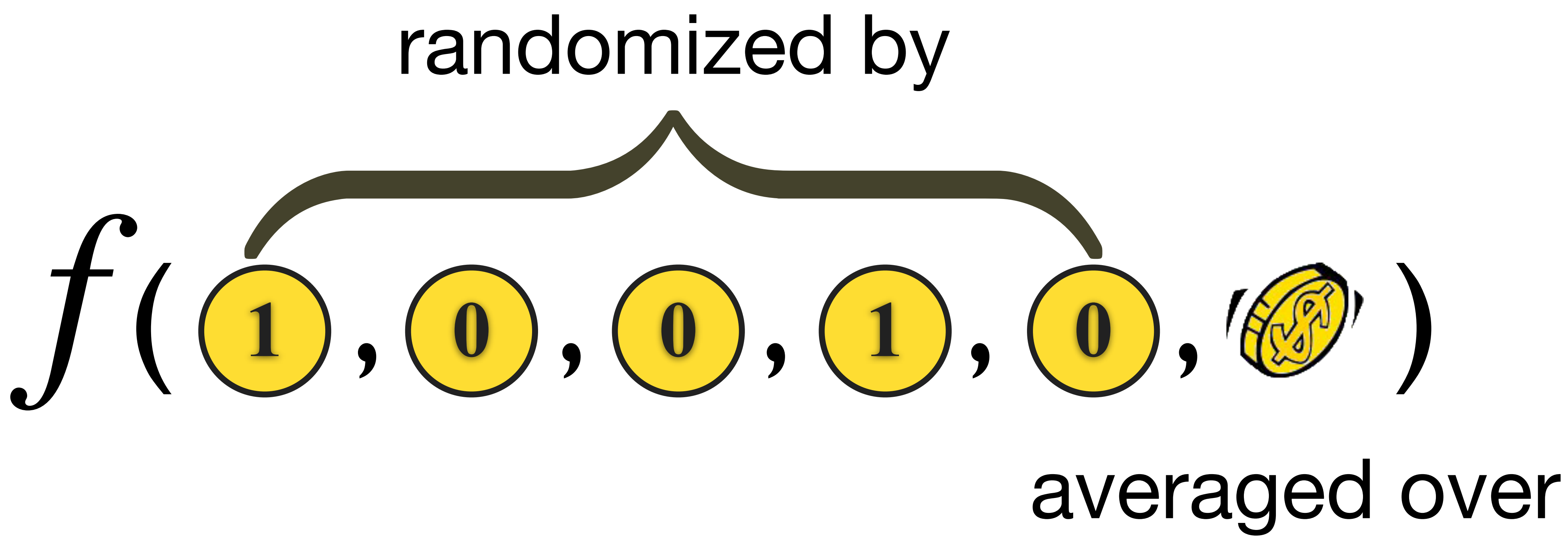


$$\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2 \to Y_3 \to Y_4$$

Doob Sequence

- The <u>**Doob sequence**</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$



$$\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2 \to Y_3 \to Y_4 \to Y_5$$

Doob Sequence

- The <u>**Doob sequence**</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$

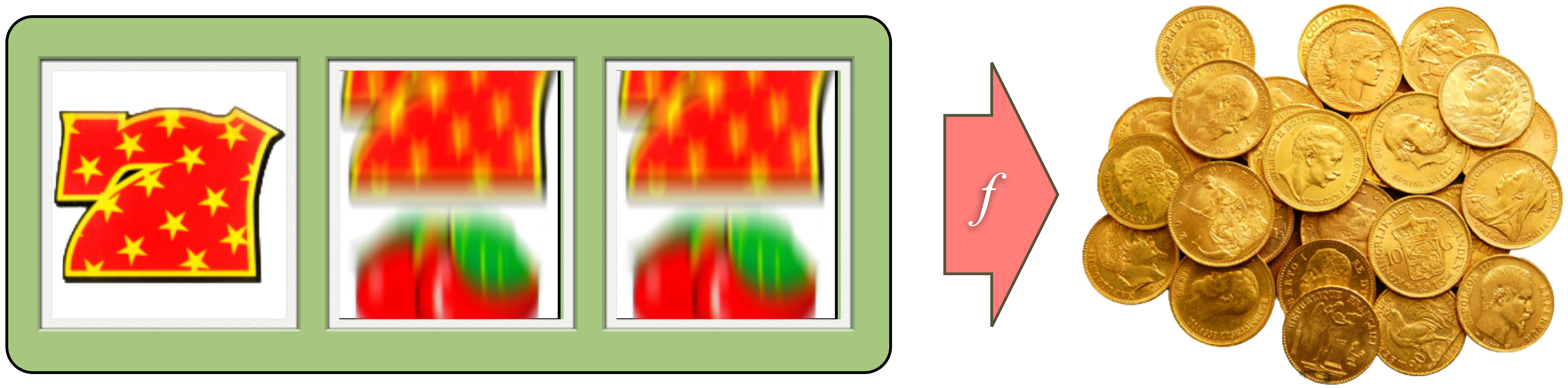
$$f(\overbrace{1, 0, 0, 1, 0, 1}^{\text{randomized by}})$$

no information $\quad \mathbb{E}[f] = Y_0 \to Y_1 \to Y_2 \to Y_3 \to Y_4 \to Y_5 \to Y_6 = f \quad$ full information

Doob Sequence

- The <u>Doob sequence</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$



Doob Sequence

- The <u>Doob sequence</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$

- <u>Martingale property</u>: $\mathbb{E}\left[Y_i \mid X_1, \ldots, X_{i-1}\right] = Y_{i-1}$

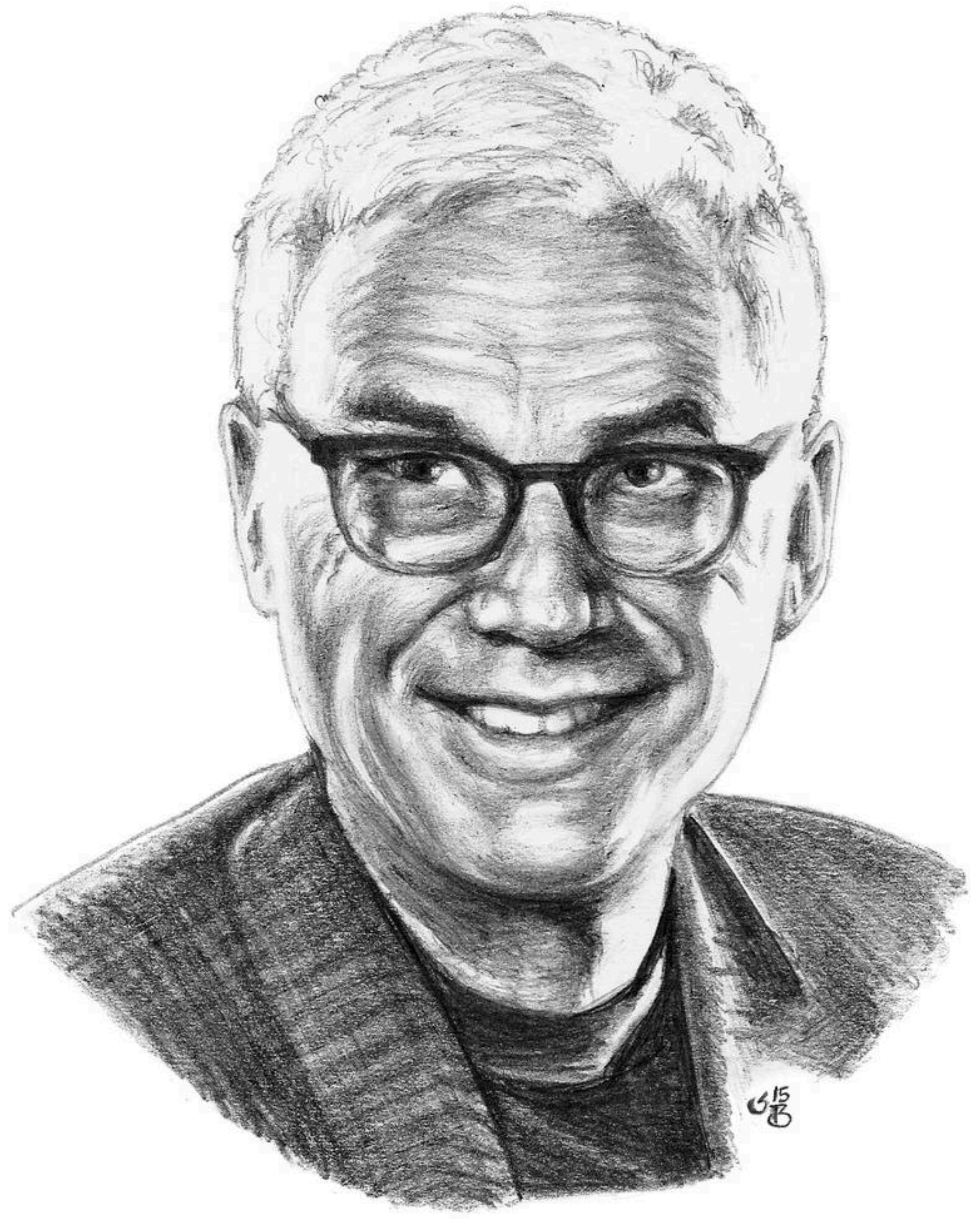
Proof: $\mathbb{E}\left[Y_i \mid X_1, \ldots, X_{i-1}\right]$

$$= \mathbb{E}\left[\mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right] \mid X_1, \ldots, X_{i-1}\right]$$

$$= \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_{i-1}\right]$$

$$= Y_{i-1}$$

because $\mathbb{E}\left[\mathbb{E}[Z \mid Y, X] \mid X\right] = \mathbb{E}\left[Z \mid X\right]$



Joseph Doob

Martingale

Martingale (鞅)

- A sequence $\{Y_n : n \geq 0\}$ of random variables is a martingale with respect to another sequence $\{X_n : n \geq 0\}$ if, for all $n \geq 0$,
  - $\mathbb{E}\left[|Y_n|\right] < \infty$
  - $\mathbb{E}\left[Y_{n+1} \mid X_0, X_1, \ldots, X_n\right] = Y_n$ (martingale property)
- By definition: Y_n is a function of $X_0, X_1, \ldots, X_n$
- Current capital Y_n in a **fair gambling game** with outcomes $X_0, X_1, \ldots, X_n$
  - Super-martingale (上鞅): $\mathbb{E}\left[Y_{n+1} \mid X_0, X_1, \ldots, X_n\right] \leq Y_n$
  - Sub-martingale (下鞅): $\mathbb{E}\left[Y_{n+1} \mid X_0, X_1, \ldots, X_n\right] \geq Y_n$

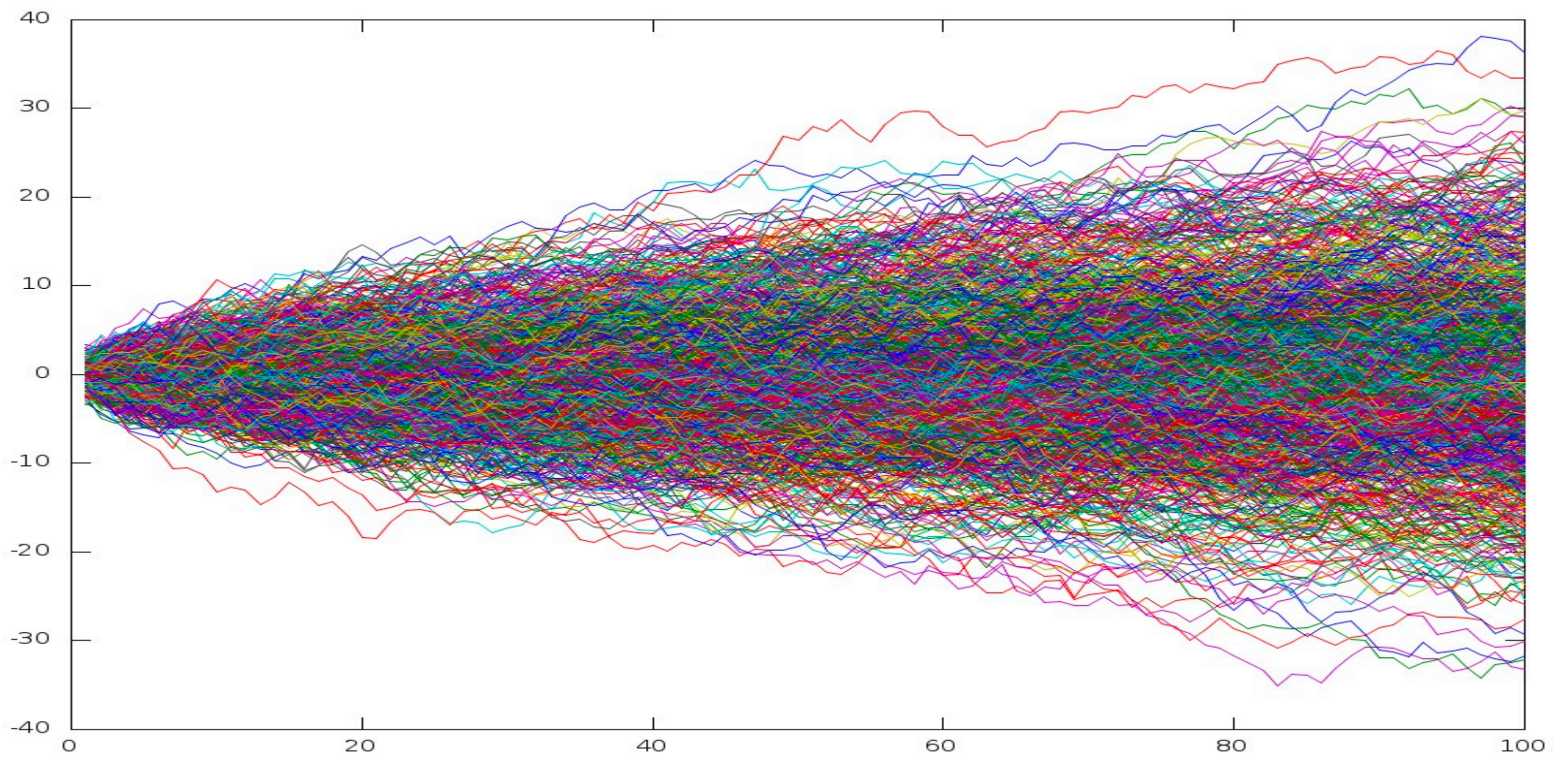
Martingale (鞅)

- A sequence $\{Y_n : n \geq 0\}$ of random variables is a **martingale** with respect to another sequence $\{X_n : n \geq 0\}$ if, for all $n \geq 0$,
  - $\mathbb{E}\left[|Y_n|\right] < \infty$
  - $\mathbb{E}\left[Y_{n+1} \mid X_0, X_1, \ldots, X_n\right] = Y_n \quad (\Longrightarrow$ Y_n is a function of $X_0, \ldots, X_n)$
- $\{X_n : n \geq 0\}$ are defined on the probability space $(\Omega, \Sigma, \Pr)$
  - $(X_0, X_1, \ldots, X_n)$ defines a sub-σ-field $\Sigma_n \subseteq \Sigma$ (the smallest σ-field s.t. $(X_0, \ldots, X_n)$ is Σ_n-measurable)
  - $\{\Sigma_n : n \geq 0\}$ is a **filtration** of Σ, i.e. $\Sigma_0 \subseteq \Sigma_1 \subseteq \cdots \subseteq \Sigma$
  - The **martingale property** is expressed as $\mathbb{E}\left[Y_{n+1} \mid \Sigma_n\right] = Y_n$

Examples of Martingale

- **Doob martingale:** $Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$
    - **vertex/edge exposure martingale** for random graph
- Capital in a **fair gambling game** (arbitrary betting strategy)
- **Unbiased 1D random walk:** $Y_n = \sum_{i=1}^{n} X_i$ with *i.i.d.* uniform $X_i \in \{-1,1\}$
- **de Moivre's martingale:** $Y_n = (p/(1-p))^{X_n}$, where $X_n = \sum_{i=1}^{n} X_i$ and $X_i \in \{-1,1\}$ are independent with $\Pr(X_i = 1) = p$
- **Polya's urn**: The urn contains marbles with different colors. At each turn, a marble is selected *u.a.r.*, and replaced with k marbles of that same color.

Studies of Martingale



- For **martingale** $\{Y_n : n \geq 0\}$ with respect to $\{X_n : n \geq 0\}$:

$$\mathbb{E}\left[\, Y_{n+1} \mid X_0, X_1, \ldots, X_n \,\right] = Y_n$$

- **Concentration of measure** (tail inequality): under what condition

$$\Pr\left(\, | Y_n - Y_0 | \geq t \right) \leq ?$$

- **Optional stopping theorem (OST)**: under what condition for a **stopping time** τ

$$\mathbb{E}[Y_\tau] = \mathbb{E}[Y_0]$$

Martingale Tail Inequality

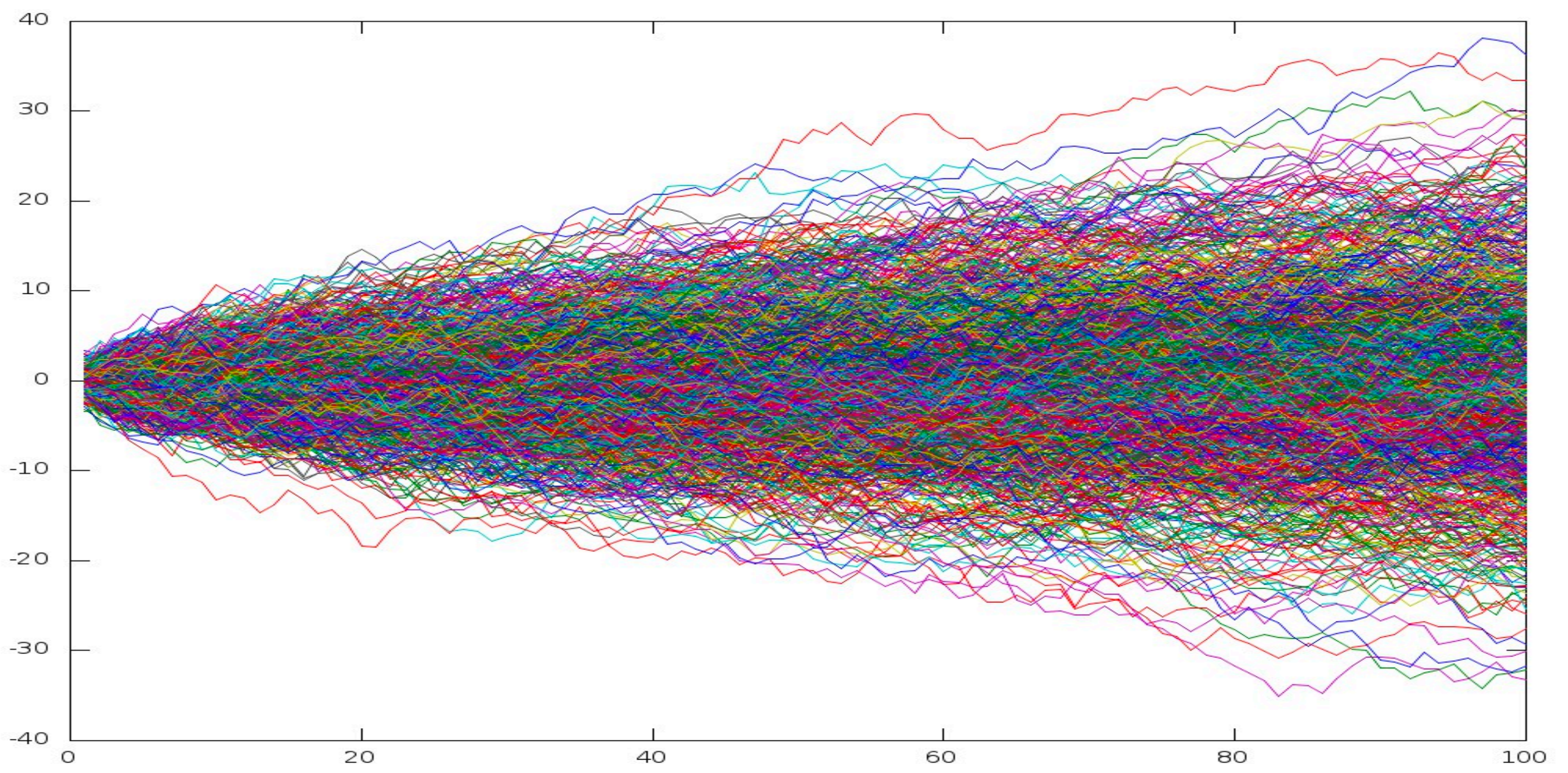
- **Azuma's inequality**: If a sequence $\{Y_n : n \geq 0\}$ is a **martingale** (with respect to some sequence $\{X_n : n \geq 0\}$), and for all $n \geq 1$,

$$\left| Y_n - Y_{n-1} \right| \leq c_n$$

then any $n \geq 1$ and any $t > 0$:

$$\Pr\left(\left| Y_n - Y_0 \right| \geq t \right) \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2} \right)$$

- **Intuition**: Your capital does not change too fast if
    - the game is fair (martingale)
    - the changes to the capital are bounded



The Method of Bounded Differences

- **Azuma's inequality**: If $\{Y_n : n \geq 0\}$ is a **martingale** with **bounded differences**

$$\forall n \geq 1: \left|Y_n - Y_{n-1}\right| \leq c_n \implies \Pr\left(\left|Y_n - Y_0\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

- **Doob sequence**: $Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right], 1 \leq i \leq n,$ is a **martingale**

- **The method of bounded differences**: If function $f(\boldsymbol{X})$ of random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$ satisfies an ***average-case*** **bounded differences property**:

$$\forall 1 \leq i \leq n, \left|\mathbb{E}\left[f(\boldsymbol{X}) \mid X_1, \ldots, X_i\right] - \mathbb{E}\left[f(\boldsymbol{X}) \mid X_1, \ldots, X_{i-1}\right]\right| \leq c_i$$

$\implies$ for any $t > 0$, $\Pr\left[\left|f(\boldsymbol{X}) - \mathbb{E}[f(\boldsymbol{X})]\right| \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$

Bounded Differences Properties

(Worst-Case $\implies$ Average-Case Bounded Differences)

- Function $f(\boldsymbol{X})$ of random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$
- If $X_i \in \mathcal{X}_i$ for all $1 \leq i \leq n$ are **independent**, and $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfies the **bounded differences property:**

$$\forall i: \quad \sup_{x_1 \in \mathcal{X}_1, \ldots, x_n \in \mathcal{X}_n, x_i' \in \mathcal{X}_i} \left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right| \leq c_i$$

then $f(\boldsymbol{X})$ satisfies the **average-case bounded differences property:**

$$\forall i: \quad \left| \mathbb{E}\left[f(\boldsymbol{X}) \mid X_1, \ldots, X_i \right] - \mathbb{E}\left[f(\boldsymbol{X}) \mid X_1, \ldots, X_{i-1} \right] \right| \leq c_i$$

$\implies$ for any $t > 0$, $\Pr\left[\left| f(\boldsymbol{X}) - \mathbb{E}[f(\boldsymbol{X})] \right| \geq t \right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2} \right)$

The Method of Bounded Differences

- <u>McDiarmid's Inequality</u>:

  Let $X_1, \ldots, X_n$ be **independent** random variables, where $X_i \in \mathcal{X}_i$ for all i.

  If $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfies the <u>bounded differences property</u>:

$$\forall i: \quad \sup_{x_1 \in \mathcal{X}_1, \ldots, x_n \in \mathcal{X}_n, x_i' \in \mathcal{X}_i} \left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right| \leq c_i$$

  then for any $t > 0$,

$$\Pr\left(\left| f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \right| \geq t \right) \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2} \right)$$

Martingale Tail Inequality

- Azuma's inequality: If a sequence $\{Y_n : n \geq 0\}$ is a martingale (with respect to some sequence $\{X_n : n \geq 0\}$), and for all $n \geq 1$,

$$\left|Y_n - Y_{n-1}\right| \leq c_n$$

then any $n \geq 1$ and $t > 0$:

$$\Pr\left(\left|Y_n - Y_0\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$$

Proof: Difference $D_i = Y_i - Y_{i-1}$ and Sum $S_n = \sum_{i=1}^{n} D_i = Y_n - Y_0$

New goal: $\Pr\left(\left|S_n\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$

Proof of Azuma's Inequality

- $\{Y_n : n \geq 0\}$ is a martingale (w.r.t. $\{X_n : n \geq 0\}$) satisfying $\left| Y_n - Y_{n-1} \right| \leq c_n$ for $n \geq 1$
- Difference $D_i = Y_i - Y_{i-1}$ and Sum $S_n = \sum_{i=1}^{n} D_i = Y_n - Y_0$
- $\{Y_n : n \geq 0\}$ is martingale w.r.t. $\{X_n : n \geq 0\}$

$$\implies \quad \mathbb{E}\left[D_n \mid X_0, \ldots, X_{n-1}\right] = \mathbb{E}\left[Y_n - Y_{n-1} \mid X_0, \ldots, X_{n-1}\right]$$
$$= \mathbb{E}\left[Y_n \mid X_0, \ldots, X_{n-1}\right] - \mathbb{E}\left[Y_{n-1} \mid X_0, \ldots, X_{n-1}\right] = Y_{n-1} - Y_{n-1} = 0$$

- Bounded difference: $\left| Y_n - Y_{n-1} \right| \leq c_n \implies D_n = Y_n - Y_{n-1} \in [-a_n, b_n]$ for $b_n - a_n = c_n$
- <u>Hoeffding's lemma</u>: $\mathbb{E}\left[D_n \mid X_0, \ldots, X_{n-1}\right] = 0$ and $D_n \in [-a_n, b_n]$ for $b_n - a_n = c_n$

$$\implies \quad \mathbb{E}\left[e^{\lambda D_n} \mid X_0, \ldots, X_{n-1}\right] \leq e^{\lambda^2 c_n^2/8}$$

Proof of Azuma's Inequality

- $\{Y_n : n \geq 0\}$ is a martingale (w.r.t. $\{X_n : n \geq 0\}$) satisfying $\left|Y_n - Y_{n-1}\right| \leq c_n$ for $n \geq 1$
- Difference $D_i = Y_i - Y_{i-1}$ and Sum $S_n = \sum_{i=1}^{n} D_i = Y_n - Y_0$

$$\mathbb{E}\left[e^{\lambda D_n} \mid X_0, \ldots, X_{n-1}\right] \leq e^{\lambda^2 c_n^2/8}$$

$$\mathbb{E}\left[e^{\lambda S_n}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\lambda S_n} \mid X_0, \ldots, X_{n-1}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[e^{\lambda(S_{n-1}+D_n)} \mid X_0, \ldots, X_{n-1}\right]\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[e^{\lambda S_{n-1}} \cdot e^{\lambda D_n} \mid X_0, \ldots, X_{n-1}\right]\right] = \mathbb{E}\left[e^{\lambda S_{n-1}} \cdot \mathbb{E}\left[e^{\lambda D_n} \mid X_0, \ldots, X_{n-1}\right]\right]$$

$$\leq \mathbb{E}\left[e^{\lambda S_{n-1}} \cdot e^{\lambda^2 c_n^2/8}\right] = e^{\lambda^2 c_n^2/8} \cdot \mathbb{E}\left[e^{\lambda S_{n-1}}\right]$$

- $\implies \mathbb{E}\left[e^{\lambda S_n}\right] \leq \exp\left(\frac{\lambda^2}{8} \sum_{i=1}^{n} c_i^2\right)$

Proof of Azuma's Inequality

- $\{Y_n : n \geq 0\}$ is a martingale (w.r.t. $\{X_n : n \geq 0\}$) satisfying $\left| Y_n - Y_{n-1} \right| \leq c_n$ for $n \geq 1$
- Difference $D_i = Y_i - Y_{i-1}$ and Sum $S_n = \sum_{i=1}^{n} D_i = Y_n - Y_0$

$$\mathbb{E}\left[e^{\lambda S_n}\right] \leq \exp\left(\frac{\lambda^2}{8}\sum_{i=1}^n c_i^2\right)$$

- (Upper tail) $\Pr(Y_n - Y_0 \geq t) = \Pr(S_n \geq t) \leq \Pr\left(e^{\lambda S_n} \geq e^{\lambda t}\right)$ (for any $\lambda > 0$)

(Markov) $\leq e^{-\lambda t} \cdot \mathbb{E}\left[e^{\lambda S_n}\right] \leq \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^n c_i^2\right) = \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$

by choosing: $\lambda = \frac{4t}{\sum_{i=1}^n c_i^2}$

- (Lower tail) $\Pr(Y_n - Y_0 \leq -t) = \Pr(S_n \leq -t) \leq \Pr\left(e^{\lambda S_n} \geq e^{-\lambda t}\right)$ (for any $\lambda < 0$)

(Markov) $\leq e^{\lambda t} \cdot \mathbb{E}\left[e^{\lambda S_n}\right] \leq \exp\left(\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^n c_i^2\right) = \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$

by choosing: $\lambda = \frac{-4t}{\sum_{i=1}^n c_i^2}$

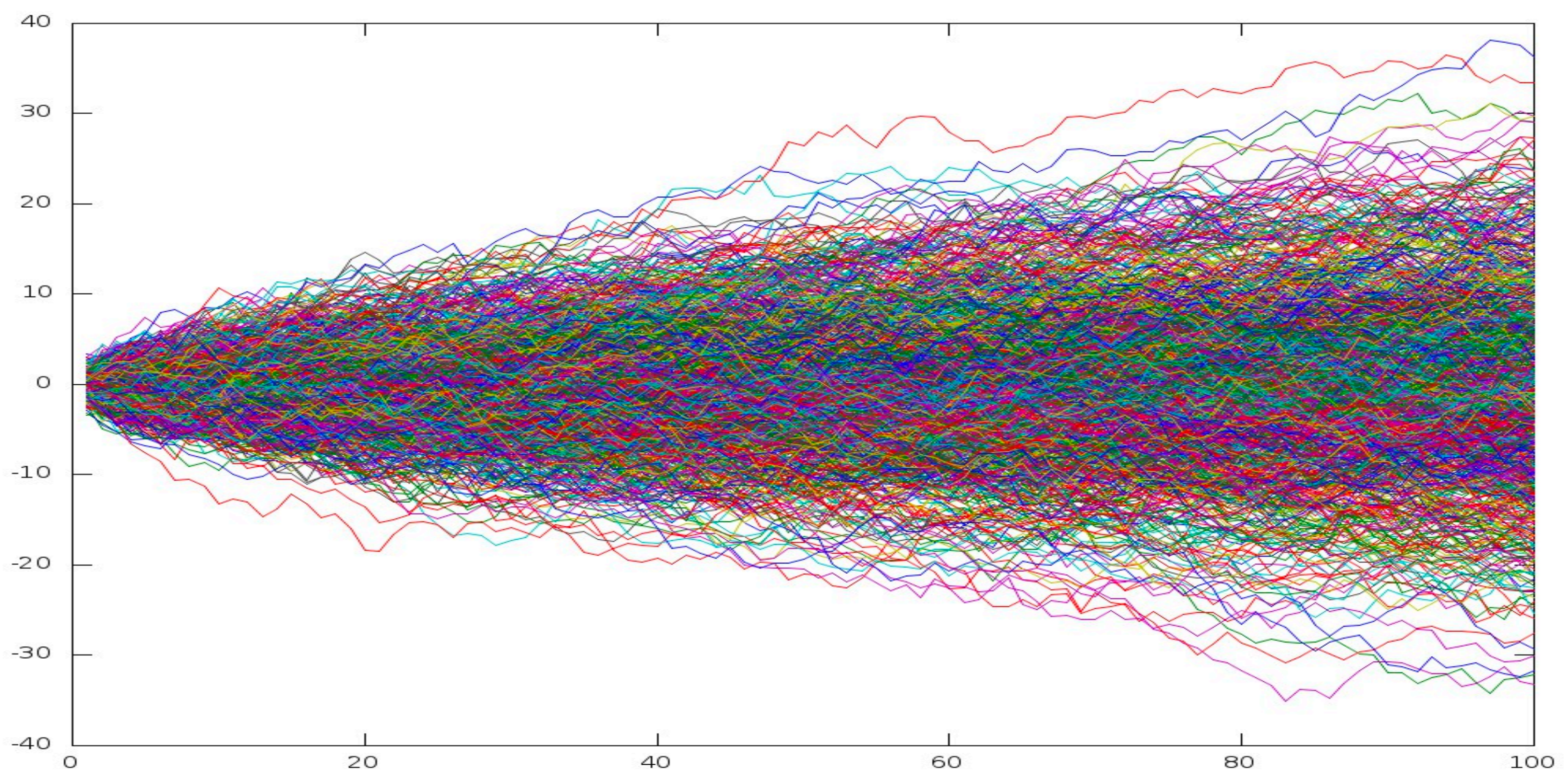
Martingale Tail Inequality

- **<u>Azuma's inequality</u>**: If a sequence $\{Y_n : n \geq 0\}$ is a martingale (with respect to some sequence $\{X_n : n \geq 0\}$), and for all $n \geq 1$,

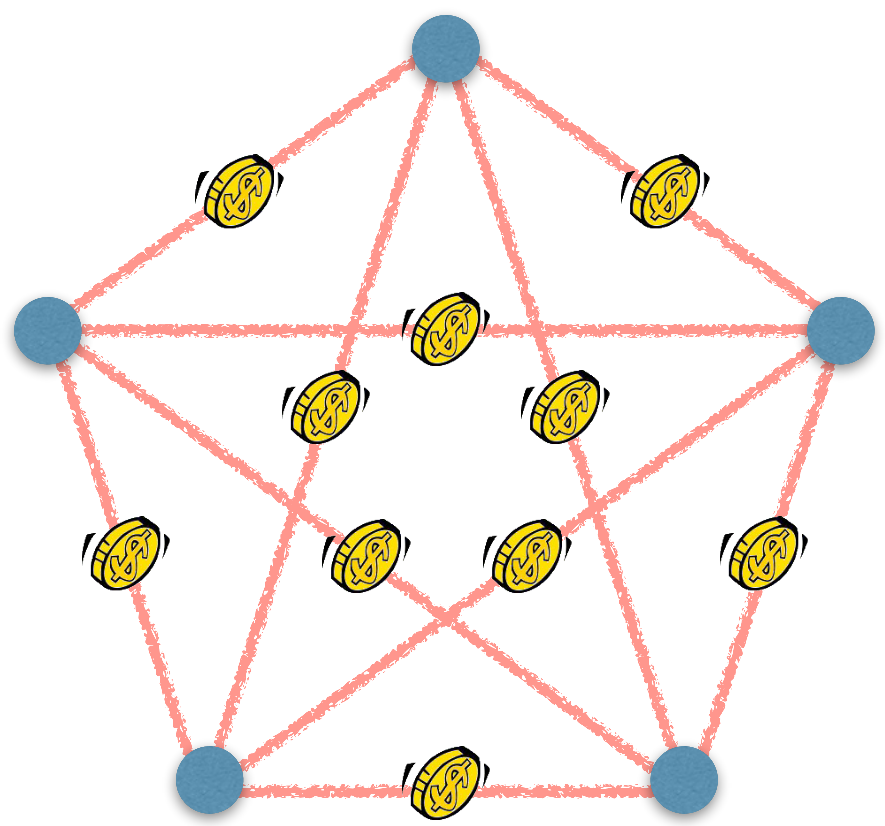
$$\left| Y_n - Y_{n-1} \right| \leq c_n$$

then any $n \geq 1$ and $t > 0$:

$$\Pr\left(\left| Y_n - Y_0 \right| \geq t \right) \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2} \right)$$



Martingales from Random Graph



- Random graph $G \sim G(n,p)$
- Graph parameter $f(G)$: chromatic number, clique number, expansion, …
- Edge exposure martingale:

$$Y_i = \mathbb{E}[f(G) \mid X_1, \ldots, X_i], \qquad 1 \le i \le \binom{n}{2}$$

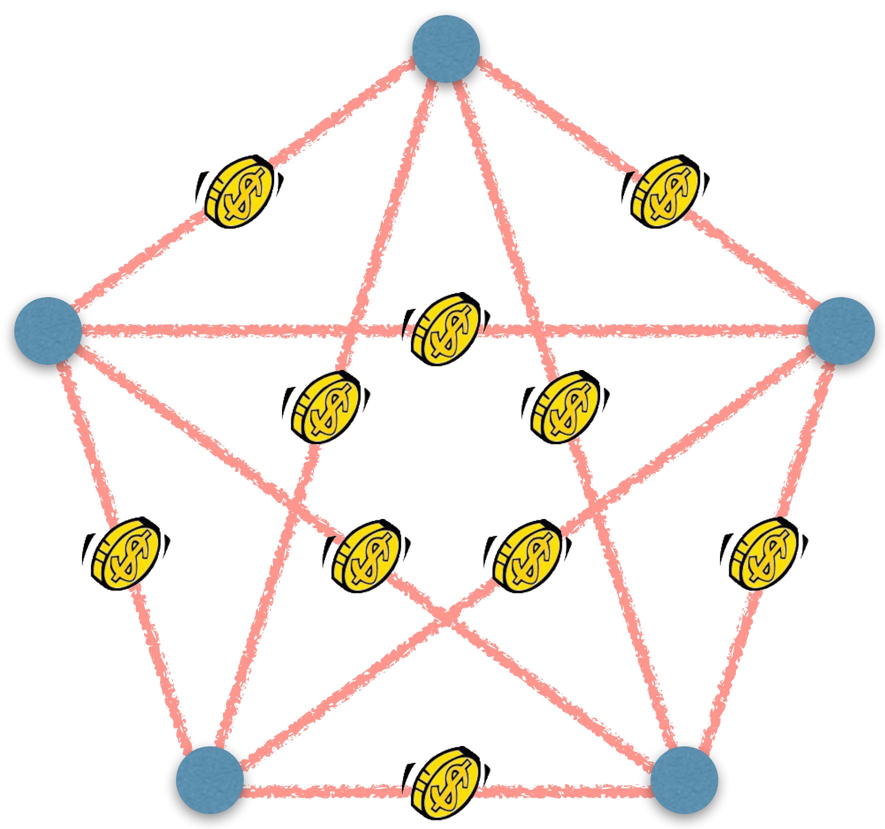
where $X_1, \ldots, X_{\binom{n}{2}}$ are i.i.d. Bernoulli(p), s.t. $X_i = I[i\text{th vertex pair is an edge in } G]$

- Vertex exposure martingale:

$$Y_i = \mathbb{E}[f(G) \mid X_1, \ldots, X_i], \qquad 1 \le i \le n$$

where $X_i = G[\{1,\ldots,i\}]$ is the subgraph of G induced by the first i vertices

Coloring Random Graph



- Random graph $G \sim G(n, p)$
- **Chromatic number** $\chi(G)$: smallest number of colors to properly color G
- <u>Vertex exposure martingale</u>:

$$Y_i = \mathbb{E}[f(G) \mid X_1, \ldots, X_i], \qquad 1 \le i \le n$$

where $X_i = G[\{1,\ldots,i\}]$ is the subgraph of G induced by the first i vertices

- **Observation**: a vertex can always be assign a new color to properly color G

$$|Y_i - Y_{i-1}| \le 1$$

- **Azuma's inequality**: $\Pr\left(|\chi(G) - \mathbb{E}[\chi(G)]| \ge \sqrt{cn}\right) \le 2\mathrm{e}^{-2c}$