Tutorial 7

31 March 2022 12:01

Concentration bounds

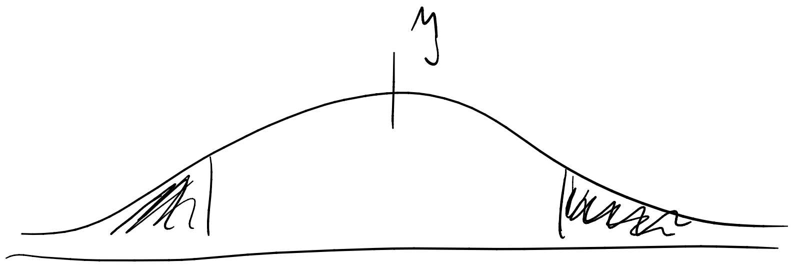
→ Chernoff bound applications

→ Polling problem

→ Algorithm to estimate π

→ Intro to routing on hypergraphs

Chernoff bounds



upper bound that r.v. X is far from it's mean μ.

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

X_i - Poisson r.v. (take values 0 and 1) with $P_r(X_i = 1) = p$

$E(X_i) = p$

$E(X) = n \cdot p = \mu$

X_i are i.i.d.

$$\boxed{\begin{aligned} &\Pr\left(X \le (1-\delta)\mu\right) \le e^{-\frac{\mu\delta^2}{2}} \\ &\Pr\left(X \ge (1+\delta)\mu\right) \le e^{-\frac{\mu\delta^2}{3}} \\ &\Pr\left(|X-\mu| \ge \delta\mu\right) \le 2\cdot e^{-\frac{\mu\delta^2}{3}} \end{aligned}} \qquad 0 \le \delta \le 1$$

Polling problem

Two presidential candidates A and B.

Estimate the number of people who will vote for A.

Let's assume everyone is decided and will vote.

The answer of i^{th} person is labeled x_i

$x_i = 1$ (vote for A)

$x_i = 0$ (vote for B)

$\Pr(X_i = 1) = p$ {this is a percentage of people voting for A}

= the number we want to estimate

After asking n people the estimate is $\boxed{X = \frac{x_1 + \dots x_n}{n}}$

$P\left(|X - p| < (0.1)\cdot p\right) > 90\%$

$$\Pr\left(|X-p| \le (0.1)\cdot p\right) > 90\%$$

$$\Pr\left(|X-p| \ge \frac{p}{10}\right) \le 10\%$$

$E(x_1) = p$

$$nX = x_1 + \dots + x_n$$

$$E(nX) = n \cdot p$$

$$\Pr\left(|nX - \underbrace{np}_{=\mu}| \ge \frac{np}{10}\right) \le 0.1 \le 2 \cdot e^{-\frac{n \cdot p}{100 \cdot 3}}$$

$\delta = 1/10$

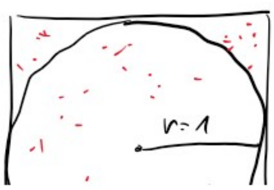
$$2 \cdot e^{-\frac{n \cdot p}{100 \cdot 3}} \le \frac{1}{10}$$

$$e^{-\frac{n \cdot p}{300}} \le \frac{1}{20} \qquad / \ln$$

$$-\frac{n \cdot p}{300} \le \ln\left(\frac{1}{20}\right)$$

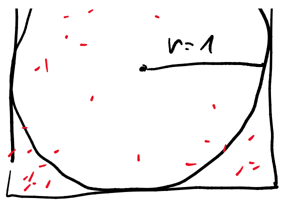
$$n \ge \frac{900}{p}$$

Algorithm to estimate π $\approx 3.1415\dots$



$z_i = 1$ if i^{th} point is inside the circle

$z_i = 0$ if i^{th} point is outside the circle



$z_i = 0$ if i^{th} point is outside the circle

$$Pr(z_i = 1) = \frac{A(\text{circle})}{A(\text{square})} = \frac{\pi r^2}{4} = \frac{\pi}{4}$$

After n samples $z = \sum_{i=1}^{n} z_i$

$$E(z) = \frac{n \cdot \pi}{4}$$

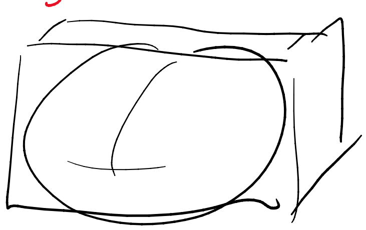
$$z' = \frac{4 \cdot z}{n} \qquad E(z') = \pi$$

$$Pr\left(|z' - \pi| \geq \varepsilon \cdot \pi \right)$$

$$Pr\left(\left| \frac{n}{4} \cdot z' - \frac{n}{4}\pi \right| \geq \frac{n}{4} \varepsilon \cdot \pi \right) \leq 2 \cdot e^{-\frac{\frac{n}{4}\pi \varepsilon^2}{3}} = 2e^{-\frac{n \pi \varepsilon^2}{12}}$$

$$|z - E(z)| \geq \varepsilon E(z)$$

Can we get better with a cube?



$$Pr(z_i = 1) = \frac{V(\text{sphere})}{V(\text{cube})} = \frac{4/3\, \pi r^3}{8} = \frac{\pi}{6}$$

$$\ldots \leq 2e^{-\frac{\frac{n}{6} \cdot \pi \cdot \varepsilon^2}{3}} = 2e^{-\frac{n \cdot \pi \cdot \varepsilon^2}{18}}$$

$\cancel{2\cdot\left(e^{-\frac{\pi\cdot\varepsilon^2}{18}}\right)^n}$

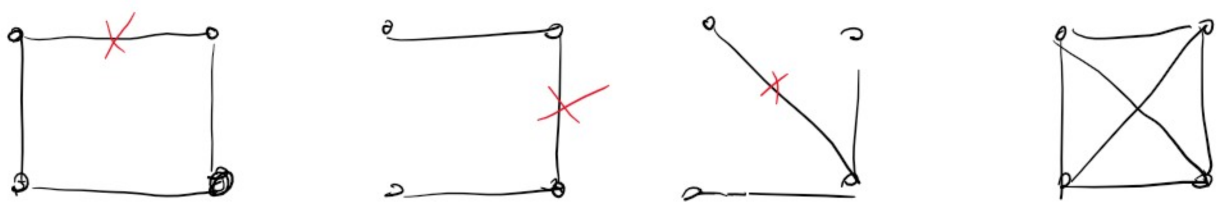
$\cancel{a = e^{-\pi\cdot\varepsilon^2} < 1}$

$\cancel{a^{\frac{1}{18}} > a^{\frac{1}{12}}}$

$\Longrightarrow$ calculate properly

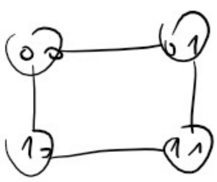
ε are not equivalent for 2D and 3D experiment

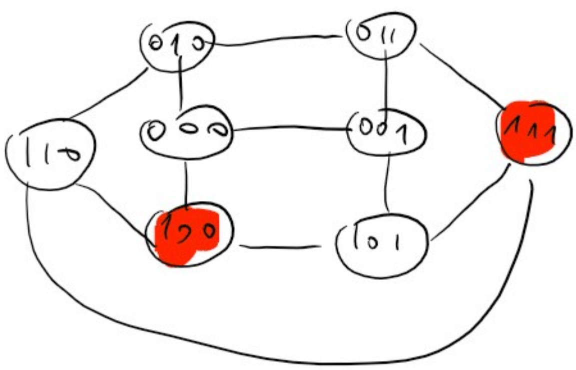
Routing on hypercubes (Introduction)



$N = 2^d$ nodes labeled by binary strings of length d

Two nodes are connected iff they differ in exactly 1 bit





Total number of edges is $\frac{N\cdot d}{2}$ $\quad O(N\cdot\log N)$

Packet $(i, X, d(i))$

i: source, $d(i)$: destination

Assumption: In each time step only 1 packet can be sent through a link.

Deterministic routing: Left to right bit-fixing

$$\begin{tikzcd}
 & \overset{LTR}{0100} \arrow[dr] & \\
1100 \arrow[ur] \arrow[dr] & & 0101 \\
 & \underset{RTL}{1101} \arrow[ur] &
\end{tikzcd}$$

($k!$ different routes of length k, where k is the Hamming distance)

Experiment to test throughput of routing algorithms

Each node gets a packet with a destination.

How long will it take to deliver all the packets?

$\boxed{\text{The worst case}}$

$x_1, \dots, x_d \longrightarrow x_d, \dots, x_1$ $\Rightarrow$ large subset of packets need to pass through $0000\cdots0$

$i = 1100 \to d(i) = 0011$

$j = 0100 \to d(j) = 0010$

$k = 1000 \to d(k) = 0001$

L→R

P_i: $1100 \to 0100 \to \boxed{0000} \to 0010 \to 0011$

P_j: $0100 \to 0000 \to 0010$

P_k: $1000 \to 0000 \to 0001$

$$\underbrace{x_1 \dots x_{d/2}}_{d/2}\ \underbrace{000 \dots}_{d/2} \to \underbrace{000 \dots 0}_{d/2}\ x_{d/2} \dots x_1$$

$101000 \to 001000 \to 000000 \to 000100$

$011000 \to 001000 \to 000000 \to 000100$

delay

$\Rightarrow$ $2^{d/2}$ strings pass through all 0 node

Node $\vec{0}$ can send at most d packets simultaneously

$= \left(2^d\right)$

$\Omega\left(\frac{2^d}{d}\right)$ time steps are needed.

$\sigma(i)$ → intermediate destination chosen at random

$i \xrightarrow{LTR} \sigma(i) \xrightarrow{LTR} d(i)$

Packet $(i, \sigma(i), d(i), X)$

W.P. at least $1-2^{-5d}$ every packet gets delivered in time $14d$.

$\boxed{\rho_i} = (\ell_{i1}, \ell_{i2}, \dots, \ell_{ik})$

$\rho_j = (\ell_{j1}, \dots, \ell_{jj'})$