

# Probability Theory & Mathematical Statistics

Concentration of Measure

# Bernoulli's Law of Large Number

## In *Ars Conjectandi* (1713)



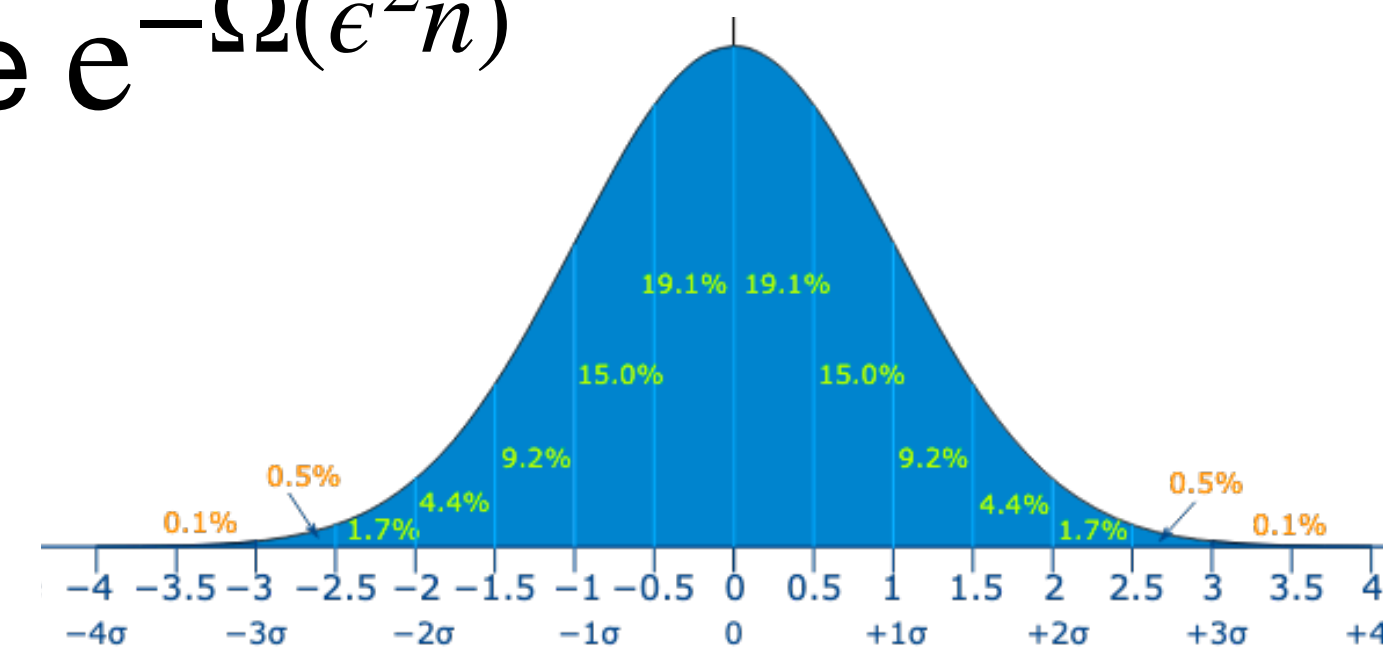
- If  $X_1, X_2, \dots \sim \text{Bernoulli}(p)$  be *i.i.d.*, then  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} p$ , i.e.  $\forall \epsilon > 0$ ,

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

- 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  $e^{-\Omega(\epsilon^2 n)}$



# Sum of Independent Trials

(Poisson binomial distribution)

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

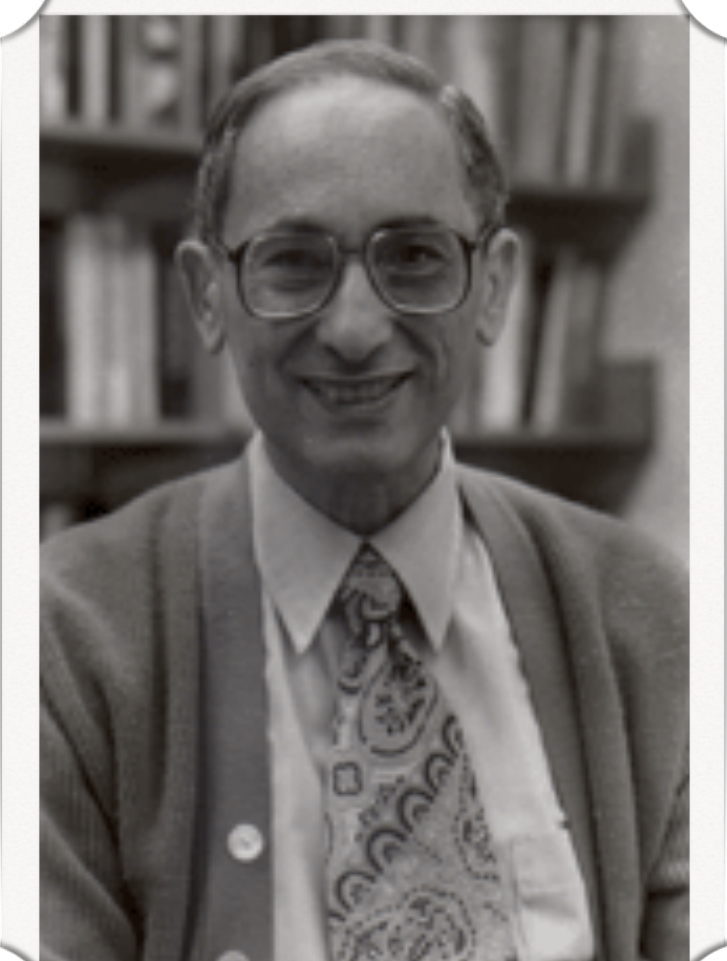
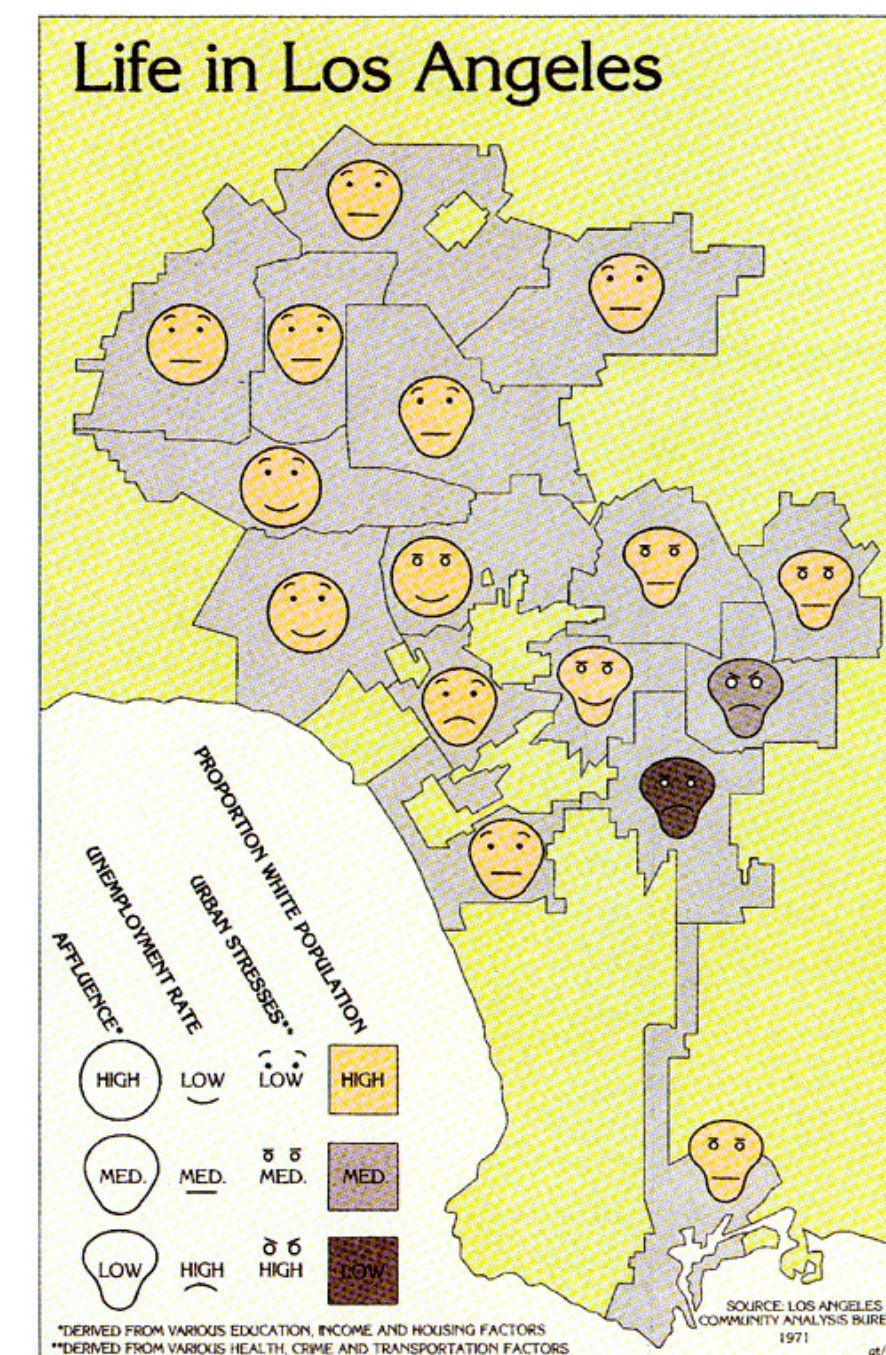
$$S_n = \sum_{i=1}^n X_i$$

(called Poisson binomial random variable)

- Deviation / concentration / tail bounds:

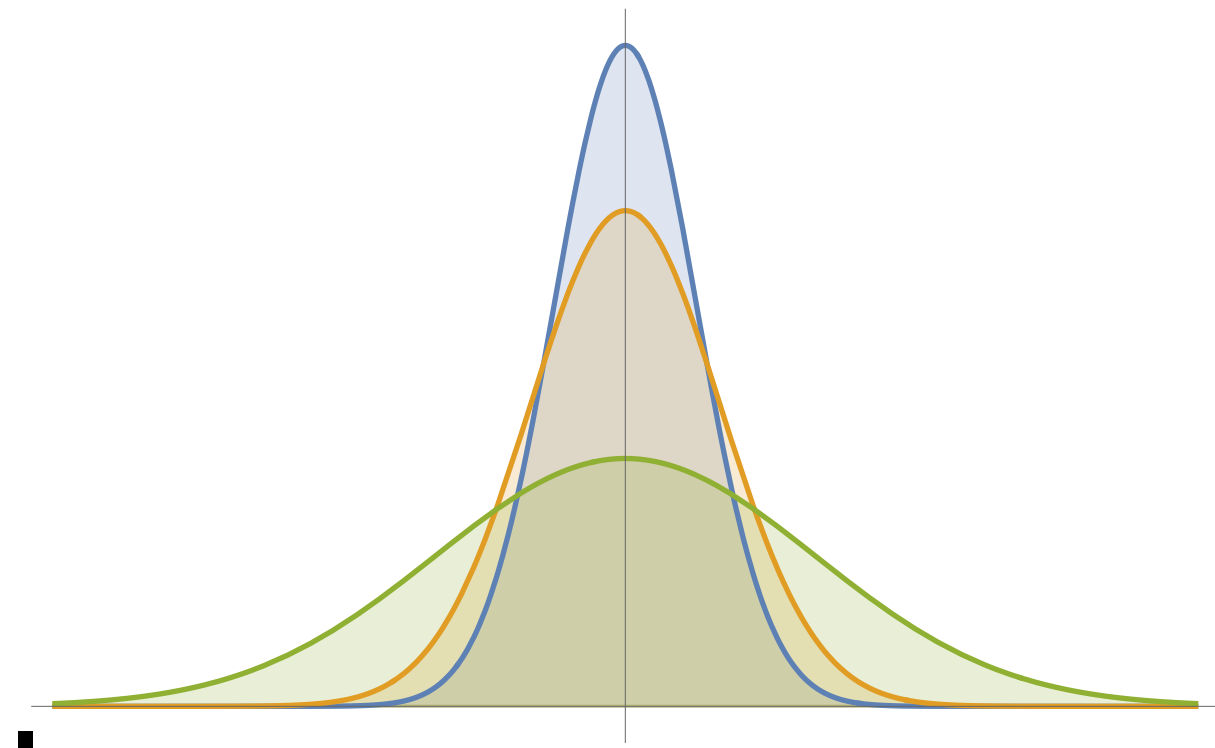
$$\Pr \left( |S_n - \mathbb{E}[S_n]| \geq ? \right) \leq ?$$

# Chernoff-Hoeffding Bounds



Herman Chernoff

# Chernoff Bound



- Chernoff bound: Let  $X_1, \dots, X_n \in \{0,1\}$  be independent trials

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

(Poisson binomial RV  
with mean  $\mu$ )

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

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

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

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



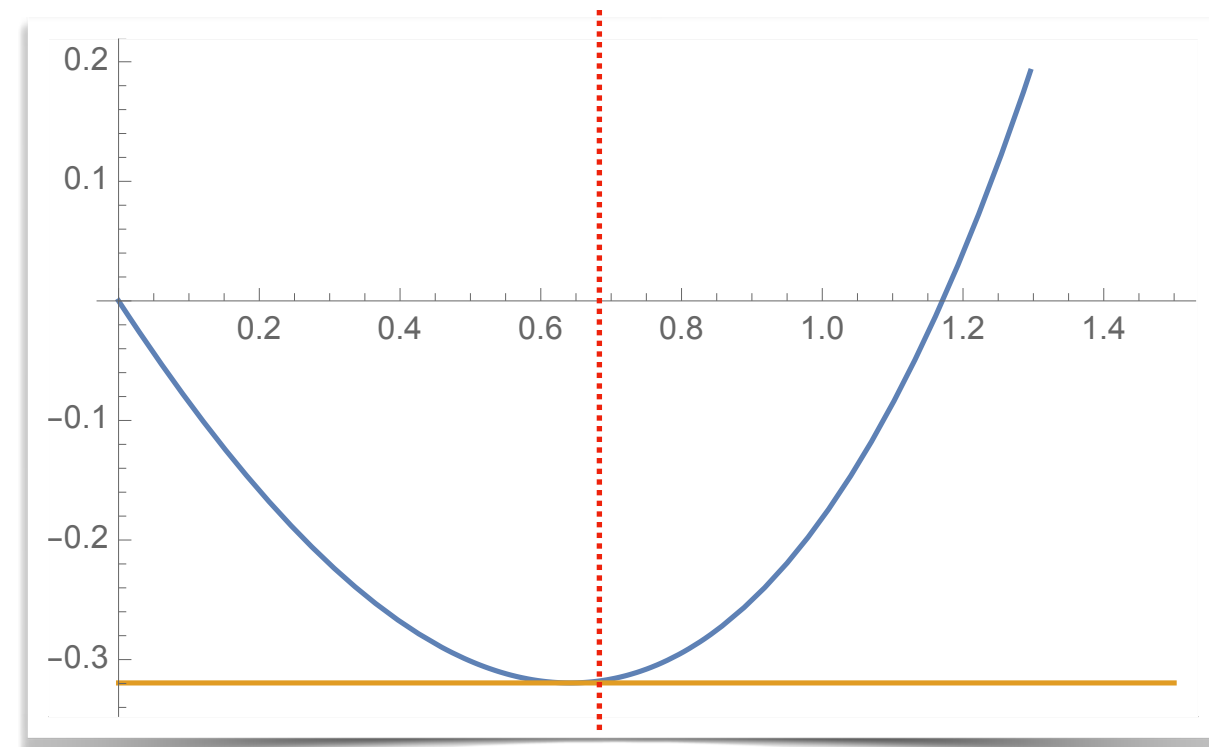
# Chernoff Bound (Upper Tail)

- Let  $X_1, \dots, X_n \in \{0, 1\}$  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:**  $\Pr(X \geq (1 + \delta)\mu) \leq \Pr(e^{tX} \geq e^{t(1+\delta)\mu})$  (for any  $t > 0$ )

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



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

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

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

# Chernoff Bound (Lower Tail)

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

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

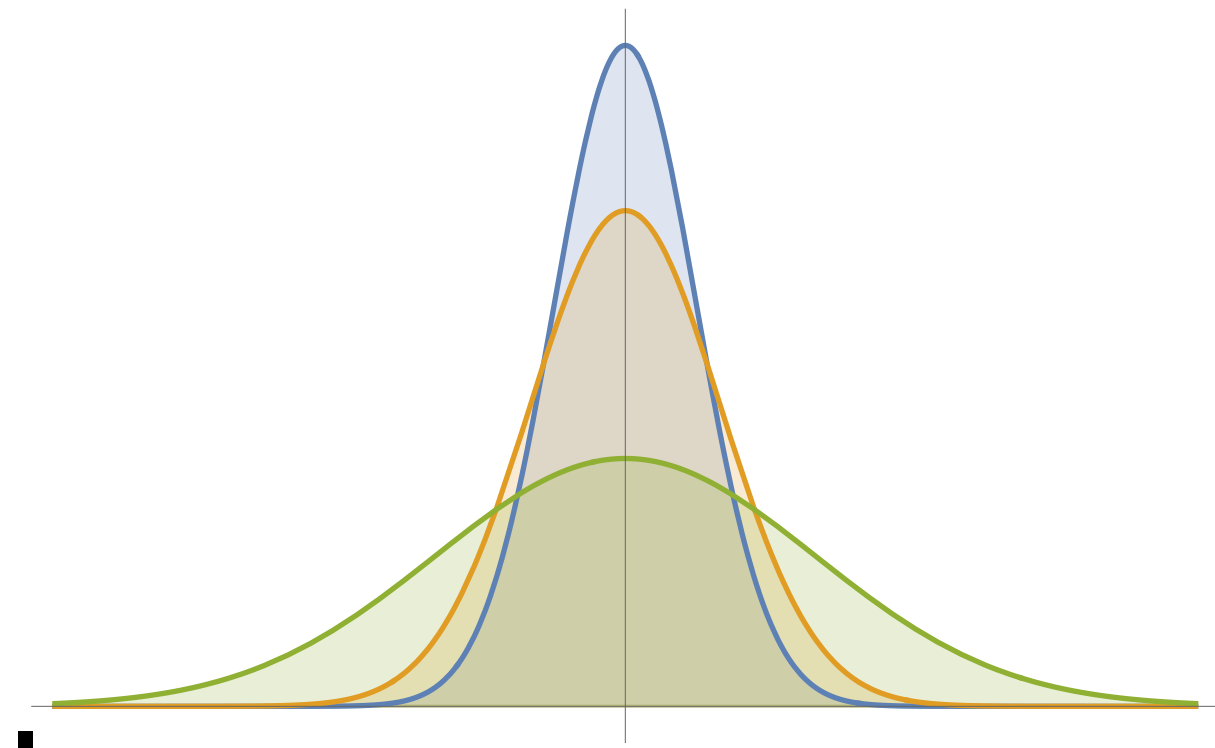
**Proof:**  $\Pr(X \leq (1 - \delta)\mu) \leq \Pr(e^{tX} \geq e^{t(1-\delta)\mu})$  (for any  $t < 0$ )

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

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

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

# Chernoff Bound



- Chernoff bound: Let  $X_1, \dots, X_n \in \{0,1\}$  be independent trials

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

(Poisson binomial RV  
with mean  $\mu$ )

- For any  $\delta > 0$ ,

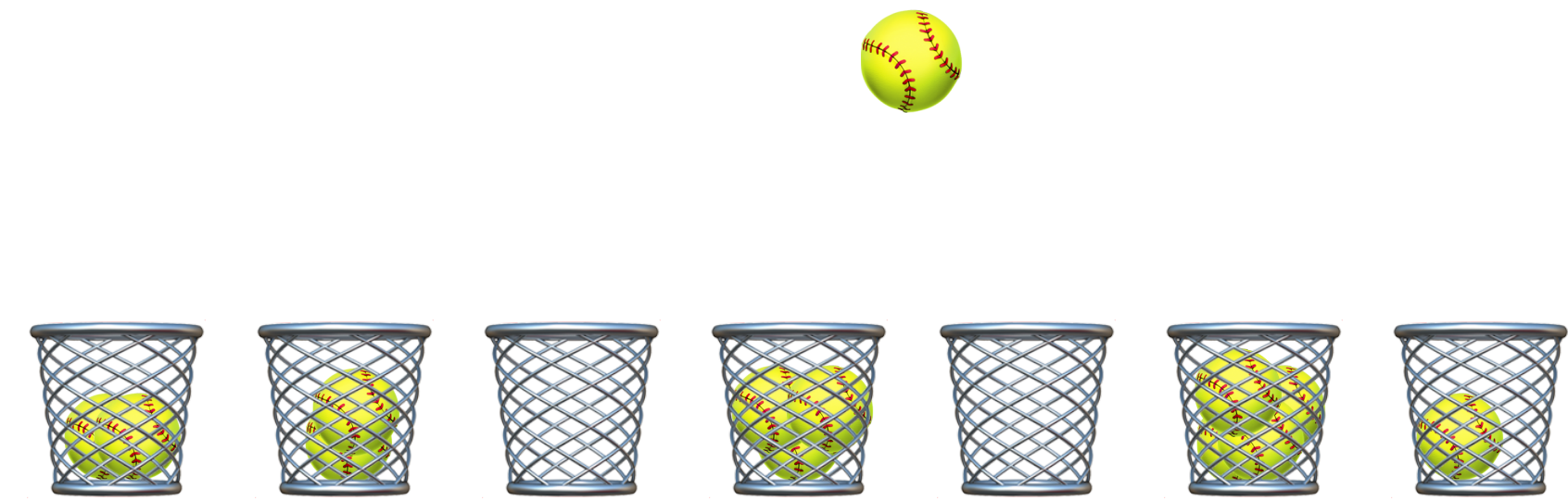
$$\Pr(X \geq (1 + \delta)\mu) \leq \left( \frac{e^\delta}{(1 + \delta)^{(1+\delta)}} \right)^\mu \leq \begin{cases} e^{-\frac{\mu\delta^2}{3}} & \text{if } 0 < \delta < 1 \\ 2^{-(1+\delta)\mu} & \text{if } (1 + \delta) \geq 2e \end{cases}$$

- 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)



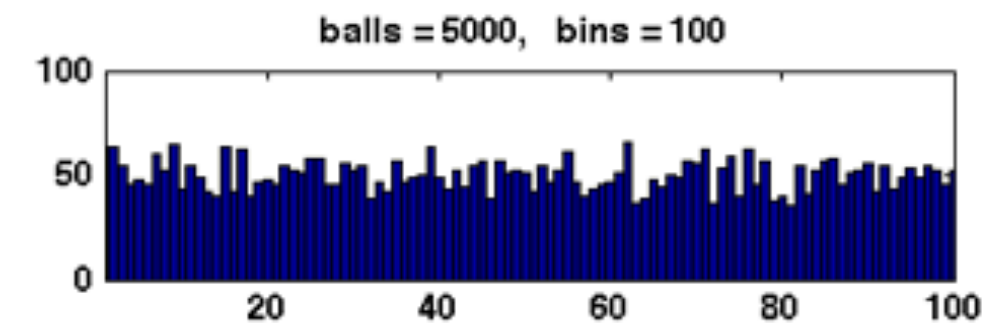
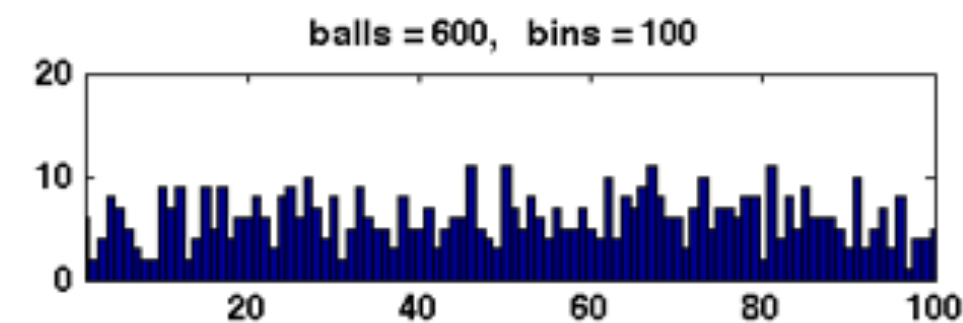
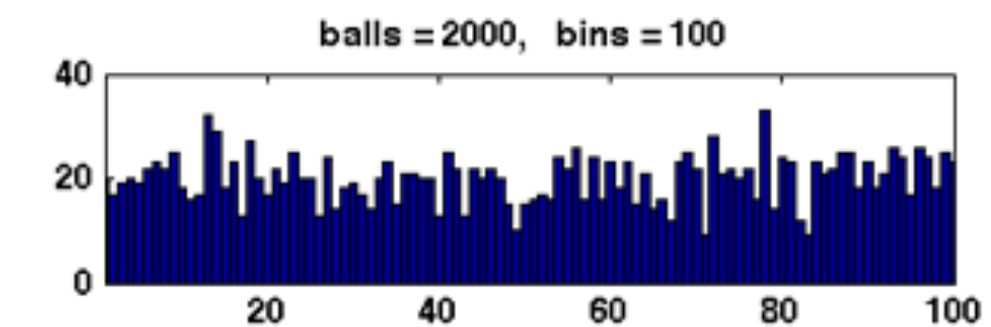
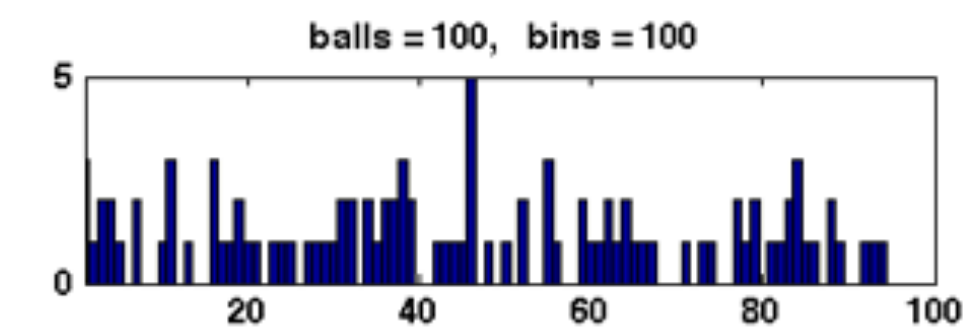
- Throw  $m$  balls into  $n$  bins *u.a.r.* Numbers of balls received in each bins:

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

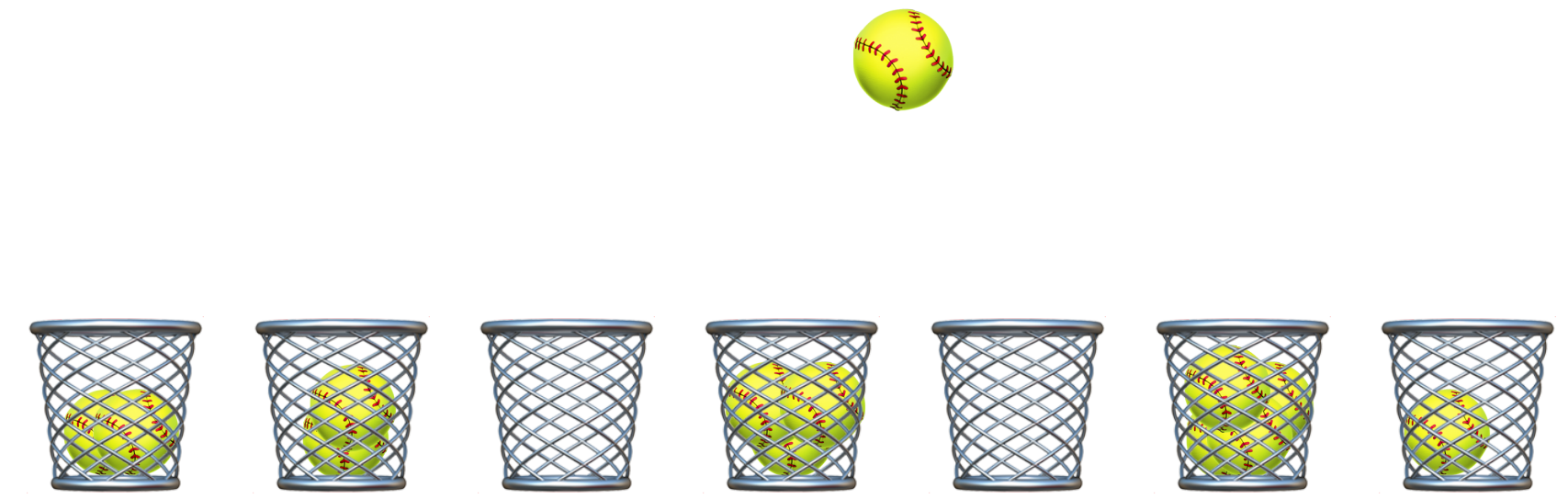
- Occupancy problem: the maximum load  $\max_{1 \leq i \leq n} X_i$  *w.h.p.* (**with high probability**)

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

*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:

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

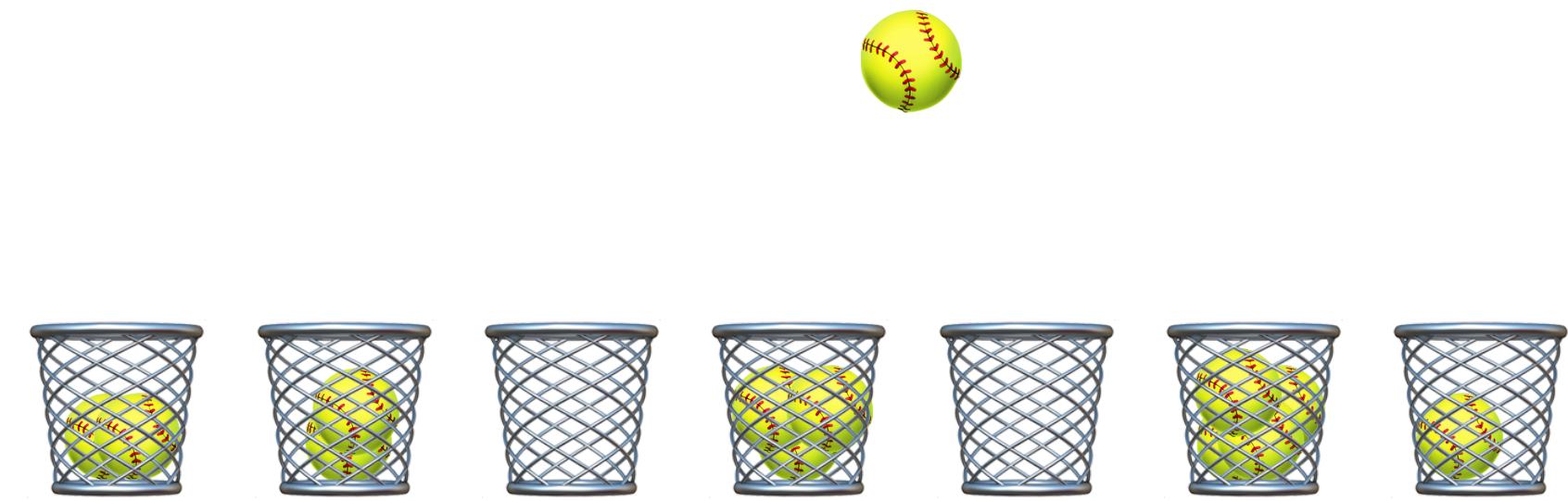
- When  $m = n$ :  $\mu = 1$

$$\Pr(X_i \geq L) = \Pr(X_i \geq L\mu) \leq \frac{e^L}{eL^L} \leq \frac{1}{n^2} \quad \text{for } L = \frac{e \ln n}{\ln \ln n}$$

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

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

# Occupancy Problem



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

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

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

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

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

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

# Chernoff-Hoeffding Bound

- Chernoff-Hoeffding bound:

If  $X_1, \dots, X_n \in \{0,1\}$  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  $X_1, \dots, X_n$  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

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

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

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

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

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

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

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

# 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  $Y = S_n - \mathbb{E}[S_n] = \sum_{i=1}^n Y_i \implies \mathbb{E}[Y] = \mathbb{E}[Y_i] = 0$

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

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

$$(\text{Hoeffding's lemma}) \leq \exp \left( -\lambda t + \frac{\lambda^2}{8} \sum_{i=1}^n (b_i - a_i)^2 \right) = \exp \left( - \frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right) \quad \text{by choosing: } \lambda = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2}$$

# 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  $Y = S_n - \mathbb{E}[S_n] = \sum_{i=1}^n Y_i \implies \mathbb{E}[Y] = \mathbb{E}[Y_i] = 0$

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

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

$$(\text{Hoeffding's lemma}) \leq \exp \left( \lambda t + \frac{\lambda^2}{8} \sum_{i=1}^n (b_i - a_i)^2 \right) = \exp \left( - \frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right) \quad \begin{array}{l} \text{by choosing:} \\ \lambda = \frac{-4t}{\sum_{i=1}^n (b_i - a_i)^2} \end{array}$$

# Chernoff-Hoeffding Bound

- Chernoff-Hoeffding bound:

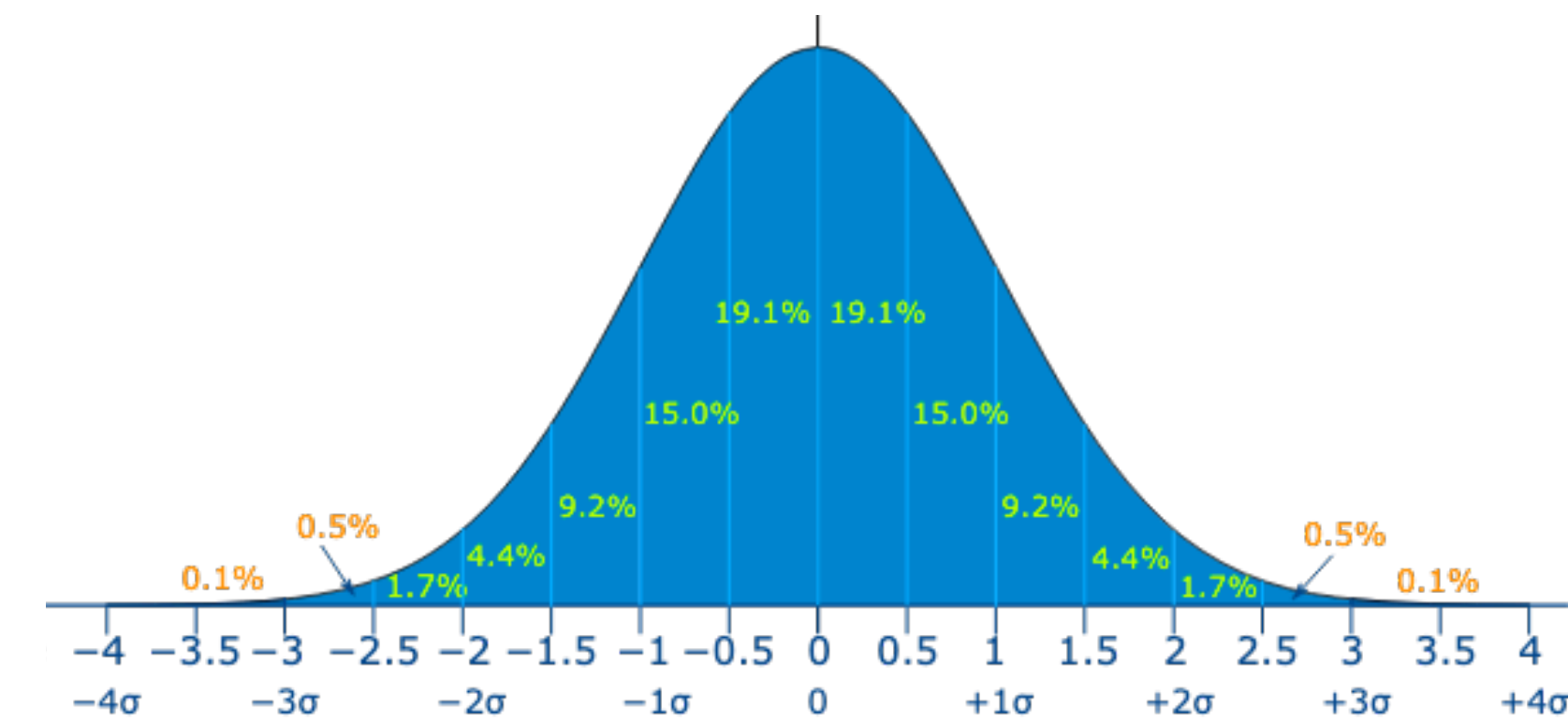
If  $X_1, \dots, X_n \in \{0,1\}$  are independent and  $S_n = \sum_{i=1}^n X_i$ , then for any  $t > 0$ ,

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

In general, if  $X_1, \dots, X_n$  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



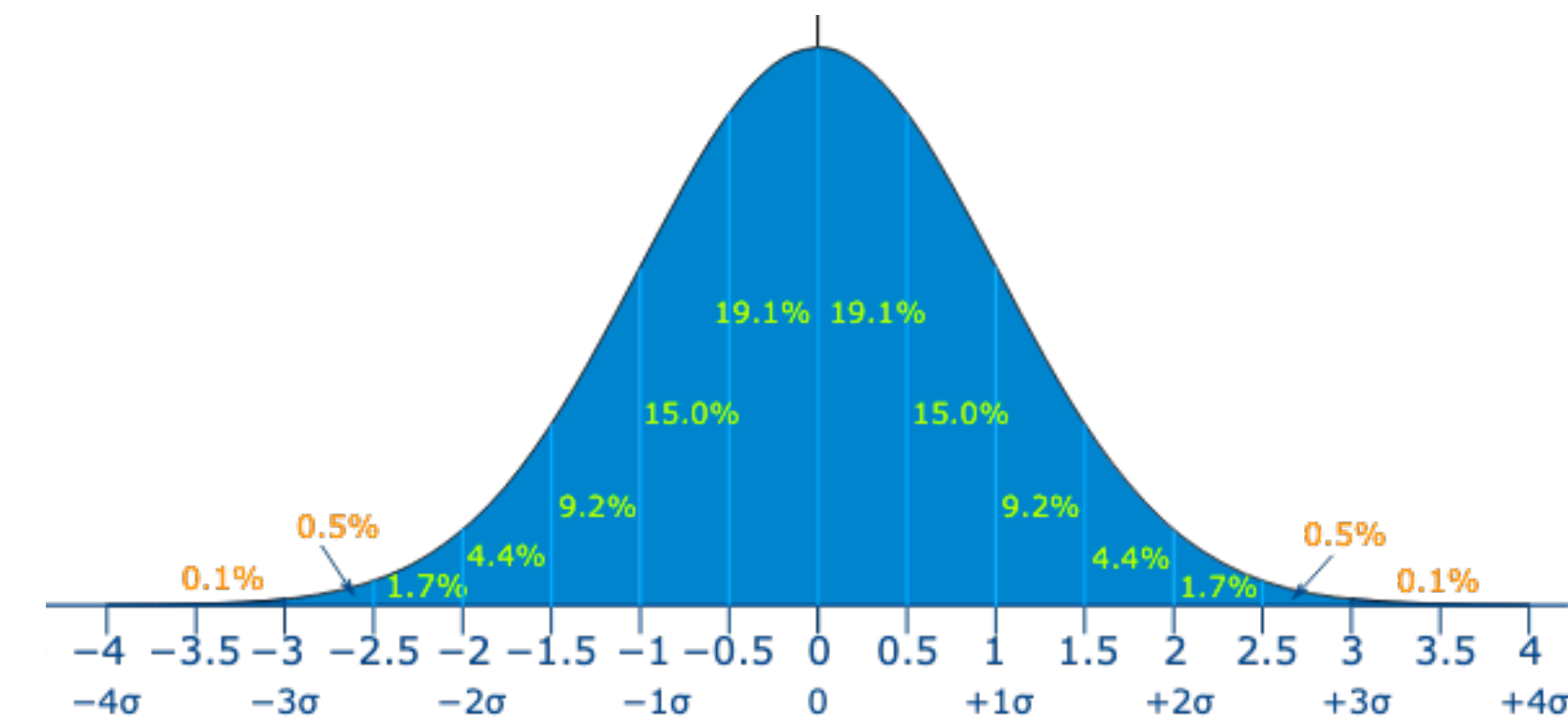
- Chernoff-Hoeffding bound:

If  $X_1, \dots, X_n \in \{0,1\}$  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 sub-Gaussian tail  $e^{-\Omega(t^2)}$

# Sub-Gaussian Tail



- Chernoff-Hoeffding bound:

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

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

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

- The “worst-case” standardized  $S_n$  has a sub-Gaussian 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  $S_n = X_1 + \dots + X_n$  for i.i.d. Bernoulli random variables  $X_1, \dots, X_n$  with parameter  $1/2$ .

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

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

# Error Reduction (two-sided case)

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

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

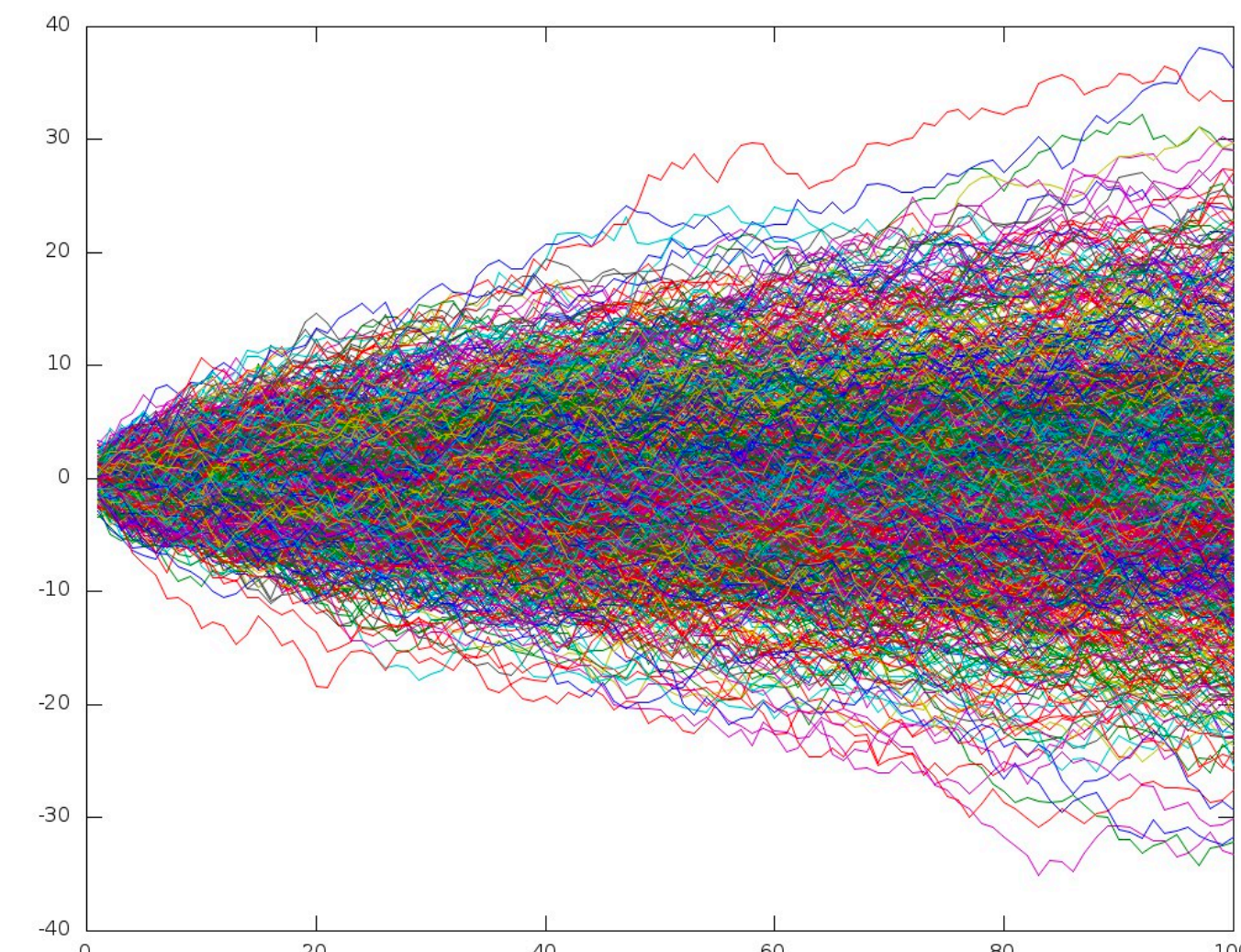
where  $S_n = X_1 + \dots + X_n$ , and  $X_i = I[\mathcal{A}(x) = f(x)]$  in  $i$ th run

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

# The Median Trick

- Computation problem  $f: \{0,1\}^* \rightarrow \mathbb{R}$ .
  - *Randomized approximation* algorithm  $\mathcal{A}: \forall x \in \{0,1\}^*$ ,
    - $\Pr(\mathcal{A}(x) \in (1 \pm \epsilon)f(x)) = \Pr((1 - \epsilon)f(x) \leq \mathcal{A}(x) \leq (1 + \epsilon)f(x)) \geq \frac{1}{2} + p$
  - $\mathcal{A}^n$ : **independently** run  $\mathcal{A}$  for  $n$  times, return **median** of the  $n$  outputs
    - Let  $X_i = I[\mathcal{A}(x) \in (1 \pm \epsilon)f(x)]$  in the  $i$ th run of  $\mathcal{A}(x)$   $\implies \mathbb{E}[X_i] \geq 1/2 + p$
    - **Observation:**  $\mathcal{A}^n(x) \in (1 \pm \epsilon)f(x)$  if  $S_n = X_1 + \dots + X_n > \frac{n}{2}$
- $$\Pr(\mathcal{A}(x) \notin (1 \pm \epsilon)f(x)) \leq \Pr\left(S_n \leq \frac{n}{2}\right) \leq \Pr(S_n \leq \mathbb{E}[S_n] - np) \leq e^{-2p^2n} \leq \delta$$
- when  $n \geq \frac{1}{2p^2} \ln \frac{1}{\delta}$

# The Method of Bounded Differences



# The Method of Bounded Differences

- McDiarmid's Inequality:

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

If  $f : \mathcal{X}_1 \times \dots \times \mathcal{X}_n \rightarrow \mathbb{R}$  satisfies the bounded differences property:

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

then for any  $t > 0$ ,

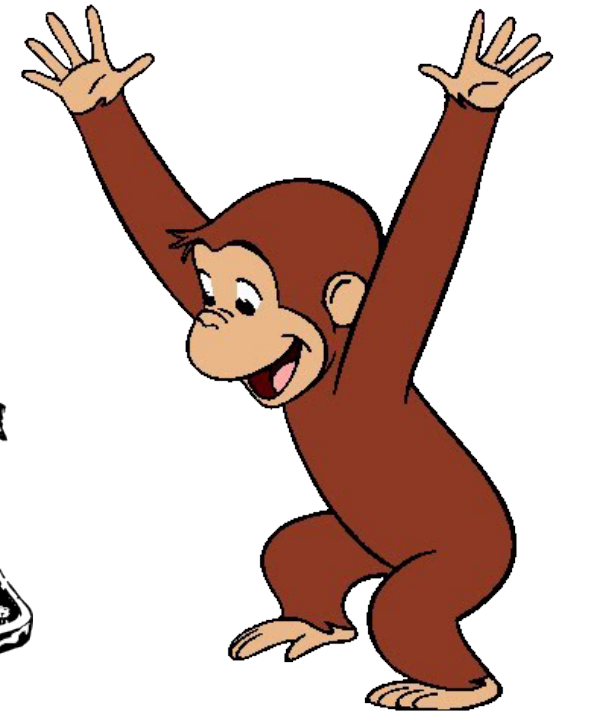
$$\Pr \left( \left| f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)] \right| \geq t \right) \leq 2 \exp \left( -\frac{2t^2}{\sum_{i=1}^n c_i^2} \right)$$

- 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

Hamlet



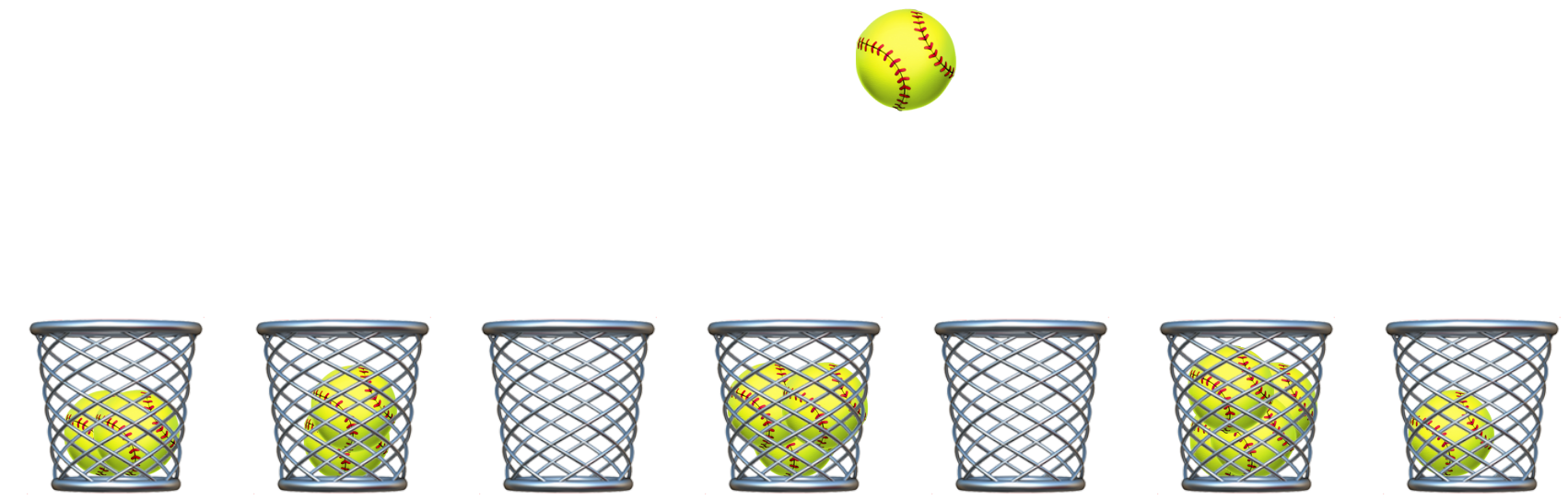
- $s = (s_1, \dots, s_n) \in Q^n$ : uniform random string of  $n$  letters from alphabet  $Q$  with  $|Q| = q$
- For pattern  $\pi \in Q^k$ , let  $X$  be the number of appearances of  $\pi$  in  $s$  as substring

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

- **Linearity of expectation:**  $\mathbb{E}[X] = \sum_{i=1}^{n-k+1} \mathbb{E}[I_i] = (n - k + 1)q^{-k}$

- **McDiarmid's Inequality:**  $\Pr\left(|X - \mathbb{E}[X]| \geq t\right) \leq 2 \exp\left(-\frac{2t^2}{nk^2}\right)$

# Empty Bins



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

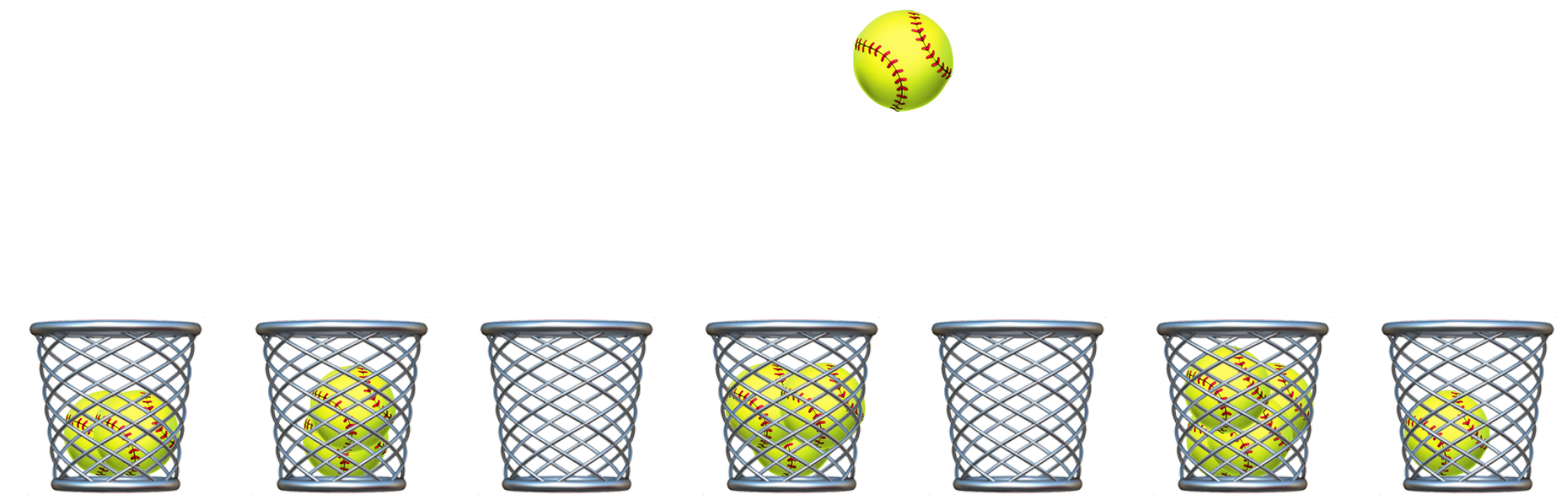
$$Y = \sum_{i=1}^n I[\textit{ith bin is empty}]$$

- **Linearity of expectation:**

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

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

# Empty Bins



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

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

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

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

has 1-bounded differences

- McDiarmid's Inequality:

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

# The Method of Bounded Differences

- McDiarmid's Inequality:

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

If  $f : \mathcal{X}_1 \times \dots \times \mathcal{X}_n \rightarrow \mathbb{R}$  satisfies the bounded differences property:

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

then for any  $t > 0$ ,

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

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

# Doob Sequence

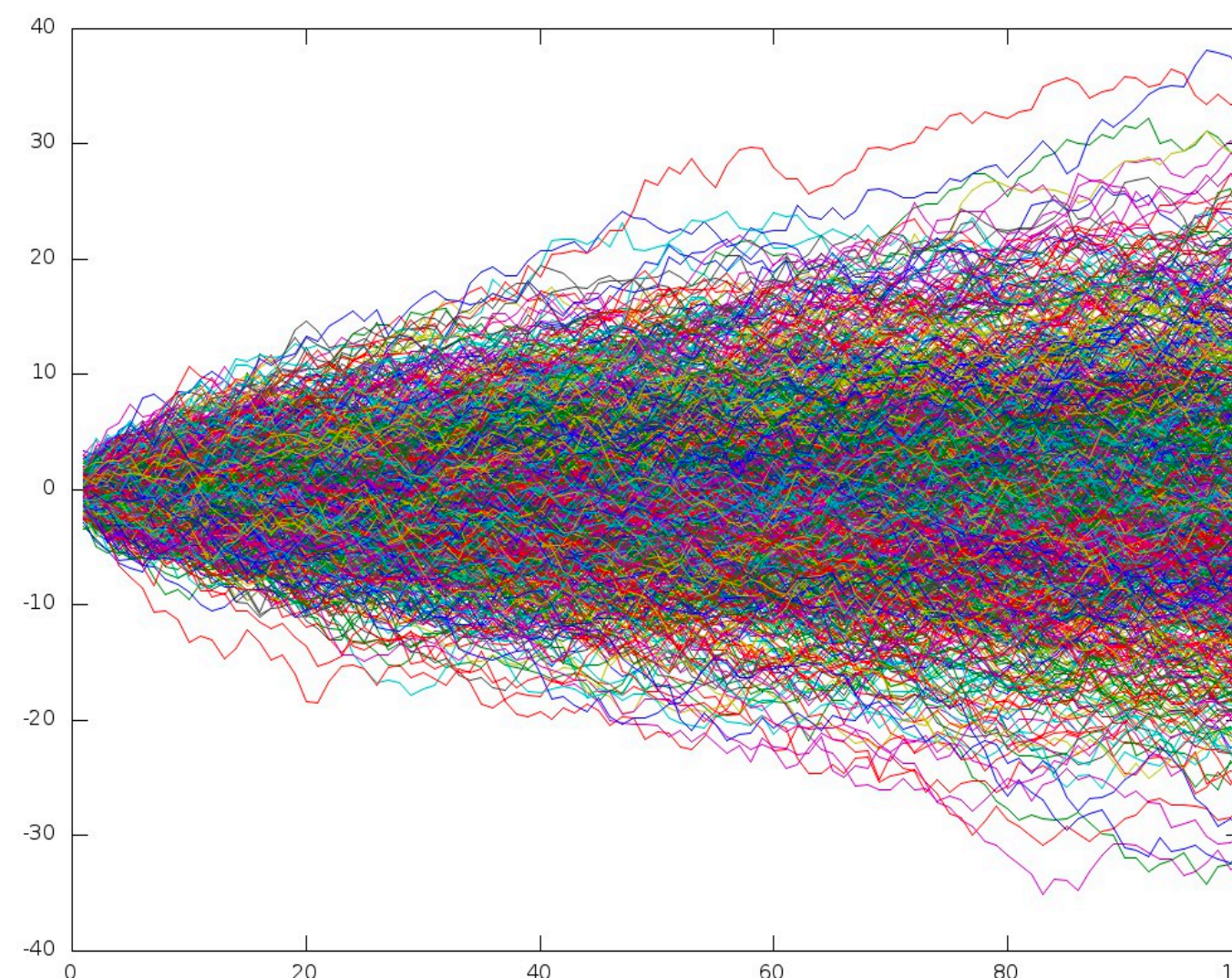
- The Doob sequence  $Y_0, Y_1, \dots, Y_n$  of  $n$ -variate function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  on random variables  $X_1, \dots, X_n$ , is given by

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

$$Y_0 = \mathbb{E} \left[ f(X_1, \dots, X_n) \right] \quad \text{-----} \rightarrow \quad f(X_1, \dots, X_n) = Y_n$$

no information

full information



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

# Doob Sequence

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

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

$$f(\underbrace{(\text{coin}, \text{coin}, \text{coin}, \text{coin}, \text{coin}, \text{coin})}_{\text{averaged over}})$$

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

# Doob Sequence

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

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

randomized by

$$f(\textcircled{1}, \textcircled{\$}, \textcircled{\$}, \textcircled{\$}, \textcircled{\$}, \textcircled{\$})$$

averaged over

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

# Doob Sequence

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

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

randomized by

$$f(\overbrace{1, 0}, \underbrace{\text{\$}, \text{\$}, \text{\$}, \text{\$}})$$

averaged over

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

# Doob Sequence

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

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

randomized by

$$f(\underbrace{1, 0, 0}_{\text{randomized by}}, \underbrace{\text{coin}, \text{coin}, \text{coin}}_{\text{averaged over}})$$

averaged over

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

# Doob Sequence

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

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

randomized by

$$f( \underbrace{1, 0, 0, 1}_{\text{randomized by}}, \underbrace{\text{coin}, \text{coin}}_{\text{averaged over}} )$$

averaged over

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

# Doob Sequence

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

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

randomized by

$$f( \textcircled{1}, \textcircled{0}, \textcircled{0}, \textcircled{1}, \textcircled{0}, \textcircled{\$} )$$

averaged over

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

# Doob Sequence

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

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

randomized by

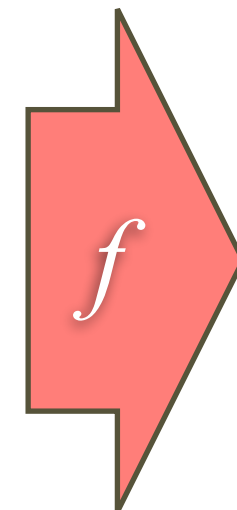
$$f(\textcircled{1}, \textcircled{0}, \textcircled{0}, \textcircled{1}, \textcircled{0}, \textcircled{1})$$

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

# Doob Sequence

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

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



# Doob Sequence

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

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

- Martingale property:  $\mathbb{E} \left[ Y_i \mid X_1, \dots, X_{i-1} \right] = Y_{i-1}$

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

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

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

$$= Y_{i-1}$$

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



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  $n \geq 0$ ,
  - $\mathbb{E} [ |Y_n| ] < \infty$
  - $\mathbb{E} [ Y_{n+1} \mid X_0, X_1, \dots, X_n ] = Y_n$  (martingale property)
- By definition:  $Y_n$  is a function of  $X_0, X_1, \dots, X_n$
- Current capital  $Y_n$  in a **fair gambling game** with outcomes  $X_0, X_1, \dots, X_n$ 
  - **Super-martingale (上鞅)**:  $\mathbb{E} [ Y_{n+1} \mid X_0, X_1, \dots, X_n ] \leq Y_n$
  - **Sub-martingale (下鞅)**:  $\mathbb{E} [ Y_{n+1} \mid X_0, X_1, \dots, X_n ] \geq Y_n$

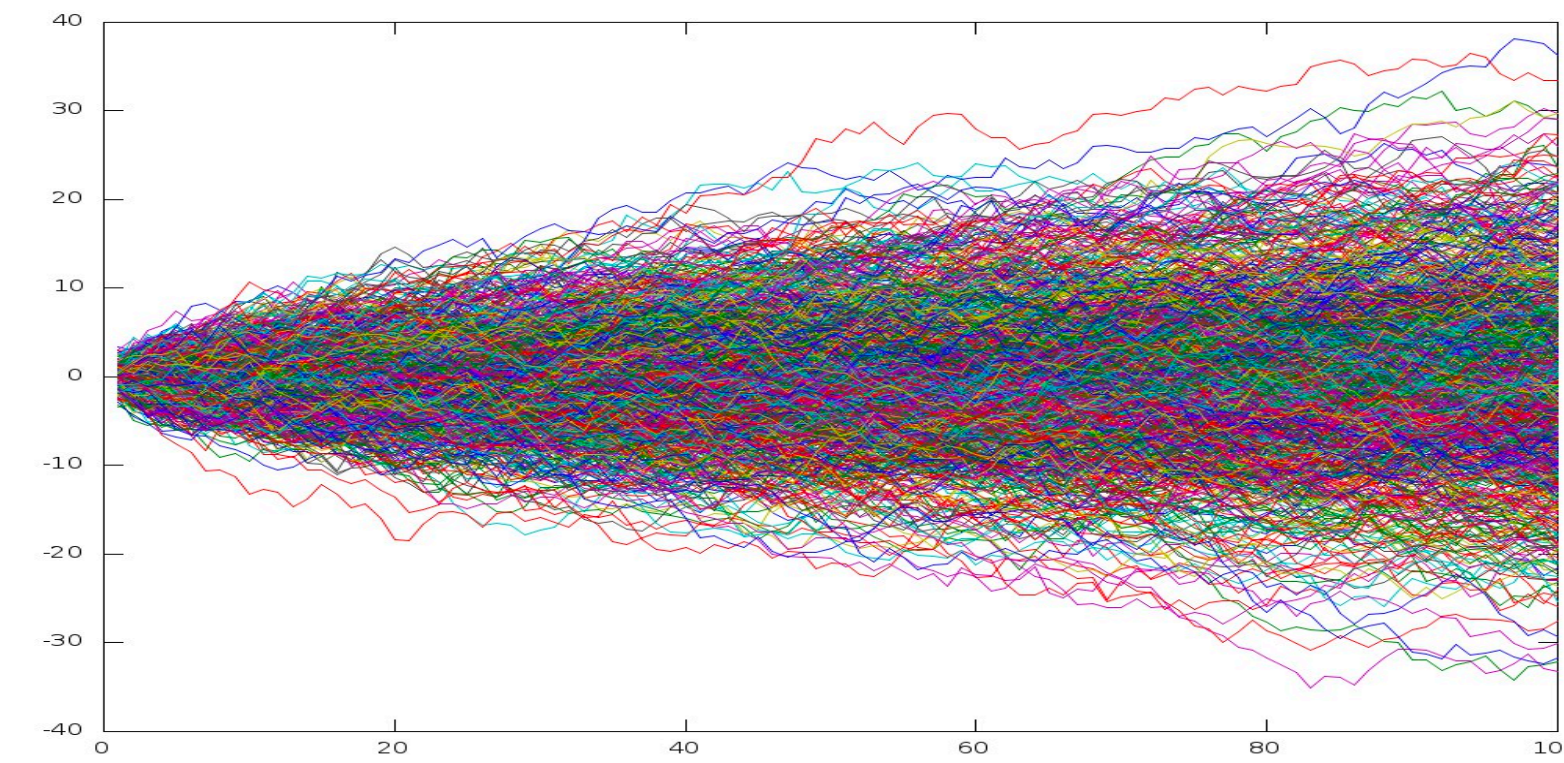
# Martingale (鞅)

- 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  $n \geq 0$ ,
  - $\mathbb{E} [ |Y_n| ] < \infty$
  - $\mathbb{E} [ Y_{n+1} \mid X_0, X_1, \dots, X_n ] = Y_n$  ( $\implies Y_n$  is a function of  $X_0, \dots, X_n$ )
- $\{X_n : n \geq 0\}$  are defined on the probability space  $(\Omega, \Sigma, \text{Pr})$ 
  - $(X_0, X_1, \dots, X_n)$  defines a sub- $\sigma$ -field  $\Sigma_n \subseteq \Sigma$  (the smallest  $\sigma$ -field s.t.  $(X_0, \dots, X_n)$  is  $\Sigma_n$ -measurable)
  - $\{\Sigma_n : n \geq 0\}$  is a **filtration** of  $\Sigma$ , i.e.  $\Sigma_0 \subseteq \Sigma_1 \subseteq \dots \subseteq \Sigma$
  - The martingale property is expressed as  $\mathbb{E} [ Y_{n+1} \mid \Sigma_n ] = Y_n$

# Examples of Martingale

- Doob martingale:  $Y_i = \mathbb{E} [f(X_1, \dots, X_n) \mid X_1, \dots, X_i]$ 
  - 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  $\{X_n : n \geq 0\}$ :

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

- 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  $\tau$

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

# Martingale Tail Inequality

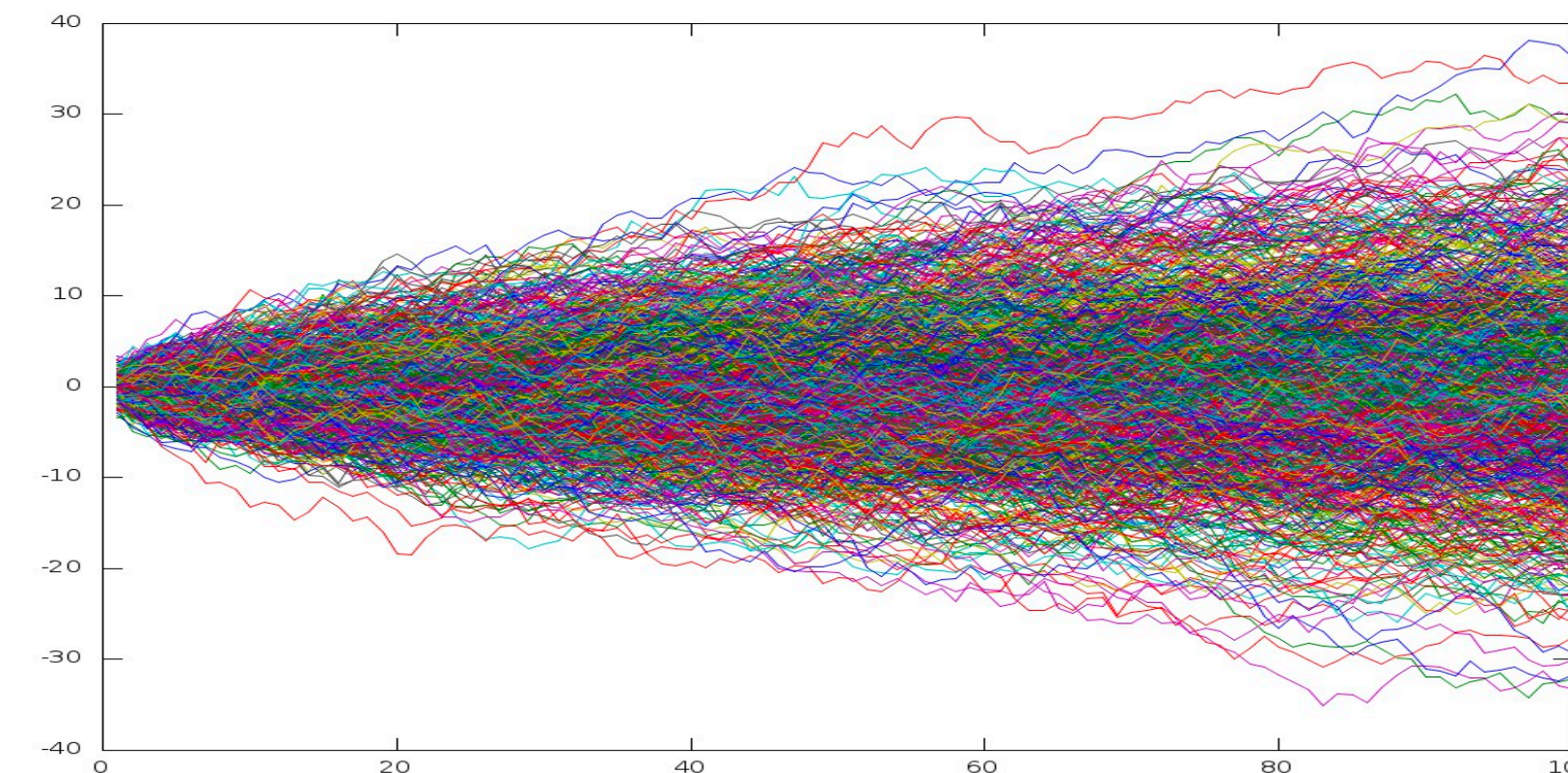
- Azuma's inequality: If a sequence  $\{Y_n : n \geq 0\}$  is a martingale (with respect to some sequence  $\{X_n : n \geq 0\}$ ), and for all  $n \geq 1$ ,

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

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

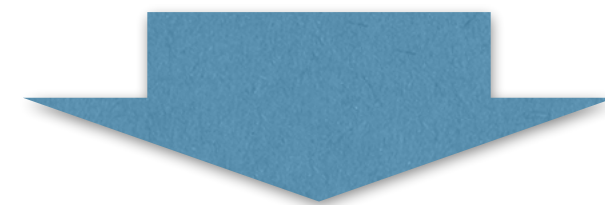


# The Method of Bounded Differences

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

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

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



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

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

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

# Bounded Differences Properties

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

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

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

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

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

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

# The Method of Bounded Differences

- McDiarmid's Inequality:

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

If  $f: \mathcal{X}_1 \times \dots \times \mathcal{X}_n \rightarrow \mathbb{R}$  satisfies the bounded differences property:

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

then for any  $t > 0$ ,

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

# 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  $n \geq 1$ ,

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

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

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

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

# Proof of Azuma's Inequality

- $\{Y_n : n \geq 0\}$  is a martingale (w.r.t.  $\{X_n : n \geq 0\}$ ) satisfying  $|Y_n - Y_{n-1}| \leq c_n$  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.  $\{X_n : n \geq 0\}$   
 $\implies \mathbb{E} [D_n | X_0, \dots, X_{n-1}] = \mathbb{E} [Y_n - Y_{n-1} | X_0, \dots, X_{n-1}]$   
 $= \mathbb{E} [Y_n | X_0, \dots, X_{n-1}] - \mathbb{E} [Y_{n-1} | X_0, \dots, X_{n-1}] = Y_{n-1} - Y_{n-1} = 0$
- Bounded difference:  $|Y_n - Y_{n-1}| \leq c_n \implies D_n = Y_n - Y_{n-1} \in [-a_n, b_n]$  for  $b_n - a_n = c_n$
- Hoeffding's lemma:  $\mathbb{E} [D_n | X_0, \dots, X_{n-1}] = 0$  and  $D_n \in [-a_n, b_n]$  for  $b_n - a_n = c_n$   
 $\implies \mathbb{E} [e^{\lambda D_n} | X_0, \dots, X_{n-1}] \leq e^{\lambda^2 c_n^2 / 8}$

# Proof of Azuma's Inequality

- $\{Y_n : n \geq 0\}$  is a martingale (w.r.t.  $\{X_n : n \geq 0\}$ ) satisfying  $|Y_n - Y_{n-1}| \leq c_n$  for  $n \geq 1$

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

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

$$\begin{aligned} \mathbb{E} \left[ e^{\lambda S_n} \right] &= \mathbb{E} \left[ \mathbb{E} \left[ e^{\lambda S_n} \mid X_0, \dots, X_{n-1} \right] \right] = \mathbb{E} \left[ \mathbb{E} \left[ e^{\lambda(S_{n-1} + D_n)} \mid X_0, \dots, X_{n-1} \right] \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ e^{\lambda S_{n-1}} \cdot e^{\lambda D_n} \mid X_0, \dots, X_{n-1} \right] \right] = \mathbb{E} \left[ e^{\lambda S_{n-1}} \cdot \mathbb{E} \left[ e^{\lambda D_n} \mid X_0, \dots, X_{n-1} \right] \right] \\ &\leq \mathbb{E} \left[ e^{\lambda S_{n-1}} \cdot e^{\lambda^2 c_n^2 / 8} \right] = e^{\lambda^2 c_n^2 / 8} \cdot \mathbb{E} \left[ e^{\lambda S_{n-1}} \right] \end{aligned}$$

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

# Proof of Azuma's Inequality

- $\{Y_n : n \geq 0\}$  is a martingale (w.r.t.  $\{X_n : n \geq 0\}$ ) satisfying  $|Y_n - Y_{n-1}| \leq c_n$  for  $n \geq 1$

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

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

- (Upper tail)  $\Pr(Y_n - Y_0 \geq t) = \Pr(S_n \geq t) \leq \Pr(e^{\lambda S_n} \geq e^{\lambda t})$  (for any  $\lambda > 0$ )  
(Markov)  $\leq e^{-\lambda t} \cdot \mathbb{E} [e^{\lambda S_n}] \leq \exp \left( -\lambda t + \frac{\lambda^2}{8} \sum_{i=1}^n c_i^2 \right) = \exp \left( -\frac{2t^2}{\sum_{i=1}^n c_i^2} \right)$  by choosing:  
 $\lambda = \frac{4t}{\sum_{i=1}^n c_i^2}$
- (Lower tail)  $\Pr(Y_n - Y_0 \leq -t) = \Pr(S_n \leq -t) \leq \Pr(e^{\lambda S_n} \geq e^{-\lambda t})$  (for any  $\lambda < 0$ )  
(Markov)  $\leq e^{\lambda t} \cdot \mathbb{E} [e^{\lambda S_n}] \leq \exp \left( \lambda t + \frac{\lambda^2}{8} \sum_{i=1}^n c_i^2 \right) = \exp \left( -\frac{2t^2}{\sum_{i=1}^n c_i^2} \right)$  by choosing:  
 $\lambda = \frac{-4t}{\sum_{i=1}^n c_i^2}$

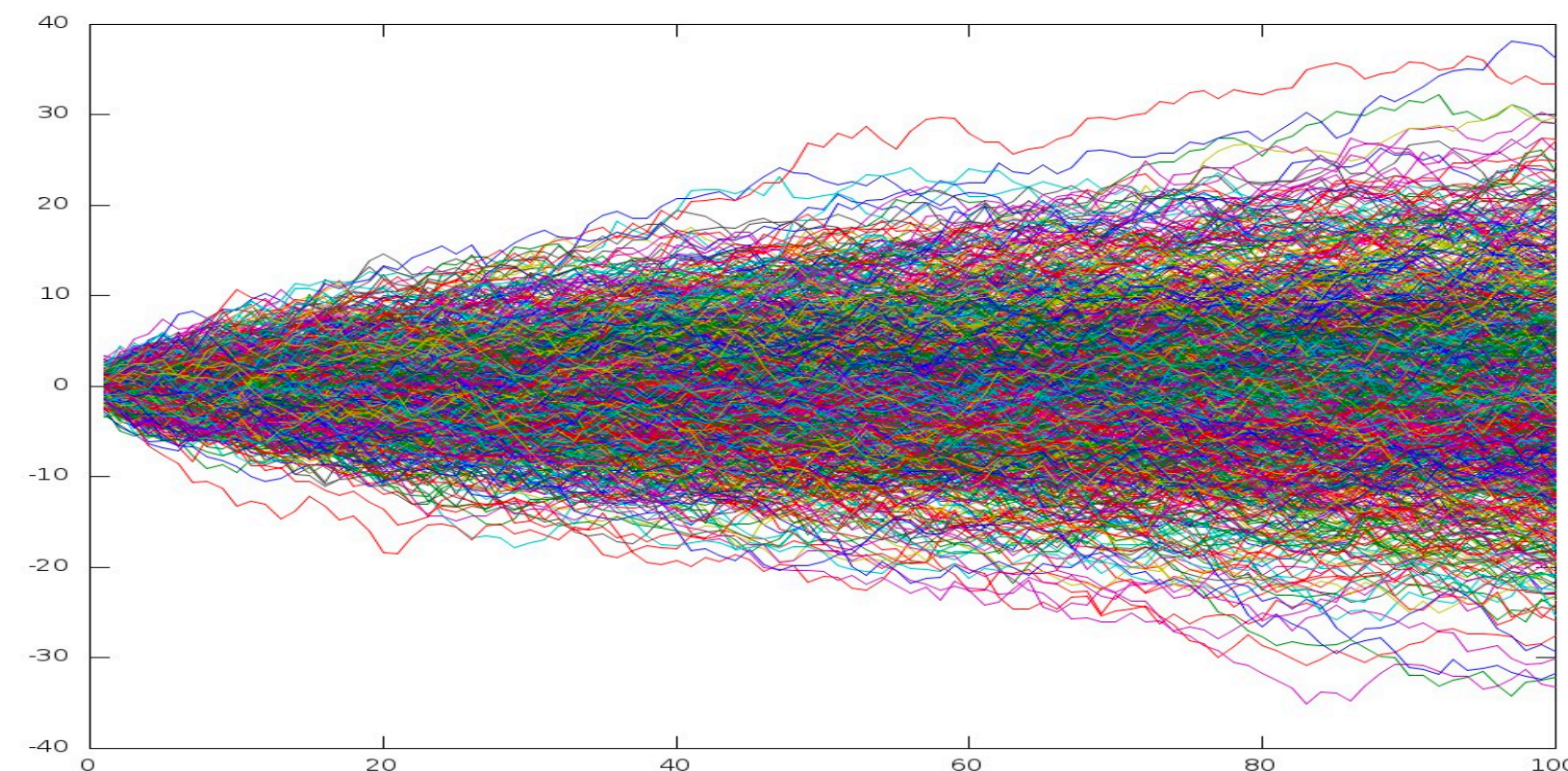
# Martingale Tail Inequality

- 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  $n \geq 1$ ,

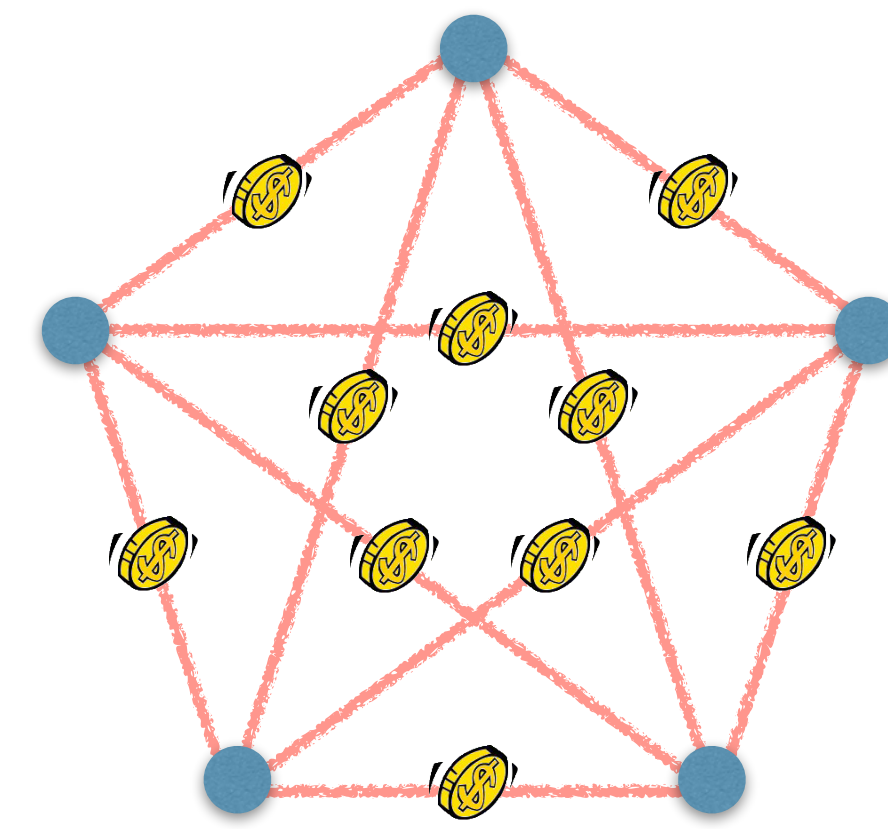
$$\left| Y_n - Y_{n-1} \right| \leq c_n$$

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

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



# Martingales from Random Graph



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

$$Y_i = \mathbb{E}[f(G) \mid X_1, \dots, X_i], \quad 1 \leq i \leq \binom{n}{2}$$

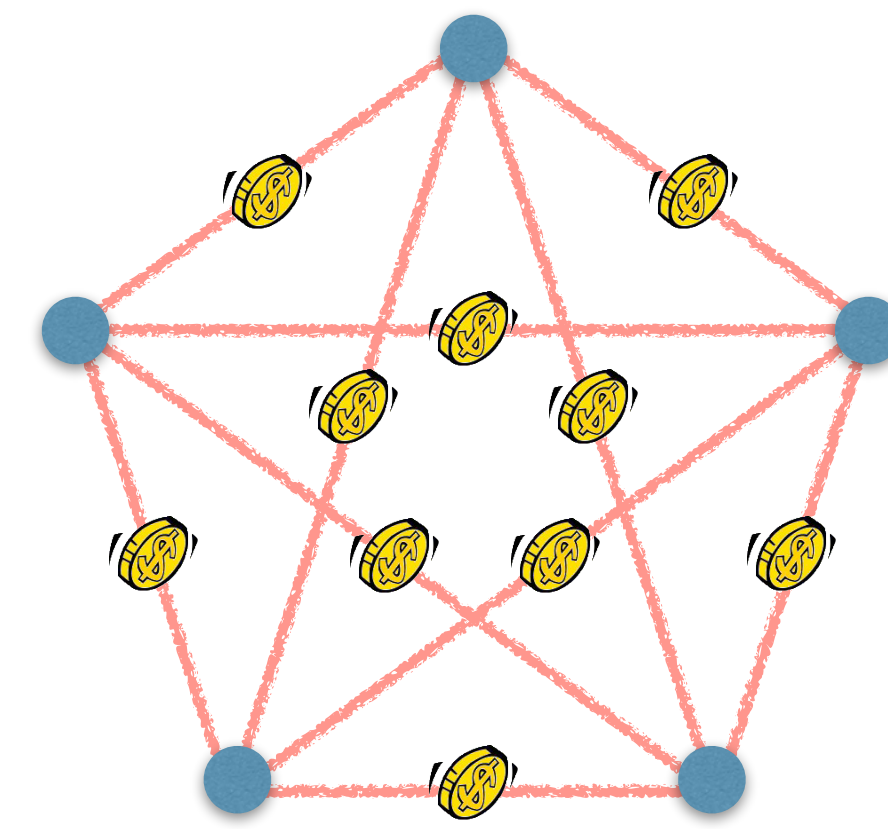
where  $X_1, \dots, X_{\binom{n}{2}}$  are i.i.d. Bernoulli( $p$ ), s.t.  $X_i = I[i\text{th vertex pair is an edge in } G]$

- Vertex exposure martingale:

$$Y_i = \mathbb{E}[f(G) \mid X_1, \dots, X_i], \quad 1 \leq i \leq n$$

where  $X_i = G[\{1, \dots, i\}]$  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  $G$
- Vertex exposure martingale:

$$Y_i = \mathbb{E}[f(G) \mid X_1, \dots, X_i], \quad 1 \leq i \leq n$$

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

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

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

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