Part I

Basic Techniques II: Concentration Bounds

Two very important inequalities

For any random variable X, any real $\lambda > 0$ and any integer $k \geq 1$ it holds:

$$Pr[|X| > \lambda] \leq \frac{\mathbf{E}(|X|^k)}{\lambda^k}$$

Case 1 $\quad k \to 1 \quad \lambda \to \lambda\mathbf{E}(|X|)$

$$Pr[|X| \geq \lambda\mathbf{E}(|X|)] \leq \frac{1}{\lambda} \qquad \textbf{Markov's inequality}$$

Case 2 $\quad k \to 2 \quad X \to X - \mathbf{E}(X), \lambda \to \lambda\sqrt{V(X)}$

$$Pr\left[|X - \mathbf{E}(X)| \geq \lambda\sqrt{V(X)}\right] \leq \frac{\mathbf{E}((X - \mathbf{E}(X))^2)}{\lambda^2 V(X)} =$$

$$= \frac{V(X)}{\lambda^2 V(X)} = \frac{1}{\lambda^2} \qquad \textbf{Chebyshev's inequality}$$

Another variant of Chebyshev's inequality:

$$\Pr[|X - \mathbf{E}(X)| \geq \lambda] \leq \frac{V(X)}{\lambda^2}$$

and this is one of the main reasons why variance is used.

Chapter 6. BASIC TECHNIQUES II: CONCENTRATION BOUNDS

Many so called **probability concentration bounds** have been already developed and broadly applied. In this chapter we derive and apply some of them - so called **tail probability bounds** - bounds on the probability that values of some random variables differ much - by some bound - from their means.

At first we determine bounds on probabilities that the random variables

$$X = \sum_{i=1}^{n} X_i,$$

differ by a fixed margin from the average, where all X_i are binary random variables with Bernoulli distribution. That is, X_i can be seen as a coin tossing with $Pr[X_i = 1] = p_i$ and $Pr[X_i = 0] = 1 - p_i$.

(Observe that as a special case $p_1 = p_2 = \dots = p_n = p$ we have a random variable X with the binomial distribution.)

At the end of the chapter we will deal with special sequences of dependent random variables, so called martingales, and also with tail bounds for martingales. That will then be applied also to the occupancy problem.

BASIC PROBLEM and METHODS - II.

- The above approach is often used to show that X lies close to $\mathbf{E}[X]$ with reasonably high probability.
- Of the large importance is the case X is the sum of random variables. For the case that these random variables are independent we derive so called Chernoff bound.
- For the case that random variables of the sum are dependent, but form so called martingale we get so called Azuma-Hoeffding bound.

Basic problem of the analysis of randomized algorithms

What is the probability of the deviation of $X = \sum_{i=1}^{n} X_i$ from its mean

$$\mathbf{E}X = \mu = \sum_{i=1}^{n} p_i$$

by a fixed factor?

Namely, let $\delta > 0$. (1) What is the probability that X is larger than $(1+\delta)\mu$?
(2) What is the probability that X is smaller than $(1-\delta)\mu$?

Notation: For a random variable X, let $\mathbf{E}\left[e^{tX}\right]$, $t > 0$ fixed, be called the **moment generating function** of X. It holds:

$$\mathbf{E}\left[e^{tX}\right] = \mathbf{E}\left[\sum_{k\geq 0} \frac{t^k X^k}{k!}\right] = \sum_{k\geq 0} t^k \frac{\mathbf{E}\left[X^k\right]}{k!}$$

Very important **Chernoff bounds** on the sum of **independent Poisson trials** are obtained when the moment generating functions of X are considered.

CHERNOFF BOUNDS - I

Theorem: Let $X_1, X_2, .., X_n$ be independent Poisson trials such that, for $1 \leq i \leq n$, $Pr\left[X_i = 1\right] = p_i$, where $0 < p_i < 1$. Then for $X = \sum_{i=1}^{n} X_i$, $\mu = E\left[X\right] = \sum_{i=1}^{n} p_i$, and any $\delta > 0$

$$Pr\left[X > (1+\delta)\mu\right] < \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} \qquad (1)$$

Proof: For any $t \in R^{>0}$

$$Pr[X > (1+\delta)\mu] = Pr\left[e^{tX} > e^{t(1+\delta)\mu}\right]$$

By applying Markov inequality to the right-hand side we get

$$Pr\left[X > (1+\delta)\mu\right] < \frac{\mathbf{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}} \qquad \text{(inequality is strict).}$$

Observe that:

$$\mathbf{E}\left[e^{tX}\right] = \mathbf{E}\left[e^{t\sum_{i=1}^{n} X_i}\right] = \mathbf{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} \mathbf{E}\left[e^{tX_i}\right],$$

$$Pr[X > (1+\delta)\mu] < \frac{\prod_{i=1}^{n} \mathbf{E}\left[e^{tX_i}\right]}{e^{t(1+\delta)\mu}}.$$

CHERNOFF BOUNDS - II.

Since $E\left[e^{tX_i}\right] = p_i e^t + (1-p_i)$, we have:

$$Pr\left[X > (1+\delta)\mu\right] < \frac{\prod_{i=1}^{n}\left[p_i e^t + 1 - p_i\right]}{e^{t(1+\delta)\mu}} = \frac{\prod_{i=1}^{n}\left[1 + p_i\left(e^t - 1\right)\right]}{e^{t(1+\delta)\mu}}.$$

By taking the inequality $1 + x < e^x$, with $x = p_i\left(e^t - 1\right)$,

$$Pr[X > (1+\delta)\mu] < \frac{\prod_{i=1}^{n} e^{p_i\left(e^t-1\right)}}{e^{t(1+\delta)\mu}} = \frac{e^{\sum_{i=1}^{n} p_i\left(e^t-1\right)}}{e^{t(1+\delta)\mu}} = \frac{e^{\left(e^t-1\right)\mu}}{e^{t(1+\delta)\mu}}.$$

Taking $t = \ln(1+\delta)$ we get our Theorem (and basic Chernoff bound), that is:

$$Pr\left[X > (1+\delta)\mu\right] < \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} \qquad (2)$$

Observe three tricks that have been used in the above proof!

COROLLARIES

From the above Chernoff bound the following corollaries can be derived

Corollary: Let $X_1, X_2, .., X_n$ be independent Poisson trials such that, for $1 \leq i \leq n$, $Pr\left[X_i = 1\right] = p_i$, where $0 < p_i < 1$. Then for

$$X = \sum_{i=1}^{n} X_i \text{ and } \mu = E\left[X\right] = \sum_{i=1}^{n} p_i,$$

it holds

1. For $0 < \delta < 1.81$
$$Pr(X > (1+\delta)\mu) \leq e^{-\mu\delta^2/3}$$
2. For $0 \leq \delta \leq 4.11$
$$Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/4}$$
3. For $R \geq 6\mu$
$$Pr(X \geq R) \leq 2^{-R} \qquad (3)$$

EXAMPLE I - SOCCER GAMES OUTCOMES

Notation: $F^+(\mu,\delta) = \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^\mu$ – the right-hand side of inequality (1) from the previous slide.

Example: A soccer team STARS wins each game with probability $\frac{1}{3}$. Assuming that outcomes of different games are independent we derive an upper bound on the probability that STARS win more than half out of n games.

Let $X_i = \begin{cases} 1, & \text{if STARS win } i\text{−th game} \\ 0, & \text{otherwise.} \end{cases}$
Let $Y_n = \sum_{i=1}^n X_i$

By applying the last theorem ($\Pr(X > (1+\delta)\mu) > F(\mu,\delta)$), we get for $\mu = \frac{n}{3}$ and $\delta = \frac{1}{2}$,

$$Pr\left[Y_n > \frac{n}{2}\right] < F^+\left(\frac{n}{3},\frac{1}{2}\right) < (0.915)^n \qquad \text{−exponentially small in } n$$

SECOND TYPE of CHERNOFF BOUNDS

Previous theorem puts an upper bound on deviations of $X = \sum X_i$ above its expectations μ, i.e. for

$$Pr\left[X > (1+\delta)\mu\right].$$

Next theorem puts a lower bound on deviations of $X = \sum X_i$ below its expectations μ, i.e. for

$$Pr\left[X < (1-\delta)\mu\right].$$

Theorem: Let $X_1, X_2, .., X_n$ be independent Poisson trials such that, for $1 \le i \le n$, $Pr[X_i = 1] = p_i$, where $0 < p_i < 1$. Then for $X = \sum_{i=1}^n X_i$, $\mu = \mathbf{E}[X] = \sum_{i=1}^n p_i$, and for $0 < \delta \le 1$

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

Proof: $Pr\left[X < (1-\delta)\mu\right] = Pr\left[-X > -(1-\delta)\mu\right] = Pr\left[e^{-tX} > e^{-t(1-\delta)\mu}\right]$ for any positive real t.
By applying Markov inequality

$$\begin{aligned} Pr[X < (1-\delta)\mu] &< \frac{\mathbf{E}\left[e^{-tX}\right]}{e^{-t(1-\delta)\mu}} = \frac{\prod_{i=1}^n \mathbf{E}\left[e^{-tX_i}\right]}{e^{-t(1-\delta)\mu}} \\ &< \frac{\prod_{i=1}^n \left[p_i e^{-t} + 1 - p_i\right]}{e^{-t(1-\delta)\mu}} = \frac{\prod_{i=1}^n \left[1 + p_i\left(e^{-t} - 1\right)\right]}{e^{-t(1-\delta)\mu}}. \end{aligned}$$

By applying the inequality $1 + x < e^x$ we get

$$Pr[X < (1-\delta)\mu] < \frac{e^{\sum_{i=1}^n p_i\left(e^{-t}-1\right)}}{e^{-t(1-\delta)\mu}} = \frac{e^{\left(e^{-t}-1\right)\mu}}{e^{-t(1-\delta)\mu}}$$

and if we take $t = \ln\frac{1}{1-\delta}$, then

$$Pr\left[X < (1-\delta)\mu\right] < \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^\mu \tag{4}$$

and then we have

$$Pr[X < (1-\delta)\mu] < e^{-\mu\frac{\delta^2}{2}}$$

From 3 and 4 it follows
Corollary: For $0 < \delta < 1$

$$Pr(|X-\mu| \ge \delta\mu) \le 2e^{-\mu\delta^2/3} \tag{5}$$

EXAMPLE - COIN TOSSING

Let X be a number of heads in a sequence of n independent fair coin flips. An application of the bound (5) gives, for $\mu = n/2$ and $\delta = \sqrt{\frac{6 \ln n}{n}}$

$$Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{1}{2}\sqrt{6n \ln n}\right) \leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{6 \ln n}{n}} = \frac{2}{n}$$

This implies that concentration of the number of heads around the mean $\frac{n}{2}$ is very tight.

Indeed, the deviations from the mean are on the order of $\mathcal{O}(\sqrt{n \ln n})$.

CHEBYSHEV versus CHERNOFF

Let X be again the number of heads in a sequence of n independent fair coin flips.

Let us consider probability of having either more than $3n/4$ or fewer than $n/4$ heads in a sequence of n independent fair coin-flips.

Chebyshev's inequality gives us

$$Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq \frac{4}{n}$$

On the other side, using Chernoff bound we have

$$Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq 2e^{-\frac{1}{3}\frac{n}{2}\frac{1}{4}} \leq 2e^{-n/24}.$$

Chernoff's method therefore gives an exponentially smaller upper bound than the upper bound obtained using Chebyshev's inequality.

SOCCER GAMES REVISITED

Notation: [For the lower tail bound function]

$$F^{-}(\mu, \delta) = e^{\frac{-\mu\delta^2}{2}}.$$

Example: Assume that the probability that STAR team wins the game is $\frac{3}{4}$. What is the probability that in n games STAR lose more than $\frac{n}{2}$ games?

In such a case $\mu = 0.75n, \delta = \frac{1}{3}$ and for $Y_n = \sum_{i=1}^{n} X_i$ we have

$$Pr\left[Y_n < \frac{n}{2}\right] < F^{-}\left(0.75n, \frac{1}{3}\right) < (0.9592)^n$$

and therefore the probability decreases exponentially fast in n.

TWO SIDED BOUNDS

By inequality (5), for $\delta < 1$,

$$Pr[|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}$$

and if we want that this bound is less than an ε, then we get

$$Pr\left[|X - \mu| \geq \sqrt{3\mu \ln(2/\varepsilon)}\right] \leq \varepsilon$$

provided $\varepsilon \geq 2e^{-\mu\delta^2/3}$.

Proof

If $\varepsilon = 2e^{-\mu\delta^2/3}$, then

$$\begin{aligned}\sqrt{3\mu \ln(2/\varepsilon)} &= \sqrt{3\mu \ln(e^{\mu\delta^2/3})} \\ &= \sqrt{3\mu \cdot \mu\delta^2/3} \\ &= \sqrt{\mu^2\delta^2} \\ &= \mu\delta\end{aligned}$$

NEW QUESTION

New question: Given ε, how large has δ be in order

$$Pr[X > (1+\delta)\mu] < \varepsilon?$$

In order to deal with such and related questions, the following definitions/notations are introduced.

Df.: $\Delta^+(\mu,\varepsilon)$ is a number such that $F^+(\mu,\Delta^+(\mu,\varepsilon)) = \varepsilon$.
$\Delta^-(\mu,\varepsilon)$ is a number such that $F^-(\mu,\Delta^-(\mu,\varepsilon)) = \varepsilon$.

In other words, a deviation of $\delta = \Delta^+(\mu,\varepsilon)$ suffices to keep $Pr[X > (1+\delta)\mu]$ bellow ε (irrespective of the values of n and p_i's).

EXAMPLE and ESTIMATIONS

There is a way to derive $\Delta^-(\mu,\varepsilon)$ explicitly.
Indeed, by taking the inequality

$$Pr[X < (1-\delta)\mu] < e^{-\frac{\mu\delta^2}{2}}$$

and setting $e^{-\frac{\mu\delta^2}{2}} = \varepsilon$ we get

$$\Delta^-(\mu,\varepsilon) = \sqrt{\frac{2\ln\frac{1}{\varepsilon}}{\mu}}. \qquad (6)$$

because $\Delta^-(\mu,\varepsilon)$ is a number such that $F^-(\mu,\Delta^-(\mu,\varepsilon)) = \varepsilon$.

Example: Let $p_i = 0.75$. How large must δ be so that $Pr[X < (1-\delta)\mu] < n^{-5}$?

From (2) it follows:

$$\delta = \Delta^-\left(0.75n, n^{-5}\right) = \sqrt{\frac{10\ln n}{0.75n}} = \sqrt{\frac{13.3\ln n}{n}}$$

SOME OTHER USEFUL ESTIMATIONS

1. $F^+(\mu,\delta) < [e/(1+\delta)]^{(1+\delta)\mu}$.
2. If $\delta > 2e-1$, then $F^+(\mu,\delta) < 2^{-(1+\delta)\mu}$,
3. $\Delta^+(\mu,\varepsilon) < \frac{\lg\frac{1}{\varepsilon}}{\mu} - 1$.
4. If $\delta \leq 2e-1$, then $F^+(\mu,\delta) < e^{-\frac{\mu\delta^2}{4}}$ and $\Delta^+(\mu,\varepsilon) < \sqrt{\frac{4\ln\frac{1}{\varepsilon}}{\mu}}$.

SUMMARY of NOTATION

Let us summarize basic relations concerning values:

$$F^+(\mu,\delta) = \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right]^\mu \text{ and } F^-(\mu,\delta) = e^{\frac{-\mu\delta^2}{2}}$$

as well as

$$\Delta^+(\mu,\varepsilon) \text{ and } \Delta^-(\mu,\varepsilon).$$

It holds

$$\Pr[X > (1+\delta)\mu] < F^+(\mu,\delta) \text{ and } \Pr[X < (1-\delta)\mu] < F^-(\mu,\delta)$$

and

$$\Pr(X > (1+\Delta^+(\mu,\varepsilon)\mu) < F^+(\mu,\Delta^+(\mu,\varepsilon)) = \varepsilon$$

$$\Pr(X < (1-\Delta^-(\mu,\varepsilon)\mu) < F^-(\mu,\Delta^-(\mu,\varepsilon)) = \varepsilon$$

EXAMPLE 2 - MONTE CARLO METHOD - I

In this example we illustrate how Chernoff bound help us to show that a simple Monte Carlo algorithm can be used to approximate number π through sampling.

The term Monte Carlo method refers to a broad collection of tools for estimating various values through sampling and simulation.

Monte Carlo methods are used extensively in all areas of physical sciences and technologies.

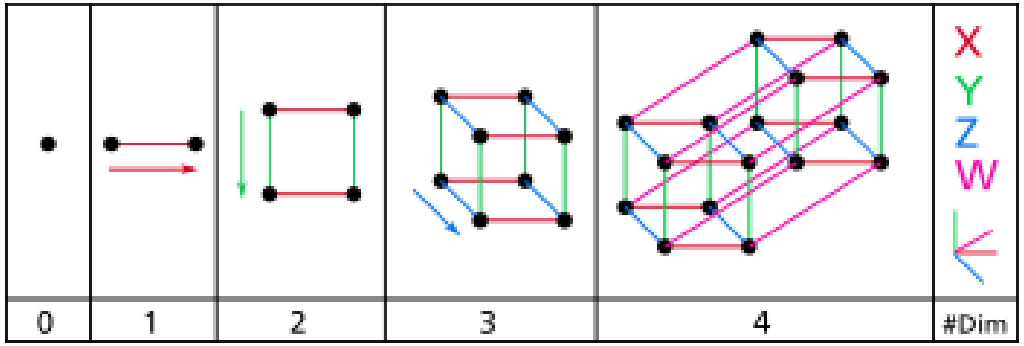
MONTE CARLO ESTIMATION OF π - I.

- Let $Z = (X, Y)$ be a point chosen randomly in a 2×2 square centered in $(0,0)$.
- This is equivalent to choosing X and Y randomly from interval $[-1,1]$.
- Let Z be considered as a random variable that has value 1 (0) if the point (X,Y) lies in the circle of radius 1 centered in the point $(0,0)$.
- Clearly
$$Pr(Z=1) = \frac{\pi}{4}$$
- If we perform such an experiment m times and Z_i be the value of Z at the ith run, and $W = \sum_{i=1}^m Z_i$, then
$$\mathbf{E}[W] = \mathbf{E}\left[\sum_{i=1}^m Z_i\right] = \sum_{i=1}^m \mathbf{E}[Z_i] = \frac{m\pi}{4}$$
and therefore $W' = (4/m)W$ is a natural estimation for π.

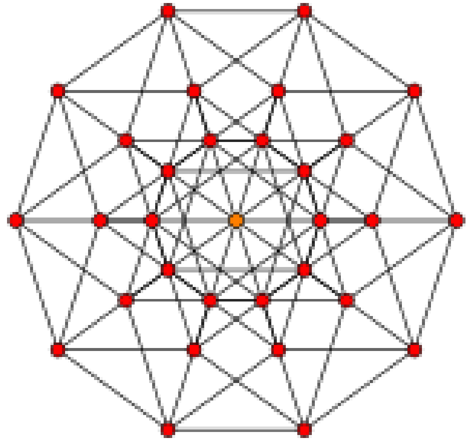
MONTE CARLO ESTIMATION OF π - II.

- How good is this estimation? An application of second Chernoff bound gives
$$\begin{aligned} Pr(|W' - \pi| \geq \varepsilon\pi) &= Pr\left(\left|W - \frac{m\pi}{4}\right| \geq \frac{\varepsilon m\pi}{4}\right) \\ &= Pr([W - \mathbf{E}[W]) \geq \varepsilon\mathbf{E}[W]) \\ &\leq 2e^{-m\pi\varepsilon^2/12} \end{aligned}$$
because $\mathbf{E}(W) = \frac{m\pi}{4}$ and for $0 < \delta < 1$
$$Pr(|X-\mu| \geq \delta\mu) \leq 2e^{-\mu\delta^2/3} \tag{7}$$
- Therefore, by taking m sufficiently large we can get an arbitrarily good approximation of π

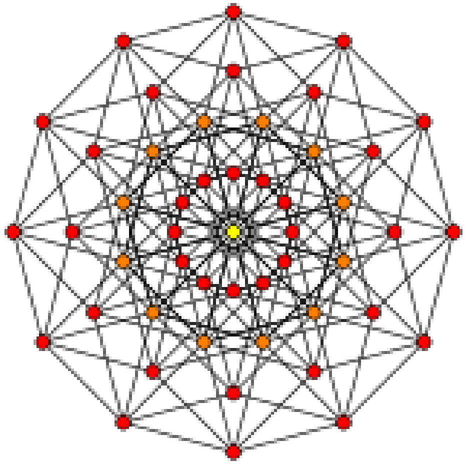
HYPERCUBES



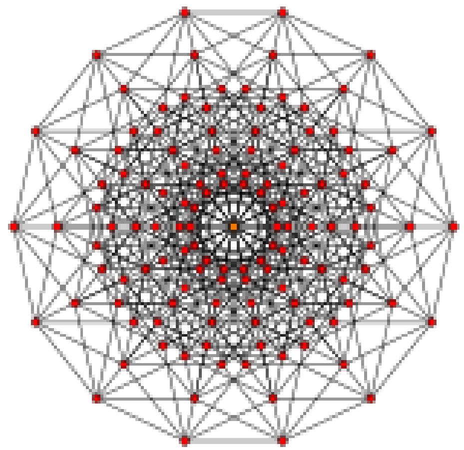
5-d hypercube



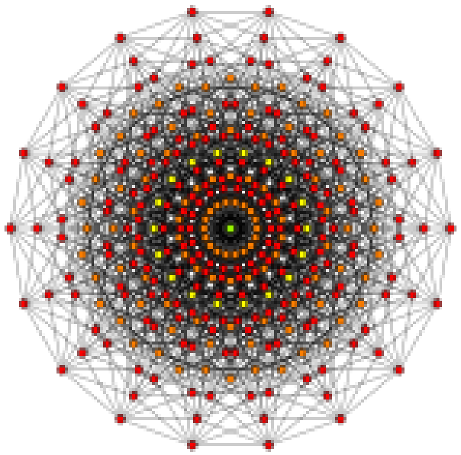
6-d hypercube



7-d hypercube



9-d hypercube



A CASE STUDY - routing on hypercubes

Networks are modeled by graphs. **Processors** by nodes and **Communication links** are represented by edges.

Principle of synchronous communication. Links (edges) can carry **packets** $(i, X, d(i))$ where i is a source node, X are data and $d(i)$ is destination node.

Permutation routing problem on an n-processor network with **nodes** $1, 2, ..., n$
Each node i wants to send, in parallel, a packet to a node $d(i)$ where $d(1), d(2), ..., d(n)$ is a permutation of $1, 2, ..., n$.

Problem: How many steps (from a node to a node) are necessary and sufficient to route an arbitrary permutation?
A **route** for a packet is a sequence of edges the packet has to follow from its source to its destination.

A **routing algorithm** for the permutation routing problem has to specify a route (in parallel) for each packet.

A packet may occasionally have to wait at a node because the next edge on its route is "busy", transmitting another packet at that moment.

We assume each node contains one **queue** for each edge. A routing algorithm must therefore specify also a **queueing discipline**.

OBLIVIOUS ROUTING ALGORITHMS

are such routing algorithms that the route followed by a packet from a source node i to a destination $d(i)$ depends on i and $d(i)$ only (and not on other $d(j)$, for $j \neq i$).

The following theorem gives a limit on the performance of oblivious algorithms.

Theorem: For any deterministic oblivious permutation routing algorithm on a network of n nodes each of the out-degree d, there is an instance of the permutation routing requiring $\Omega\left(\sqrt{\frac{n}{d}}\right)$ steps.
Example:
Consider any d-dimensional hypercube H_d and the **left-to-right routing**.

Any packet with the destination node $d(i)$ is sent from any current node n_i to the node n_j such that binary representation of n_j differs from the binary representation of n_i in the leftmost bit in which n_i and $d(i)$ differ.
Example Consider the permutation routing in H_{10} given by the "reverse" mapping $b_1 \ldots b_{10} \rightarrow b_{10} \ldots b_1$

Observe that if the left-to-right routing strategy is used, then all messages from nodes $b_1 b_2 b_3 b_4 b_5 00000$ have to go through the node 0000000000.

Left-to-right routing on hypercube H_d requires sometimes $\Omega\left(\sqrt{\frac{2^d}{d}}\right)$ steps.

RANDOMIZED ROUTING

We show now a **randomized (oblivious) routing algorithm** with expected number of steps smaller, asymptotically, than $\sqrt{\frac{2^d}{d}}$. **Notation** : $N = 2^d$

Phase 1: Pick a random intermediate destination $\sigma(i)$ from $\{1, ..., N\}$. Let the packet v_i is to travel first to the node $\sigma(i)$.

Phase 2: Let the packet v_i to travel next from $\sigma(i)$ to its final destination $D(i)$.

(In both phases the deterministic left-to-right oblivious routing is used.)

Queueing discipline: FIFO for each edge.
(Actually any queueing discipline is good provided at each step there is a packet ready to travel.)

Question: How many steps are needed before a packet v_i reaches its destination? (Let us consider at first only the Phase 1).

Let ρ_i denote the route for a packet v_i. It clearly holds:

The number of steps taken by v_i is equal to the length of ρ_i, which is at most d, plus the number of steps for which v_i is queued at intermediate nodes of ρ_i.

Fact: For any two packets v_i, v_j there is at most one queue q such that v_i and v_j are in the queue q at the same time.

Lemma: Let the route of a packet v_i follow the sequence of edges $\rho_i = (e_1, e_2, \ldots, e_k)$. Let S be the set of packets (other than v_i), whose routes pass through at least one of the edges $\{e_1, \ldots, e_k\}$. Then the delay the packet v_i makes is at most $|S|$.

Proof: A packet in S is said to leave ρ_i at that time step at which it traverses an edge of ρ_i for the last time.

If a packet is ready to follow an edge e_j at time t we define its **lag** at time t to be $t - j$.

Clearly, the lag of a packet v_i is initially 0, and the total delay of v_i is its lag when it traverses the last edge e_k of the route ρ_i.

We show now that at each step at which the lag of v_i increases by 1, the lag can be charged to a distinct member of S.

If the lag of v_i reaches a number $l+1$, some packet in S leaves ρ_i with lag l. (When the lag of v_i increases from l to $l+1$, there must be at least one packet (from S) that wishes to traverse the same edge as v_i at that time step.) Thus, S contains at least one packet whose lag is l.

Let t' be the last step any packet in S has the lag l. Thus there is a packet $v \in S$ ready to follow an edge $e_{j'}$, at $t' = l + j'$. We show that some packet of S leaves ρ_i at t'. This establish Lemma by the Fact from the slide before the previous slide.

Since v is ready to follow $e_{j'}$ at t', some packet ω (which may be v itself) in S follow edge $e_{j'}$, at t'. Now ω has to leave ρ_i at t'. We charge to ω the increase in the lag of v_i from l to $l+1$; since ω leaves ρ_i it will never be charged again.

Thus, each member of S whose route intersects ρ_i is charged for at most one delay, what proves the lemma.

PROOF CONTINUATION - I.

Let H_{ij} be the random variable defined as follows

$$H_{ij} = \begin{cases} 1 & \text{if routes } \rho_i \text{ and } \rho_j \text{ share an edge} \\ 0 & \text{otherwise} \end{cases}$$

The total delay a packet v_i makes is at most $\sum_{j=1}^{N} H_{ij}$.

Since the routes of different packets are chosen independently and randomly, the H_{ij}'s are independent Poisson trials for $j \neq i$.

Thus, to bound the delay of the packet v_i from above, using the Chernoff bound, it suffices to obtain an upper bound on $\sum_{j=1}^{N} H_{ij}$. **At first we find a bound for $\mathbf{E}\left[\sum_{j=1}^{N} H_{ij}\right]$.**

For any edge e of the hypercube let the random variable $T(e)$ denote the number of routes that pass through e.

Fix any route $\rho_i = (e_{i,1}, e_{i,2}, \ldots, e_{i,k})$, $k \leq d$. Then

$$\sum_{j=1}^{N} H_{ij} \leq \sum_{j=1}^{k} T(e_{i,j}) \Rightarrow \mathbf{E}\left[\sum_{j=1}^{N} H_{ij}\right] \leq \sum_{j=1}^{k} \mathbf{E}\left[T(e_{i,j})\right]$$

PROOF CONTINUATION - II.

It can be shown that $\mathbf{E}[T(e_{i,j})] = \mathbf{E}[T(e_{i,m})]$ for any two edges.

The expected length of ρ_i is $\frac{d}{2}$. An expectation of the total route length, summed over all packets, is therefore $\frac{Nd}{2}$. The number of edges in the hypercube is Nd and therefore $\Rightarrow \mathbf{E}[T(e_{ij})] \leq \frac{Nd/2}{Nd} = \frac{1}{2}$ for any i, j.) Therefore

$$\mathbf{E}\left[\sum_{j=1}^{N} H_{ij}\right] \leq \frac{k}{2} \leq \frac{d}{2}.$$

By the Chernoff bound (for $\delta > 2e - 1$), see page 7,

$$Pr[X > (1+\delta)\mu] < 2^{-(1+\delta)\mu}$$

with $X = \sum_{j=1}^{N} H_{ij}$, $\delta = 11$, $\mu = \frac{d}{2}$, we get that probability that $\sum_{j=1}^{N} H_{ij}$ exceeds $6d$ is less than 2^{-6d}.

The total number of packets is $N = 2^d$.

The probability that any of the N packets experiences a delay exceeding $6d$ is less than $2^d \times 2^{-6d} = 2^{-5d}$.

PROOF CONTINUATION - III.

By adding the length of the route to the delay we get:

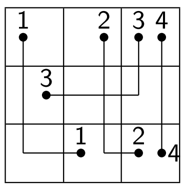
Theorem: With probability at least $1 - 2^{-5d}$ every packet reaches its intermediate destination in Phase 1 in $7d$ or fewer steps.

The routing scheme for Phase 2 can be seen as the scheme for Phase 1, which "runs backwards". Therefore the probability that any packet fails to reach its target in either phase is less than $2 \cdot 2^{-5d}$. To summarize:

Theorem: With probability at least $1 - \frac{1}{2^{5d}}$ every packet reaches its destination in $14d$ or fewer steps.

WIRING PROBLEM - I.

Global wiring in gate arrays
Gate-array: is $\sqrt{n} \times \sqrt{n}$ array of n gates.
Net - is a pair of gates to be connected by a wire.
Manhattan wiring - wires can run vertically and horizontally only.



Global wiring problem I (GWP$_W$): given a set of nets and an integer W we need to specify, if possible, the set of gates each wire should pass through in connecting its end-points, in such a way that no more than W nets pass through any boundary.

Global wiring problem II: For a boundary b between two gates in the array, let $W_S(b)$ be the number of wires that pass through b in a solution S to the global wiring problem.

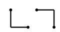
Notation: $W_S = \max_b W_S(b)$

Goal: To find S such that W_S is minimal.

WIRING PROBLEM - II.

We will consider only so called **one-bend Manhattan routing** at which direction is changed at most once.

Problem; how to decide for each net which of the following connections to use:



(that is vertical first and horizontal (right or left) next or vice verse) in order to get wiring S with minimal W_S.

REFORMULATION of the WIRING PROBLEM

Global wiring problem can be reformulated as a 0-1 linear programming problem.

For the net i we use two binary variables x_{i0}, x_{i1}

$x_{i0} = 1 \Leftrightarrow$ i-th route step has the form (horiz.=vert.) ┐
$x_{i1} = 1 \Leftrightarrow$ i-th route step has the form (vert+hori.) └

Notation:

$$T_{b0} = \{\ i \mid \text{net } i \text{ passes through } b \text{ and } x_{i0} = 1\ \}$$

and

$$T_{b1} = \{\ i \mid \text{net } i \text{ passes through } b \text{ and } x_{i1} = 1\ \}.$$

The corresponding 0-1 linear programming problem

minimize W,

where
$x_{i0}, x_{i1} \in \{0, 1\}$ for each net i (3)
$x_{i0} + x_{i1} = 1$ for each net i (4)
$\sum_{i \in T_{b0}} x_{i0} + \sum_{i \in T_{b1}} x_{i1} \leq W$ for all b. (5)

Last condition requires that at most W wires cross the boundary b

Denote W_0 the minimum obtained this way.

TRICK - I.

1. Solve the corresponding **rational linear programming problem** with

$$x_{i0}, x_{i1} \in [0, 1]$$

instead of (3).

This trick is called **linear relaxation**.

Denote $\widehat{x}_{i0}, \widehat{x}_{i1}$ solutions of the above rational linear programming problem, $1 \leq i \leq n$, and let $\widehat{W}$ be the value of the objective function for this solution. Obviously,

$$W_0 \geq \widehat{W}.$$

2. Apply the technique called **randomized rounding**.

Independently for each i, set $\overline{x}_{i0}$ to 1 with probability $\widehat{x}_{i0}$
to 0 " $\widehat{x}_{i1}$
and set $\overline{x}_{i1}$ to 0 " $\widehat{x}_{i0}$
to 1 " $\widehat{x}_{i1}$

The idea of randomized rounding is to interpret the fractional solutions provided by the linear program as probabilities for the rounding process.

TRICK - II.

A nice property of randomized rounding is that if the fractional value of a variable is close to 0 (or to 1), then this variable is likely to be set to 0 (or 1).

Theorem: If $0 < \varepsilon < 1$, then with probability $1 - \varepsilon$ the global wiring S produced by randomized rounding satisfies the inequalities:

$$W_S \leq \widehat{W}\left(1 + \Delta^+\left(\widehat{W}, \frac{\varepsilon}{2n}\right)\right) \leq W_0\left(1 + \Delta^+\left(W_0, \frac{\varepsilon}{2n}\right)\right)$$

Proof: We show that following the rounding process, with probability at least $1 - \varepsilon$, no more than $\widehat{W}\left(1 + \Delta^+\left(\widehat{W}, \frac{\varepsilon}{2n}\right)\right)$ wires pass through any boundary.

This will be done by showing, for any boundary b, that the probability that $W_S(b) > \widehat{W}\left(1 + \Delta^+\left(\widehat{W}, \frac{\varepsilon}{2n}\right)\right)$ is at most $\frac{\varepsilon}{2n}$.

Since a $\sqrt{n} \times \sqrt{n}$ array has at most $2n$ boundaries, one has to sum the above probability of failure over all boundaries b to get an upper bound of ε on the failure probability.

TRICK - III.

Let b be a boundary. The solution of the rational linear program satisfy its constrains, therefore we have

$$\sum_{i \in T_{b0}} \widehat{x}_{i0} + \sum_{i \in T_{b1}} \widehat{x}_{i1} \leq \widehat{W}.$$

The number of wires passing through b in the solution S is

$$W_S(b) = \sum_{i \in T_{b0}} \overline{x}_{i0} + \sum_{i \in T_{b1}} \overline{x}_{i1}.$$

$\overline{x}_{i0}$ and $\overline{x}_{i1}$ are Poisson trials with probabilities
$\widehat{x}_{i0}$ and $\widehat{x}_{i1}$

In addition, $\overline{x}_{i0}$ and $\overline{x}_{i1}$ are each independent of $\overline{x}_{j0}$ and $\overline{x}_{j1}$ for $i \neq j$.

Therefore $W_S(b)$ is the sum of independent Poisson trials.

$$E[W_S(b)] = \sum_{i \in T_{bo}} E\left[\overline{x}_{i0}\right] + \sum_{i \in T_{b1}} E\left[\overline{x}_{i1}\right] = \sum_{i \in T_{b0}} \widehat{x}_{i0} + \sum_{i \in T_{b1}} \widehat{x}_{i1} \leq \widehat{W}$$

Since $\Delta^+\left(\widehat{W}, \frac{\varepsilon}{2n}\right)$ is such that

$$Pr\left[W_S(b) > \widehat{W}\left(1 + \Delta^+\left(\widehat{W}, \frac{\varepsilon}{2n}\right)\right)\right] \leq \frac{\varepsilon}{2n}$$

the theorem follows.

HOEFFDING INEQUALITY

The problem with Chernoff bounds is that they work only for 0-1 random variables. **Hoeffding inequality** is another concentration bound based on the moment generating functions that applies to any sum of independent random variables with mean 0.

Theorem Let $X_1 \ldots, X_n$ be independent random variables with $\mathbf{E}[X_i] = 0$ and $|X_i| \leq c_i$ for all i and some constants c_i. Then for all t,

$$\Pr\left[\sum_{i=1}^{n} X_i \geq t\right] \leq e^{-\frac{t^2}{2\sum_{i=1}^{n} c_i^2}}$$

In the case x_i are dependent, but form so called **martingale** Hoeffding inequality can be generalized and we get so called **Azuma-Hoeffding inequality**.

MARTINGALES

MARTINGALES

MARTINGALES

Martingales are very special sequences of random variables that arise at numerous applications, such as at gambling or at random walks.

These sequences have various interesting properties and for them powerful techniques exist to derive special Chernoff-like tail bounds.

Martingales can be very useful in showing that values of a random variable V are sharply concentrated around its expectation $\mathbf{E}[V]$.

Martingales originally referred to systems of betting in which a player increases his stake (usually by doubling) each time he lost a bet.

For analysis of randomized algorithms of large importance is that, as a general rule of thumb says, most things that work for sums of independent random variables work also for martingales.

MARTINGALES - MAIN DEFINITION

Definition: A sequence of random variables $Z_0, Z_1, \ldots$ is a **martingale with respect to a sequence** of rand. variabl., $X_0, X_1, \ldots$, if, for all $n \geq 0$, the following conditions hold:

- Z_n is a function of $X_0, X_1, \ldots, X_n$
- $\mathbf{E}[|Z_n|] < \infty$;
- $\mathbf{E}[Z_{n+1}|X_0, \ldots, X_n] = Z_n$;

A sequence of random variables $Z_0, Z_1, \ldots$ is called **martingale** if it is mrtngl with respect to itself. That is $\mathbf{E}[|Z_n|] < \infty$ and $\mathbf{E}[Z_{n+1}|Z_0, \ldots, Z_n] = Z_n$

EXAMPLE

- Let us have a gambler who plays a sequence of fair games.
- Let X_i be the amount the gambler wins in the ith game.
- Let Z_i be the gambler's total winnings at the end of the ith game.
- Because each game is fair we have $\mathbf{E}[X_i] = 0$
- $\mathbf{E}[Z_{i+1}|X_1, X_2, \ldots, X_i] = Z_i + \mathbf{E}[X_{i+1}] = Z_i$

Thus $Z_1, Z_2, \ldots, Z_n$ is martingale with respect to the sequence $X_1, X_2, \ldots, X_n$.

DOOB MARTINGALES

A **Doob martingale** is a martingale constructed using the following general scheme:

Let $X_0, X_1, \ldots, X_n$ be a sequence of random variables, and let Y be another random variable with $\mathbf{E}[|Y|] < \infty$.
The sequence

$$Z_i = \mathbf{E}[Y \mid X_0, \ldots, X_i], i = 1, \ldots, n$$

for $i = 0, 1, 2, \ldots$ is a martingale with respect to $X_0, X_1, \ldots, X_n$.
Indeed,

$$\begin{aligned} \mathbf{E}[Z_{i+1} \mid X_0, \ldots, X_i] &= \mathbf{E}[\mathbf{E}[Y \mid X_0, \ldots, X_{i+1}] \mid X_0, \ldots, X_i] \\ &= \mathbf{E}[Y \mid X_0, \ldots, X_i] = Z_i \end{aligned}$$

Here we have used the fact that $\mathbf{E}[V \mid W] = \mathbf{E}[\mathbf{E}[V \mid U, W] \mid W]$ for any r.v. U, V, W.

REMAINDER - CONDITIONAL EXPECTATION

Definition: It is natural and useful to define conditional expectation of a random variable Y, conditioned on an event E, by

$$\mathbf{E}[Y|E] = \sum y Pr(Y = y|E).$$

Example: Let we roll independently two perfect dice and let X_i be the number that shows on the ith dice and let X be sum of numbers on both dice.

$$\mathbf{E}[X|X_1 = 3] = \sum_x x Pr(X = x|X_1 = 3) = \sum_{x=4}^{9} x\frac{1}{6} = \frac{13}{2}$$

$$\mathbf{E}[X_1|X = 5] = \sum_{x=1}^{4} x Pr(X_1 = x|X = 5) = \sum_{x=1}^{4} x\frac{Pr(X_1 = x \cap X = 5)}{Pr(X = 5)} = \frac{5}{2}$$

Definition: For two random variables Y and Z, $\mathbf{E}[Y|Z]$ is defined to be a random variable $f(Z)$ that takes on the value $\mathbf{E}[Y|Z = z]$ when $Z = z$.

Theorem For any random variables Y, Z it holds

$$\mathbf{E}[Y] = \mathbf{E}[\mathbf{E}[Y|Z]].$$

A USEFUL FACT

For random variables X, Y it holds

$$\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[X]$$

In words: what you will expect to expect X to be after learning Y is the same as what you now expect X to be.

Proof:

$$\mathbf{E}[X, Y = y] = \sum_x x Pr[X = x, Y = y] = \sum_x x \frac{Pr[x, y]}{Pr_Y[y]}$$

and therefore

$$\mathbf{E}[\mathbf{E}[X|Y = y]] = \sum_y Pr_Y[y] \sum_x x \frac{Pr[x, y]}{Pr_Y[y]} = \sum_x \sum_y x Pr[x, y] = \mathbf{E}[X]$$

STOPPING TIME

A stopping time corresponds to such a strategy to stop a sequence of steps (say at a gambling), that is based only on the outcomes seen so far.

Examples of such rules at which the decision to stop gambling is a stopping time:

1. First time the gambler wins 5 games in total;
2. First time the gambler either wins or looses 1000 dolars;
3. First time the gambler wins 4 times in a row.

The rule "Last time the gambler wins 4 times in a row" is not a stopping time.

MARTINGALE STOPPING THEOREM

Theorem: If $Z_0, Z_1, \ldots,$ is a martingale with respect to $X_1, X_2, \ldots$ and if T is a stopping time for $X_1, X_2, \ldots$, then

$$\mathbf{E}[Z_T] = \mathbf{E}[Z_0]$$

whenever one of the following conditions holds:

- there is a constant c such that, for all i, $|Z_i| \leq c$ - that is all Z_i are bounded;
- T is bounded;
- $\mathbf{E}[T] < \infty$ and there is a constant c such that

$$\mathbf{E}[|Z_{i+1} - Z_i| \mid X_1, \ldots, X_i] < c;$$

EXAMPLE - GAMBLER's PROBLEM

- Consider a sequence of independent fair games, where in each round each player either wins or looses one euro with probability $\frac{1}{2}$.
- Let $Z_0 = 0$, let X_i be the amount won at the ith game and let Z_i be the total amount won after i games.
- Let us assume that the player quits the game when he either looses l_1 euro or wins l_2 euro (for given l_1, l_2).
- What is the probability that the player wins l_2 euro before losing l_1 euro?

GAMBLER's PROBLEM - ANSWER

- Let T be the time when the gambler for the first time either won l_2 or lost l_1 euro. T is stopping time for the sequence $X_1, X_2, \ldots$.
- Sequence $Z_0, Z_1, \ldots$ is martingale. Since values of Z_i are bounded, the martingale stopping theorem can be applied. Therefore, we have:

$$\mathbf{E}[Z_T] = 0$$

- Let now p be probability that the gambler quits playing after winning l_2 euro. Then

$$\underline{E}[Z_T] = l_2 p - l_1(1 - p) = 0$$

and therefore

$$p = \frac{l_1}{l_1 + l_2}$$

ELECTION PROBLEM

- Suppose candidates A and B run for elections and at the end A gets v_A votes and B gets v_B votes and $v_B < v_A$.
- Let us assume that votes are counted at random. What is the probability that the candidate A will be always ahead during the counting process?
- Let $n = v_A + v_B$ and let S_k be the number of votes by which A is leading after k votes were counted. Clearly $S_n = v_A - v_B$.
- For $0 \leq k \leq n-1$ we define

$$X_k = \frac{S_{n-k}}{n-k}$$

- It can be shown, after some calculations, that the sequence $X_0, X_1, \ldots, X_n$ forms a martingale.
- Note that the sequence $X_0, X_1, \ldots, X_n$ relates to the counting process in a backward order - X_0 is a function of S_n,....

ELECTION PROBLEM - RESULT

- Let T be the minimum k such that $X_k = 0$ if such a k exists, and $T = n-1$ otherwise.
- T is a bounded stopping time and therefore the martingale stopping theorem gives

$$\mathbf{E}[X_T] = \mathbf{E}[X_0] = \frac{\mathbf{E}[S_n]}{n} = \frac{v_A - v_B}{v_A + v_B}$$

- **Case 1:** Candidate A leads through the count. In such a case all S_{n-k} and therefore all X_k are positive, $T = n-1$ and $X_T = X_{n-1} = S_1 = 1$.
- **Case 2:** Candidate A does not lead through the count. For some $k < n-1$ $X_k = 0$. If candidate B ever leads it has to be a k where $S_k = X_k = 0$. In this case $T == k < n-1$ and $X_T = 0$..
- We have therefore

$$\mathbf{E}[X_T] = \frac{v_A - v_B}{v_A + v_B} = 1 \cdot \Pr(\text{Case 1}) + 0 \cdot \Pr(\text{Case 2})$$

- Therefore the probability of Case 1, in which candidate A leads through the account, is

$$\frac{v_A - v_B}{v_A + v_B}$$

AZUMA-HOEFFDING INEQUALITY

Perhaps the main importance of the martingale concept for the analysis of randomized algorithms is due to various special Chernoff-type inequalities that can be applied even in case random variables are not independent.

Theorem Let $X_0, X_1, \ldots, X_n$ be a martingale such that for any k

$$|X_k - X_{k-1}| \leq c_k.$$

for some c_k.

Then, for all $t \geq 0$ and any $\lambda > 0$

$$Pr(|X_t - X_0| \geq \lambda) \leq 2e^{-\lambda^2/(2\sum_{i=1}^{t} c_i^2)}$$

EXAMPLE - PATTERN MATCHING - I.

- Let $S = (s_1, \ldots, s_n)$ be a string of n characters chosen randomly from an s-nary alphabet Σ. Let $P = (p_1, \ldots, p_k)$ be a string (pattern) of k characters from Σ.
- Let $F_{P,S}$ be the number of occurrences of P in S. Clearly

$$\mathbf{E}[F_{P,S}] = (n-k+1)\left(\frac{1}{s}\right)^k$$

- We use now a Doob martingale and Azuma-Hoeffding inequality to show that, if k is relatively small with respect to n, then the number of occurrences of the pattern P in S is highly concentrated around its mean.
- Let $Z_0 = \mathbf{E}[F_{P,S}]$ and, for $1 \leq i \leq n$ let

$$Z_i = \mathbf{E}[F_{P,S} \mid s_1, \ldots, s_i].$$

- The sequence $Z_0, \ldots, Z_n$ is a Doob martingale, and $Z_n = F_{P,S}$.

EXAMPLE - PATTERN MATCHING - II.

- Since each character in the pattern P can participate in no more than k possible matches, for any $0 \leq i \leq n$ we have

$$|Z_{i+1} - Z_i| \leq k.$$

In other word, the value of s_{i+1} can affect the value of F by at most k. Hence

$$|\mathbf{E}[F_{P,S} \mid s_1, \ldots, s_{i+1}] - \mathbf{E}[F_{P,S} \mid s_1, \ldots, s_i]| = |Z_{i+1} - Z_i| \leq k.$$

- By Azuma-Hoeffding inequality/theorem,

$$Pr(|F_{P,S} - \mathbf{E}[F_{P,S}]| \geq \varepsilon) = Pr(|(Z_n - Z_0)| \geq \varepsilon) \leq 2e^{-\varepsilon^2/2nk^2}.$$

WAITING TIMES for PATTERNS PROBLEM

Problem: Let us suppose that we flip coins until we see some pattern to appear. What is the expected number of coin-flips until this happens?

Example: We flip coins until we see HTHH.

Suppose that $x_1 x_2 \ldots x_n$ is the pattern we want to get.

Let us imagine we have an army of gamblers, and let one new shows up before each new coin flip.

Let each gambler start by borrowing 1$ and betting that the next coin-flip will be x_1. If she wins, she takes her 2$ and bets 2$ that next coin-flip will be x_2, continuing to play double-or-nothing until either she loses (and is out of her initial 1$) or wins her last bet on x_k (and is up $2^k - 1$ dollars).

Because each gambler's winnings form a martingale, so does their sum, and so the expected total return of all gamblers up to the **stopping time** τ at which our pattern occurs for the first time is 0.

The above facts will now be used to compute $\mathbf{E}[\tau]$.

When we stop at time τ we have one gambler who has won $2^k - 1$. Other gamblers may still play.

For each i with $x_1 \ldots x_k = x_{k-i+1} \ldots x_k$ there will be a gambler with net winnings $2^i - 1$. All remaining gamblers will all be at -1.

Let $\chi_i = 1$ if $x_1 \ldots x_i = x_{k-i+1} \ldots x_k$, and 0 otherwise. Then, using the stopping time theorem,

$$\mathbf{E}[X_\tau] = \mathbf{E}\left[-\tau + \sum_{i=1}^{k} \chi_i 2^i\right] = -\mathbf{E}[\tau] + \sum_{i=1}^{k} \chi_i 2^i = 0$$

and therefore

$$\mathbf{E}[\tau] = \sum_{i=1}^{k} \chi_i 2^i.$$

Examples: if pattern is HTHH (HHHH) [THHH], then $\mathbf{E}[\tau]$ equals 18 (30) [16].

EXAMPLE - OCCUPANCY PROBLEM

Suppose that m balls are thrown randomly into n bins and let Z denote the number of bins that remain empty at the end.

For $0 \leq t \leq m$ let Z_t be the expectation at time t of the number of bins that are empty at time m.
The sequence of random variables

$$Z_0, Z_1, \ldots, Z_m$$

is a martingale, $Z_0 = \mathbf{E}[Z]$ and $Z_m = Z$.

SOME ESTIMATIONS

Kolmogorov-Doob inequality Let $X_0, X_1, \ldots$ be a martingale. Then for any $\lambda > 0$

$$\Pr[\max_{0 \le i \le n} X_i \ge \lambda] \le \frac{\mathbf{E}[|X_n|]}{\lambda}.$$

Azuma inequality Let $X_0, X_1, \ldots$ be a martingale sequence such that for each k

$$|X_k - X_{k-1}| \le c_k,$$

then for all $t \ge 0$ and any $\lambda > 0$

$$\Pr[|X_t - X_0| \ge \lambda] \le 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^{t} c_k^2}\right).$$

Corollary Let $X_0, X_1, \ldots$ be a martingale sequence such that for each k

$$|X_k - X_{k-1}| \le c$$

where c is independent of k. Then, for all $t \ge 0$ and any $\lambda > 0$

$$\Pr[|X_t - X_0| \ge \lambda c\sqrt{t}] \le 2e^{-\lambda^2/2},$$

OCCUPANCY PROBLEM REVISITED

Let us have m balls thrown randomly into n bins and let Z denote the number of bins that remain empty.

Azuma inequality allows to show:

$$\mu = \mathbf{E}[Z] = n(1 - \frac{1}{n})^m \approx ne^{-m/n}$$

and for $\lambda > 0$

$$\Pr[|Z - \mu| \ge \lambda] \le 2e^{-\frac{\lambda^2(n-1/2)}{n^2-\mu^2}}.$$

APPENDIX

APPENDIX

EXERCISES

1. What is larger, e^π or π^e, for the basis e of natural logarithms
2. Hint 1: There exists one-line proof of the correct relation.

EXERCISES

1. What is larger, e^{π} or π^{e}, for the basis e of natural logarithms
2. Hint 1: There exists one-line proof of correct relation.
3. Hint 2: Solution: use inequality $e^{x} > 1 + x$ with $x = \pi/e - 1$.

EXERCISES

1. What is larger, e^{π} or π^{e}, for the basis e of natural logarithms
2. Hint 1: There exists one-line proof of correct relation.
3. Hint 2: Use the inequality $e^{x} > 1 + x$ with $x = \pi/e - 1$.
4. Solution:

$$e^{\pi/e-1} > 1 + \pi/e - 1$$

implies:

$$e^{\pi/e-1} > \pi/e ==> e^{\pi/e} > \pi ==> e^{\pi} > \pi^{e}$$