Tail Inequalities

Mohsen Alambardar Meibodi

Department of Computer Science
Yazd University

November 17, 2014

Outline

Outline

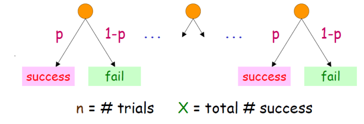
Bernoulli Distribution

Definition

independent Bernoulli tails: random variables $X_1, X_2, \dots, X_n$ such that

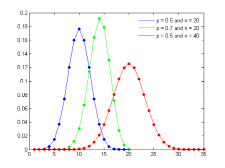
$$Pr[X_i = 1] = p$$

$$Pr[X_i = 0] = 1 - p$$



Binomial distribution: sum of i.i.d. Bernoulli trials, i.e. $\mathbb{X} = \sum_{i=1}^{n} X_i$

$$Pr[\mathbb{X} = k] = \binom{n}{k} . p^k . (1-p)^{n-k}$$



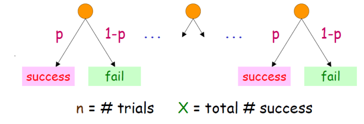
Bernoulli Distribution

Definition

independent Bernoulli tails: random variables $X_1, X_2, \ldots, X_n$ such that

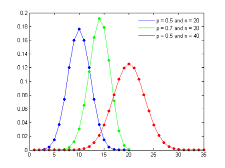
$$Pr[X_i = 1] = p$$

$$Pr[X_i = 0] = 1 - p$$



Binomial distribution: sum of i.i.d. Bernoulli trials, i.e. $\mathbb{X} = \sum_{i=1}^{n} X_i$

$$Pr[\mathbb{X} = k] = \binom{n}{k}.p^k.(1-p)^{n-k}$$

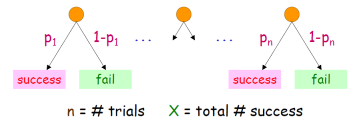


Poisson trial

Definition

Poisson trial: random variables $X_1, X_2, \ldots, X_n$ such that

$$Pr[X_i = 1] = p_i$$
$$Pr[X_i = 0] = 1 - p_i$$



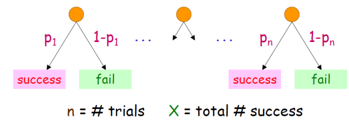
Sum of Poisson trial: $\mathbb{X} = X_1 + X_2 + \cdots + X_n$ such that $X_1, X_2, \ldots, X_n$ be a sequence of independent Poisson trials, then $E[\mathbb{X}] = E[X_1 + X_2 + \cdots + X_n] = p_1 + p_2 + \cdots + p_n$

Poisson trial

Definition

Poisson trial: random variables $X_1, X_2, \ldots, X_n$ such that

$$Pr[X_i = 1] = p_i$$
$$Pr[X_i = 0] = 1 - p_i$$



Sum of Poisson trial: $\mathbb{X} = X_1 + X_2 + \cdots + X_n$ such that $X_1, X_2, \ldots, X_n$ be a sequence of independent Poisson trials, then

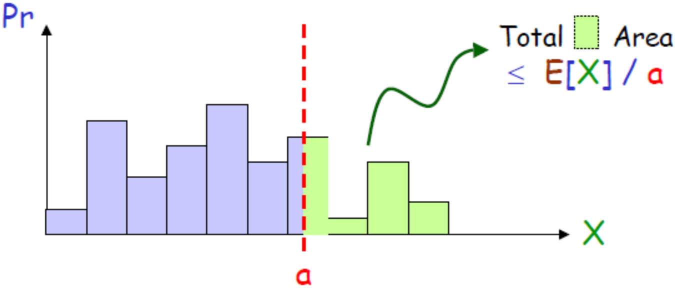
$$E[\mathbb{X}] = E[X_1 + X_2 + \cdots + X_n] = p_1 + p_2 + \cdots + p_n$$

Markov Inequality

Theorem

Let X be a random variable that takes on non-negative values only. Then, for any positive number a,

$$Pr(X \geq a) \leq \frac{E[X]}{a}$$

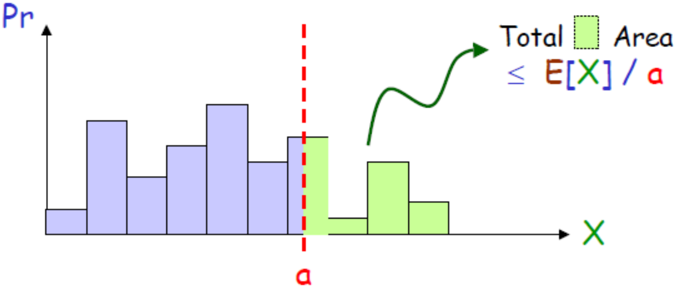


Markov Inequality

Theorem

Let X be a random variable that takes on non-negative values only. Then, for any positive number a,

$$Pr(X \geq a) \leq \frac{E[X]}{a}$$

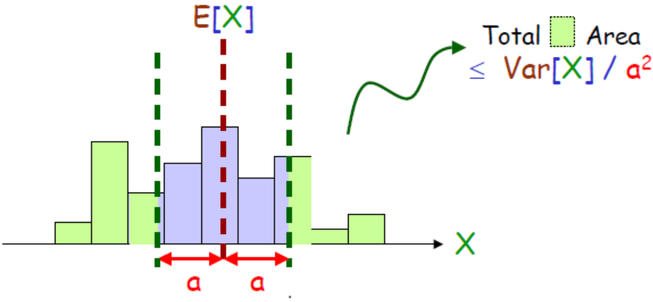


Chebyshev Inequality

Theorem

Let X be a random variable that takes on non-negative values only. Then, for any positive number a,

$$Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}$$



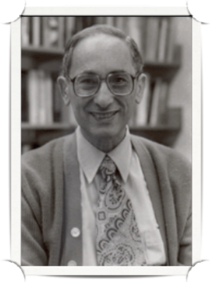
Definition

Question1: For a real number δ, what is the probability that $\mathbb{X}$ exceeds $(1+\delta)\mu$?

Question2: How large must δ be in order that the tail probability is less than a prescribed value ϵ?

Tight answer can be obtained using technique known as the

Chernoff Bound

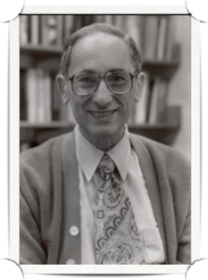


Definition

- Question1: For a real number δ, what is the probability that $\mathbb{X}$ exceeds $(1+\delta)\mu$?
- Question2: How large must δ be in order that the tail probability is less than a prescribed value ϵ?

Tight answer can be obtained using technique known as the

Chernoff Bound

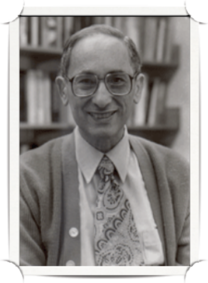


Definition

Question1: For a real number δ, what is the probability that $\mathbb{X}$ exceeds $(1+\delta)\mu$?

Question2: How large must δ be in order that the tail probability is less than a prescribed value ϵ?

Tight answer can be obtained using technique known as the

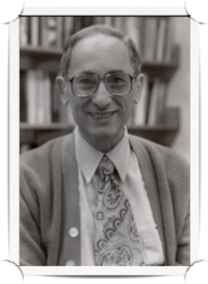


Definition

Question1: For a real number δ, what is the probability that $\mathbb{X}$ exceeds $(1+\delta)\mu$?

Question2: How large must δ be in order that the tail probability is less than a prescribed value ϵ?

Tight answer can be obtained using technique known as the



Outline

Chernoff Bound

Moment Generating Function

The moment generating function of X is

$$M(t) = E(e^{tX})$$

Taylor expanation:

$$E(e^{tX}) = E(\sum_{k=0}^{\infty} \frac{t^k}{k!} X^k) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k)$$

Theorem

Let $X_1, X_2, \dots, X_n$ be independent Poisson trials such that $Pr[X_i = 1] = p_i$. Let $\mathbb{X} = X_1 + X_2 + \cdots + X_n$ and $\mu = E[\mathbb{X}]$. Then, for all $\delta > 0$,

$$Pr[\mathbb{X} > (1+\delta)\mu] \leq [\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu}$$

in other words

$$Pr[\mathbb{X} - \mu > \delta\mu] \leq [\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu}$$

$$Pr[\mathbb{X} > (1+\delta)\mu] \leq [\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu}$$

Proof strategy:

- Markovs inequality + moment generating function.
- Bound the moment generating function.
- Parameterization + optimization.

Proof.

For any positive real number t,

$$Pr[\mathbb{X} > (1+\delta)\mu] = Pr[e^{t\mathbb{X}} > e^{t(1+\delta)\mu}]$$
$$\leq \frac{E(e^{t\mathbb{X}})}{e^{t(1+\delta)\mu}} \text{ (Markov inequality)}$$

$$\mathrm{E}[\mathrm{e}^{t\mathbb{X}}] = E[e^{(t\sum_{i=1}^{n} X_i)}] = E(\prod_{i=1}^{n} e^{tX_i}) = \prod_{i=1}^{n} E(e^{tX_i})$$

$$Pr[\mathbb{X} > (1+\delta)\mu] \leq [\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu}$$

Proof (Contd.).

$$e^{tX_i} = \begin{cases} e^t & X_i = 1 \text{ (prob } p_i). \\ 1 & X_i = 0 \text{ (prob } 1-p_i). \end{cases} \Rightarrow E[e^{tX_i}] = p_i e^t + 1 - p_i$$

$$\begin{aligned} Pr[\mathbb{X} > (1+\delta)\mu] &= Pr[e^{t\mathbb{X}} > e^{t(1+\delta)\mu}] \\ &\leq \frac{\prod_{i=1}^{n} E[e^{tX_i}]}{e^{t(1+\delta)\mu}} \\ &\leq \frac{\prod_{i=1}^{n} [p_i e^t + 1 - p_i]}{e^{t(1+\delta)\mu}} \\ &= \frac{\prod_{i=1}^{n} [1 + p_i(e^t - 1)]}{e^{t(1+\delta)\mu}} \end{aligned}$$

$$Pr[\mathbb{X} > (1+\delta)\mu] \leq [\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu}$$

Proof (Contd.).

We use the fact $1 + x \leq e^x$ with $x = p_i(e^t - 1)$

$$\begin{aligned} Pr[\mathbb{X} > (1+\delta)\mu] &\leq \frac{\prod_{i=1}^{n} e^{p_i(e^t-1)}}{e^{t(1+\delta)\mu}} \\ &= \frac{e^{\sum_{i=1}^{n} p_i(e_t-1)}}{e^{t(1+\delta)\mu}} \\ &= \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} \end{aligned}$$

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

$$[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu}$$

Definition

$$F^+(\mu, \delta) = [\frac{e^\delta}{(1+\delta)^{(1+\delta)}}]^\mu$$

Example

- Player win each game with probability $\frac{1}{3}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is greater than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^n X_i$
- $Pr(Y_n > \frac{n}{2}) < F^+(\frac{n}{3}, \frac{1}{2}) < (0.965)^n$

Definition

$$F^+(\mu,\delta) = [\frac{e^\delta}{(1+\delta)^{(1+\delta)}}]^\mu$$

Example

- Player win each game with probability $\frac{1}{3}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is greater than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^n X_i$
- $Pr(Y_n > \frac{n}{2}) < F^+(\frac{n}{3}, \frac{1}{2}) < (0.965)^n$

Definition

$$F^+(\mu, \delta) = [\frac{e^\delta}{(1+\delta)^{(1+\delta)}}]^\mu$$

Example

- Player win each game with probability $\frac{1}{3}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is greater than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^n X_i$
- $Pr(Y_n > \frac{n}{2}) < F^+(\frac{n}{3}, \frac{1}{2}) < (0.965)^n$

Definition

$$F^+(\mu, \delta) = [\frac{e^\delta}{(1+\delta)^{(1+\delta)}}]^\mu$$

Example

- Player win each game with probability $\frac{1}{3}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is greater than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n > \frac{n}{2}) < F^+(\frac{n}{3}, \frac{1}{2}) < (0.965)^n$

Definition

$$F^{+}(\mu, \delta) = [\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}]^{\mu}$$

Example

- Player win each game with probability $\frac{1}{3}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is greater than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n > \frac{n}{2}) < F^{+}(\frac{n}{3}, \frac{1}{2}) < (0.965)^n$

Definition

$$F^+(\mu, \delta) = [\frac{e^\delta}{(1+\delta)^{(1+\delta)}}]^\mu$$

Example

- Player win each game with probability $\frac{1}{3}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is greater than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n > \frac{n}{2}) < F^+(\frac{n}{3}, \frac{1}{2}) < (0.965)^n$

Chernoff Bound

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent Poisson trials such that $Pr(X_i = 1) = p_i$. Let $\mathbb{X} = X_1 + X_2 + \cdots + X_n$ and $\mu = E[\mathbb{X}]$. Then, for all $\delta > 0$,

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

In other words,

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

Definition

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

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

Proof.

$$\begin{aligned} Pr[\mathbb{X} \leq (1-\delta)\mu] &= Pr[-\mathbb{X} > -(1-\delta)\mu] \\ &= Pr[e^{(-t\mathbb{X})} \geq e^{(-t(1-\delta)\mu)}] \\ &\leq \frac{\prod_{i=1}^{n} E[e^{(-t\mathbb{X})}]}{e^{(-t(1-\delta)\mu)}} \\ &\leq \frac{e^{(\mu(e^{-t}-1))}}{e^{(-t(1-\delta)\mu)}} \end{aligned}$$

Let $t = \ln(\frac{1}{1-\delta}) \Rightarrow Pr[\mathbb{X} \leq (1-\delta)\mu] \leq [\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}]^{\mu}$.
We simplify this using the McLaurin expanasion for $\ln(1-\delta)$ for $\delta \in (0,1]$

$$(1-\delta)^{(1-\delta)} > e^{(-\delta+\frac{\delta^2}{2})}$$

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

Proof.

$$\begin{aligned} Pr[\mathbb{X} \leq (1-\delta)\mu] &= Pr[-\mathbb{X} > -(1-\delta)\mu] \\ &= Pr[e^{(-t\mathbb{X})} \geq e^{(-t(1-\delta)\mu)}] \\ &\leq \frac{\prod_{i=1}^{n} E[e^{(-t\mathbb{X})}]}{e^{(-t(1-\delta)\mu)}} \\ &\leq \frac{e^{(\mu(e^{-t}-1))}}{e^{(-t(1-\delta)\mu)}} \end{aligned}$$

Let $t = \ln(\frac{1}{1-\delta}) \Rightarrow Pr[\mathbb{X} \leq (1-\delta)\mu] \leq [\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}]^{\mu}$.

We simplify this using the McLaurin expanasion for $\ln(1-\delta)$ for $\delta \in (0,1]$

$$(1-\delta)^{(1-\delta)} > e^{(-\delta+\frac{\delta^2}{2})}$$

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

Proof.

$$
\begin{aligned}
Pr[\mathbb{X} \leq (1-\delta)\mu] &= Pr[-\mathbb{X} > -(1-\delta)\mu] \\
&= Pr[e^{(-t\mathbb{X})} \geq e^{(-t(1-\delta)\mu)}] \\
&\leq \frac{\prod_{i=1}^{n} E[e^{(-t\mathbb{X})}]}{e^{(-t(1-\delta)\mu)}} \\
&\leq \frac{e^{(\mu(e^{-t}-1))}}{e^{(-t(1-\delta)\mu)}}
\end{aligned}
$$

Let $t = \ln(\frac{1}{1-\delta}) \Rightarrow Pr[\mathbb{X} \leq (1-\delta)\mu] \leq [\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}]^{\mu}$.
We simplify this using the McLaurin expanasion for $\ln(1-\delta)$ for $\delta \in (0,1]$

$$(1-\delta)^{(1-\delta)} > e^{(-\delta+\frac{\delta^2}{2})}$$

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

Proof.

$$\begin{aligned} Pr[\mathbb{X} \le (1-\delta)\mu] &= Pr[-\mathbb{X} > -(1-\delta)\mu] \\ &= Pr[e^{(-t\mathbb{X})} \ge e^{(-t(1-\delta)\mu)}] \\ &\le \frac{\prod_{i=1}^{n} E[e^{(-t\mathbb{X})}]}{e^{(-t(1-\delta)\mu)}} \\ &\le \frac{e^{(\mu(e^{-t}-1))}}{e^{(-t(1-\delta)\mu)}} \end{aligned}$$

Let $t = \ln(\frac{1}{1-\delta}) \Rightarrow Pr[\mathbb{X} \le (1-\delta)\mu] \le [\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}]^{\mu}$.
We simplify this using the McLaurin expanasion for $\ln(1-\delta)$ for $\delta \in (0,1]$

$$(1-\delta)^{(1-\delta)} > e^{(-\delta+\frac{\delta^2}{2})}$$

Example

- Player win each game with probability $\frac{3}{4}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is less than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n < \frac{n}{2}) < F^{-}(\frac{3}{4}n, \frac{1}{3}) < (0.9592)^n$

Example

- Player win each game with probability $\frac{3}{4}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is less than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n < \frac{n}{2}) < F^{-}(\frac{3}{4}n, \frac{1}{3}) < (0.9592)^n$

Example

- Player win each game with probability $\frac{3}{4}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is less than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n < \frac{n}{2}) < F^{-}(\frac{3}{4}n, \frac{1}{3}) < (0.9592)^n$

Example

- Player win each game with probability $\frac{3}{4}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is less than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n < \frac{n}{2}) < F^{-}(\frac{3}{4}n, \frac{1}{3}) < (0.9592)^n$

Example

- Player win each game with probability $\frac{3}{4}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is less than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n < \frac{n}{2}) < F^{-}(\frac{3}{4}n, \frac{1}{3}) < (0.9592)^n$

Example

- Player win each game with probability $\frac{3}{4}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is less than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n < \frac{n}{2}) < F^{-}(\frac{3}{4}n, \frac{1}{3}) < (0.9592)^n$

Example

- Player win each game with probability $\frac{3}{4}$.
- Outcomes of each games are independent.
- Find an upper bound for the probability that the number of games he wins is less than $n/2$.
- X_i is 1 if he wins the ith game and 0 otherwise.
- $Y_n = \sum_{i=1}^{n} X_i$
- $Pr(Y_n < \frac{n}{2}) < F^{-}(\frac{3}{4}n, \frac{1}{3}) < (0.9592)^n$

Second Question

How large must δ be in order that the tail probability is less than a prescribed value ϵ ?

Definition

For any positive μ and ϵ, $\Delta^{+}(\mu, \epsilon)$ is the value of δ that satisfies

$$F^{+}(\mu, \Delta^{+}(\mu, \epsilon)) = \epsilon$$

Similarly, $\Delta^{-}(\mu, \epsilon)$ is the value of δ that satisfies

$$F^{-}(\mu, \Delta^{-}(\mu, \epsilon)) = \epsilon$$

Deriving $\Delta^{-}(\mu, \epsilon)$ is easy:

$$\Delta^{-}(\mu, \epsilon) = \sqrt{\frac{2\ln(\frac{1}{\epsilon})}{\mu}}$$

Second Question

How large must δ be in order that the tail probability is less than a prescribed value ϵ ?

Definition

For any positive μ and ϵ, $\Delta^+(\mu, \epsilon)$ is the value of δ that satisfies

$$F^+(\mu, \Delta^+(\mu, \epsilon)) = \epsilon$$

Similarly, $\Delta^-(\mu, \epsilon)$ is the value of δ that satisfies

$$F^-(\mu, \Delta^-(\mu, \epsilon)) = \epsilon$$

Deriving $\Delta^-(\mu, \epsilon)$ is easy:

$$\Delta^-(\mu, \epsilon) = \sqrt{\frac{2\ln(\frac{1}{\epsilon})}{\mu}}$$

Example

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

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

What if we wanted that $\Pr[X < (1-\delta)\mu]$ be less than $e^{-1.5n}$?

$$\Delta^{-}(0.75n, e^{-1.5n}) = \sqrt{\frac{3n}{0.75n}} = 2$$

Exercise

Prove that

$$F^+(\mu, \delta) \leq [\frac{e}{(1+\delta)}]^{(1+\delta)\mu}$$

Hence infer that if $\delta > 2e - 1$

$$F^+(\mu, \delta) \leq 2^{-(1+\delta)\mu}$$

Theorem

Let X be defined as in Chernoff theorem. Then

$$P[\mathbb{X} \geq (1+\delta)\mu] \leq \begin{cases} e^{-\mu\delta^2/4} & 0 < \delta \leq 1 \\ e^{\frac{-\mu\delta \ln n}{2}} & \delta > 1. \end{cases}$$

Proof.

For the interval $0 < \delta \leq 1$, we prove the weaker inequality

$$(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}})^{\mu} \leq e^{\mu\delta^2/4}$$

□

Outline

Set Balancing

Definition

- Suppose we have a group of m students

Set Balancing

Definition

- Suppose we have a group of m students
- We try to classify them by checking whether they have a particular feature or not

Set Balancing

Definition

- Suppose we have a group of m students
- We try to classify them by checking whether they have a particular feature or not
- Let n be the number of features

Set Balancing

Definition

- Suppose we have a group of m students
- We try to classify them by checking whether they have a particular feature or not
- Let n be the number of features
- Can you try to divide the m students into two groups $G1$ and $G2$, such that for each k,

$$\text{no. of students with feature } k \text{ in } G1 = \text{no. of students with feature } k \text{ in } G2.$$

	subject 1	subject 2	subject 3	subject 4
gender	boy	girl	boy	girl
age	≥ 3	≥ 3	<3	<3
teeth	un-healthy	un-healthy	healthy	un-healthy
bodyfat	normal	over-weight	normal	normal
allergy	no	no	yes	yes

Set Balancing

It is desirable to find a partition such that minimizes $\max_k\{$difference in no. of students with feature k in $G1$ and $G2.\}$.

	subject 1	subject 2	subject 3	subject 4
gender	boy	girl	boy	girl
age	≥3	≥3	<3	<3
teeth	un-healthy	un-healthy	healthy	un-healthy
bodyfat	normal	over-weight	normal	normal
allergy	no	no	yes	yes

	$subject_1$	$subject_2$	$subject_3$	$subject_4$
$feature_1$	1	0	1	0
$feature_2$	1	1	0	0
$feature_3$	0	0	1	0
$feature_4$	1	0	1	1
$feature_5$	0	1	0	1

Set Balancing

0-1 matrix:

$$
\begin{array}{r c c c}
 & \begin{array}{cccc} \text{sub } 1 & \text{sub } 2 & \cdots & \text{sub } m \end{array} & \begin{array}{c} \{+1,-1\} \\ \text{vector} \end{array} & \\
\begin{array}{r} \text{feature 1:} \\ \text{feature 2:} \\ \vdots \\ \text{feature n:} \end{array} &
\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} &
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = &
\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}
\end{array}
$$

Q: Given an $n \times n$ matrix A all of whose entries are 0 or 1, find a column vector $b \in \{-1, 1\}^n$ minimizing $\|Ab\|_\infty$.

Set Balancing

0-1 matrix:

$$
\begin{array}{r c c}
 & \begin{array}{cccc} \text{sub } 1 & \text{sub } 2 & \cdots & \text{sub } m \end{array} & \begin{array}{c} \{+1,-1\} \\ \text{vector} \end{array} \\
\begin{array}{r} \text{feature 1:} \\ \text{feature 2:} \\ \vdots \\ \text{feature n:} \end{array} & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}
\end{array}
$$

Q: Given an $n \times n$ matrix A all of whose entries are 0 or 1, find a column vector $b \in \{-1, 1\}^n$ minimizing $\|Ab\|_\infty$.

Theorem

Suppose that $X_i, 1 \le i \le n$, be n independent and identically distributed $\{-1, +1\}$ valued random variables such that $Pr[X_i = +1] = P[X_i = -1] = 1/2$. Let the random variable X be defined by $X = \sum_{i=1}^{n} X_i$. Then for any $\delta > 0$

$$Pr[X \ge \delta] = Pr[X \le -\delta] \le e^{-\delta^2/2n}$$

Olivious Routing

Consider the ith row of A. By previous theorem

$$Pr[(Ab)_i \geq \sqrt{4n\ln n}] \leq n^{-2}$$

and

$$Pr[(Ab)_i \leq -\sqrt{4n\ln n}] \leq n^{-2}$$

So

$$Pr[|(Ab)_i| \geq \sqrt{4n\ln n}] \leq 2n^{-2}$$

There are n rows, so

$$Pr[\|(Ab)\|_\infty \geq \sqrt{4n\ln n}] \leq 2/n$$

In other words,

With probability at least $1-2/n$, we find a vector b for witch $\|Ab\|_\infty \leq \sqrt{4n\ln n}$.

Outline

Routing


Sister Cities
Eilat
Jakarta
St. Pete
Giza
Athens
Berlin 5759 mi
Lusaka
Bordeaux
Split
7406 mi Makati

Network of Parallel Processors

Definition

1. N processors are connected by a network.

Network of Parallel Processors

Definition

1. N processors are connected by a network.
2. Each processor may send packets to others.

Network of Parallel Processors

Definition

1. N processors are connected by a network.
2. Each processor may send packets to others.
3. Each link can carry a unit packet in a step.

Network of Parallel Processors

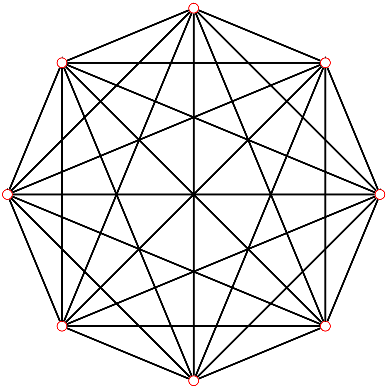
Definition

1. N processors are connected by a network.
2. Each processor may send packets to others.
3. Each link can carry a unit packet in a step.
4. During a step, a processor can send at most one packet to each of its neighbors.

Network of Parallel Processors

Definition

1. N processors are connected by a network.
2. Each processor may send packets to others.
3. Each link can carry a unit packet in a step.
4. During a step, a processor can send at most one packet to each of its neighbors.
5. Each processor has a unique identifying number between 1 and N.



nodes: processors
edges: links

Permutation Routing

Definition

1. N nodes: $[N]$.

Permutation Routing

Definition

1. N nodes: $[N]$.
2. Each node sends one packet to a distinct destination.

Permutation Routing

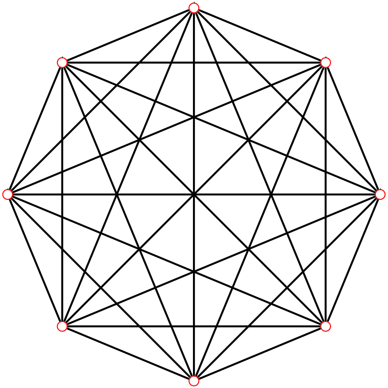
Definition

1. N nodes: $[N]$.
2. Each node sends one packet to a distinct destination.
3. Destinations: specified by a permutation of [N].

Permutation Routing

Definition

1. N nodes: $[N]$.
2. Each node sends one packet to a distinct destination.
3. Destinations: specified by a permutation of [N].
4. origin: i, dest.: $d(i)$.



Complete graph: one step to route all N packets without congestion.

Olivious Routing

Definition

An oblivious algorithm for the permutation routing problem satisfies the following property:
The route followed by v_i depends on $d(i)$ alone, and not on $d(j)$ for any $j \neq i$.

Olivious Routing

Definition

An oblivious algorithm for the permutation routing problem satisfies the following property:
The route followed by v_i depends on $d(i)$ alone, and not on $d(j)$ for any $j \neq i$.

Theorem

For any deterministic oblivious permutation routing algorithm on a network of N nodes of out-degree d, there is an instance of permutation routing $\Omega(\sqrt{N}/d)$ steps.

Proof.

Cf. (Kaklamanis, Krizanc, Tsantilas, 1991) based on (Borodin, Hopcroft, 1985) □

Boolean Hypercube

Definition

1. n dimension hypercube

Boolean Hypercube

Definition

1. n dimension hypercube
2. Nodes: $\{0,1\}^n$
   $N = 2^n$

Boolean Hypercube

Definition

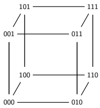
1. n dimension hypercube
2. Nodes:$\{0,1\}^n$
   $N = 2^n$
3. Edges: $\forall u, v \in \{0,1\}^n$ $u \sim v$ if
   $H(u,v) = 1$

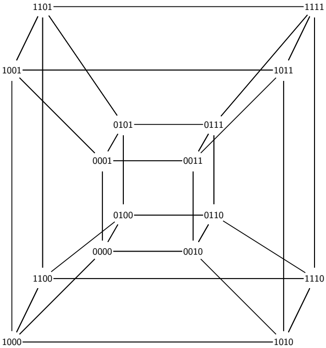
Boolean Hypercube

Definition

1. n dimension hypercube
2. Nodes: $\{0,1\}^n$
   $N = 2^n$
3. Edges: $\forall u, v \in \{0,1\}^n$ $u \sim v$ if $H(u,v) = 1$
4. Degree: n

Boolean Hypercube

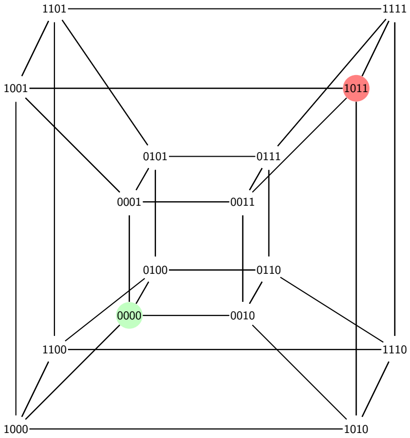




Definition

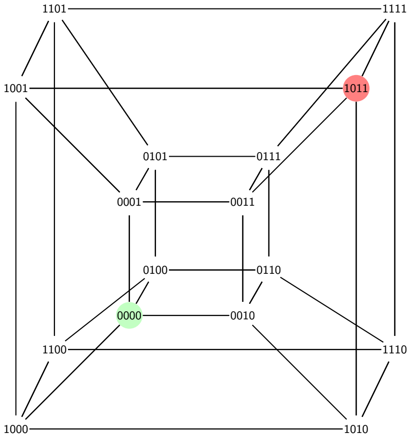
1. n dimension hypercube
2. Nodes:$\{0,1\}^n$
   $N = 2^n$
3. Edges: $\forall u, v \in \{0,1\}^n$ $u \sim v$ if $H(u,v) = 1$
4. Degree: n
5. Number of edges: $N \times n$ (for a directed hypercube).

bit-fixing strategy



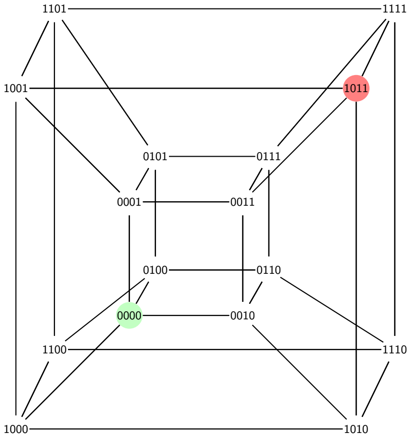
bit-fixing strategy

- 0000



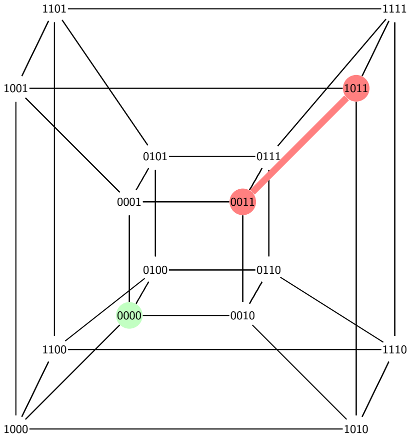
bit-fixing strategy

- 0000
- 1011



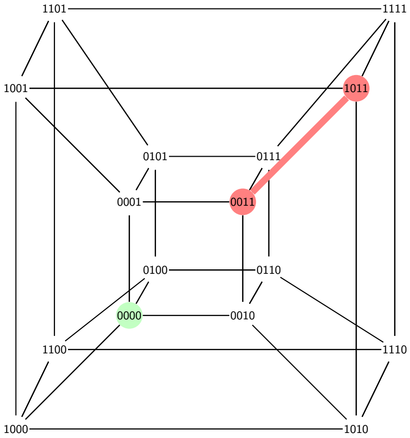
bit-fixing strategy

- 0000
- 0011



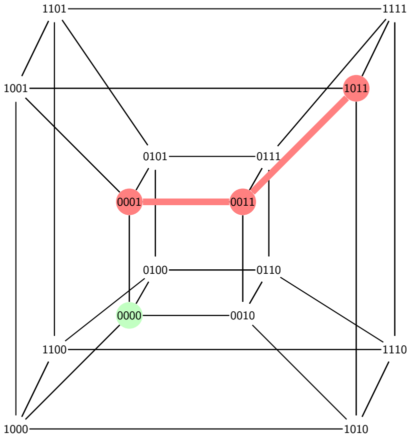
bit-fixing strategy

- 0000
- 0011



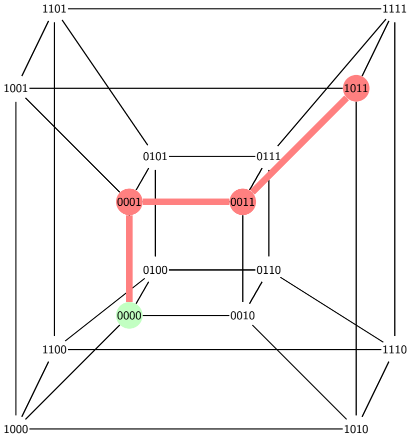
bit-fixing strategy

- 0000
- 0001



bit-fixing strategy

- 0000



Randomized Oblivious Routing Algorithm

Algorithm

Phase1: Pick a random intermediate destination $\sigma(i)$ from $\{1, 2, \ldots, N\}$. Packet v_i travels to node $\sigma(i)$.

Phase2: Packet v_i travels from node $\sigma(i)$ on to its destination $d(i)$.

Randomized Oblivious Routing Algorithm

Algorithm

Phase1: Pick a random intermediate destination $\sigma(i)$ from $\{1, 2, \ldots, N\}$. Packet v_i travels to node $\sigma(i)$.

Phase2: Packet v_i travels from node $\sigma(i)$ on to its destination $d(i)$.

How many steps does elapse before packet v_i reaches its destination?

Randomized Oblivious Routing Algorithm

Algorithm

Phase1: Pick a random intermediate destination $\sigma(i)$ from $\{1, 2, \ldots, N\}$. Packet v_i travels to node $\sigma(i)$.

Phase2: Packet v_i travels from node $\sigma(i)$ on to its destination $d(i)$.

How many steps does elapse before packet v_i reaches its destination?

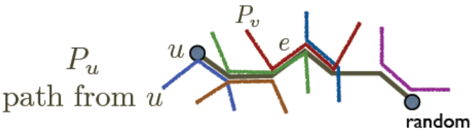
First Consider Phase1

CVCXVCV

For every $u \in \{0,1\}^n$ destination is a uniform and independent $v \in \{0,1\}^n$

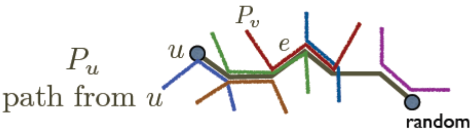
Exercise

view each route in phase1 as a directed path in the hypercube from the source to the intermediate destination. Prove that once two routes separate, they do not rejoin.



Lemma

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



Definition

- $H_{ij} = \begin{cases} 1 & \rho_i \text{ and } \rho_j \text{ share at least one edge.} \\ 0 & \text{otherwise.} \end{cases}$

Definition

- $H_{ij} = \begin{cases} 1 & \rho_i \text{ and } \rho_j \text{ share at least one edge.} \\ 0 & \text{otherwise.} \end{cases}$
- Total delay incurred by v_i is at most $\sum_{i=1}^{N} H_{ij}$

Definition

- $H_{ij} = \begin{cases} 1 & \rho_i \text{ and } \rho_j \text{ share at least one edge.} \\ 0 & \text{otherwise.} \end{cases}$
- Total delay incurred by v_i is at most $\sum_{i=1}^{N} H_{ij}$
- H_{ij}'s are independent Poisson trials for $i \neq j$

Definition

- $H_{ij} = \begin{cases} 1 & \rho_i \text{ and } \rho_j \text{ share at least one edge.} \\ 0 & \text{otherwise.} \end{cases}$
- Total delay incurred by v_i is at most $\sum_{i=1}^{N} H_{ij}$
- H_{ij}'s are independent Poisson trials for $i \neq j$ **(Why?)**

Definition

- $H_{ij} = \begin{cases} 1 & \rho_i \text{ and } \rho_j \text{ share at least one edge.} \\ 0 & \text{otherwise.} \end{cases}$
- Total delay incurred by v_i is at most $\sum_{i=1}^{N} H_{ij}$
- H_{ij}'s are independent Poisson trials for $i \neq j$ **(Why?)**
- To bound the delay of packet v_i with Chernoff's bound, it suffices to obtain an upper bound on $\sum_{i=1}^{N} H_{ij}$.

Definition

- $H_{ij} = \begin{cases} 1 & \rho_i \text{ and } \rho_j \text{ share at least one edge.} \\ 0 & \text{otherwise.} \end{cases}$
- Total delay incurred by v_i is at most $\sum_{i=1}^{N} H_{ij}$
- H_{ij}'s are independent Poisson trials for $i \neq j$ **(Why?)**
- To bound the delay of packet v_i with Chernoff's bound, it suffices to obtain an upper bound on $\sum_{i=1}^{N} H_{ij}$.
- For an edge e, $T(e)$ denotes the number of routes that through e.

Definition

- $H_{ij} = \begin{cases} 1 & \rho_i \text{ and } \rho_j \text{ share at least one edge.} \\ 0 & \text{otherwise.} \end{cases}$
- Total delay incurred by v_i is at most $\sum_{i=1}^{N} H_{ij}$
- H_{ij}'s are independent Poisson trials for $i \neq j$ **(Why?)**
- To bound the delay of packet v_i with Chernoff's bound, it suffices to obtain an upper bound on $\sum_{i=1}^{N} H_{ij}$.
- For an edge e, $T(e)$ denotes the number of routes that through e.

$$\sum_{j=1}^{N} H_{ij} \leq \sum_{l=1}^{k} T(e_l).$$

Global Wiring

Given $\sum_{j=1}^{N} H_{ij} \leq \sum_{l=1}^{k} T(e_l)$, then

$$E[\sum_{j=1}^{N} H_{ij}] \leq E[\sum_{l=1}^{k} T(e_l)].$$

Global Wiring

Given $\sum_{j=1}^{N} H_{ij} \leq \sum_{l=1}^{k} T(e_l)$, then

$$E[\sum_{j=1}^{N} H_{ij}] \leq E[\sum_{l=1}^{k} T(e_l)].$$

$E(l_{\rho_j})$ is $\frac{n}{2}$ for all j, so the total route length is $Nn/2$.

$$E[T(e)] = \frac{Nn/2}{Nn} = 1/2.$$

Global Wiring

Given $\sum_{j=1}^{N} H_{ij} \leq \sum_{l=1}^{k} T(e_l)$, then

$$E[\sum_{j=1}^{N} H_{ij}] \leq E[\sum_{l=1}^{k} T(e_l)].$$

$E(l_{\rho_j})$ is $\frac{n}{2}$ for all j, so the total route length is $Nn/2$.

$$E[T(e)] = \frac{Nn/2}{Nn} = 1/2.$$

$$E[\sum_{j=1}^{N} H_{ij}] \leq E[\sum_{l=1}^{k} T(e_l)] = \frac{k}{2} \leq \frac{n}{2}.$$

By Chernoff Bound $Pr[X > (1+\delta)\mu] \leq 2^{-(1+\delta)\mu}$, we have

$$Pr[\sum_{j=1}^{N} H_{ij} > 6n] \leq 2^{-6n}.$$

As the number of packets is $N = 2^n$, then the probability that at least one packet has delay more than $6n$ is less than $2^n \times 2^{-6n} = 2^{-5n}$.

Delay and length of path for each packet are at most $6n$ and n, respectively. Then, the number of steps for phase1 is at most $7n$.

Theorem

With probability at least $1 - 2^{-5n}$,every packet reaches its intermediate destination in phase1 in $7n$ or fewer steps.

The situation for phase2 ia similar to phase1.
Therefore, the probability that a packet cannot reach its destination is

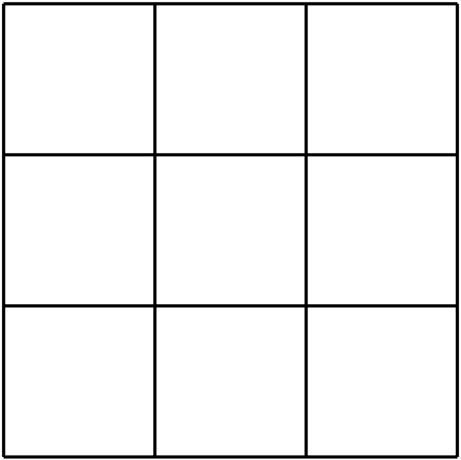
$$2 \times 2^{-5n} \leq 2^{-n} = \frac{1}{N}$$

Theorem

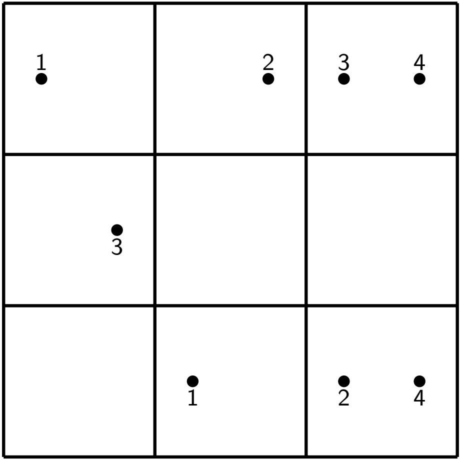
With probability at least $1 - 1/N$, every packet reaches its destination in $14n$ or fewer steps.

Outline

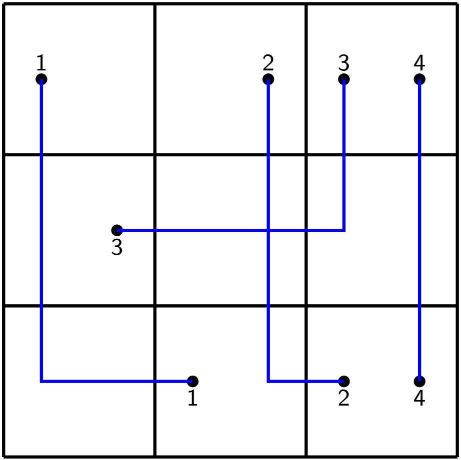
Global Wiring



Global Wiring



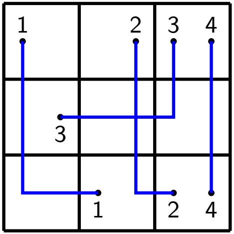
Global Wiring



Global Wiring

Definition

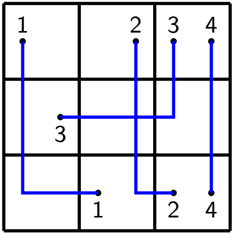
1. $x_{i0} = \begin{cases} 1 & \text{net i goes horizantally} \\ & \text{first from the left end-point.} \\ 0 & \text{otherwise.} \end{cases}$



Global Wiring

Definition

1. $x_{i0} = \begin{cases} 1 & \text{net i goes horizantally} \\ & \text{first from the left end-point.} \\ 0 & \text{otherwise.} \end{cases}$



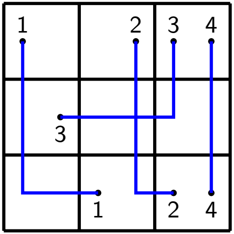
Global Wiring

Definition

1. $x_{i0} = \begin{cases} 1 & \text{net i goes horizantally} \\ & \text{first from the left end-point.} \\ 0 & \text{otherwise.} \end{cases}$
2. $x_{i1} = \begin{cases} 1 & \text{net i goes vertically} \\ & \text{from the left end-point.} \\ 0 & \text{otherwise.} \end{cases}$

Example

$$x_{11} = 0 \text{ And } x_{31} = 1$$



Wiring Problem

Integer program

For each boundary b,

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

and

$$T_{b1} = \{i | \text{net i passes through b if } x_{i1} = 1\}$$

Integer program

For each boundary b,

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

and

$$T_{b1} = \{i|\text{net i passes through b if } x_{i1} = 1\}$$

Definition

$$\min w$$

s.t.

$$\sum_{i \in T_{b0}} x_{i0} + \sum_{i \in T_{b1}} x_{i1} \leq w (\forall \text{ boundaries } b),$$

$$x_{i0} + x_{i1} = 1 (\forall \text{ net } i),$$

$$x_{i0}, x_{i1} \in \{0, 1\} (\forall \text{ net } i)$$

Integer program

For each boundary b,

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

and

$$T_{b1} = \{i | \text{net i passes through b if } x_{i1} = 1\}$$

Definition

$$\min w$$

s.t.

$$\sum_{i \in T_{b0}} x_{i0} + \sum_{i \in T_{b1}} x_{i1} \leq w (\forall \text{ boundaries } b),$$

$$x_{i0} + x_{i1} = 1 (\forall \text{ net } i),$$

$$x_{i0}, x_{i1} \in \{0, 1\} (\forall \text{ net } i)$$

Zero-One Linear Programming is $\mathcal{NP}$-hard.

Relaxed L.P. Model

Definition

min w

s.t.

$$\sum_{i \in T_{b0}} x_{i0} + \sum_{i \in T_{b1}} x_{i1} \leq w(\forall \text{ boundaries } b),$$

$$x_{i0} + x_{i1} = 1(\forall \text{ net } i),$$

$$x_{i0}, x_{i1} \in \{0, 1\}(\forall \text{ net } i)$$

Relaxed L.P. Model

Definition

$$\min w$$

s.t.

$$\sum_{i \in T_{b0}} x_{i0} + \sum_{i \in T_{b1}} x_{i1} \leq w (\forall \text{ boundaries } b),$$

$$x_{i0} + x_{i1} = 1 (\forall \text{ net } i),$$

$$x_{i0}, x_{i1} \in \{0, 1\} (\forall \text{ net } i)$$

Definition

$$\min w$$

s.t.

$$\sum_{i \in T_{b0}} x_{i0} + \sum_{i \in T_{b1}} x_{i1} \leq w (\forall \text{ boundaries } b),$$

$$x_{i0} + x_{i1} = 1 (\forall \text{ net } i),$$

$$x_{i0}, x_{i1} \in [0, 1] (\forall \text{ net } i)$$

Relaxed L.P. Model

- In order to find an optimal solution to the above 0-1 linear programming, we need to solve an $\mathcal{NP} - hard$ problem!
- Therefore, we seek for a near optimal solution by the following approach:
  - Relax the $0-1$ constraint to $x_{i0}, x_{i1} \in [0,1]$ for each i,
  - The obtained model is a linear programming model which can be easily solved using Simplex method, for example.
  - Let $\hat{x_{i0}}$ and $\hat{x_{i1}}$ be the solutions to the relaxed model obtained in the previous step and $\hat{w}$ be the value of the objective function for these solutions,
- ...

Randomized Rounding

- ...
- $\hat{x_{i0}}$ and $\hat{x_{i1}}$ might be fractional numbers, then they cannot be appropriate answers for our $0-1$ model. Hence we round them to 0 or 1 in a probabilistic fashion, called *randomized rounding*, as follow

$$Pr[\overline{x_{i0}} = 1] = \hat{x_{i0}}$$
$$Pr[\overline{x_{i1}} = 1] = \hat{x_{i1}}$$

  where $\overline{x_{i1}}$ and $\overline{x_{i1}}$ denote the rounded value of $\hat{x_{i1}}$ and $\hat{x_{i1}}$, respectively.

Randomized rounding

Theorem

Let ϵ be a real number such that $0 < \epsilon < 1$. Then with probability $1 - \epsilon$, the global wiring S produced by randomized rounding satisfies

$$w_S \leq \hat{w}(1 + \Delta^+(\hat{w}, \epsilon/2n)) \leq w_o(1 + \Delta^+(w_o, \epsilon/2n)))$$

Proof.

- $\sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}$

Randomized rounding

Theorem

Let ϵ be a real number such that $0 < \epsilon < 1$. Then with probability $1 - \epsilon$, the global wiring S produced by randomized rounding satisfies

$$w_S \leq \hat{w}(1 + \Delta^+(\hat{w}, \epsilon/2n)) \leq w_o(1 + \Delta^+(w_o, \epsilon/2n)))$$

Proof.

- $\sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}$
- $w_S(b) = \sum_{i \in T_{b0}} \overline{x}_{i0} + \sum_{i \in T_{b1}} \overline{x}_{i1}$

Randomized rounding

Theorem

Let ϵ be a real number such that $0 < \epsilon < 1$. Then with probability $1 - \epsilon$, the global wiring S produced by randomized rounding satisfies

$$w_S \leq \hat{w}(1 + \Delta^+(\hat{w}, \epsilon/2n)) \leq w_o(1 + \Delta^+(w_o, \epsilon/2n)))$$

Proof.

- $\sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}$
- $w_S(b) = \sum_{i \in T_{b0}} \overline{x}_{i0} + \sum_{i \in T_{b1}} \overline{x}_{i1}$
- $E[w_S(b)] = \sum_{i \in T_{b0}} E[\overline{x}_{i0}] + \sum_{i \in T_{b1}} E[\overline{x}_{i1}] = \sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}$.

Randomized rounding

Theorem

Let ϵ be a real number such that $0 < \epsilon < 1$. Then with probability $1 - \epsilon$, the global wiring S produced by randomized rounding satisfies

$$w_S \leq \hat{w}(1 + \Delta^+(\hat{w}, \epsilon/2n)) \leq w_o(1 + \Delta^+(w_o, \epsilon/2n)))$$

Proof.

- $\sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}$
- $w_S(b) = \sum_{i \in T_{b0}} \overline{x}_{i0} + \sum_{i \in T_{b1}} \overline{x}_{i1}$
- $E[w_S(b)] = \sum_{i \in T_{b0}} E[\overline{x}_{i0}] + \sum_{i \in T_{b1}} E[\overline{x}_{i1}] = \sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}$.
- $Pr[w_S(b) > \hat{w}(1 + \Delta^+(\hat{w}, \epsilon/2n))] \leq \epsilon/2n$.

Randomized rounding

Theorem

Let ϵ be a real number such that $0 < \epsilon < 1$. Then with probability $1 - \epsilon$, the global wiring S produced by randomized rounding satisfies

$$w_S \leq \hat{w}(1 + \Delta^+(\hat{w}, \epsilon/2n)) \leq w_o(1 + \Delta^+(w_o, \epsilon/2n)))$$

Proof.

- $\sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}$
- $w_S(b) = \sum_{i \in T_{b0}} \overline{x}_{i0} + \sum_{i \in T_{b1}} \overline{x}_{i1}$
- $E[w_S(b)] = \sum_{i \in T_{b0}} E[\overline{x}_{i0}] + \sum_{i \in T_{b1}} E[\overline{x}_{i1}] = \sum_{i \in T_{b0}} \hat{x}_{i0} + \sum_{i \in T_{b1}} \hat{x}_{i1} \leq \hat{w}.$
- $Pr[w_S(b) > \hat{w}(1 + \Delta^+(\hat{w}, \epsilon/2n))] \leq \epsilon/2n.$
- since the number of boundaries is less than 2n

$$Pr[w_S > \hat{w}(1 + \Delta^+(\hat{w}, \epsilon/2n))] \leq \epsilon.$$

□

Consider $w_0 = n^\gamma$

$$w_S \leq n^\gamma (1 + \sqrt{\frac{4 \ln 2n/\epsilon}{n^\gamma}})$$

Outline

Outline