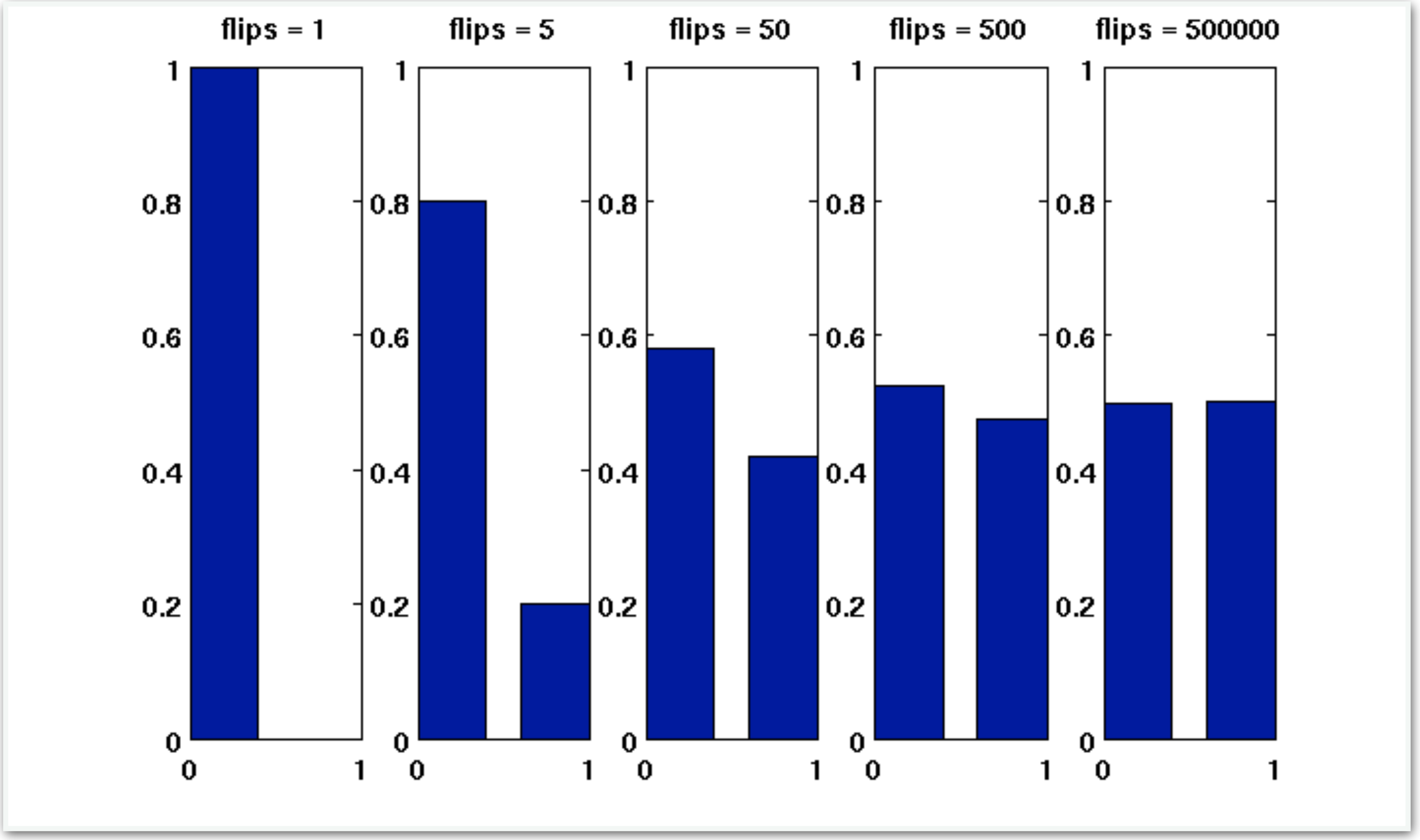
Advanced Algorithms

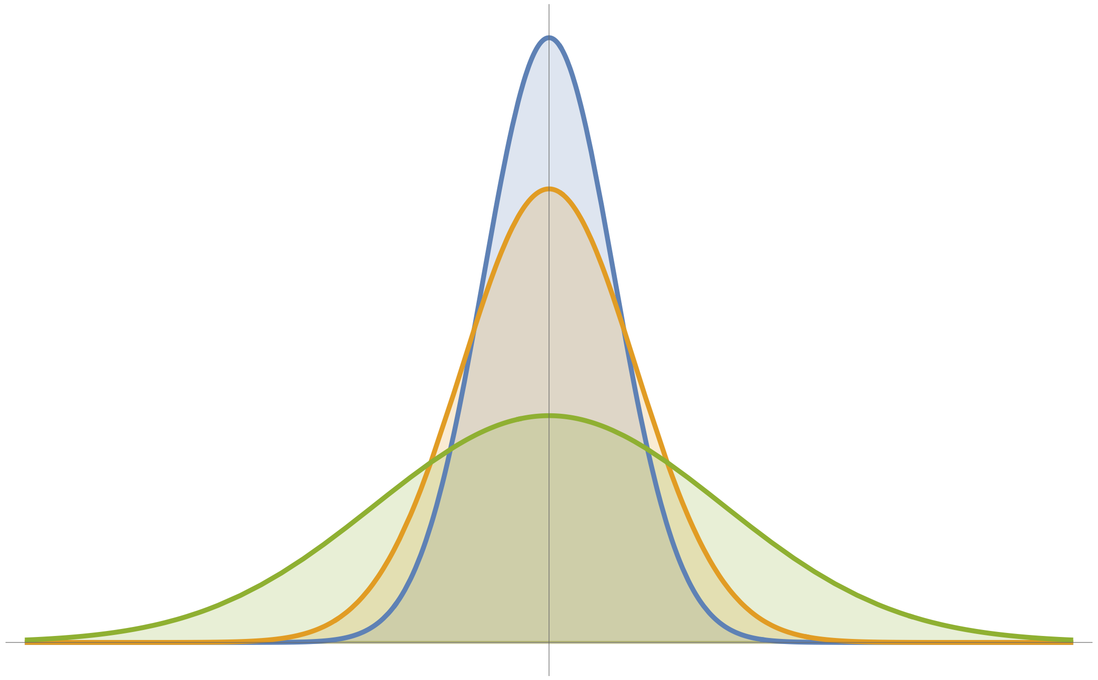
Concentration of Measure

尹一通 Nanjing University, 2022 Fall

Measure Concentration

- Flip a coin for many times:

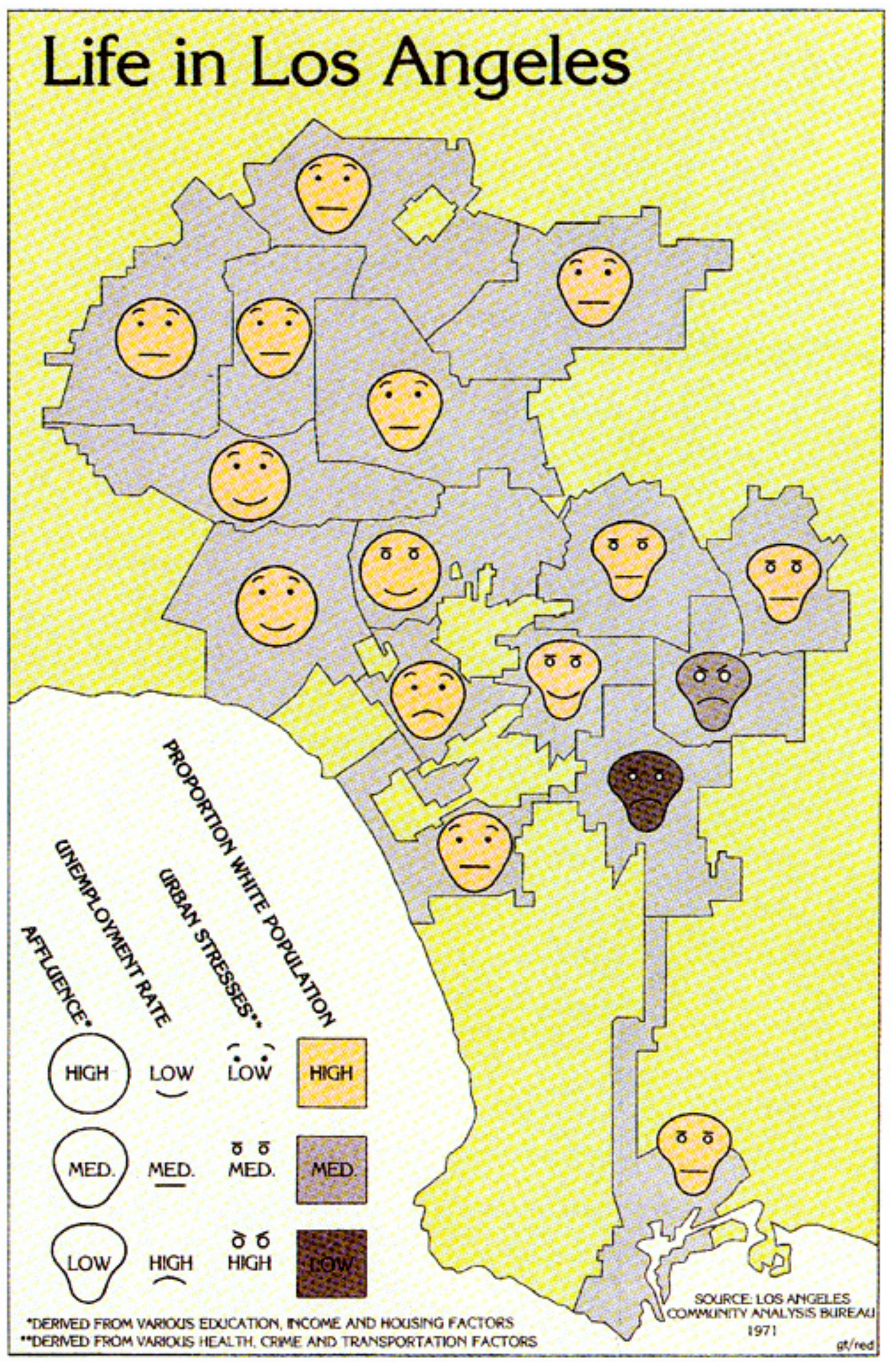




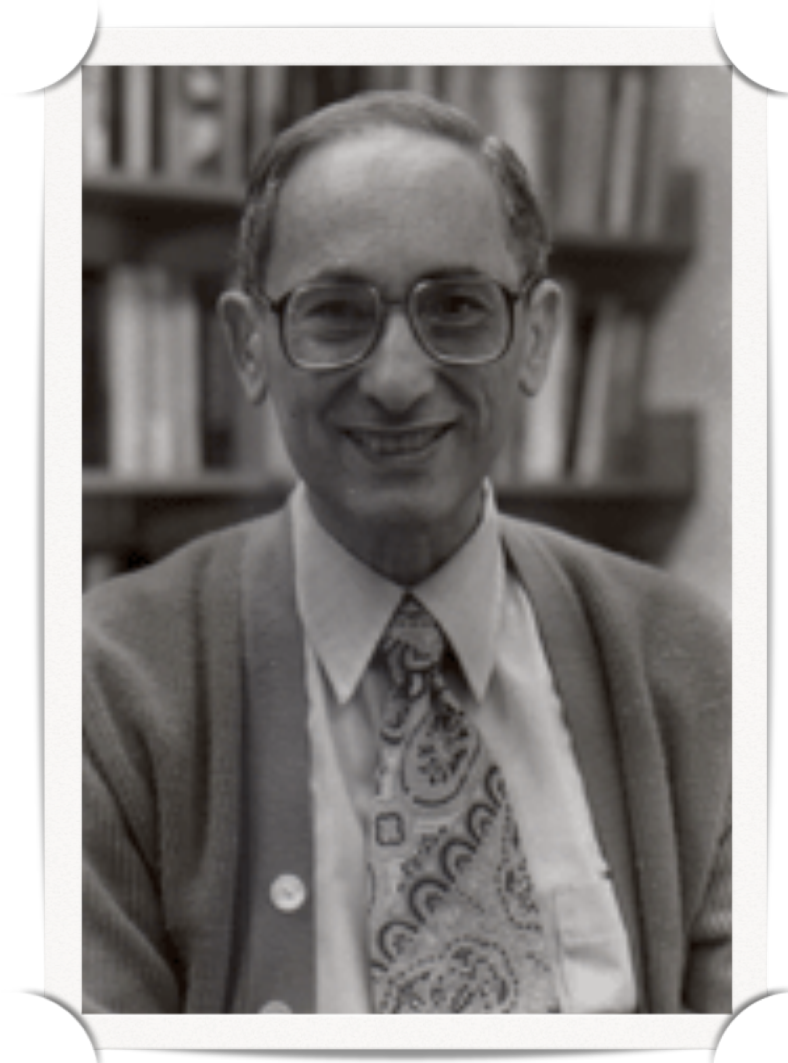
Chernoff-Hoeffding Bounds

Chernoff Bound

(Bernstein Inequalities)



Chernoff face



Herman Chernoff

Chernoff Bound

Chernoff Bound:

For *independent* $X_1, \ldots, X_n \in \{0,1\}$ with

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

For any $\delta > 0$,

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

For any $0 < \delta < 1$,

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

Chernoff Bound

Chernoff Bound:

For *independent* $X_1, \ldots, X_n \in \{0,1\}$ with

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

For any $0 < \delta < 1$,

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

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

For $t \geq 2\mathrm{e}\mu$:

$$\Pr[X \geq t] \leq 2^{-t}$$

Balls into Bins

m balls are thrown into n bins.

X_i : number of balls in the i-th bin

$$X_i = \sum_{j=1}^{m} X_{ij} \quad \text{where } X_{ij} = \begin{cases} 1 & \text{with prob. } \frac{1}{n} \\ 0 & \text{with prob. } 1 - \frac{1}{n} \end{cases}$$

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

Chernoff Bound: For $\delta > 0$,

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

m balls are thrown into n bins.

X_i : number of balls in the i-th bin

$$\mu = \mathbb{E}[X_i] = \frac{m}{n}$$

Chernoff Bound: For $\delta > 0$,

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

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

$$\Pr\left[X_i \geq L\right] \leq \frac{\mathrm{e}^{L}}{\mathrm{e}L^{L}} \leq \frac{1}{n^2} \qquad \text{for } L = \frac{\mathrm{e}\ln n}{\ln\ln n}$$

- **union bound:**

$$\Pr\left[\max_{1\leq i\leq n} X_i \geq L\right] \leq n\Pr\left[X_i \geq L\right] \leq \frac{1}{n}$$

Max load is $O\left(\frac{\log n}{\log\log n}\right)$ w.h.p.

m balls are thrown into n bins.

X_i : number of balls in the i-th bin

$$\mu = \mathbb{E}[X_i] = \frac{m}{n}$$

Chernoff Bound: For $L \geq 2e\mu$,

$$\Pr\left[X_i \geq L\right] \leq 2^{-L}$$

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

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

- **union bound:**

$$\Pr\left[\max_{1\leq i\leq n} X_i \geq \frac{2em}{n}\right] \leq n \Pr\left[X_i \geq \frac{2em}{n}\right] \leq \frac{1}{n}$$

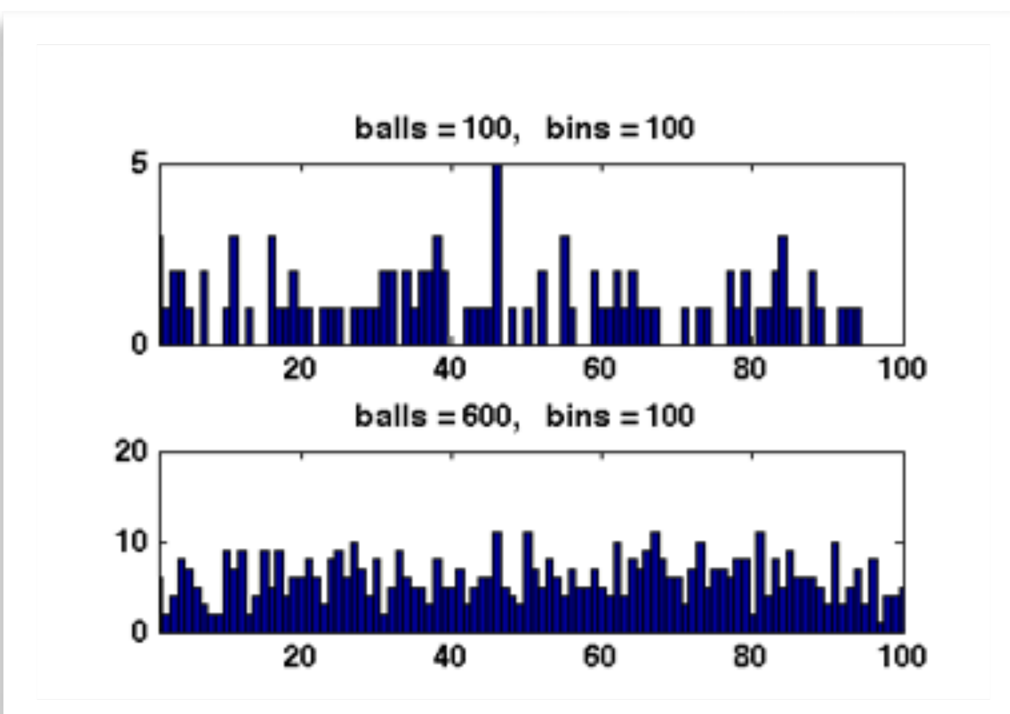
Max load is $O\left(\frac{m}{n}\right)$ w.h.p.

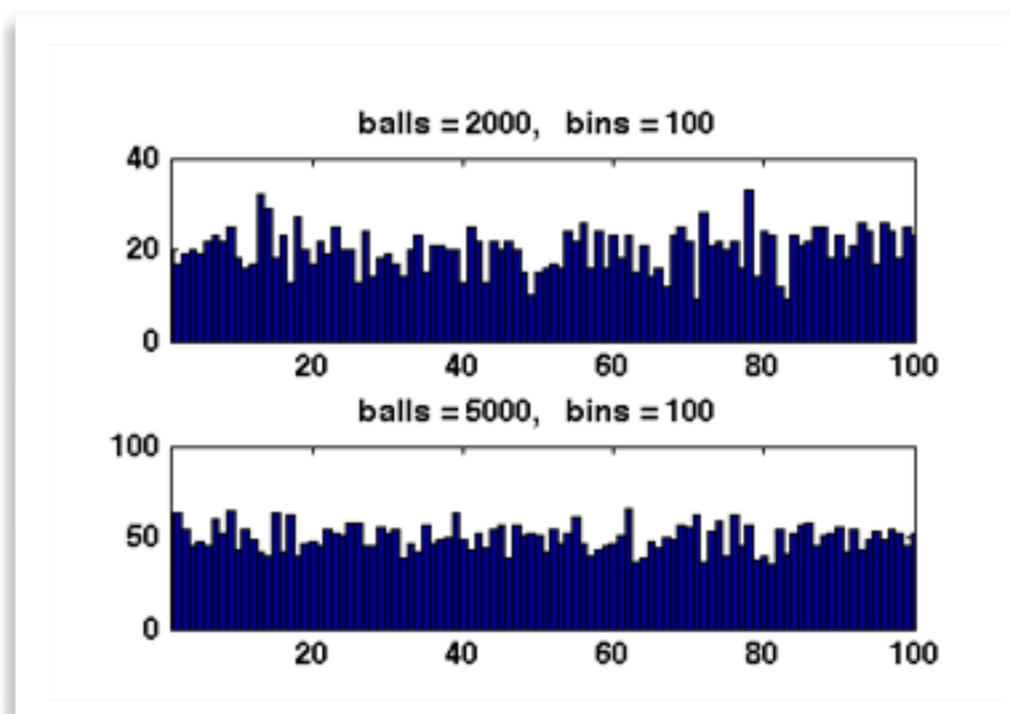
Balls into Bins

- m balls are thrown into n bins uniformly and independently:

Theorem: With high probability, the maximum load is

$$\begin{cases} O\left(\dfrac{\log n}{\log\log n}\right) & \text{when } m = n \\ O\left(\dfrac{m}{n}\right) & \text{when } m \geq n \ln n \end{cases}$$





Chernoff Bound

Chernoff Bound:

For *independent* $X_1, \ldots, X_n \in \{0,1\}$ with

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

For any $\delta > 0$,

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

For any $0 < \delta < 1$,

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

Markov's Inequality

Markov's Inequality

For *nonnegative* random variable X, for any $t > 0$,

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

Corollary

For random variable X and ***nonnegative-valued*** function f, for any $t > 0$,

$$\Pr[f(X) \geq t] \leq \frac{\mathbb{E}[f(X)]}{t}$$

Moment Generating Function

Moment generating function (MGF):

The MGF of a random variable X is defined as

$$M(\lambda) = \mathbb{E}\left[\mathrm{e}^{\lambda X}\right].$$

- Taylor's expansion:

$$\mathbb{E}\left[\mathrm{e}^{\lambda X}\right] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} X^k\right] = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathbb{E}\left[X^k\right]$$

- Independent $X_1, \ldots, X_n \in \{0,1\}$ with

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

- **Markov for MGF**: (for any $\lambda > 0$) (Markov's inequality)

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

- **Bound MGF**:

(independence) $\left(\mu = \sum_{i=1}^{n} p_i\right)$

$$\mathbb{E}\left[\mathrm{e}^{\lambda X}\right] = \mathbb{E}\left[\prod_{i=1}^{n} \mathrm{e}^{\lambda X_i}\right] = \prod_{i=1}^{n} \mathbb{E}\left[\mathrm{e}^{\lambda X_i}\right] \leq \prod_{i=1}^{n} \mathrm{e}^{(\mathrm{e}^{\lambda}-1)p_i} = \mathrm{e}^{(\mathrm{e}^{\lambda}-1)\mu}$$

$$\mathbb{E}\left[\mathrm{e}^{\lambda X_i}\right] = p_i \cdot \mathrm{e}^{\lambda \cdot 1} + (1-p_i)\mathrm{e}^{\lambda \cdot 0} = 1 + (\mathrm{e}^{\lambda} - 1)p_i \leq \mathrm{e}^{(\mathrm{e}^{\lambda}-1)p_i}$$

(where $p_i = \Pr[X_i = 1]$)

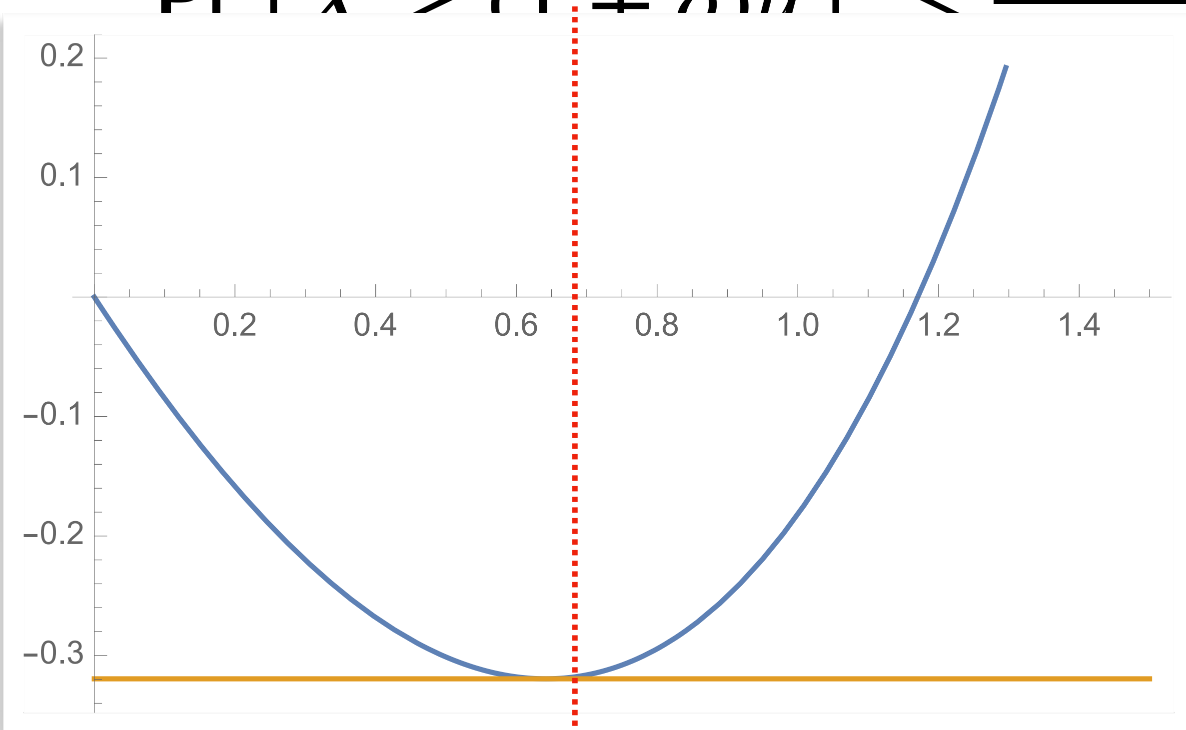
- Independent $X_1, \ldots, X_n \in \{0,1\}$ with

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

- **Markov for MGF**: (for any $\lambda > 0$)

$$\Pr\left[X \geq (1+\delta)\mu\right] \leq \frac{\mathbb{E}\left[e^{\lambda X}\right]}{\cdots\delta)\mu} \leq e^{\left(e^{\lambda}-1-\lambda(1+\delta)\right)\mu} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

(when $\lambda = \ln(1+\delta)$)



$$= \prod_{i=1}^{n} \mathbb{E}\left[e^{\lambda X_i}\right] \leq e^{(e^{\lambda}-1)\mu}$$

- **Optimization**:

$\left(e^{\lambda} - 1 - \lambda(1+\delta)\right)$ achieves Min at stationary point $\lambda = \ln(1+\delta)$

- Independent $X_1, \ldots, X_n \in \{0,1\}$ with

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

- **Markov for MGF**: (for any $\lambda > 0$)

$$\Pr\left[X \geq (1+\delta)\mu\right] \leq \frac{\mathbb{E}\left[e^{\lambda X}\right]}{e^{\lambda(1+\delta)\mu}} \leq e^{\left(e^{\lambda}-1-\lambda(1+\delta)\right)\mu} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

(when $\lambda = \ln(1+\delta)$)

- **Bound MGF**:

$$\mathbb{E}\left[e^{\lambda X}\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{\lambda X_i}\right] \leq e^{(e^{\lambda}-1)\mu}$$

- **Optimization**:

$\left(e^{\lambda} - 1 - \lambda(1+\delta)\right)$ achieves Min at stationary point $\lambda = \ln(1+\delta)$

Chernoff Bound

Chernoff Bound:

For *independent* $X_1, \ldots, X_n \in \{0,1\}$ with

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

For any $\delta > 0$,

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

Chernoff Bound

Chernoff Bound:

For *independent* $X_1, \ldots, X_n \in \{0,1\}$ with

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

For any $0 < \delta < 1$,

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

(for any $\lambda < 0$)

$$\Pr\left[X \le (1-\delta)\mu\right] \le \Pr\left[e^{\lambda X} \ge e^{\lambda(1-\delta)\mu}\right] \le \frac{\mathbb{E}\left[e^{\lambda X}\right]}{e^{\lambda(1-\delta)\mu}} \le e^{(e^{\lambda}-1-\lambda(1-\delta))\mu}$$

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

Chernoff Bound:

For *independent* or *negatively associated* $X_1, \ldots, X_n \in \{0,1\}$

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

For any $\delta > 0$,

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

For any $0 < \delta < 1$,

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

For *negatively associated* $X_1, \ldots, X_n \in \{0,1\}$:

$$\mathbb{E}\left[\mathrm{e}^{\lambda X}\right] = \mathbb{E}\left[\prod_{i=1}^{n} \mathrm{e}^{\lambda X_i}\right] \leq \prod_{i=1}^{n} \mathbb{E}\left[\mathrm{e}^{\lambda X_i}\right]$$

Chernoff-Hoeffding Bound

Chernoff Bound:

For $X = \sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n \in \{0,1\}$ are *independent* (or *negatively associated*),

for any $t > 0$:

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

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

A party of $O(\sqrt{n \log n})$ can manipulate a vote w.h.p. against n voters who neither care (uniform) nor communicate (independent).

Chernoff-Hoeffding Bound

Hoeffding Bound:

For $X = \sum_{i=1}^{n} X_i$, where $X_i \in [a_i, b_i]$, $1 \le i \le n$, are *independent* (or *negatively associated*),

for any $t > 0$:

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

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

$X = \sum_{i=1}^{n} X_i$, where $X_i \in [a_i, b_i]$ for every $1 \leq i \leq n$

let $\begin{array}{l} Y = X - \mathbb{E}[X] \\ Y_i = X_i - \mathbb{E}[X_i] \end{array} \implies \begin{cases} \mathbb{E}[Y] = \mathbb{E}[Y_i] = 0 \\ Y = \sum_{i=1}^{n} Y_i \end{cases}$

(for $\lambda > 0$) (neg. assoc.)

$$\Pr[\,X - \mathbb{E}[X] \geq t\,] = \Pr[Y \geq t] \leq \mathrm{e}^{-\lambda t}\mathbb{E}\left[\mathrm{e}^{\lambda Y}\right] \leq \mathrm{e}^{-\lambda t}\prod_{i=1}^{n}\mathbb{E}\left[\mathrm{e}^{\lambda Y_i}\right]$$

Hoeffding's Lemma: For any $Y \in [a, b]$ with $\mathbb{E}[Y] = 0$,

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

$$\leq \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

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

$X = \sum_{i=1}^{n} X_i$, where $X_i \in [a_i, b_i]$ for every $1 \le i \le n$

let $\begin{matrix} Y = X - \mathbb{E}[X] \\ Y_i = X_i - \mathbb{E}[X_i] \end{matrix} \implies \begin{cases} \mathbb{E}[Y] = \mathbb{E}[Y_i] = 0 \\ Y = \sum_{i=1}^{n} Y_i \end{cases}$

$$\Pr[X - \mathbb{E}[X] \le -t] = \Pr[Y \le -t] \underset{(\text{for } \lambda < 0)}{\le} e^{\lambda t}\mathbb{E}\left[e^{\lambda Y}\right] \underset{(\text{neg. assoc.})}{\le} e^{\lambda t}\prod_{i=1}^{n}\mathbb{E}\left[e^{\lambda Y_i}\right]$$

Hoeffding's Lemma: For any $Y \in [a, b]$ with $\mathbb{E}[Y] = 0$,

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

$$\le \exp\left(\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right) \le \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

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

Chernoff-Hoeffding Bound

Hoeffding Bound:

For $X = \sum_{i=1}^{n} X_i$, where $X_i \in [a_i, b_i]$, $1 \le i \le n$, are *independent* (or *negatively associated*),

for any $t > 0$:

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

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

Sub-Gaussian Random Variables

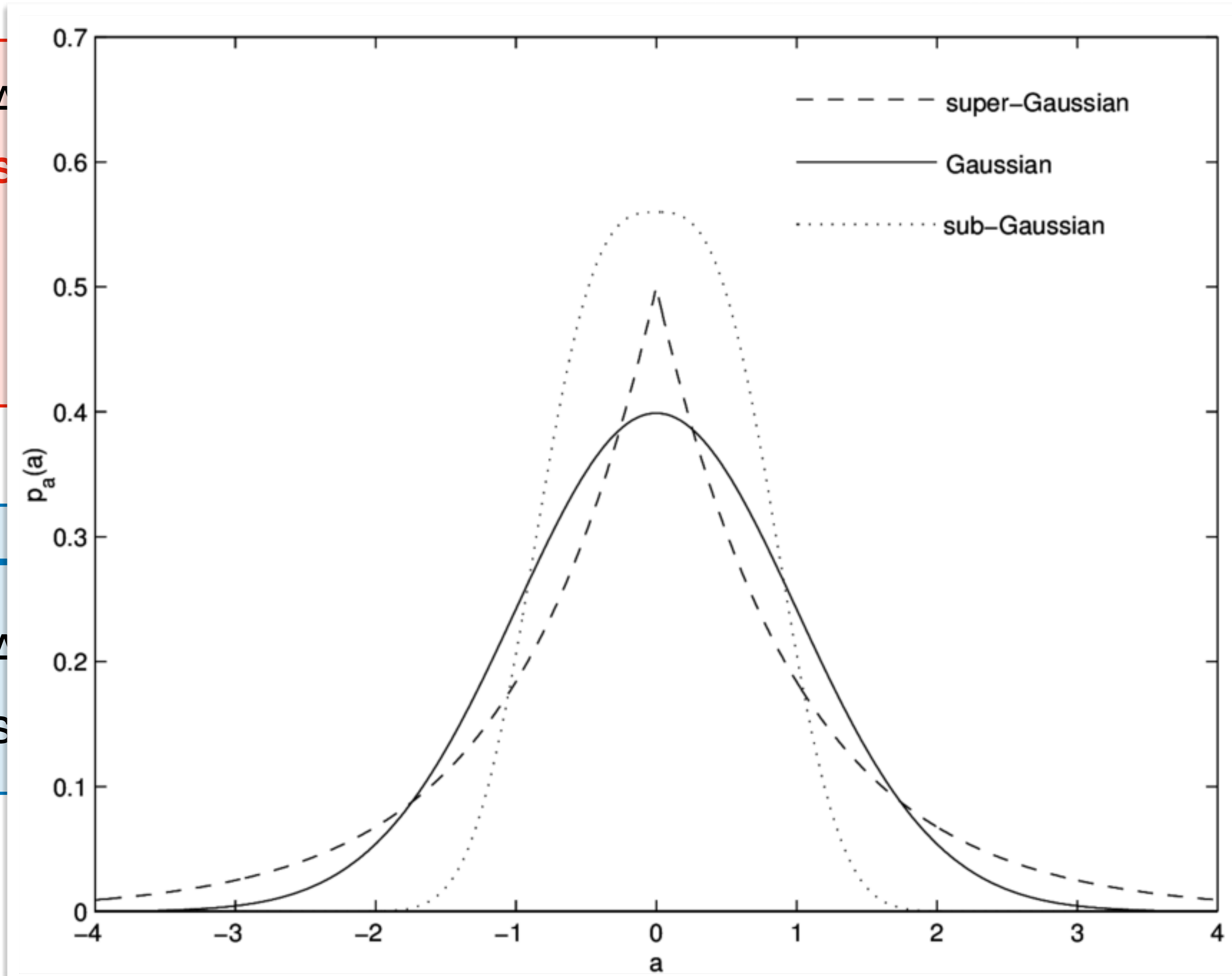
A ***centered*** ($\mathbb{E}[Y] = 0$) random variable Y is said to be **sub-Gaussian with variance factor ν** (denoted $Y \in \mathcal{G}(\nu)$) if

$$\mathbb{E}\left[\mathrm{e}^{\lambda Y}\right] \leq \exp\left(\frac{\lambda^2 \nu}{2}\right)$$

Hoeffding's Lemma:

Any centered bounded random variable $Y \in [a, b]$ is sub-Gaussian with variance factor $(b-a)^2/4$.

Sub-Gaussian Random Variables



Sub-Gaussian Random Variables

Chernoff-Hoeffding:

For $Y = \sum_{i=1}^{n} Y_i$, where $Y_i \in \mathcal{G}(\nu_i)$, $1 \leq i \leq n$, are *independent* (or ***negatively associated***) and centered (i.e. $\mathbb{E}[Y_i] = 0$)

for any $t > 0$:

$$\Pr[\, Y \geq t \,] \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} \nu_i}\right)$$

$$\Pr[\, Y \leq -t \,] \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} \nu_i}\right)$$

The Method of Bounded Differences

The Method of Bounded Differences

McDiarmid's Inequality:

For **independent** $X_1, X_2 \ldots, X_n$, if n-variate function f satisfies the **Lipschitz condition**: for every $1 \leq i \leq n$,

$$\left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) \right| \leq c_i$$

for any possible $x_1, \ldots, x_n$ and y_i,

then for any $t > 0$:

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

- **Chernoff**: sum of Boolean variables, 1-Lipschitz
- **Hoeffding**: sum of $[a_i, b_i]$-bounded variables, $(b_i - a_i)$-Lipschitz

Balls into Bins

m balls are thrown into n bins
Y : number of empty bins

$$Y_i = \begin{cases} 1 & \text{bin } i \text{ is empty} \\ 0 & \text{otherwise} \end{cases}$$

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

linearity of expectation:

$$\mathbb{E}[Y] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = n\left(1 - \frac{1}{n}\right)^m$$

$$\Pr\Big[\,\big|Y - \mathbb{E}[Y]\big| \geq t\,\Big] < ?$$

Y_i's are dependent

Balls into Bins

m balls are thrown into n bins
Y : number of empty bins

$$\mathbb{E}[Y] = n\left(1 - \frac{1}{n}\right)^m$$

X_j : the index of the bin into which the j-th ball is thrown

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

$Y = f(X_1, \ldots, X_m) = n - \left|\{X_1, \ldots, X_m\}\right|$ is 1-Lipschitz

$$\Pr\left[\,\left|Y - \mathbb{E}[Y]\right| \geq t\,\right] = \Pr\left[\,\left|f(X_1, \ldots, X_m) - \mathbb{E}[f(X_1, \ldots, X_m)]\right| \geq t\,\right]$$

(McDiarmid's inequality) $\leq 2\exp\left(-\frac{t^2}{2m}\right)$

Pattern Matching

uniform random string $X \in \Sigma^n$ with alphabet size $|\Sigma| = m$

fixed pattern $\pi \in \Sigma^k$

Y : number of substrings of X matching the pattern π

$$Y_i = \begin{cases} 1 & \text{if } X_i X_{i+1} \cdots X_{i+k-1} = \pi \\ 0 & \text{otherwise} \end{cases} \qquad Y = \sum_{i=1}^{n-k+1} Y_i$$

$$\mathbb{E}[Y_i] = \Pr[\, X_i X_{i+1} \cdots X_{i+k-1} = \pi \,] = \frac{1}{m^k}$$

linearity of expectation:

$$\mathbb{E}[Y] = \sum_{i=1}^{n-k+1} \mathbb{E}[Y_i] = \frac{n-k+1}{m^k}$$

Pattern Matching

uniform random string $X \in \Sigma^n$ with alphabet size $|\Sigma| = m$

fixed pattern $\pi \in \Sigma^k$

Y : number of substrings of X matching the pattern π

$$\mathbb{E}[Y] = \frac{n-k+1}{m^k}$$

$X_1, \ldots, X_n \in \Sigma$ are independent

$$Y = f_\pi(X_1, \ldots, X_n) = \sum_{i=1}^{n-k+1} I\left[X_i X_{i+1} \cdots X_{i+k-1} = \pi\right]$$

is k-Lipschitz

$$\Pr\left[\,\left|Y - \mathbb{E}[Y]\right| \geq t\,\right] = \Pr\left[\,\left|f_\pi(X_1, \ldots, X_n) - \mathbb{E}[f_\pi(X_1, \ldots, X_n)]\right| \geq t\,\right]$$

(McDiarmid's inequality)

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

Sprinkling Points on Hypercube

uniform random point $X \in \{0,1\}^n$ in hypercube

fixed subset $S \subseteq \{0,1\}^n$

Y : shortest *Hamming* distance from X to S

Hamming distance: $H(x, y) = \sum_{i=1}^{n} |x_i - y_i|$ for $x, y \in \{0,1\}^n$

$$Y = \min_{y \in S} H(X, y) \;\; = f_S(X_1, \ldots, X_n)$$ is 1-Lipschitz

$$\Pr\Big[\, \big| Y - \mathbb{E}[Y] \big| \geq t \,\Big] = \Pr\Big[\, \big| f_S(X_1, \ldots, X_n) - \mathbb{E}[f_S(X_1, \ldots, X_n)] \big| \geq t \,\Big]$$

(McDiarmid's inequality) $$\leq 2 \exp\left(-\frac{t^2}{2n} \right)$$

Sprinkling Points on Hypercube

uniform random point $X \in \{0,1\}^n$ in hypercube

fixed subset $S \subseteq \{0,1\}^n$

Y : shortest *Hamming* distance from X to S

Hamming distance: $H(x, y) = \sum_{i=1}^{n} |x_i - y_i|$ for $x, y \in \{0,1\}^n$

$$\Pr\left[\,\left|Y - \mathbb{E}[Y]\right| \geq \sqrt{2cn\ln n}\,\right] \leq 2n^{-c}$$

the distance to S is pretty much the same from pretty much everywhere

(unless S is very big)

The Method of Bounded Differences

McDiarmid's Inequality:

For independent $X_1, X_2 \ldots, X_n$, if n-variate function f satisfies the Lipschitz condition: for every $1 \le i \le n$,

$$\left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) \right| \le c_i$$

for any possible $x_1, \ldots, x_n$ and y_i,

then for any $t > 0$:

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

Every Lipschitz function is well approximated by a constant function under product measures.

Martingale

Martingale:

A sequence of random variables $X_0, X_1, \ldots$ is a martingale if for all $t > 0$,

$$\mathbb{E}\left[X_t \mid X_0, X_1, \ldots, X_{t-1}\right] = X_{t-1}.$$

- For random variable X and event A: (discrete probability)

$$\mathbb{E}[\, X \mid A \,] = \sum_x x \Pr[\, X = x \mid A \,]$$

- For random variables X and Y (not necessarily independent):

$f(y) = \mathbb{E}[\, X \mid Y = y \,]$ is well-defined

$\mathbb{E}[\, X \mid Y \,] = f(Y)$ is a random variable

Martingale:

A sequence of random variables $X_0, X_1, \ldots$ is a martingale if for all $t > 0$,

$$\mathbb{E}\left[X_t \mid X_0, X_1, \ldots, X_{t-1}\right] = X_{t-1}.$$

- **Fair gambling game**: Given the capitals up until time $t-1$, the expected change to the capital after the t-th bet is 0.

A sequence of random variables $X_0, X_1, \ldots$ is:

a super-martingale if for all $t > 0$,

$$\mathbb{E}\left[X_t \mid X_0, X_1, \ldots, X_{t-1}\right] \leq X_{t-1}$$

a sub-martingale if for all $t > 0$,

$$\mathbb{E}\left[X_t \mid X_0, X_1, \ldots, X_{t-1}\right] \geq X_{t-1}$$

Martingale (Generalized)

Martingale (Generalized Version):

A sequence of random variables $Y_0, Y_1, \ldots$ is a **martingale with respect to** $X_0, X_1, \ldots$ if for all $t \geq 0$,

- Y_t is a function of $X_0, \ldots, X_t$
- $\mathbb{E}\left[Y_{t+1} \mid X_0, X_1, \ldots, X_t\right] = Y_t$

- A **fair gambling game:**
  - X_i : outcome (win/loss) of the i-th betting
  - Y_i : capital after the i-th betting

Martingale (Generalized)

Martingale (Generalized Version):

A sequence of random variables $Y_0, Y_1, \ldots$ is a **martingale with respect to** $X_0, X_1, \ldots$ if for all $t \geq 0$,

- Y_t is a function of $X_0, \ldots, X_t$
- $\mathbb{E}\left[Y_{t+1} \mid X_0, X_1, \ldots, X_t\right] = Y_t$

- A probability space: $(\Omega, \mathscr{F}, \Pr)$
- A **filtration** of σ-algebras $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots$ s.t. for all $t \geq 0$:
  - Y_t is $\mathscr{F}_t$-measurable
  - $\mathbb{E}\left[Y_{t+1} \mid \mathscr{F}_t\right] = Y_t$

Azuma's Inequality

Azuma's Inequality:

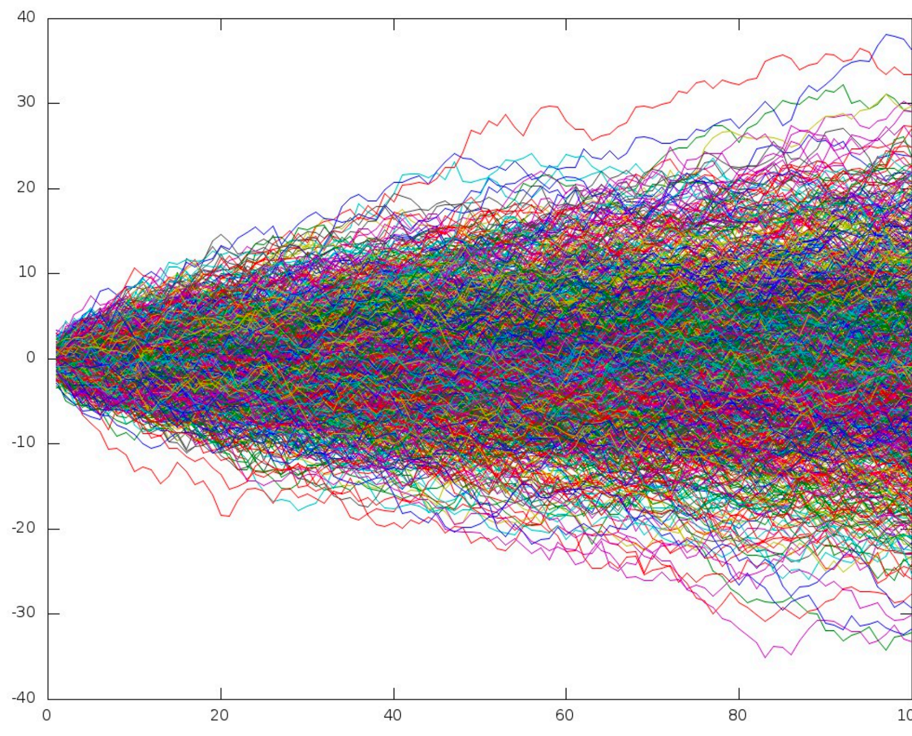
For martingale $Y_0, Y_1, \ldots$ (with respect to $X_0, X_1, \ldots$) satisfying:

$$\forall i \geq 0, \left| Y_i - Y_{i-1} \right| \leq c_i$$

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

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

- Your capital does not change too fast if:
  - the game is fair (martingale)
  - payoff for each gambling is bounded



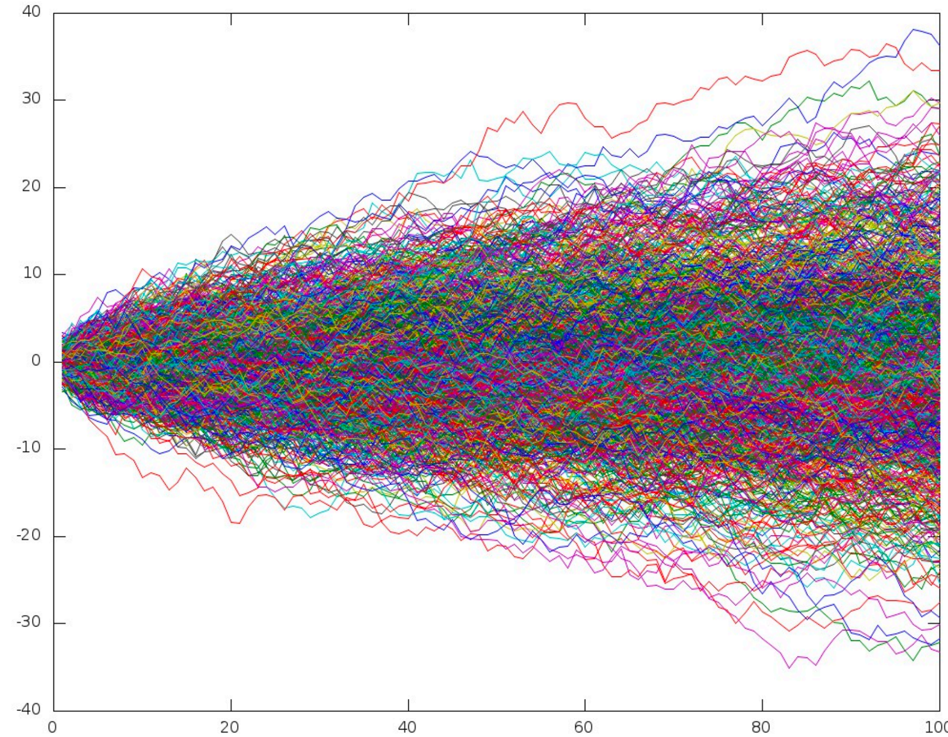
Doob Martingale

A **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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

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

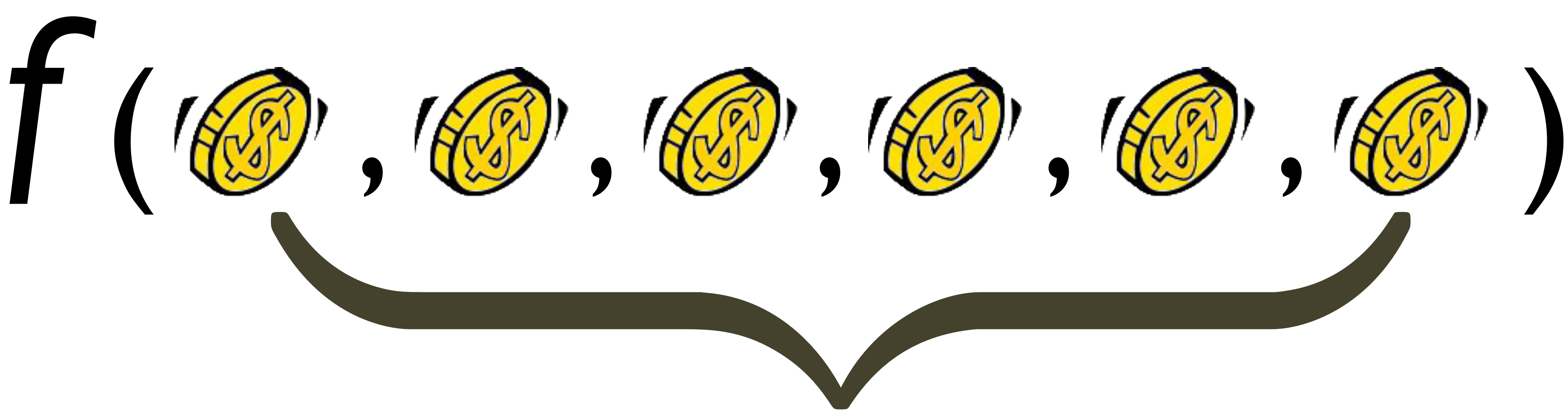
no information full information



Doob Martingale

A **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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



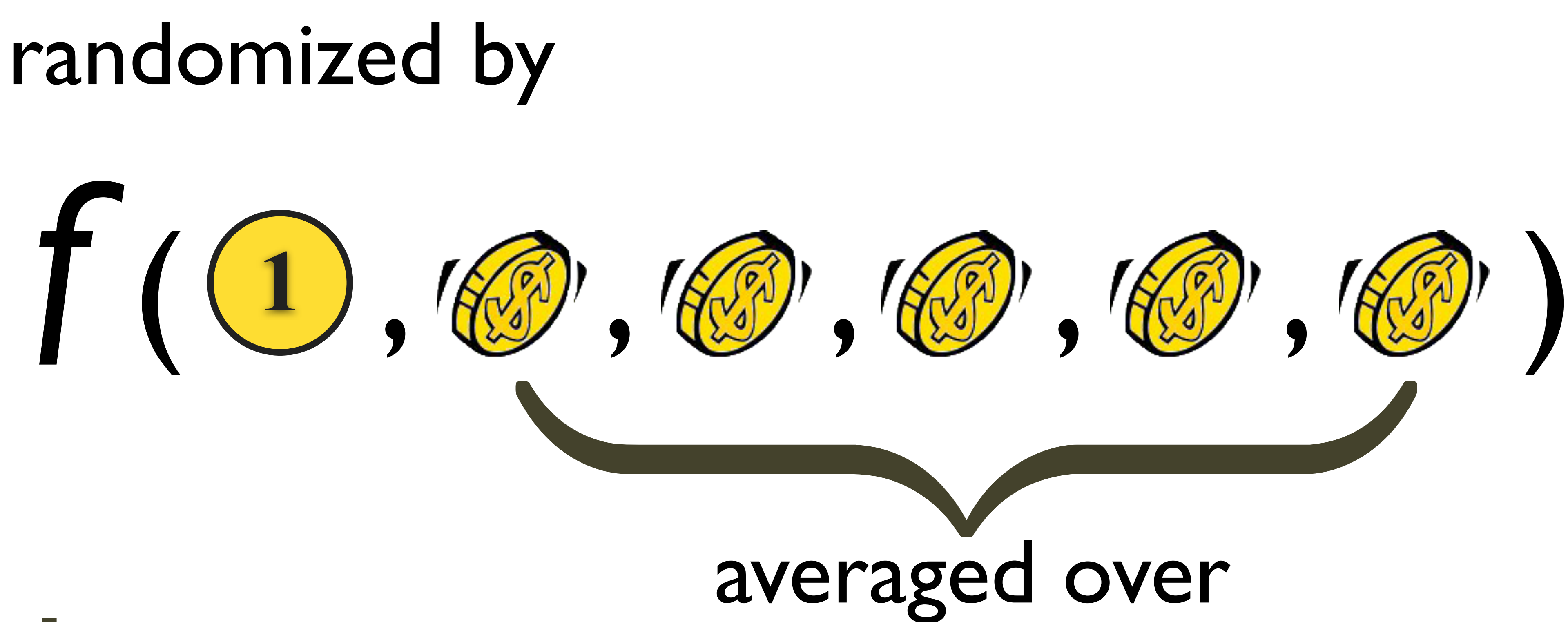
averaged over

$\mathbf{E}[f] = Y_0,$

Doob Martingale

A **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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

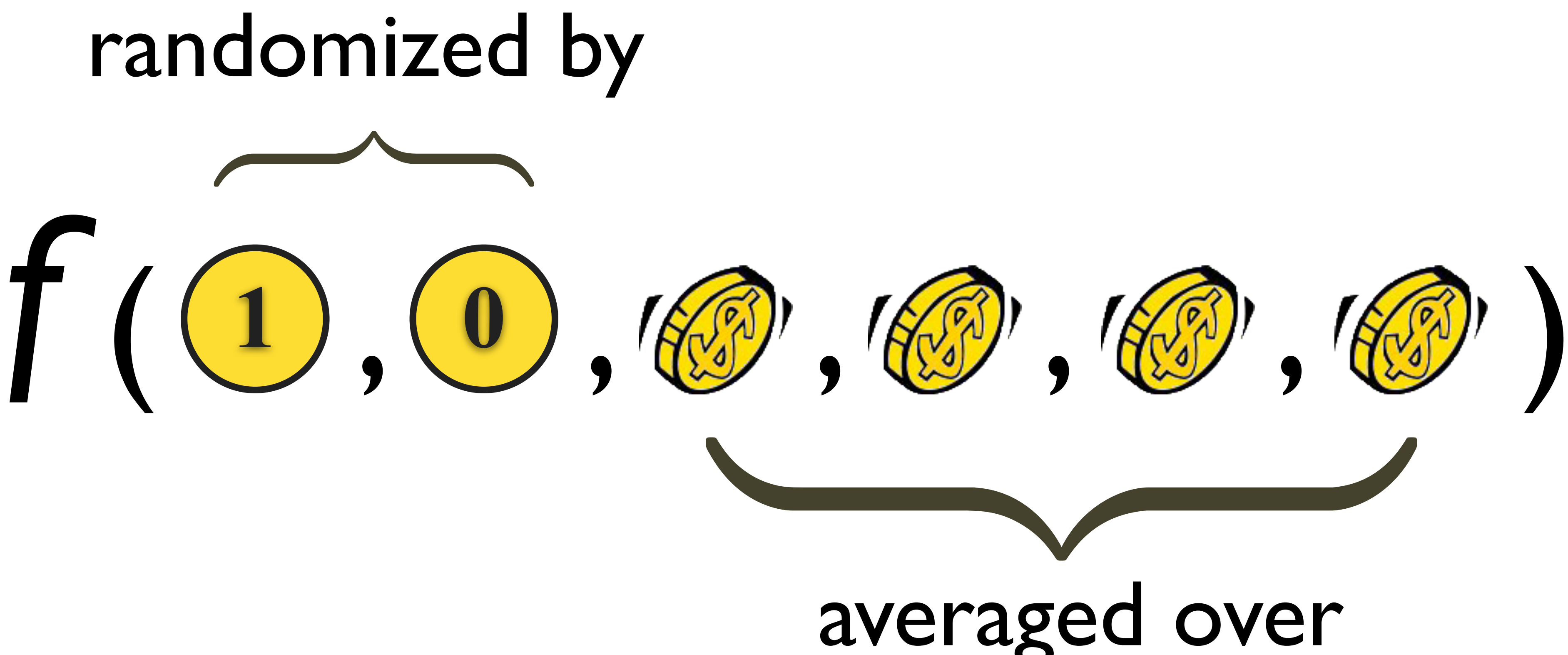


$$\mathbf{E}[f] = Y_0, \quad Y_1,$$

Doob Martingale

A **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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

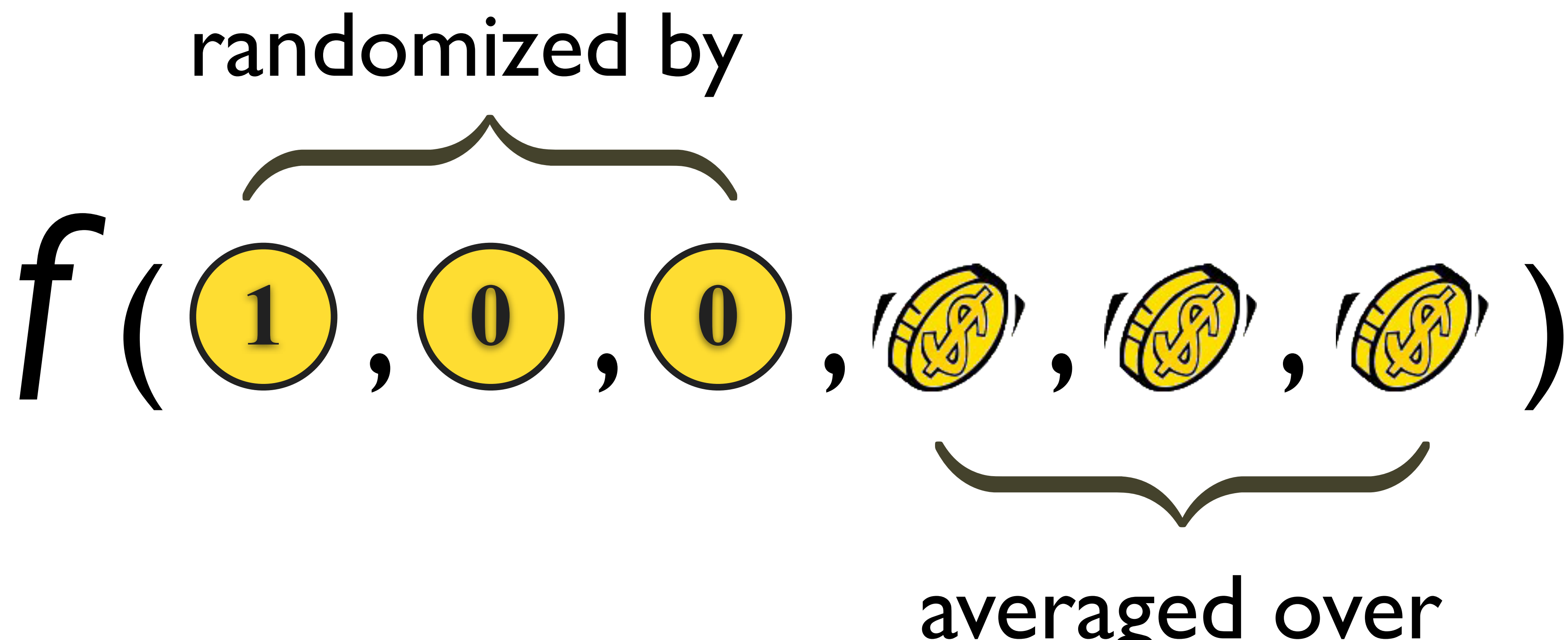


$$\mathbf{E}[f] = Y_0, \quad Y_1, \quad Y_2,$$

Doob Martingale

A **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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

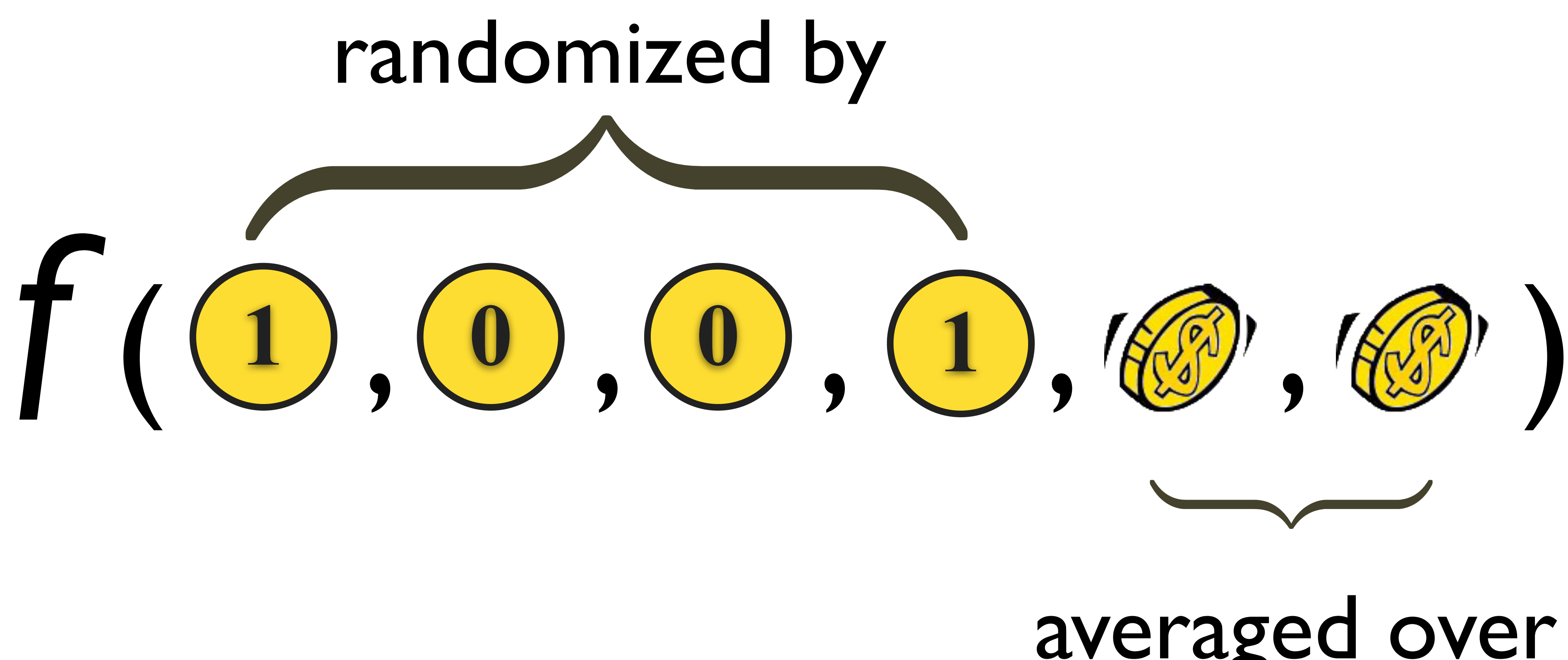


$$\mathbf{E}[f] = Y_0, \quad Y_1, \quad Y_2, \quad Y_3,$$

Doob Martingale

A Doob sequence $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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

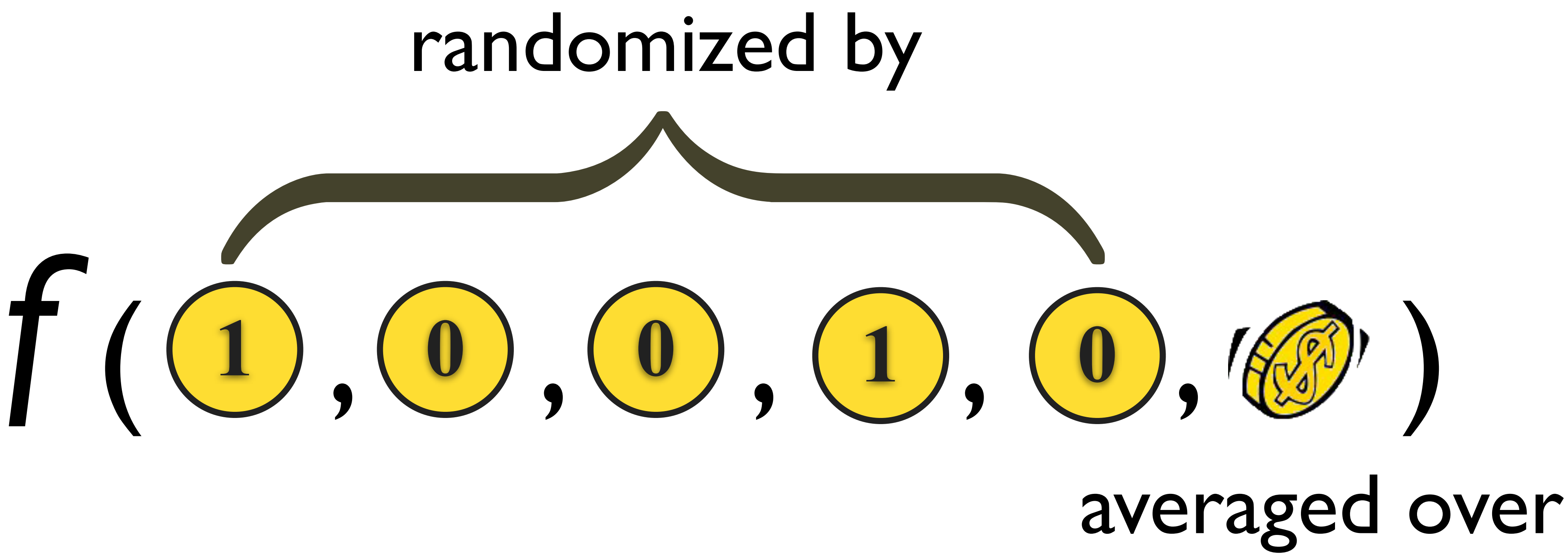


$$\mathbf{E}[f] = Y_0, \quad Y_1, \quad Y_2, \quad Y_3, \quad Y_4,$$

Doob Martingale

A **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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



$$\mathbf{E}[f] = Y_0, \quad Y_1, \quad Y_2, \quad Y_3, \quad Y_4, \quad Y_5,$$

Doob Martingale

A **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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

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

no information

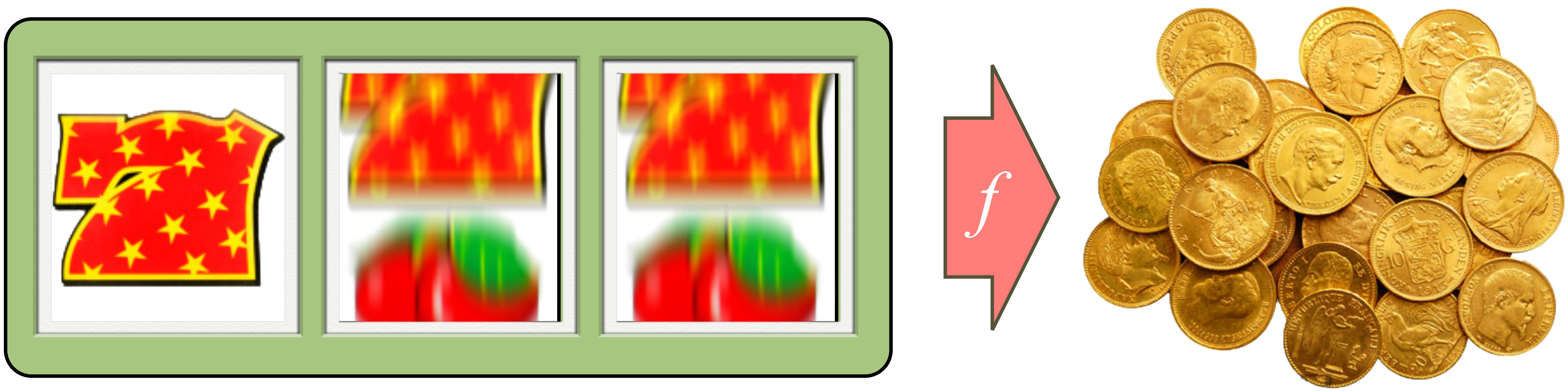
full information

$$\mathbf{E}[f] = Y_0, \quad Y_1, \quad Y_2, \quad Y_3, \quad Y_4, \quad Y_5, \quad Y_6 \quad = f$$

Doob Martingale

A **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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



Doob Martingale

A Doob sequence $Y_0, Y_1, \ldots, Y_n$ of an n-variate function f with respect to a random vector $(X_1, \ldots, X_n)$ is:

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

Theorem:

The Doob sequence $Y_0, Y_1, \ldots, Y_n$ is a martingale w.r.t. $X_1, \ldots, X_n$.

- $\forall 0 \leq i \leq n$, Y_i is a function of $X_1, \ldots, X_i$
- $\mathbb{E}\left[Y_i \mid X_1, \ldots, X_{i-1}\right]$

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

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

The Method of Bounded Differences

The Method of Bounded Differences:

For n-variate function f on random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ satisfying the Lipschitz condition: for every $1 \leq i \leq n$

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

for any $t > 0$:

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

The Method of Bounded Differences

The Method of Bounded Differences:

For n-variate function f on random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ satisfying the Lipschitz condition: for every $1 \le i \le n$

$$\left| \underbrace{\mathbb{E}\left[f(\boldsymbol{X}) \mid X_1, \ldots, X_i \right]}_{Y_i} - \underbrace{\mathbb{E}\left[f(\boldsymbol{X}) \mid X_1, \ldots, X_{i-1} \right]}_{Y_{i-1}} \right| \le c_i$$

for any $t > 0$:

$$\Pr\left[\left| \underbrace{f(\boldsymbol{X})}_{Y_n} - \underbrace{\mathbb{E}[f(\boldsymbol{X})]}_{Y_0} \right| \ge t \right] \le \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} c_i^2} \right)$$

Doob martingale: $Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$

Azuma's Inequality

Azuma's Inequality:

For martingale $Y_0, Y_1, \ldots$ (with respect to $X_0, X_1, \ldots$) satisfying:

$$\forall i \geq 0, \left| Y_i - Y_{i-1} \right| \leq c_i$$

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

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

The Method of Bounded Differences

The Method of Bounded Differences:

For n-variate function f on random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ satisfying the Lipschitz condition: for every $1 \le i \le n$

$$\Bigg| \underbrace{\mathbb{E}\left[f(\boldsymbol{X}) \mid X_1, \ldots, X_i\right]}_{Y_i} - \underbrace{\mathbb{E}\left[f(\boldsymbol{X}) \mid X_1, \ldots, X_{i-1}\right]}_{Y_{i-1}} \Bigg| \le c_i$$

for any $t > 0$:

(Azuma) $$\Pr\left[\left|\underbrace{f(\boldsymbol{X})}_{Y_n} - \underbrace{\mathbb{E}[f(\boldsymbol{X})]}_{Y_0}\right| \ge t\right] \le \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} c_i^2}\right)$$

Doob martingale: $Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$

The Method of Bounded Differences

The Method of Bounded Differences:

For n-variate function f on random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ satisfying the average-case Lipschitz condition: for every $1 \le i \le n$

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

usually difficult to verify

for any $t > 0$:

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

worst-case Lipschitz: for every $1 \le i \le n$,

$$\left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) \right| \le c_i$$

for any possible $x_1, \ldots, x_n$ and y_i

The Method of Bounded Differences

McDiarmid's Inequality:

For independent $X_1, X_2 \ldots, X_n$, if n-variate function f satisfies the Lipschitz condition: for every $1 \le i \le n$,

$$\left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n) \right| \le c_i$$

for any possible $x_1, \ldots, x_n$ and y_i,

then for any $t > 0$:

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

$$\left.\begin{array}{c} \text{worst-case Lipschitz condition} \\ + \\ \text{independent } X_1, \ldots, X_n \end{array}\right\} \Longrightarrow \begin{array}{c} \text{average-case} \\ \text{Lipschitz condition} \end{array}$$

Martingale Concentration

Azuma's Inequality:

For martingale $Y_0, Y_1, \ldots$ (with respect to $X_0, X_1, \ldots$) satisfying:

$$\forall i \geq 0, \left| Y_i - Y_{i-1} \right| \leq c_i$$

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

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

Difference: $D_i = Y_i - Y_{i-1}$

$$S_n = \sum_{i=1}^{n} D_i = Y_n - Y_0$$

Martingale difference: D_i is a function of $X_0, \ldots, X_i$

$$\begin{aligned} \mathbb{E}\left[D_i \mid X_0, \ldots, X_{i-1} \right] &= \mathbb{E}\left[Y_i - Y_{i-1} \mid X_0, \ldots, X_{i-1} \right] \\ &= \mathbb{E}\left[Y_i \mid X_0, \ldots, X_{i-1} \right] - \mathbb{E}\left[Y_{i-1} \mid X_0, \ldots, X_{i-1} \right] \\ &= Y_{i-1} - Y_{i-1} \quad = 0 \end{aligned}$$

- Martingale property:
  - D_i is a function of $X_0, \ldots, X_i$ and $\mathbb{E}\left[D_i \mid X_0, \ldots, X_{i-1}\right] = 0$
- Bounded differences: $\forall i \geq 1, \left|D_i\right| \leq c_i$

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

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

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

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

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

- Martingale property:
  - D_i is a function of $X_0, \ldots, X_i$ and $\mathbb{E}\left[D_i \mid X_0, \ldots, X_{i-1} \right] = 0$
- Bounded differences: $\forall i \geq 1, \left| D_i \right| \leq c_i$

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

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

Hoeffding's Lemma: For any $Z \in [a, b]$ with $\mathbb{E}[Z] = 0$,

$$\mathbb{E}\left[\mathrm{e}^{\lambda Z}\right] \leq \mathrm{e}^{\lambda^2 (b-a)^2/8}$$

$$Z = \left(D_n \mid X_0, \ldots, X_{n-1}\right) \qquad \Longrightarrow \mathbb{E}\left[\mathrm{e}^{\lambda D_n} \mid X_0, \ldots, X_{n-1}\right] \leq \mathrm{e}^{\lambda^2 c_n^2/2}$$

- Martingale property:
  - D_i is a function of $X_0, \ldots, X_i$ and $\mathbb{E}\left[D_i \mid X_0, \ldots, X_{i-1} \right] = 0$
- Bounded differences: $\forall i \geq 1, \left| D_i \right| \leq c_i$

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

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

(for $\lambda > 0$)

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

when $\lambda = \dfrac{t}{\sum_{i=1}^{n} c_i^2}$

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

- **Martingale property:**
  - D_i is a function of $X_0, \ldots, X_i$ and $\mathbb{E}\left[D_i \mid X_0, \ldots, X_{i-1}\right] = 0$
- **Bounded differences:** $\forall i \geq 1, \left|D_i\right| \leq c_i$

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

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

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

$$\leq \exp\left(\frac{\lambda^2}{2}\sum_{i=1}^{n} c_i^2 + \lambda t\right) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} c_i^2}\right)$$

when $\lambda = \frac{-t}{\sum_{i=1}^{n} c_i^2}$

$$\mathbb{E}\left[\mathrm{e}^{\lambda S_n}\right] \leq \mathrm{e}^{\lambda^2 c_n^2/2} \cdot \mathbb{E}\left[\mathrm{e}^{\lambda S_{n-1}}\right] \leq \exp\left(\frac{\lambda^2}{2}\sum_{i=1}^{n} c_i^2\right)$$