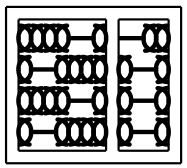


UNIVERSITY OF CAMPINAS - UNICAMP
INSTITUTE OF COMPUTING - IC

Randomized Algorithms

Flávio Keidi Miyazawa

Campinas, 2018

Randomized algorithms are in general:

- simpler;
- faster;
- avoid pathological cases;
- can give interesting deterministic informations;

But (hopefully, with small probability) can

- give wrong answers;
- take too long.

Verifying Polynomials

- Given polynomials $F(x)$ and $G(x)$ as:
- $F(x) = (x+1)(x-2)(x+3)(x-4)(x+5)(x-6)$
- $G(x) = x^6 - 7x^3 + 25$.
- How to verify if $F(x) \equiv G(x)$.

- Natural solution in $O(d^2)$.

Consider the following randomized algorithm:

ALGORITHM VP

1. Choose $r \in \{1, \ldots, 100d\}$ randomly.
2. Verify if $F(r)$ is equal to $G(r)$, in time $O(d)$.
3. If $F(r) = G(r)$ then return YES;
4. otherwise, return NO.

The algorithm:

- Has time complexity $O(d)$.
- Fails if $r \in \{1, \ldots, 100d\}$ is root of $H(x) = 0$, where $H(x) = F(x) - G(x)$.

As $H(x)$ has maximum degree d,

- $H(x)$ has at most d roots
- The probability that algorithm *VP* fail is at most

$$\Pr(VP\ fail) \leq \frac{d}{100d} = \frac{1}{100}$$

And how to decrease the probability to fail to $\frac{1}{1000000}$?

Axioms of Probability

Def.: *A probability space has 3 components:*

- A sample space Ω
- A family $\mathcal{F}$ of events, each $E \in \mathcal{F}$ is s.t. $E \subseteq \Omega$.
- A probability function $\Pr : \mathcal{F} \to \mathbb{R}^+$

$E \in \mathcal{F}$ is called *simple* or *elementary* if $|E| = 1$

Def.: *A probability function is any function* $\Pr : \mathcal{F} \to \mathbb{R}^+$ *s.t.*

- $\forall E \in \mathcal{F}$ we have $0 \leq \Pr(E) \leq 1$
- $\Pr(\Omega) = 1$
- For any finite or enumerable sequence of events mutually disjoint $E_1, E_2, \ldots,$ we have

$$\Pr(E_1 \cup E_2 \ldots) = \Pr(E_1) + \Pr(E_2) + \ldots$$

Example: *In the polynomial verification*

- $\Omega = \{1, \ldots, 100d\}$
- Each simple event E_i is an event of choosing $r = i$, for $i = 1, \ldots, 100d$
- E_i is chosen uniformly at random $\Rightarrow \Pr(E_i) = \Pr(E_j), \forall i, j$.
- $\Pr(\Omega) = 1 \Rightarrow \Pr(E_i) = \frac{1}{100d}$.

Example: *Consider tossing a die (of 6 sides).*

- $\Omega = \{1, \ldots, 6\}$

Examples of events we may consider,

- E' = Event of having an even number.
- E'' = Event of having a number at most 3.
- E''' = Event of having prime number.

Lemma: *For events E_1 and E_2 we have*

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$$

Proof.

$$\begin{aligned}
\Pr(E_1) &= \Pr(E_1 - (E_1 \cap E_2)) + \Pr(E_1 \cap E_2) \\
\Pr(E_2) &= \Pr(E_2 - (E_1 \cap E_2)) + \Pr(E_1 \cap E_2)
\end{aligned}$$

$$\begin{aligned}
\Pr(E_1 \cup E_2) = &\Pr(E_1 - (E_1 \cap E_2)) + \Pr(E_2 - (E_1 \cap E_2)) + \Pr(E_1 \cap E_2) \\
= &\Pr(E_1) - (E_1 \cap E_2)) + \\
&\Pr(E_2) - (E_1 \cap E_2)) + \Pr(E_1 \cap E_2) \\
= &\Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)
\end{aligned}$$

□

Corollary: *For events E_1 and E_2 we have*

$$\Pr(E_1 \cup E_2) \leq \Pr(E_1) + \Pr(E_2)$$

Lemma: *For events $E_1, E_2, \ldots$ we have*

$$\Pr(\bigcup_{i \geq 1} E_i) \leq \sum_{i \geq 1} \Pr(E_i)$$

Lemma: *For events $E_1, E_2, \ldots$ we have*

$$\begin{aligned}
\Pr\Big(\bigcup_{i\geq 1} E_i\Big) &= \sum_{i\geq 1} \Pr(E_i) \\
&\quad - \sum_{i<j} \Pr(E_i \cap E_j) \\
&\quad + \sum_{i<j<k} \Pr(E_i \cap E_j \cap E_k) \\
&\quad \vdots \\
&\quad (-1)^{l+1} \sum_{i_1<i_2<\ldots<i_l} \Pr(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_l})
\end{aligned}$$

Proof. Exercise. □

Def.: *Two events E and F are said to be independent iff*

$$\Pr(E \cap F) = \Pr(E) \cdot \Pr(F)$$

and events $E_1, \ldots, E_k$ are mutually independent iff $\forall I \subseteq \{1, \ldots, k\}$ we have

$$\Pr(\bigcap_{i \in I} E_i) = \Pi_{i \in I} \Pr(E_i).$$

Example: *Strengthening algorithm VP with $k \geq 2$ runnings*

ALGORITHM VP_k

1. *Execute algorithm VP k times (possibly with repetitions).*
2. *Return* NO *if one of the k executions of VP returns* NO*;*
3. *otherwise, return* YES.

- *Let E_i be the event the algorithm choose a root of $F(x) - G(x) = 0$ in the i-th execution of VP.*
- *The events E_i are mutually independent.*
- *The probability the algorithm fail is:*

$$\Pr(E_1 \cap E_2 \cap \ldots \cap E_k) = \Pi_{i=1}^{k} \Pr(E_i) \leq \Pi_{i=1}^{k} \frac{d}{100d} \leq \left(\frac{1}{100}\right)^k$$

Def.: *The conditional probability of event E given that F occurred is given by*

$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)},$$

given that $\Pr(F) > 0$.

Proposition: *If E and F are events, with $\Pr(F) > 0$, then*

$$\Pr(E \cap F) = \Pr(E|F) \cdot \Pr(F) = \Pr(F|E) \cdot \Pr(E).$$

Proposition: *If E and F are independent events, with $\Pr(F) > 0$, then*

$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)} = \frac{\Pr(E) \cdot \Pr(F)}{\Pr(F)} = \Pr(E).$$

Lemma: *If $E_1, \ldots, E_k$ are events, then*

$$\Pr(E_1 \cap \ldots \cap E_k) = \Pr(E_1) \cdot \Pr(E_2|E_1) \cdot \Pr(E_3|E_1 \cap E_2) \cdots \Pr(E_k|E_1 \cap E_2 \cap \ldots \cap E_{k-1}).$$

Proof.

$$\begin{aligned}
& \Pr(E_1 \cap \ldots \cap E_k) \\
= \; & \Pr((E_1 \cap \ldots \cap E_{k-1}) \cap E_k) \\
= \; & \Pr(E_1 \cap \ldots \cap E_{k-1}) \cdot \Pr(E_k|E_1 \cap E_2 \cap \ldots \cap E_{k-1}) \\
= \; & \Pr(E_1 \cap \ldots \cap E_{k-2}) \cdot \Pr(E_{k-1}|E_1 \cap E_2 \cap \ldots \cap E_{k-2}) \\
& \qquad \cdot \Pr(E_k|E_1 \cap E_2 \cap \ldots \cap E_{k-1}) \\
& \vdots \\
= \; & \Pr(E_1) \cdot \Pr(E_2|E_1) \cdot \Pr(E_3|E_1 \cap E_2) \cdots \Pr(E_k|E_1 \cap E_2 \cap \ldots \cap E_{k-1})
\end{aligned}$$

□

Example: *Consider a modification of algorithm VP_k:*

ALGORITHM VP2$_k$

1. *For* $i \leftarrow 1$ *to* k *do*
2. *Choose* r_i *uniformly at random in* $\{1, \ldots, 100d\} \setminus \{r_1, \ldots, r_{i-1}\}$.
3. *If* $F(r_i) \neq G(r_i)$ *return* NO.
4. *Return* YES.

This algorithm can be implemented to run in $O(k \cdot d)$.

What is the probability that $VP2_k$ fails ?

- Let E_i be the event of choosing a root of $F(x) - G(x) = 0$ in the i-th iteration of *VP2*.
- The probability the algorithm fail is:

As $\Pr(E_j|E_1 \cap \ldots \cap E_j - 1)$ is the probability to choose a root of $F(x) - G(x) = 0$ considering we obtained $j-1$ roots, $j-1 < d$ it remains $d - (j-1)$ roots. That is,

$$\Pr(E_j|E_1 \cap \ldots \cap E_{j-1}) \leq \frac{d-(j-1)}{100d-(j-1)}.$$

So,

$$\Pr(E_1 \cap \ldots \cap E_k) \leq \Pi_{j=1}^{k} \frac{d-(j-1)}{100d-(j-1)} \leq \left(\frac{1}{100}\right)^k.$$

Note that $VP2_{d+1}$ is an exact algorithm, but with time $\Theta(d^2)$.

VERIFYING MATRIX MULTIPLICATION

Example: *Given matrices A, B and C verify if* $A \cdot B = C$.

By simplicity, consider matrices of order n and integers $\mod 2$.

- Trivial Algorithm: Time complexity $O(n^3)$
- Sofisticated Algorithm: Time complexity $O(n^{2.37})$
- We will see a randomized algorithm $O(n^2)$ that fails with probability $\leq \frac{1}{2}$.

Algorithm $VMM(A, B, C)$

1. Choose $r \in \{0,1\}^n$
2. If $A \cdot (B \cdot r) = C \cdot r$ return YES.
3. Otherwise, return NO.

Time Complexity: $O(n^2)$.

Theorem: *If* $A \cdot B \neq C$ *then* $\Pr(A \cdot B \cdot r = C \cdot r) \leq \frac{1}{2}$.

Proof. Supose that $D = A \cdot B - C \neq 0$ and $D \cdot r = 0$ (i.e., $A \cdot B \cdot r = C \cdot r$).
If $D = (d_{ij}) \neq 0$ there exists ij such that $d_{ij} \neq 0$.

W.L.O.G., let $d_{11} \neq 0$. As $D \cdot r = 0$ we have

$$\sum_{j=1}^{n} d_{1j} r_j = 0$$

and therefore

$$r_1 = \frac{-\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}$$

Choosing a random vector $r \in \{0,1\}^n$ is the same to choose r_n, then $r_{n-1}, \ldots$, and then choose r_1.
Suppose that $r_n, \ldots, r_2$ have already been chosen and r_1 not.
At this point, $\sum_{j=2}^{n} d_{1j} r_j$ is determined.
Consider the choose of r_1. The probability that

$$r_1 = \frac{-\sum_{j=2}^{n} d_{1j} r_j}{d_{11}}$$

is valid is at most $\frac{1}{2}$. So, $\Pr(A \cdot B \cdot r = C \cdot r) \leq \frac{1}{2}$. □

This technique is called *Principle of Deferred Decisions*: Consider some of the variables defined and let others open (or deferred).

Theorem: *(Law of Total Probabilities)*
Let $E_1, \ldots, E_n$ *mutually disjoint events and* $\cup_{i=1}^{n} E_i = \Omega$.
Then,

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B|E_i) \cdot \Pr(E_i).$$

Example: *Applying the law of total probabilities in the example of matrix multiplication:*

$$\begin{aligned}
\Pr(\overbrace{ABr=Cr}^{\text{Event } B}) &= \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr\Big((\overbrace{ABr=Cr}^{\text{Event } B}) \cap (\overbrace{(r_2,\ldots,r_n)=(x_2,\ldots,x_n)}^{\text{Event } E_i})\Big) \\
&\leq \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr\Big((r_1=\frac{-\sum_{j=2}^{n} d_{1j}r_j}{d_{11}} \cap ((r_2,\ldots,r_n)=(x_2,\ldots,x_n))\Big) \\
&= \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr(r_1=\frac{-\sum_{j=2}^{n} d_{1j}r_j}{d_{11}}) \cdot \Pr((r_2,\ldots,r_n)=(x_2,\ldots,x_n)) \\
&\leq \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \frac{1}{2}\Pr((r_2,\ldots,r_n)=(x_2,\ldots,x_n)) \leq \frac{1}{2}
\end{aligned}$$

Repeating the algorithm k times, the probability it fails is $\leq \left(\frac{1}{2}\right)^k$

Let us analyse the change as the algorithm returns YES in each execution.

Theorem: *(Bayes Law) Given disjoint events $E_1, \ldots, E_n$ such that $\cup_{i=1}^{n} E_i = \Omega$ and event B, we have*

$$\begin{aligned}
\Pr(E_j|B) &= \frac{\Pr(E_j \cap B)}{\Pr(B)} \\
&= \frac{\Pr(E_j \cap B)}{\sum_{i=1}^{n} \Pr(B \cap E_i)} \\
&= \frac{\Pr(B|E_j)\Pr(E_j)}{\sum_{i=1}^{n} \Pr(B|E_i)\Pr(E_i)}
\end{aligned}$$

Example: *Consider 3 coins, such that 2 are fair and 1 is biased, that show up head, H, with probability* $\frac{2}{3}$ *and tail, T, with probability* $\frac{1}{3}$.

Suppose the coins are given in a random order. Call the coins as 1, 2 e 3.

Suppose we toss the 3 coins and we obtain $(1 = H, 2 = H, 3 = T)$.
What is the probability that coin 1 is biased ?

Let B be the event that the tossing obtain $(1 = H, 2 = H, 3 = T)$.
Let E_i be the event that coin i is biased.

We can calculate $\Pr(E_1|B)$, as follows:

$$\Pr(E_1|B) = \frac{\Pr(B|E_1) \cdot \Pr(E_1)}{\sum_{i=1}^{3} \Pr(B|E_i) \cdot \Pr(E_i)} = \frac{2}{5}$$

Now, let's see how the confidence of the algorithm Verify Matrix Multiplication increases as the iterations returns YES.

Let E be the event that $A \cdot B = C$ is valid.

If we do not know anything about the identity $A \cdot B = C$ then it is reasonable to suppose that $\Pr(E) = \Pr(\overline{E}) = \frac{1}{2}$

Let E be the event that $A \cdot B = C$ is valid.
Let B be the event that algorithm return YES in the first call (that is, $A \cdot B \cdot r = C \cdot r$).
And how would $\Pr(E|B)$ be ?
We can calculate $\Pr(B|E)$ and $\Pr(B|\overline{E})$:

$$\begin{aligned}
\Pr(B|E) &= 1 \\
\Pr(B|\overline{E}) &\leq \frac{1}{2} \\
\Pr(E|B) &= \frac{\Pr(B|E) \cdot \Pr(E)}{\Pr(B|E) \cdot \Pr(E) + \Pr(B|\overline{E}) \cdot \Pr(\overline{E})} \\
&\geq \frac{1 \cdot \frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}} = \frac{2}{3} = 0.66\ldots
\end{aligned}$$

We increased the confidence!!

Suppose we execute *VMM* again and we obtain YES. What is the new confidence ?

Updating the probabilities of the events, we have:

$$\begin{aligned} \Pr(E) &\geq \frac{2}{3} \\ \Pr(\overline{E}) &\leq \frac{1}{3} \\ \Pr(B|E) &= 1 \\ \Pr(B|\overline{E}) &\leq \frac{1}{2} \end{aligned}$$

So,

$$\begin{aligned} \Pr(E|B) &= \frac{\Pr(B|E)\cdot\Pr(E)}{\Pr(B|E)\cdot\Pr(E)+\Pr(B|\overline{E})\cdot\Pr(\overline{E})} \\ &\geq \frac{1\cdot\frac{2}{3}}{1\cdot\frac{2}{3}+\frac{1}{2}\cdot\frac{1}{3}} = \frac{4}{5} = 0.8 \end{aligned}$$

Following the same reasoning for i iterations, we have

$$\begin{aligned}
\Pr(E) &\geq \frac{2^i}{2^i+1} \\
\Pr(\overline{E}) &\leq 1 - \frac{2^i}{2^i+1} = \frac{1}{2^i+1} \\
\Pr(B|E) &= 1 \\
\Pr(B|\overline{E}) &\leq \frac{1}{2}
\end{aligned}$$

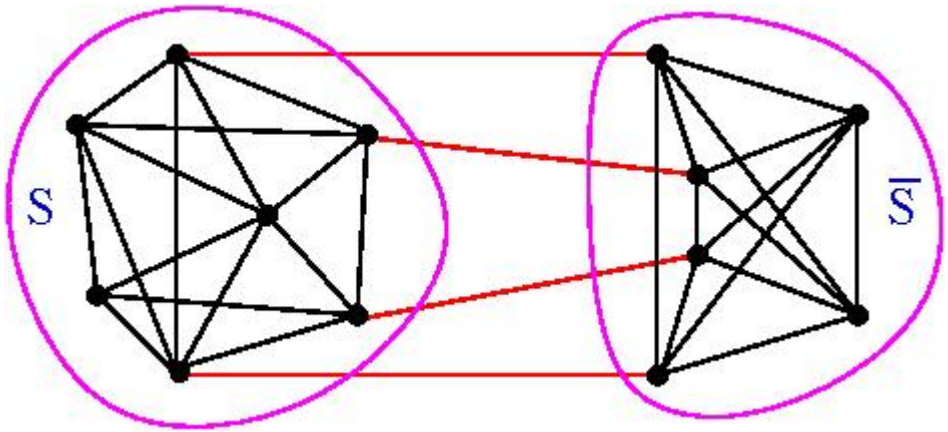
So,

$$\Pr(E|B) \geq \frac{1 \cdot \frac{2^i}{2^i+1}}{1 \cdot \frac{2^i}{2^i+1} + \frac{1}{2} \cdot \frac{1}{2^i+1}} = 1 - \frac{1}{2^i+1}$$

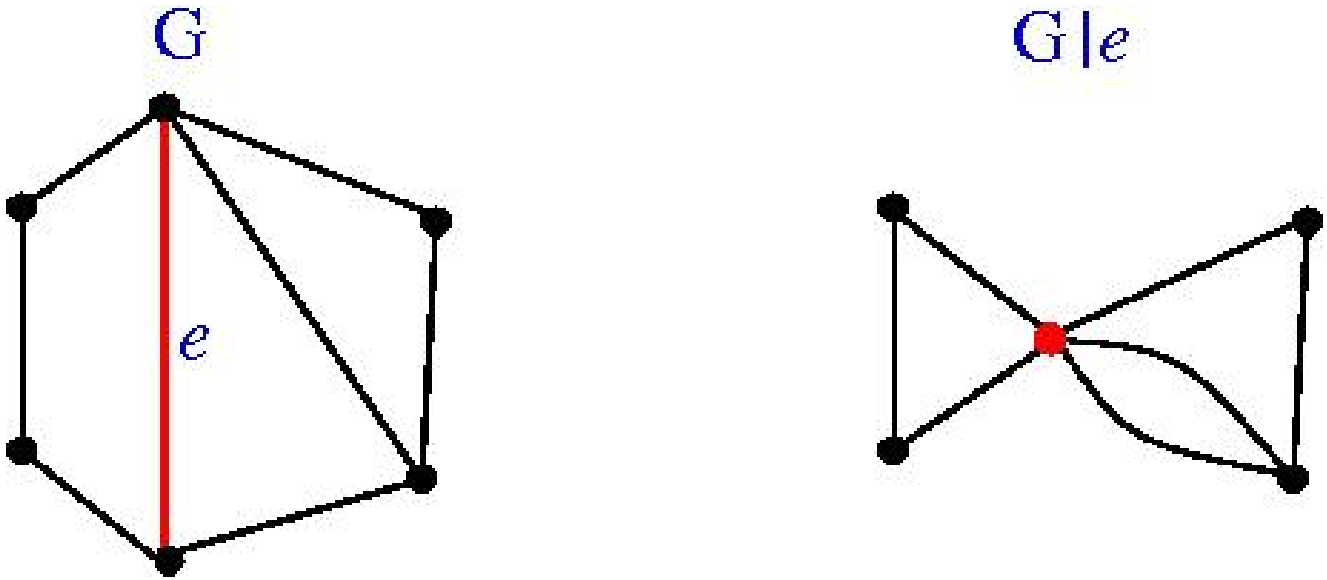
That is (a bit) better than $1 - \frac{1}{2^i}$

Minimum Cut Problem

Problem: *Given a graph $G = (V, E)$, non-oriented, find $\emptyset \neq S \subset V$ such that the number of edges in $(S, \overline{S})$ is minimum.*



Def.: *Given a graph $G = (V, E)$ and an edge $e = \{u, v\} \in E$, we define by $G|e$ the graph obtained from G joining the nodes u and v into only one node and maintaining parallel edges.*



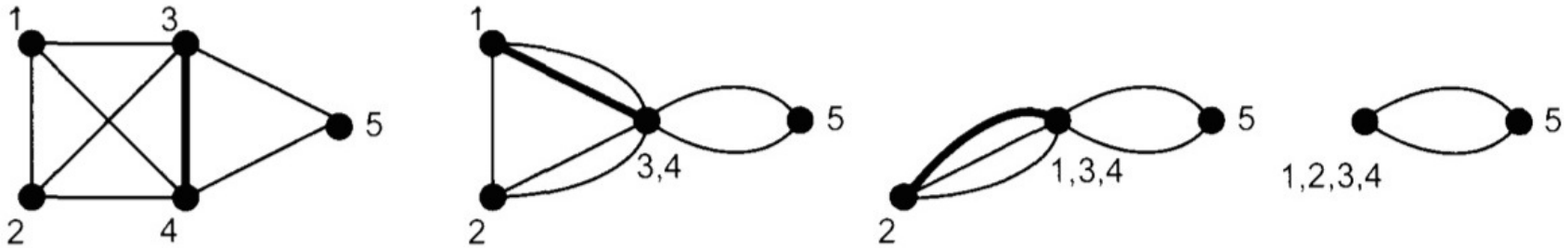
Obs.: *Note that a cut in $G|e$ is also a cut in G, with the same cardinality.*

Algorithm $RandMinCut(G), \quad G = (V, E)$

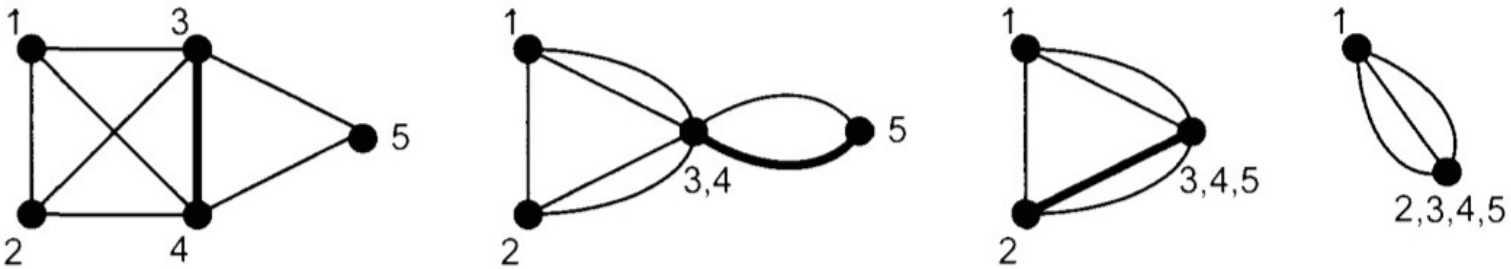
1. $n \leftarrow |V|$.
2. $G_0 \leftarrow G$.
3. For $i \leftarrow 1$ to $n - 2$ do
4. choose edge $e \in G_{i-1}$
5. $G_i \leftarrow G_{i-1}|e$
6. Let C_A be the set of edges in G_{n-2}.
7. Return C_A.

Idea: In each iteration, the chance to choose an edge of the minimum cut is "small".

Execution of the algorithm *RandMinCut*



(a) A successful run of min-cut.



(b) An unsuccessful run of min-cut.

Lemma: *Let $C_{\min}$ be a minimum cut in G and C_A the cut obtained by the algorithm. Then* $\Pr(C_{\min} = C_A) \geq \frac{2}{n(n-1)}$.

Proof. Let k be the number of edges in $C_{\min}$.

Then, $|\delta(v)| \geq k$ for each $v \in G$, where $\delta(v)$ is the set of edges incidents to v.

So,

$$|E| = \frac{\sum_{v \in V} |\delta(v)|}{2} \geq \frac{n \cdot k}{2}.$$

Let E_i be the event that algorithm do not select $e \in C_{\min}$ in iteration i.

$$\begin{aligned}
\Pr(E_1) &= 1 - \frac{k}{|E|} \\
&\geq 1 - \frac{k}{\frac{kn}{2}} = \frac{n-2}{n} \\
\Pr(E_2|E_1) &\geq 1 - \frac{k}{\frac{k(n-1)}{2}} = \frac{n-3}{n-1} \\
\Pr(E_3|E_1 \cap E_2) &\geq 1 - \frac{k}{\frac{k(n-2)}{2}} = \frac{n-4}{n-2} \\
&\vdots \\
\Pr(E_{n-2}|\bigcap_{i=1}^{n-3} E_i) &\geq 1 - \frac{k}{\frac{k(n-(n-3))}{2}} = \frac{1}{3}
\end{aligned}$$

$$\begin{aligned}\Pr\left(\bigcap_{i=1}^{n-2} E_i\right) &\geq \frac{1}{3}\cdot\frac{2}{4}\cdot\frac{3}{5}\cdot\frac{4}{6}\cdots\frac{n-4}{n-2}\cdot\frac{n-3}{n-1}\cdot\frac{n-2}{n} \\ &= \frac{2}{n(n-1)}\end{aligned}$$

□

Let $RandMinCut^t$ be the strengthened algorithm that executes $RandMinCut$ t times and returns the smallest cut.

Proposition: *Let C_A and $C_{\min}$ the smallest cut returned by algorithm $RandMinCut^{n^2}$ and a minimum cut, respectively. Then,*

$$\Pr(C_{\min} \neq C_A) \leq \frac{1}{e^2} \approx 0.135.$$

Proof. We have

$$\Pr(\textit{RandMinCut} \text{ do not obtain } C_{\min}) \leq 1 - \frac{2}{n^2}$$

therefore,

$$\Pr(\textit{RandMinCut}^{n^2} \text{ do not obtain } C_{\min}) \leq (1 - \frac{2}{n^2})^{n^2}$$

It is valid that $(1 + \frac{t}{m})^m \leq e^t$, for $m \geq 1$ and $|t| \leq m$.

Setting $t = -2$ and $m = n^2$, we have

$$\Pr(\textit{RandMinCut}^{n^2} \text{ do not obtain } C_{\min}) \leq \frac{1}{e^2}.$$

□

Proposition: *Let C_A and $C_{\min}$ the cut obtained by algorithm $RandMinCut^{n^2 \ln(n)}$ and a minimum cut, respectively. Then,*

$$\Pr(C_{\min} = C_A) \geq 1 - \frac{1}{n^2}.$$

Proof.

$$\Pr(RandMinCut^{n^2 \ln(n)} \text{ do not obtain } C_{\min}) \leq \left(\frac{1}{e^2}\right)^{\ln(n)} = \frac{1}{n^2}$$

So,

$$\Pr(C_{\min} = C_A) \geq 1 - \frac{1}{n^2}.$$

□

I.e., $RandMinCut^{n^2 \ln(n)}$ obtain a minimum cut with high probability.

Def.: *A random variable (r.v.) X on a random space Ω is a function* $X : \Omega \to \mathbb{R}$.

Def.: *A r.v. X is said to be discrete if it takes finite or countably infinite values.*

We only consider discrete random variables, and we state when it is not the case.

Given r.v. X and a real value a, the event "$X = a$" represents the set $\{e \in \Omega : X(e) = a\}$.

So,

$$\Pr(X = a) = \sum_{e \in \Omega:\ X(e) = a} \Pr(e).$$

Example: *Let X be a r.v. that is the sum of two dice.*

- *X can have 11 possible values:* $X \in \{2, 3, \ldots, 12\}$
- *There are 36 possibilities for the dice:* $\{(1,1), (1,2), (2,1) \ldots, (6,6)\}$

Event $X = 4$ *has 3 basic events:* $\{(1,3), (2,2), (3,1)\}$

Therefore

$$\Pr(X = 4) = \frac{3}{36} = \frac{1}{12}$$

Def.: *Two r.v. X and Y are independent if and only if*

$$\Pr(\ (X = x) \cap (Y = y)\ ) \quad = \quad \Pr(X = x) \cdot \Pr(Y = y) \quad \forall x, y$$

Def.: *The r.v. $X_1, X_2, \ldots, X_k$ are mutually independent if and only if $\forall I \subseteq \{1, \ldots, k\}$ and all x_i, $i \in I$,*

$$\Pr(\bigcap_{i \in I} X_i = x_i) = \Pi_{i \in I} \Pr(X_i = x_i)$$

Def.: *The expectation of a discrete random variable (d.r.v.) X is given by*

$$E[X] = \sum_i i \cdot \Pr(X = i),$$

where the summation is over all values in the range of X.

Def.: *The expectation is finite if $E[X]$ converges; otherwise is said to be unbounded. In this case, we use the notation $E[X] = \infty$.*

Example: *Let X be a r.v. that is the sum of two dice.*

$$E[X] = 2\frac{1}{36} + 3 \cdot \frac{2}{36} + \cdots + 12 \cdot \frac{1}{36} = 7$$

Obs.: *In the above sum, we have to know the number of basic events for each value of X.*

Linearity of Expectation

Theorem: *For each finite collection of discrete r.v. $X_1, \ldots, X_n$ with finite expectations*

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1} E[X_i]$$

Obs.: *Note that there is no restrictions on the independente of $X_1, \ldots, X_n$.*

Proof. We prove that $E[X+Y] = E[X] + E[Y]$ for r.v. X and Y.

$$
\begin{aligned}
E[X+Y] &= \sum_i \sum_j (i+j)\Pr(\ (X=i) \cap (Y=j)\ ) \\
&= \sum_i \sum_j i \cdot \Pr(\ (X=i) \cap (Y=j)\ ) + \\
&\quad \sum_j \sum_i j \cdot \Pr(\ (X=i) \cap (Y=j)\ ) \\
&= \sum_i i \cdot \sum_j \Pr(\ (X=i) \cap (Y=j)\ ) + \\
&\quad \sum_j j \cdot \sum_i \Pr(\ (X=i) \cap (Y=j)\ ) \\
&= \sum_i i \cdot \Pr(X=i) + \sum_j j \cdot \Pr(Y=j) = E[X] + E[Y]
\end{aligned}
$$

Exercise: Complete the proof for more r.v. (sug. by induction). □

Lemma: *Given a r.v. X and constant c, we have $E[c \cdot X] = c \cdot E[X]$.*

Proof. The lemma is straighforward for $c = 0$.

$$\begin{aligned} E[c \cdot X] &= \sum_i i \cdot \Pr(c \cdot X = i) \\ &= \sum_i i \cdot \Pr(X = \frac{i}{c}) \\ &= c \cdot \sum_i \frac{i}{c} \cdot \Pr(X = \frac{i}{c}) \\ &= c \cdot E[X] \end{aligned}$$

□

Example: *Consider the example of two dice.*

Let X_1 be the r.v. of the value of the first die.
Let X_2 be the r.v. of the value of the second die.
Let X be the r.v. of the sum of the values of the two dice.

Note that $X = X_1 + X_2$. So,

$$\begin{aligned} E[X] &= E[X_1 + X_2] \\ &= E[X_1] + E[X_2] \\ &= 2 \cdot \sum_{i=1}^{6} i \cdot \frac{1}{6} \\ &= 7 \end{aligned}$$

JENSEN'S INEQUALITY

Example: *Let X be the length of a side of a square chosen uniformly at random in $[1, 99]$.*

What is the expectation of $E[X^2]$ of the area of the square ?

It is tempting to thing that is equal to $E[X]^2 = 2500$.

But, the true value is $E[X^2] = \frac{9950}{3} \approx 3316.6 > 2500$.

Lemma: *Let X be a r.v., then $E[X^2] \geq E[X]^2$.*

Proof. Let Y be a r.v. such that $Y = (X - E[X])^2$.

$$\begin{aligned} 0 &\leq E[Y] = E[\,(X - E[X])^2\,] \\ &= E[X^2] - 2E[X \cdot E[X]] + E[X]^2 \\ &= E[X^2] - 2E[X]^2 + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

So,

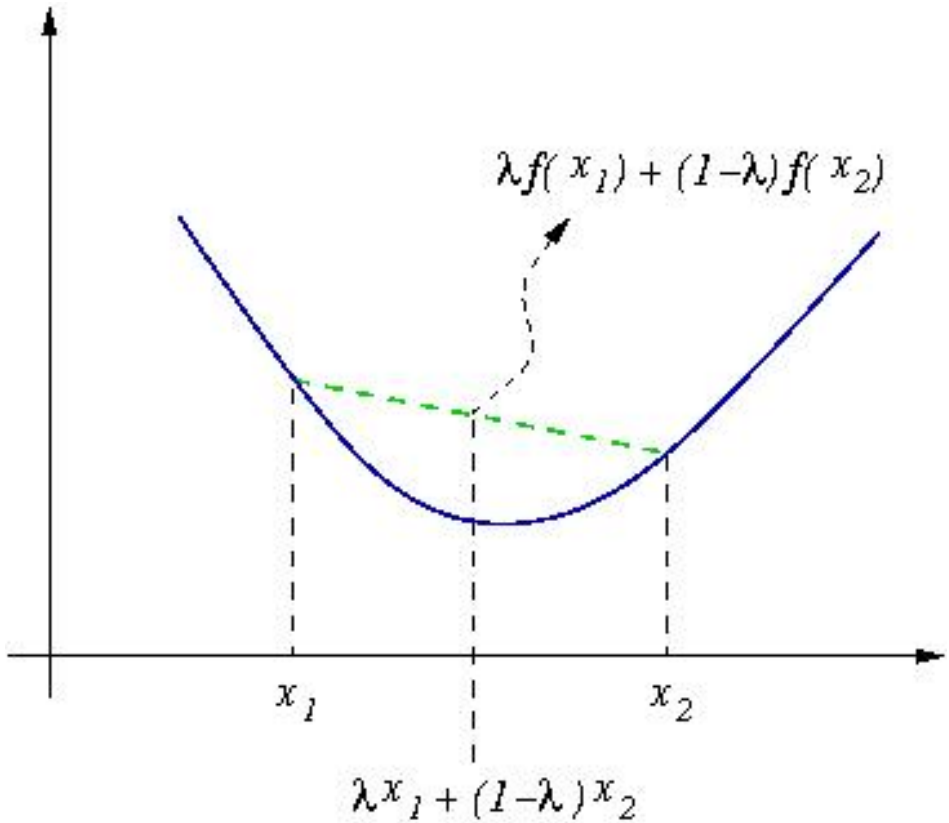
$$E[X^2] \geq E[X]^2$$

□

This is valid for any convex function.

Def.: *A function* $f : \mathbb{R} \to \mathbb{R}$ *is said to be convex if* $\forall x_1, x_2$ *and* $0 \leq \lambda \leq 1$ *we have*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$



Lemma: *If* f *is a twice differentiable function, then* f *is convex if and only if* $f''(x) \geq 0$

Theorem: *(Jensen's Inequality) If f is a convex function, then*

$$E[f(X)] \geq f(E[X]).$$

Proof. We suppose that f has Taylor expansion and let $\mu = E[X]$. By Taylor's Theorem, there exists c such that

$$\begin{aligned} f(X) &= f(\mu) + f'(\mu) \cdot (X - \mu) + f''(c) \cdot \frac{(X-\mu)^2}{2} \\ &\geq f(\mu) + f'(\mu) \cdot (X - \mu) \end{aligned}$$

Applying the expectation in both sides, we have

$$\begin{aligned} E[f(X)] &\geq E[f(\mu) + f'(\mu)(X - \mu)] \\ &= E[f(\mu)] + f'(\mu) \cdot (E[X] - \mu) \\ &= f(\mu) \\ &= f(E[X]) \end{aligned}$$

□

Bernoulli and Binomial Random Variables

Consider an experiment that has probability of success p and fail of $1-p$.

Let Y be a r.v. such that

$$Y = \begin{cases} 1 & \text{if the experiment has success} \\ 0 & \text{otherwise} \end{cases}$$

Then, Y is said to be a Bernoulli r.v. or an indicator r.v.

Lemma: *If Y is a Bernoulli r.v. with $\Pr(Y=1)=p$, then $E[Y]=p$.*

Consider a sequence of n independent experiments, each one with probability of success equal to p.

If X is the number of success in the n experiments, we say that X has binomial distribution.

Def.: *A binomial r.v. X with parameters n and p, denoted by $B(n,p)$ is defined by the following probability distribution with $j = 0, 1, \ldots, n$:*

$$\Pr(X = j) = \binom{n}{j} p^j (1-p)^{n-j}$$

Exercise: *Show that* $\sum_{j=0}^{n} \Pr(X = j) = 1$.

Lemma: *If X is a binomial r.v. $B(n,p)$, then $E[X] = n \cdot p$.*

Proof.

$$\begin{aligned} E[X] &= \sum_{j=0}^{n} j \cdot \binom{n}{j} p^j (1-p)^{n-j} \\ &\vdots \quad \text{(Exercise)} \\ &= n \cdot p \end{aligned}$$

□

Suggestion: use the fact that $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \cdot y^{n-k}$.

Another proof:

Lemma: *If X is a binomial r.v. $B(n,p)$, then $E[X] = n \cdot p$.*

Proof. Let X_i be a bernoulli r.v. of the i-th experiment.

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

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} E[X_i] \\ &= n \cdot p \end{aligned}$$

□

Conditional Expectation

Def.:

$$E[Y|Z=z] = \sum_{y} y \cdot \Pr(Y=y|Z=z),$$

where the summation is over all values y that Y can assume.

Example: *Consider the example of r.v. X that is the sum of two dice, where* $X = X_1 + X_2$.

$$E[X|X_1=2] = \sum_{x} x \cdot \Pr(X=x|X_1=2) = \sum_{x=3}^{8} x \cdot \frac{1}{6} = \frac{11}{2}$$

Example: *Consider the example of a r.v. X that is the sum of two dice, where $X = X_1 + X_2$.*

$$
\begin{aligned}
E[X_1|X=5] &= \sum_{x=1}^{4} x \cdot \Pr(X_1 = x | X = 5) \\
&= \sum_{x=1}^{4} x \cdot \frac{\Pr(X_1 = x \cap X = 5)}{\Pr(X=5)} \\
&= \sum_{x=1}^{4} x \cdot \frac{1/36}{4/36} = \frac{5}{2}
\end{aligned}
$$

Lemma: *For any r.v. X and Y,*

$$E[X] = \sum_y \Pr(Y = y) E[X|Y = y],$$

where the summation is over all possible values of Y and suppose that all expectations exist.

Proof.

$$\begin{aligned}
\sum_y \Pr(Y = y) E[X|Y = y] &= \sum_y \Pr(Y = y) \cdot \sum_x x \cdot \Pr(X = x|Y = y) \\
&= \sum_x \sum_y x \cdot \Pr(X = x|Y = y) \cdot \Pr(Y = y) \\
&= \sum_x x \cdot \sum_y \Pr(X = x \cap Y = y) \\
&= \sum_x x \cdot \Pr(X = x) \;=\; E[X]
\end{aligned}$$

□

The linearity of expectation also extend to conditional expectations.

Lemma: *For any finite collection of r.v. $X_1, \ldots, X_n$ with finite expectations and any r.v. Y,*

$$E[\sum_{i=1}^{n} X_i | Y = y] = \sum_{i=1}^{n} E[X_i | Y = y]$$

Proof. Exercise. □

Def.: *The expression $E[Y|Z]$ is a r.v. $f(Z)$ with value $E[Y|Z=z]$, when $Z=z$.*

Example: *Consider the example of the sum of two dice, with $X = X_1 + X_2$.*

$$\begin{aligned} E[X|X_1] &= \sum_x x \cdot \Pr(X = x|X_1) \\ &= \sum_{x=X_1+1}^{X_1+6} x \cdot \frac{1}{6} \\ &= X_1 + \frac{7}{2} \end{aligned}$$

As $E[X|X_1]$ is a r.v., it makes sense to calculate $E[E[X|X_1]]$.

Example: *In the previous example:*

$$E[E[X|X_1]] = E[X_1 + \frac{7}{2}] = \frac{1}{6}\frac{(1+6)6}{2} + \frac{7}{2} = 7 = E[X]$$

Theorem: *If Y and Z are r.v. then, $E[Y] = E[E[Y|Z]]$.*

Proof. Let $E[Y|Z]$ be a function $f(Z)$ that has value $E[Y|Z=z]$ when $Z=z$,

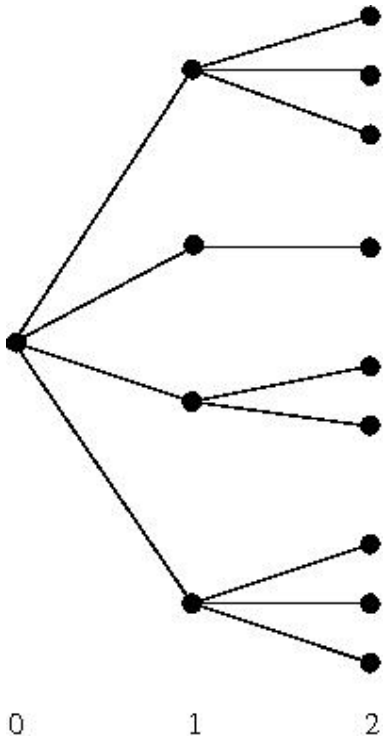
$$\begin{aligned} E[E[Y|Z]] &= \sum_z E[Y|Z=z] \cdot \Pr(Z=z) \\ &= E[Y] \end{aligned}$$

□

Example: *Consider the following probabilistic recursive algorithm:*

Algorithm $R(n,p)$

1. *Repeat n times*
2. *with probability p call* $R(n,p)$.



Generations of calls of $R(n,p)$

What is the number of recursive calls ?

Let Y_i be the total number of generations i.
Suppose that we know the number of calls y_{i-1} in generation $i-1$.
Let Z_k, for $k = 1, \ldots, y_{i-1}$, the number of calls given in the k-th call of generation $i-1$.
Each Z_k is a binomial r.v.
We show that

$$E[Y_i | Y_{i-1} = y_{i-1}] = y_{i-1} \cdot n \cdot p$$

If $y_{i-1} = 0$ the equality is trivially valid.

Now we show that $E[Y_i|Y_{i-1} = y_{i-1}] = y_{i-1} \cdot n \cdot p$ when $y_{i-1} > 0$.

$$\begin{aligned}
E[Y_i|Y_{i-1} = y_{i-1}] &= E[\sum_{k=1}^{y_{i-1}} Z_k | Y_{i-1} = y_{i-1}] \\
&= \sum_{k=1}^{y_{i-1}} E[Z_k|Y_{i-1} = y_{i-1}] \\
&= \sum_{k=1}^{y_{i-1}} \sum_{j\geq 0} j \cdot \Pr(Z_k = j|Y_{i-1} = y_{i-1}) \\
&= \sum_{k=1}^{y_{i-1}} \sum_{j\geq 0} j \cdot \Pr(Z_k = j) \\
&= \sum_{k=1}^{y_{i-1}} E[Z_k] = \sum_{k=1}^{y_{i-1}} n \cdot p = y_{i-1} \cdot n \cdot p
\end{aligned}$$

So,

$$
\begin{aligned}
E[Y_i] &= E[E[Y_i|Y_{i-1}]] \\
&= \sum_{y_{i-1}\geq 0} E[Y_i|Y_{i-1}=y_{i-1}]\cdot \Pr(Y_{i-1}=y_{i-1}) \\
&= \sum_{y_{i-1}\geq 0} y_{i-1}\cdot n\cdot p\cdot \Pr(Y_{i-1}=y_{i-1}) \\
&= n\cdot p\cdot \sum_{y_{i-1}\geq 0} y_{i-1}\cdot \Pr(Y_{i-1}=y_{i-1}) \\
&= n\cdot p\cdot E[Y_{i-1}] \\
&= (n\cdot p)^i
\end{aligned}
$$

and therefore

$$
\begin{aligned}
E[\sum_{i\geq 0} Y_i] &= \sum_{i\geq 0} E[Y_i] \\
&= \sum_{i\geq 0} (n\cdot p)^i \\
&= \begin{cases} \infty & \text{if } n\cdot p \geq 1 \\ \frac{1}{1-n\cdot p} & \text{if } n\cdot p < 1 \end{cases}
\end{aligned}
$$

Geometric Distribution

Def.: *A r.v. X is said to be geometric with parameter p if has distribution*

$$\Pr(X = n) = (1-p)^{n-1} \cdot p.$$

I.e., the probability to toss a coin $n-1$ times with tail (or fail) and in the n-th obtain head (success).

Geometric random variables are said to be memoryless.

Lemma:

$$\Pr(X = n + k | X > k) = \Pr(X = n).$$

Proof.

$$\begin{aligned}
\Pr(X = n + k | X > k) &= \frac{\Pr((X = n + k) \cap (X > k))}{\Pr(X > k)} \\
&= \frac{\Pr(X = n + k)}{\Pr(X > k)} \\
&= \frac{(1-p)^{n+k-1} \cdot p}{\sum_{i=k}^{\infty} (1-p)^i \cdot p} \\
&= \frac{(1-p)^{n+k-1} \cdot p}{(1-p)^k} \\
&= (1-p)^{n-1} \cdot p = \Pr(X = n)
\end{aligned}$$

□

Lemma: *Let X be a d.r.v. that have only non-negative integers. Then*

$$E[X] = \sum_{i=1}^{\infty} \Pr(X \geq i).$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{\infty} \Pr(X \geq i) &= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \Pr(X = j) \\
&= \sum_{j=1}^{\infty} \sum_{i=1}^{j} \Pr(X = j) \\
&= \sum_{j=1}^{\infty} j \cdot \Pr(X = j) \\
&= E[X]
\end{aligned}$$

□

Corollary: *If X is a geometric r.v. with parameter p, then*

$$E[X] = \frac{1}{p}.$$

Proof. Note that if X is a geometric r.v. then

$$\Pr(X \geq i) = \sum_{n=i}^{\infty} (1-p)^{n-1} p = (1-p)^{i-1}.$$

Therefore,

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty} \Pr(X \geq i) = \sum_{i=1}^{\infty} (1-p)^{i-1} \\ &= \frac{1}{1-(1-p)} = \frac{1}{p} \end{aligned}$$

□

Example: *In a cereal box, there is one coupon of a total of n different coupons. How many boxes of cereal we need to by to have at least one different coupon of each type?*

Let X be the number of boxes that you have to buy.

Let X_i be the number of boxes you bought while you have $i-1$ different coupons.

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

If we have $i-1$ different coupons, the probability to obtain a new different coupon is $1-\frac{i-1}{n}=\frac{n-i+1}{n}$.
As X_i is a geometric r.v., we have

$$E[X_i]=\frac{1}{p}=\frac{1}{(n-i+1)/n}=\frac{n}{n-i+1}.$$

So,

$$\begin{aligned}
E[X] &= \sum_{i=1}^{n} E[X_i] \\
&= \sum_{i=1}^{n} \frac{n}{n-i+1} \\
&= n\cdot\left(\frac{1}{n}+\frac{1}{n-1}+\cdots+\frac{1}{1}\right) \\
&= n\cdot H_n = n\ln n+\Theta(1)
\end{aligned}$$

Expected Time of QuickSort

Aplication: *Algorithm QuickSort*(S),
where $S = (x_1, \ldots, x_n)$ *distinct elements*

1. *If* $n = 1$ *or* $n = 0$ *return* S.
2. *else*
3. *Choose a pivo* $x \in S$ *uniformly*
4. $S_1 \leftarrow (y \in S : \ y < x)$.
5. $S_2 \leftarrow (y \in S : \ y > x)$.
6. *Sort* S_1 *using QuickSort*(S_1).
7. *Sort* S_2 *using QuickSort*(S_2).
8. *Return* $(S_1 \| x \| S_2)$.

Theorem: *The expected number of comparisons made by QuickSort is* $2n \ln n + O(n)$.

Proof. Let $(y_1 < y_2 < \ldots < y_n)$ be the sorted elements of S.

For $i < j$ let X_{ij} r.v. that indicate that y_i was compared with y_j.

So, the number of comparisons X is

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}.$$

We have

$$\begin{aligned} E[X_{ij}] &= 0 \cdot \Pr(y_i \text{ is not compared to } y_j) + \\ &\quad 1 \cdot \Pr(y_i \text{ is compared to } y_j) \\ &= \Pr(y_i \text{ ser comparado com } y_j) \end{aligned}$$

Consider the choice of the pivot and comparison between y_i and y_j:

$$\underbrace{y_1, y_2, \ldots, y_{i-1}}_{\text{postpone}}, \mathbf{y_i}, \underbrace{y_{i+1}, \ldots, y_{j-1}}_{\text{not comp.}}, \mathbf{y_j}, \underbrace{y_{j+1}, \ldots, y_n}_{\text{postpone}}$$

So,

$$\Pr(y_i \text{ is compared with } y_j) = \frac{2}{j-i+1}.$$

So,

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[X_{ij}] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} = \sum_{k=2}^{n} \sum_{i=1}^{n+1-k} \frac{2}{k} \\
&= \sum_{k=2}^{n} (n+1-k)\frac{2}{k} \\
&= 2 \cdot (n+1) \cdot \sum_{k=1}^{n} \frac{1}{k} - 2 \cdot (n-1) - 2 \cdot (n+1) \\
&= 2 \cdot (n+1) \cdot H_n - 4n \\
&= 2 \cdot n \cdot \ln n + \Theta(n)
\end{aligned}
$$

□

Theorem: *Consider the deterministic QuickSort that uses the first element as pivot. So, the expected number of comparisons for a uniformly chosen input between all possible permutations is* $2n \ln n + O(n)$.

Proof. The proof is basically the same as done for probabilistic *QuickSort*. Exercise. □

Moments and Deviations

Techniques to bound the tail distribution — probability that a r.v. obtain a value distant from the expectation.

MARKOV INEQUALITY

Theorem: *Let X be a r.v. that have non-negative values. Then,*

$$\Pr(X \geq a) \leq \frac{E[X]}{a} \qquad \forall a > 0.$$

Proof. For $a > 0$, let I r.v.

$$I = \begin{cases} 1 & \text{if } X \geq a, \\ 0 & \text{otherwise,} \end{cases}$$

As $X \geq 0$ we have $I \leq \frac{X}{a}$ and as I is Indicator r.v.,

$$E[I] = \Pr(I = 1) = \Pr(X \geq a)$$

That is,

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

□

Corollary: *Let X be a r.v. that have positive values. Then,*

$$\Pr(X \geq \lambda E[X]) \leq \frac{1}{\lambda} \qquad \forall \lambda > 0.$$

Proof. Let $a = \lambda E[X]$.

$$\begin{aligned} \Pr(X \geq \lambda E[X]) = \Pr(X \geq a) &\leq \frac{E[X]}{a} \\ &= \frac{E[X]}{\lambda E[X]} \\ &= \frac{1}{\lambda} \end{aligned}$$

□

Example: *Suppose we toss n fair coins and let X be the number of heads. Use the Markov Inequality to bound the probability to obtain at least* $\frac{3n}{4}$ *heads.*

$$\text{Let } X_i = \begin{cases} 1 & \text{if } i\text{-th coin is head.,} \\ 0 & \text{otherwise.} \end{cases}$$

We have $X = \sum_{i=1}^{n} X_i$. We have that $E[X] = \frac{n}{2}$. So,

$$\begin{aligned} \Pr(X \geq \frac{3n}{4}) &= \Pr(X \geq \frac{3}{2} \cdot \frac{n}{2}) \\ &= \Pr(X \geq \frac{3}{2} \cdot E[X]) \\ &\leq \frac{1}{3/2} = \frac{2}{3} \end{aligned}$$

Variance and Moments of a Random Variable

- Markov Inequality is the best we can do when we know only the expectation.
- But we can obtain better results if we have more information about the distribution of the r.v.

Def.: *The k-th moment of r.v. X is* $E[X^k]$.

So, $E[X]$ is the first moment.

Def.: *The variance of a r.v. X is defined as:*

$$\text{var}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

Def.: *The Standard Deviation of r.v. X is defined as*

$$\sigma(X) = \sqrt{\text{var}[X]}.$$

Example:

- *If X is constant:*

$$\text{var}(X) = E[(X - E[X])^2] = 0$$

and

$$\sigma(X) = 0.$$

Example:

- *Let X be a r.v. such that* $X = \begin{cases} k \cdot E[X] & \text{with probability } \frac{1}{k} \\ 0 & \text{with probability } 1 - \frac{1}{k} \end{cases}$

  Calculate the variance and standard deviation

$$\begin{aligned} \text{var}(X) &= E[(X - E[X])^2] \\ &= \frac{1}{k}(k \cdot E[X] - E[X])^2 + (1 - \frac{1}{k}) \cdot (0 - E[X])^2 \\ &= \frac{(k-1)^2 \cdot E[X]^2}{k} + \frac{(k-1)}{k} \cdot E[X]^2 \\ &= (\frac{(k-1)^2 + (k-1)}{k}) \cdot E[X]^2 \\ &= (k-1) \cdot E[X]^2 \end{aligned}$$

and

$$\sigma(X) = \sqrt{(k-1) \cdot E[X]^2} = \sqrt{k-1} \cdot E[X].$$

Def.: *The covariance of two variables X and Y is*

$$\operatorname{cov}(X,Y) = E[(X - E[X]) \cdot (Y - E[Y])].$$

Theorem: *For r.v. X and Y we have*

$$\operatorname{var}[X+Y] = \operatorname{var}[X] + \operatorname{var}[Y] + 2 \cdot \operatorname{cov}(X,Y).$$

Proof.

$$\begin{aligned}
\operatorname{var}[X+Y] &= E[((X+Y)-E[X+Y])^2] \\
&= E[((X+Y-E[X]-E[Y])^2] \\
&= E[(X-E[X])^2+(Y-E[Y])^2+2(X-E[X])(Y-E[Y])] \\
&= E[(X-E[X])^2]+E[(Y-E[Y])^2]+2E[(X-E[X])(Y-E[Y])] \\
&= \operatorname{var}(X)+\operatorname{var}(Y)+2\operatorname{cov}(X,Y)
\end{aligned}$$

□

Exercise: *Suppose X, Y, W, V, X_i are random variables and a, b, c and d constants, prove that*

- $\operatorname{cov}(X, a) = 0$
- $\operatorname{cov}(X, X) = \operatorname{var}(X)$
- $\operatorname{cov}(X, Y) = \operatorname{cov}(Y, X)$
- $\operatorname{cov}(aX, bY) = a\,b\operatorname{cov}(X, Y)$
- $\operatorname{cov}(X + a, Y + b) = \operatorname{cov}(X, Y)$
- $\operatorname{cov}(aX + bY, cW + dV) =$
  $a\,c\operatorname{cov}(X, W) + a\,d\operatorname{cov}(X, V) + b\,c\operatorname{cov}(Y, W) + b\,d\operatorname{cov}(Y, V)$
- $\operatorname{var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \operatorname{var}(X_i) + 2 \cdot \sum_{i=1}^{n} \sum_{j>i} \operatorname{cov}(X_i, Y_j)$.

Theorem: *If X and Y are independent r.v. then*

$$E[X \cdot Y] = E[X] \cdot E[Y].$$

Proof.

$$\begin{aligned}
E[X \cdot Y] &= \sum_i \sum_j (i \cdot j) \cdot \Pr(X = i \cap Y = j) \\
&= \sum_i \sum_j (i \cdot j) \cdot \Pr(X = i) \cdot \Pr(Y = j) \\
&= \left(\sum_i i \cdot \Pr(X = i) \right) \cdot \left(\sum_j j \cdot \Pr(Y = j) \right) \\
&= E[X] \cdot E[Y]
\end{aligned}$$

□

Corollary: *If X and Y are independent r.v. then*

$$\operatorname{cov}(X, Y) = 0$$

and

$$\operatorname{var}[X + Y] = \operatorname{var}[X] + \operatorname{var}[Y]$$

Proof.

$$\begin{aligned}
\operatorname{cov}(X, Y) &= E[(X - E[X]) \cdot (Y - E[Y])] \\
&= E[X - E[X]] \cdot E[Y - E[Y]] \\
&= (E[X] - E[X]) \cdot (E[Y] - E[Y])] \\
&= 0
\end{aligned}$$

□

Example: *If X and Y are non-independent r.v. then the expectation of the product can be different from the product of expectations.*

Suppose that X and Y are coins with value 1 for head and 0 for tail, and suppose that X and Y are tossed and they show the same value (they are welded).

I.e.,

$$E[X] = E[Y] = \frac{1}{2}$$

and

$$E[X \cdot Y] = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}.$$

So,

$$E[X \cdot Y] \neq E[X] \cdot E[Y]$$

Example: *Variance of a Bernoulli variable.*

Let X be a Bernoulli r.v. with parameter p.

Then,

$$\begin{aligned}
\text{var}[X] &= E[(X - E[X])^2] \\
&= E[(X - p)^2] \\
&= (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) \\
&= p - p^2 \\
&= p \cdot (1-p)
\end{aligned}$$

Example: *Variance of a Binomial variable.*

Let X be a Binomial r.v. with parameters n and p.
Then, X can be considered as $X = \sum_{i=1}^{n} X_i$, where X_i is a Bernoulli r.v. with parameter p.
Note that $X_1, \ldots, X_n$ are independent.
So,

$$\begin{aligned} \text{var}[X] &= \sum_{i=1}^{n} \text{var}[X_i] \\ &= \sum_{i=1}^{n} p \cdot (1-p) \\ &= n \cdot p \cdot (1-p) \end{aligned}$$

The variance of a Binomial r.v. can also be calculated obtaining $E[X^2]$.

Example: *If X is a Binomial r.v.,*

$$\begin{aligned} E[X^2] &= \sum_{j=0}^{n} j^2 \binom{n}{j} p^j (1-p)^{n-j} \\ &\vdots \quad (exercise) \\ &= n(n-1)p^2 + np \end{aligned}$$

Chebyshev's Inequality

Using the variance, we can obtain a better bound than using only Markov Inequality.

Theorem: *For any* $a > 0$

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

Proof. Note that

$$\Pr(|X - E[X]| \geq a) = \Pr(|X - E[X]|^2 \geq a^2)$$

From Markov Inequality,

$$\begin{aligned} \Pr(|X - E[X]| \geq a) &= \Pr(|X - E[X]|^2 \geq a^2) \\ &\leq \frac{E[|X - E[X]|^2]}{a^2} \\ &= \frac{\text{var}[X]}{a^2} \end{aligned}$$

□

Corollary: *For any* $t > 0$

$$\Pr(|X - E[X]| \geq t \cdot \sigma[X]) \leq \frac{1}{t^2}$$

and

$$\Pr(|X - E[X]| \geq t \cdot E[X]) \leq \frac{\text{var}[X]}{t^2 \cdot E[X]^2}$$

Proof. We only need to replace a in the previous theorem. □

Example: *Consider the example of tossing n coins and bound the probability to obtain at least* $\frac{3n}{4}$ *heads.*

Applying the Chebyshev's Inequality:

$$\begin{aligned}
\Pr(X \geq \frac{3n}{4}) &\leq \Pr(|X - \frac{n}{2}| \geq \frac{n}{4}) \\
&= \Pr(|X - E[X]| \geq \frac{n}{4}) \\
&\leq \frac{\text{var}[X]}{(n/4)^2} \\
&= \frac{n \cdot \frac{1}{2}(1 - \frac{1}{2})}{(n/4)^2} \\
&= \frac{4}{n}
\end{aligned}$$

This bound can be improved to $\frac{2}{n}$. Why ?

Better than the bound of $\frac{2}{3}$ obtained using only Markov Inequality.

Lemma: *The variance of a geometric r.v. Y with parameter p is $\frac{1-p}{p^2}$.*

Proof. As $\text{var}[Y] = E[Y^2] - E[Y]^2$, we can calculate $E[Y^2]$.
Let Y_1 be the r.v. of the first toss.

$$\begin{aligned}
E[Y^2] &= \sum_{i=0}^{1} \Pr(Y_1 = i) \cdot E[Y^2 | Y_1 = i] \\
&= (1-p) \cdot E[Y^2 | Y_1 = 0] + p \cdot E[Y^2 | Y_1 = 1] \\
&= (1-p) \cdot E[(Z+1)^2] + p \cdot 1, \quad \text{where } Z \text{ is a geom.r.v. with param. } p \\
&= (1-p) \cdot (E[Z^2] + 2 \cdot E[Z] + 1) + p \cdot 1.
\end{aligned}$$

Using the fact that a geometric r.v. is memoryless, we have

$$E[Y^2] = (1-p) \cdot (E[Y^2] + 2 \cdot E[Y] + 1) + p$$

$$E[Y^2] = (1-p) \cdot (E[Y^2] + 2 \cdot E[Y] + 1) + p$$

I.e.,

$$(1-(1-p))E[Y^2] = 2 \cdot (1-p) \cdot E[Y] + (1-p) + p$$

Isolating $E[Y^2]$ we have

$$E[Y^2] = \frac{2-2 \cdot p}{p^2} + \frac{p}{p^2} = \frac{2-p}{p^2}$$

Therefore,

$$\begin{aligned} \text{var}[Y] &= E[Y^2] - (E[Y])^2 \\ &= \frac{2-p}{p^2} - \frac{1}{p^2} \\ &= \frac{1-p}{p^2} \end{aligned}$$

□

Example: *Consider the coupon collector problem.*

Let X be the number of boxes to obtain the n coupons.
By the Markov Inequality, we have

$$\Pr(X \geq 2 \cdot n \cdot H_n) \leq \frac{1}{2}.$$

Now, we use the Chebyshev's Inequality.
Let X_i be the number of boxes bought to obtain the i-th different coupon, having $i-1$ different coupons.
Let $X = \sum_{i=1}^{n} X_i$. Note that X_i is a geometric r.v. with parameter $\frac{n-(i-1)}{n}$.
The variables X_i are independent. So,

$$\text{var}[X] = \sum_{i=1}^{n} \text{var}[X_i].$$

$$
\begin{aligned}
\operatorname{var}[X] &= \sum_{i=1}^{n} \operatorname{var}[X_i]. \\
&= \sum_{i=1}^{n} \frac{1-p_i}{(p_i)^2}, \qquad \text{where } p_i = \frac{n-i+1}{n} \\
&\leq \sum_{i=1}^{n} \frac{1}{(p_i)^2} = \sum_{i=1}^{n} \left(\frac{n}{n-i+1}\right)^2 \\
&= n^2 \cdot \sum_{i=1}^{n} \frac{1}{i^2} \\
&\leq n^2 \cdot \frac{\pi^2}{6}, \qquad \text{because} \quad \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}
\end{aligned}
$$

Using Chebyshev's Inequality:

$$\begin{aligned}
\Pr(X \geq 2 \cdot n \cdot H_n) &\leq \Pr(|X - n \cdot H_n| \geq n \cdot H_n) \\
&\leq \frac{\frac{n^2 \cdot \pi^2}{6}}{(n \cdot H_n)^2} \\
&= \frac{\pi^2}{6 \cdot (H_n)^2} \\
&= \Theta(\frac{1}{\ln^2 n}).
\end{aligned}$$

Better than the bound $\frac{1}{2}$ obtained with Markov Inequality.

Example: *It is possible to obtain a better bound than the one obtained with Chebyshev's Inequality for the coupon collector problem.*

Consider the different coupons as the set $\{1, 2, \ldots, n\}$.
Let E_i be the event to not obtain the i-th coupon after $n \ln n + \alpha n$ boxes. So, the event $E = \cup_{i=1}^{n} E_i$ is the event to not obtain some different coupon after $n \ln n + \alpha n$ boxes.

$$\begin{aligned}
\Pr(E_i) &= \left(1 - \frac{1}{n}\right)^{n \ln n + \alpha n} \\
&= \left[\left(1 - \frac{1}{n}\right)^{n}\right]^{\ln n + \alpha} \\
&\leq e^{-(\ln n + \alpha)} \\
&\leq \frac{1}{n \cdot e^{\alpha}}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr(E) &= \Pr(\cup_{i=1}^{n} E_i) \\
&\leq \sum_{i=1}^{n} \Pr(E_i) \\
&\leq \sum_{i=1}^{n} \frac{1}{n \cdot e^{\alpha}} \\
&= \frac{1}{e^{\alpha}}
\end{aligned}$$

Placing $\alpha = \ln n$, we have $\Pr(E) \leq \frac{1}{n}$. That is better then the bound obtained with Chebyshev's Inequality.

ALGORITHM TO COMPUTE THE MEDIAN

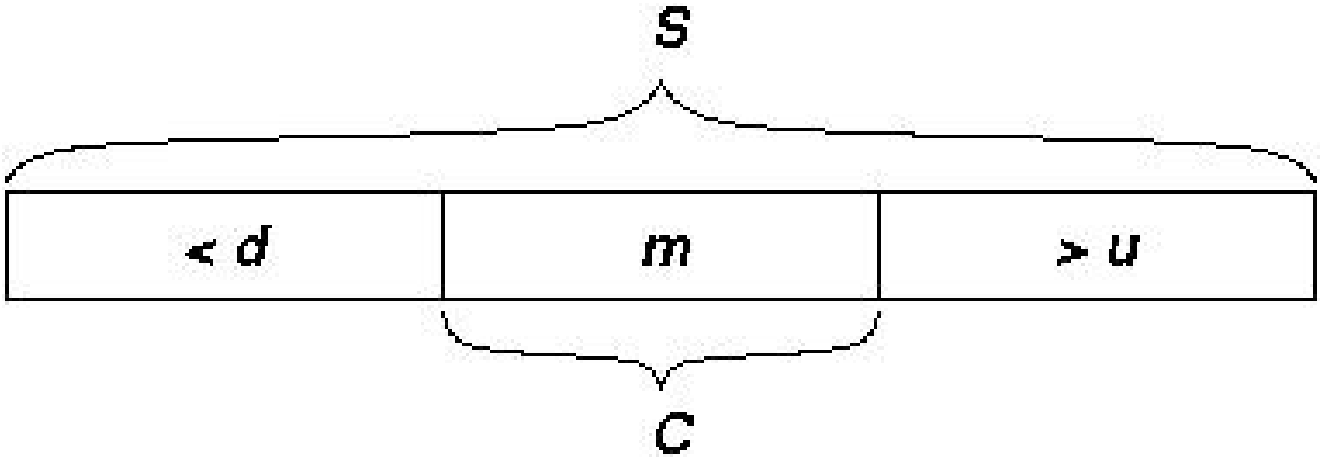
Problem: *Given a set S with n elements and total order, find $m \in S$ such that $\lfloor n/2 \rfloor$ elements are smaller than or equal to m and $\lfloor n/2 \rfloor + 1$ are larger than or equal to m.*

- Trivial algorithm: Sort and pick the median: $O(n \lg n)$.
- There exists a linear time algorithm, but complicated.
- We will se a randomized linear time algorithm that is correct with high probability.

Idea:

Given set S with n distinct elements, n odd.

1. Find $d, u \in S$ such that $d \leq m \leq u$
2. Let $C = \{s \in S : \ d \leq s \leq u\}$, where $|C| = o(n/\lg n)$.



3. Sort C and return the median m of S.

Algorithm RMedian(S), where $|S| = n$.

1. Let R be a multi-set with $\lfloor n^{3/4} \rfloor$ elements of S chosen uniformly at random, with replacement.
2. Sort R.
3. Let d the $(\frac{1}{2}n^{3/4} - \sqrt{n})$-th element of R.
4. Let u the $(\frac{1}{2}n^{3/4} + \sqrt{n})$-th element of R.
5. Let $\begin{aligned} C^- &= \{s \in S : s < d\} \\ C &= \{s \in S : d \leq s \leq u\} \\ C^+ &= \{s \in S : s > d\} \end{aligned}$
6. If $|C^-| > \frac{n}{2}$ or $|C^+| > \frac{n}{2}$ then return *FAIL*.
7. If $|C| > 4 \cdot n^{3/4}$ then return *FAIL*.
8. Sort C
9. Return the $(\lfloor \frac{n}{2} \rfloor - |C^-| + 1)$-th element of C.

Idea:

Let

$Y_1 = |\{r \in R : r \leq m\}|$ (number of elements in the sample $\leq$ median)

$Y_2 = |\{r \in R : r \geq m\}|$ (number of elements in the sample $\geq$ median)

As R has $n^{3/4}$ elements, we have

$$E[Y_1] \approx \tfrac{1}{2}n^{3/4} \quad \text{and} \quad E[Y_2] \approx \tfrac{1}{2}n^{3/4}$$

When we insert a gap of $\sqrt{n}$, we expect that

$$\Pr(Y_1 < E[Y_1] - \sqrt{n}) \quad \text{is small and}$$

$$\Pr(Y_2 < E[Y_2] - \sqrt{n}) \quad \text{is small.}$$

Analysis of the Algorithm: For convenience, we consider that $\sqrt{n}$ and $n^{3/4}$ are integers.

Theorem: *The algorithm RMedian has linear time complexity and if it does not fail, it returns the correct answer.*

Proof. Exercise. □

Now we show that the probability that the algorithm fail is smaller than $\frac{1}{n^{1/4}}$.

Consider the events E_1, E_2 and E_3 as follows:

$$\begin{aligned} E_1 &= \text{Event}\left(Y_1 = |\{r \in R : r \leq m\}| < \frac{1}{2}n^{3/4} - \sqrt{n}\right) \\ E_2 &= \text{Event}\left(Y_2 = |\{r \in R : r \geq m\}| < \frac{1}{2}n^{3/4} - \sqrt{n}\right) \\ E_3 &= \text{Event}\left(|C| > 4n^{3/4}\right) \end{aligned}$$

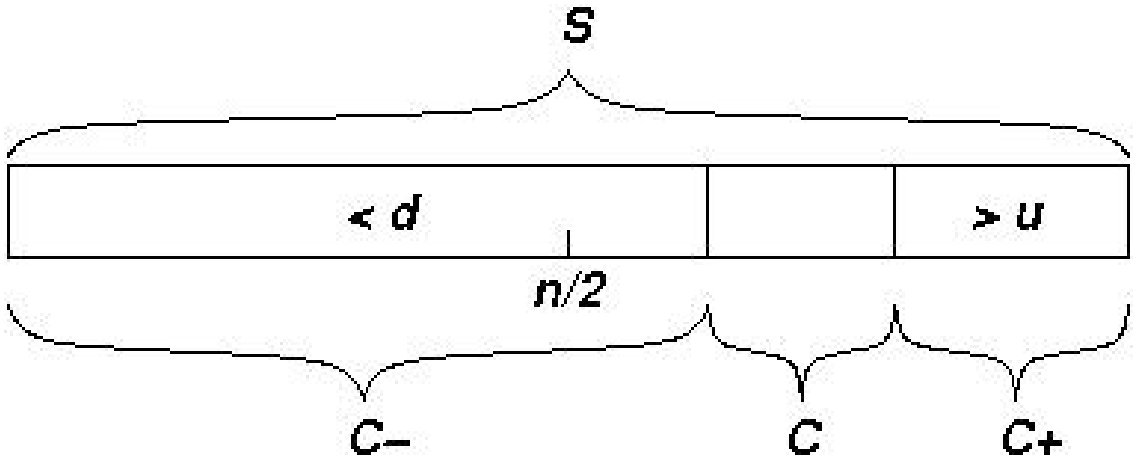
Lemma: *The algorithm RMedian fail if and only if one of the events E_1, E_2 or E_3 occur.*

Proof.

- There is a direct correspondence between event E_3 and stopping case.

We analyse other cases.

Suppose that RMedian fail because $|C^-| > \frac{n}{2}$.



In this case, $m < d$ and as d is the $(\lfloor \frac{1}{2}n^{3/4} - \sqrt{n} \rfloor)$-th element of R we have

$$\begin{aligned} Y_1 &= |\{r \in R : r \leq m\}| \\ &\leq |\{r \in R : r < d\}| \\ &= \left\lfloor \frac{1}{2}n^{3/4} - \sqrt{n} \right\rfloor - 1 \\ &< \frac{1}{2}n^{3/4} - \sqrt{n} \end{aligned}$$

So, $|C^-| > \frac{n}{2} \Rightarrow$ event E_1 occur.

And, if event E_1 occur then

$$Y_1 = |\{r \in R : r \leq m\}| < \frac{1}{2}n^{3/4} - \sqrt{n} = \text{position of } d$$

This implies that $m < d$. Therefore, $|C^-| > \frac{n}{2}$ (fail case in E_1).

The corresponding proof for event E_2 when RMedian fail with $|C^+| > \frac{n}{2}$ is analogous. □

Lemma:

$$\Pr(E_1) \leq \frac{1}{4 \cdot n^{1/4}}.$$

Proof. Let X_i be a Bernoulli r.v. such that

$$X_i = \begin{cases} 1 & \text{if } i\text{-th chosen element in } C \text{ is less than or equal to } m \\ 0 & \text{otherwise} \end{cases}$$

As, there are $\frac{n-1}{2} + 1 = \frac{n+1}{2}$ elements smaller than or equal to m,

$$\Pr(X_i = 1) = \frac{\frac{n+1}{2}}{n} = \frac{1}{2} + \frac{1}{2n}$$

With this, the event E_1 is equivalent to

$$\left(Y_1 = \sum_{i=1}^{n^{3/4}} X_i < \frac{1}{2} n^{3/4} - \sqrt{n} \right).$$

Note that Y_1 is a Binomial r.v. with parameters $n^{3/4}$ and $\frac{1}{2} + \frac{1}{2n}$.

Let us calculate $E[Y_1]$ and $\text{var}[Y_1]$.

$$\begin{aligned} E[Y_1] &= \left(n^{3/4}\right) \cdot \left(\frac{1}{2} + \frac{1}{2n}\right) \\ &= \frac{1}{2} n^{3/4} + \frac{1}{2 \cdot n^{1/4}} \end{aligned}$$

$$
\begin{aligned}
\mathrm{var}[Y_1] &= \left(n^{3/4}\right) \cdot \left(\frac{1}{2} + \frac{1}{2n}\right) \cdot \left(1 - \left(\frac{1}{2} + \frac{1}{2n}\right)\right) \\
&= n^{3/4}\left(\frac{1}{2} + \frac{1}{2n}\right)\left(\frac{1}{2} - \frac{1}{2n}\right) \\
&= \frac{1}{4}n^{3/4} - \frac{1}{4 \cdot n^{5/4}} \\
&< \frac{1}{4}n^{3/4}
\end{aligned}
$$

Applying Chebyshev's Inequality

$$
\begin{aligned}
\Pr(E_1) &= \Pr(Y_1 < \frac{1}{2}n^{3/4} - \sqrt{n}) \\
&= \Pr(\frac{1}{2}n^{3/4} - Y_1 > \sqrt{n}) \\
&\leq \Pr(\left(\frac{1}{2}n^{3/4} + \frac{1}{2 \cdot n^{1/4}}\right) - Y_1 > \sqrt{n}) \\
&= \Pr(E[Y_1] - Y_1 > \sqrt{n}) \\
&\leq \Pr(|Y_1 - E[Y_1]| > \sqrt{n}) \\
&\leq \frac{\mathrm{var}[Y_1]}{(\sqrt{n})^2} \\
&= \frac{\frac{1}{4}n^{3/4}}{n} = \frac{1}{4 \cdot n^{1/4}}
\end{aligned}
$$

Lemma:

$$\Pr(E_3) \leq \frac{1}{2 \cdot n^{1/4}}.$$

Proof. If E_3 occur, then $|C| > 4 \cdot n^{3/4}$.
Then, one of the events occur:

Event E_3': at least $2 \cdot n^{3/4}$ elements of C are larger than m.

Event E_3'': at least $2 \cdot n^{3/4}$ elements of C are smaller than m.

So,

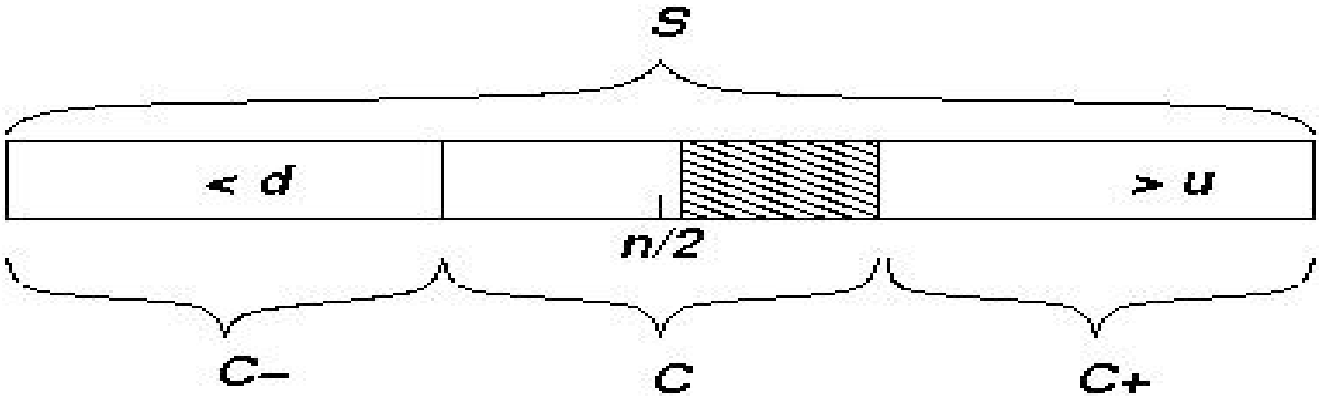
$$E_3 \subseteq E_3' \cup E_3''$$

and

$$\Pr(E_3) \leq \Pr(E_3' \cup E_3'').$$

We will bound $\Pr(E_3')$.

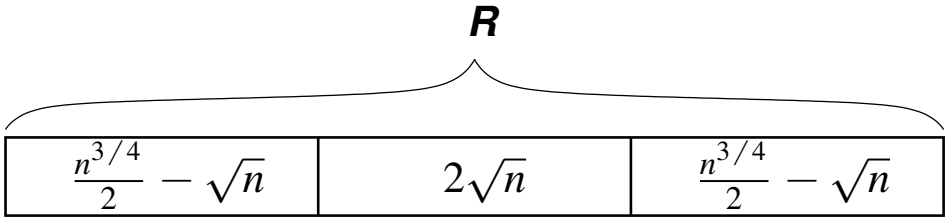
If there are $2 \cdot n^{3/4}$ elements of C larger than m the position of u in S is $\geq \frac{n}{2} + 2 \cdot n^{3/4}$.



I.e.,

$$|C^+| \leq n - (\frac{n}{2} + 2 \cdot n^{3/4}) = \frac{n}{2} - 2 \cdot n^{3/4}.$$

The $\frac{n^{3/4}}{2} - \sqrt{n}$ large elements from R were taken from C^+.



Let X be the number of choices R in C^+. Then,

$$X = \sum_{i=1}^{n^{3/4}} X_i, \text{ where } X_i = \begin{cases} 1 & \text{if } i\text{-th choice is in the} \\ & \frac{n}{2} - 2n^{3/4} \text{ largest elements of } S \\ 0 & \text{otherwise} \end{cases}$$

As X is a Binomial r.v. with parameters $n^{3/4}$ and p, where

$$p = \left(\frac{\frac{n}{2} - 2 \cdot n^{3/4}}{n} \right) = \frac{1}{2} - 2 \cdot n^{-1/4}.$$

So,

$$\begin{aligned} E[X] &= n^{3/4} \cdot p \\ &= n^{3/4} \cdot (\frac{1}{2} - 2 \cdot n^{-1/4}) \\ &= \frac{n^{3/4}}{2} - 2 \cdot \sqrt{n}, \end{aligned}$$

$$
\begin{aligned}
\mathrm{var}[X] &= n^{3/4} \cdot p \cdot (1-p) \\
&= n^{3/4} \cdot (\frac{1}{2} - 2 \cdot n^{-1/4}) \cdot (\frac{1}{2} + 2 \cdot n^{-1/4}) \\
&= n^{3/4} (\frac{1}{4} - 4 \cdot n^{-1/2}) \\
&< \frac{n^{3/4}}{4}
\end{aligned}
$$

Applying Chebyshev's inequality,

$$\begin{aligned}
\Pr(E_3') &= \Pr(X \geq \frac{1}{2}n^{3/4} - \sqrt{n}) \\
&= \Pr(X - (\frac{1}{2}n^{3/4} - 2\sqrt{n}) \geq \sqrt{n}) \\
&\leq \Pr(|X - E[X]| \geq \sqrt{n}) \\
&\leq \frac{\mathrm{var}[X]}{(\sqrt{n})^2} \\
&= \frac{n^{3/4}}{4 \cdot n} = \frac{1}{4 \cdot n^{1/4}}
\end{aligned}$$

Analogously,

$$\Pr(E_3'') \leq \frac{1}{4 \cdot n^{1/4}}$$

So,

$$\Pr(E_3) = \Pr(E_3' \cup E_3'') \leq \Pr(E_3') + \Pr(E_3'') = \frac{1}{2 \cdot n^{1/4}}$$

Theorem: *The probability the algorithm fail is smaller than* $\frac{1}{n^{1/4}}$.

Proof.

$$\begin{aligned} \Pr(E_1 \cup E_2 \cup E_3) &\leq \Pr(E_1) + \Pr(E_2) + \Pr(E_3) \\ &\leq \frac{n^{-1/4}}{4} + \frac{n^{-1/4}}{4} + \frac{n^{-1/4}}{2} \\ &= \frac{1}{n^{1/4}} \end{aligned}$$

□

Chernoff Bounds

- Exponentially decreasing bounds
- Derived by using Markov's Inequality on
- Moment Generating Functions that captures all the moments

Def.: *A moment generating function (m.g.f.) of a r.v. X is*

$$M_X(t) = E[e^{tX}].$$

The function $M_X(t)$ captures all moments of X.

Theorem: *Let X be a r.v. with m.g.f. $M_X(t)$. Under the assumption that exchanging expectation and differentiation is valid, for all $n > 1$ we have*

$$E[X^n] = M_X^{(n)}(0),$$

where $M_X^{(n)}(0)$ is the n-th derivative of $M_X(t)$ in $t = 0$.

Proof. Deriving, we have

$$M_X^{(n)}(t) = E[X^n \cdot e^{tX}].$$

Calculating at $t = 0$ we have

$$M_X^{(n)}(0) = E[X^n].$$

□

Lemma: *Given a geometric r.v. X with parameter p, we have*

$$E[X] = \frac{1}{p} \quad e \quad E[X^2] = \frac{2-p}{p^2}.$$

Proof. Proof with m.g.f. We first obtain $M_X(t)$ and then its derivatives.

$$\begin{aligned}
M_X(t) &= E[e^{tX}] = \sum_{k=1}^{\infty} \Pr(X=k) \cdot e^{tk} \\
&= \sum_{k=1}^{\infty} (1-p)^{k-1} \cdot p \cdot e^{tk} \\
&= \frac{p}{1-p} \sum_{k=1}^{\infty} \left((1-p) \cdot e^t\right)^k \\
&= \frac{p}{1-p} \cdot \frac{(1-p) \cdot e^t}{1-(1-p) \cdot e^t} \quad \text{when} \quad (1-p) \cdot e^t < 1
\end{aligned}$$

$$\begin{aligned} M_X(t) &= \frac{p}{1-p} \cdot \frac{(1-p)\cdot e^t}{1-(1-p)\cdot e^t} \\ &= \frac{p}{1-p} \cdot \left(\frac{1}{1-(1-p)e^t} - 1 \right) \\ &= \frac{p}{1-p} \cdot \left((1-(1-p)e^t)^{-1} - 1 \right) \end{aligned}$$

Obtaining the first and second derivatives, we have

$$\begin{aligned} M_X^{(1)}(t) &= \frac{p}{1-p}(-1)(1-(1-p)e^t)^{-2} \cdot (-(1-p)e^t) \\ &= p(1-(1-p)e^t)^{-2} \cdot e^t. \end{aligned}$$

and

$$M_X^{(2)}(t) = 2p(1-p)(1-(1-p)e^t)^{-3}e^{2t} + p(1-(1-p)e^t)^{-2}e^t.$$

So,

$$\begin{aligned} E[X] &= M_X^{(1)}(0) \\ &= p(1-(1-p)e^0)^{-2} \cdot e^0 \\ &= p(1-1-p)^{-2} \\ &= \frac{1}{p} \end{aligned}$$

and

$$\begin{aligned} E[X^2] &= M_X^{(2)}(0) \\ &= 2p(1-p)(1-(1-p)e^0)^{-3}e^0 + p(1-(1-p)e^0)^{-2} \cdot e^0 \\ &= \frac{2-p}{p^2} \end{aligned}$$

□

Theorem: *If X and Y are r.v. such that $M_X(t) = M_Y(t)$, for any $t \in (-\delta, \delta)$ and some $\delta > 0$, then X and Y has the same distribution.*

Proof outside of the scope of the course.

Theorem: *If X and Y are independent r.v. then*

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$$

Proof.

$$\begin{aligned} M_{X+Y}(t) &= E[e^{t(X+Y)}] \\ &= E[e^{tX} \cdot e^{tY}] \\ &= E[e^{tX}] \cdot E[e^{tY}] \\ &= M_X(t) \cdot M_Y(t) \end{aligned}$$

□

Generalization for sum of several independent r.v. is straightforward.

Deriving Chernoff Bounds

Theorem: *If X is a r.v. and $t > 0$, then*

$$\Pr(X \geq a) \leq \min_{t>0} \frac{E[e^{tX}]}{e^{ta}}.$$

Proof.

$$\begin{aligned} \Pr(X \geq a) &= \Pr(tX \geq ta) \\ &= \Pr(e^{tX} \geq e^{ta}) \\ &\leq \frac{E[e^{tX}]}{e^{ta}} \qquad \text{applying Markov Inequality} \end{aligned}$$

As the above inequality is valid for any $t > 0$, we have

$$\Pr(X \geq a) \leq \min_{t>0} \frac{E[e^{tX}]}{e^{ta}}.$$

Deriving Chernoff Bounds

Theorem: *If X is a r.v. and $t < 0$, then*

$$\Pr(X \leq a) \leq \min_{t<0} \frac{E[e^{tX}]}{e^{ta}}.$$

Proof. Let X be a r.v. and $t < 0$

$$\begin{aligned} \Pr(X \leq a) &= \Pr(tX \geq ta) \\ &= \Pr(e^{tX} \geq e^{ta}) \\ &\leq \frac{E[e^{tX}]}{e^{ta}} \qquad \text{applying Markov} \end{aligned}$$

As the above inequality is valid for any $t < 0$, we have

$$\Pr(X \leq a) \leq \min_{t<0} \frac{E[e^{tX}]}{e^{ta}}.$$

□

Bounds obtained from this approach are called Chernoff Bounds.

CHERNOFF BOUNDS FOR SUM OF POISSON TRIALS

Def.: *Poisson Trials: Sequence of independent binary r.v. non-necessarily with the same probability.*

Let $X_1, \ldots, X_n$ be a sequence of poisson trials, with $\Pr(X_i = 1) = p_i$.
Let $X = \sum_{i=1}^{n} X_i$.

We use Chernoff Bounds to calculate the probability of X to deviate by more than $(1 \pm \delta)E[X]$.

Note that

$$\mu = E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p_i.$$

Fact: *For the binary r.v. X_i, we have*

$$M_{X_i}(t) \leq e^{p_i(e^t-1)}.$$

Proof.

$$\begin{aligned} M_{X_i} &= E[e^{tX_i}] \\ &= p_i \cdot e^t + (1-p_i) \cdot e^0 \\ &= 1 + p_i(e^t - 1) \\ &\leq e^{p_i(e^t-1)}. \end{aligned}$$

The last inequality is valid because $1 + y \leq e^y$ for any y. □

Fact: *For the r.v. X, it is valid that*

$$M_X(t) \leq e^{(e^t-1)\mu}$$

Proof.

$$\begin{aligned} M_X(t) &= M_{X_1+\cdots+X_n}(t) \\ &= \prod_{i=1}^n M_{X_i}(t) \\ &\leq \prod_{i=1}^n e^{p_i(e^t-1)} \\ &= e^{\sum_{i=1}^n p_i(e^t-1)} \\ &= e^{(e^t-1)\mu} \end{aligned}$$

□

Theorem: *Let $X_1, \ldots, X_n$ be a sequence of poisson trials such that $\Pr(X_i = 1) = p_i$. If $X = \sum_{i=1}^{n} X_i$, then,*

a) For any $\delta > 0$: $\Pr(X \geq (1+\delta)\mu) < \left[\dfrac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}$.

b) For any $0 < \delta \leq 1$: $\Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\mu\delta^2}{3}}$.

c) For any $R \geq 6\mu$: $\Pr(X \geq R) \leq \dfrac{1}{2^R}$.

Proof. The item a) is stronger. The items b) and c) derive from a).

a) For any $\delta > 0$: $\Pr(X \geq (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right)^{\mu}$.

Applying Markov Inequality for $t > 0$:

$$\begin{aligned} \Pr(X \geq (1+\delta)\mu) &= \Pr(e^{tX} \geq e^{t(1+\delta)\mu}) \\ &\leq \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}} = \frac{M_X(t)}{e^{t(1+\delta)\mu}} \\ &\leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} \qquad (*) \end{aligned}$$

For $\delta > 0$ and using $t = \ln(1+\delta) > 0$ in $(*)$ we have

$$\Pr(X \geq (1+\delta)\mu) \leq \frac{e^{(e^{\ln(1+\delta)}-1)\mu}}{e^{\ln(1+\delta)(1+\delta)\mu}} = \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right)^{\mu}$$

Proof of b). For $0 < \delta \leq 1$ we will prove that

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq e^{\frac{-\delta^2}{3}}$$

Applying $\ln(\cdot)$ in both sides, the above inequality is equivalent to

$$f(\delta) = \delta - (1+\delta)\ln(1+\delta) + \frac{\delta^2}{3} \leq 0$$

Deriving $f(\delta)$ we have

$$\begin{aligned} f'(\delta) &= 1 - \left[1 \cdot \ln(1+\delta) + (1+\delta) \cdot \frac{1}{1+\delta} \cdot 1\right] + \frac{2\delta}{3} \\ &= -\ln(1+\delta) + \frac{2\delta}{3} \end{aligned}$$

So, $f'(\delta) = -\ln(1+\delta) + \frac{2\delta}{3}$

Deriving again, we obtain

$$f''(\delta) = -\frac{1}{1+\delta} + \frac{2}{3}$$

Note that $f''(\delta) < 0$ for $\delta \in [0, \frac{1}{2}]$ and $f''(\delta) > 0$ for $\delta \in [\frac{1}{2}, 1]$.

I.e., $f'(\delta)$ decrease from $0 \to \frac{1}{2}$ and increase from $\frac{1}{2} \to 1$.

As $f'(0) = 0$ e $f'(1) = -\ln(2) + \frac{2}{3} < 0$, then $f'(\delta) < 0$ for any $0 < \delta \leq 1$ and so, $f(\delta)$ is decreasing in this interval.

As $f(0) = 0$, we finish the proof of b).

Proof of c). Let $R = (1+\delta)\mu$. Isolating δ, we have $\delta = \frac{R}{\mu} - 1$.
For $R \geq 6\mu$ we have

$$\delta = \frac{R}{\mu} - 1 \geq \frac{6\mu}{\mu} - 1 = 5$$

So,

$$\begin{aligned}
\Pr(X \geq R) &= \Pr(X \geq (1+\delta)\mu) \\
&\leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \\
&\leq \left(\frac{e^{1+\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \\
&\leq \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \\
&\leq \left(\frac{3}{1+5}\right)^{R} = \frac{1}{2^R}
\end{aligned}$$

Theorem: *Let $X_1, \ldots, X_n$ be a sequence of poisson trials such that $\Pr(X_i = 1) = p_i$. If $X = \sum_{i=1}^{n} X_i$ and $U \geq \mu$, then ,*

a) For any $\delta > 0$: $\Pr(X \geq (1+\delta)U) < \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{U}$.

b) For any $0 < \delta \leq 1$: $\Pr(X \geq (1+\delta)U) \leq e^{-\frac{U\delta^2}{3}}$.

Proof. Exercise (follows from the previous theorem with U in the place of μ). □

Theorem: *Let $X_1, \ldots, X_n$ be a sequence of poisson trials such that $\Pr(X_i = 1) = p_i$. If $X = \sum_{i=1}^{n} X_i$, then for $0 < \delta < 1$ we have*

a) $\Pr(X \leq (1-\delta)\mu) \leq \left(\dfrac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{\mu}$.

b) $\Pr(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$.

Proof. The proof is analogous (exercise). □

Corollary: *Let $X_1, \ldots, X_n$ be a sequence of r.v. of poisson trials such that $\Pr(X_i = 1) = p_i$ and $X = \sum_{i=1}^{n} X_i$. If $0 < \delta < 1$ then,*

$$\Pr(|X - \mu| \geq \delta\mu) \leq \frac{2}{e^{\mu\delta^2/3}}$$

Proof.

$$\begin{aligned}
\Pr(|X - \mu| \geq \delta\mu) &\leq \Pr((X - \mu) \geq \delta\mu) + \Pr((\mu - X) \geq \delta\mu) \\
&= \Pr(X \geq (1+\delta)\mu) + \Pr(X \leq (1-\delta)\mu) \\
&\leq e^{-\frac{\delta^2\mu}{3}} + e^{-\frac{\delta^2\mu}{2}} \\
&\leq 2e^{-\frac{\delta^2\mu}{3}}
\end{aligned}$$

□

Example: *Let X be the number of heads in a sequence of n fair coin flips.*
Applying the previous corollary, we have

$$\begin{aligned}
\Pr(|X - E[X]| \geq \frac{1}{2}\sqrt{6n\ln n}) &= \Pr(|X - E[X]| \geq \frac{\sqrt{6n\ln n}}{n} \cdot \frac{n}{2}) \\
&\leq 2 \cdot e^{-\frac{n}{2}\cdot\left(\frac{\sqrt{6n\ln n}}{n}\right)^2 \Big/ 3} \\
&= 2 \cdot e^{-\frac{n}{2}\cdot\frac{6n\ln n}{n^2} \Big/ 3} \\
&= 2 \cdot e^{-\ln n} \\
&= \frac{2}{n}.
\end{aligned}$$

Note that $\sqrt{6n\ln n}$ is assintotically smaller than n.

Let us compare the corollary with one obtained with Chebyshev Inequality.

$$\begin{aligned}\Pr(|X - E[X]| \geq \frac{n}{4}) &\leq \frac{\text{var}[X]}{(n/4)^2} \\ &= \frac{n \cdot \frac{1}{2} \cdot (1 - \frac{1}{2})}{n^2/16} \\ &= \frac{16}{4n} = \frac{4}{n}\end{aligned}$$

From Chernoff, we have

$$\begin{aligned}\Pr(|X - E[X]| \geq \frac{n}{4}) &\leq 2 \cdot e^{-\frac{n}{2} \cdot \left(\frac{1}{2}\right)^2 / 3} \\ &= \frac{2}{e^{n/24}}\end{aligned}$$

I.e., the bound obtained using Chernoff is asymptotically better.

Application: Estimating a parameter

Consider a gene mutation that occurs in the population with probability p.

But we only know p if we analyse all the population and unfortunately the test to verify the mutation is expensive...

So, we can choose randomly (only) n individuals to do the exam and obtain an estimation $\tilde{p}$ of p.

We would like to know

$$\Pr(p \in [\tilde{p} - \delta, \tilde{p} + \delta]) \geq 1 - \gamma, \quad \text{for small values of } \delta \text{ and } \gamma.$$

Let $X = n \cdot \tilde{p}$ be the number of mutations found in n experiments. We can calculate the probability that p stay outside the interval $[\tilde{p} - \delta, \tilde{p} + \delta]$:

$$\Pr(p \notin [\tilde{p} - \delta, \tilde{p} + \delta]) \leq \Pr(p < \tilde{p} - \delta) + \Pr(p > \tilde{p} + \delta).$$

$$\begin{aligned}
\Pr(p < \tilde{p} - \delta) &= \Pr(\tilde{p} > p + \delta) \\
&= \Pr(n\tilde{p} > n(p + \delta)) \\
&= \Pr(n\tilde{p} > (1 + \frac{\delta}{p})np) \\
&= \Pr(X > (1 + \frac{\delta}{p})E[X]) \\
&\leq e^{-np(\delta/p)^2/2} \\
&= e^{-\frac{n\delta^2}{2p}}
\end{aligned}$$

$$
\begin{aligned}
\Pr(p > \tilde{p} + \delta) &= \Pr(\tilde{p} < p - \delta) \\
&= \Pr(n\tilde{p} < n(p - \delta)) \\
&= \Pr(n\tilde{p} < (1 - \frac{\delta}{p})np) \\
&= \Pr(X < (1 - \frac{\delta}{p})E[X]) \\
&\leq e^{-np(\delta/p)^2/3} \\
&= e^{-\frac{n\delta^2}{3p}}
\end{aligned}
$$

Therefore

$$\begin{aligned} \Pr(p \notin [\tilde{p}-\delta, \tilde{p}+\delta]) &\leq \Pr(p < \tilde{p}-\delta) + \Pr(p > \tilde{p}+\delta) \\ &\leq e^{-\frac{n\delta^2}{2p}} + e^{-\frac{n\delta^2}{3p}} \end{aligned}$$

For example, using the fact that $p \leq 1$,
$n = 13000$ and $\delta = 2\%$ we have

$$\Pr(p \notin [\tilde{p}-\delta, \tilde{p}+\delta]) \leq 2.32\%.$$

and for $n = 1000$ and $\delta = 10\%$ we have

$$\Pr(p \notin [\tilde{p}-\delta, \tilde{p}+\delta]) \leq 4.25\%.$$

BETTER BOUNDS FOR SOME SPECIAL CASES

Theorem: *Let r.v. $X = X_1 + \cdots + X_n$ such that X_i's assume value 1 or -1, with probability $\frac{1}{2}$, independently. Then*

$$\Pr(X \geq a) \leq e^{\frac{-a^2}{2n}}, \qquad \text{for any } a > 0.$$

Proof.

Note that $E[X] = 0$ and therefore Chernoff give us a bound of 1.

For any $t > 0$,

$$E[e^{tX_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t}.$$

The Taylor expansion of e^x is

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^i}{i!} + \cdots$$

$$
\begin{aligned}
E[e^{tX_i}] &= \frac{1}{2}e^t + \frac{1}{2}e^{-t} \\
&= \frac{1}{2}\sum_{i\geq 0}\left(\frac{t^i}{i!} + \frac{(-t)^i}{i!}\right) = \sum_{i\geq 0}\frac{t^{2i}}{(2i)!} \\
&= \sum_{i\geq 0}\frac{t^{2i}}{2i\cdot(2i-1)\cdot(2i-2)\cdot\ldots\cdot(i+1)\cdot i!} \\
&\leq \sum_{i\geq 0}\frac{t^{2i}}{\underbrace{2\cdot 2\cdot\ldots\cdot 2}_{i \text{ times}}\cdot i!} \\
&= \sum_{i\geq 0}\frac{(t^2/2)^i}{i!} \\
&= e^{t^2/2}
\end{aligned}
$$

Using this bound, we have

$$\begin{aligned} E[e^{tX}] &= E[e^{t(X_1+\cdots+X_n)}] \\ &= \Pi_{i=1}^{n} E[e^{tX_i}] \\ &\leq (e^{t^2/2})^n \\ &= e^{t^2 n/2} \end{aligned}$$

So,

$$\begin{aligned} \Pr(X \geq a) &= \Pr(e^{tX} \geq e^{ta}) \\ &\leq \frac{E[e^{tX}]}{e^{ta}} \\ &\leq \frac{e^{\frac{t^2 n}{2}}}{e^{ta}} \\ &= e^{\frac{t^2 n}{2} - ta} \end{aligned}$$

Setting $t = \frac{a}{n}$ we have

$$\begin{aligned} \Pr(X \geq a) &\leq e^{\frac{t^2 n}{2} - ta} \\ &= e^{\frac{a^2}{n^2}\frac{n}{2} - \frac{a}{n}a} \\ &= e^{\frac{-a^2}{2n}} \end{aligned}$$

By symmetry, we can also prove that

Theorem: *Let r.v.* $X = X_1 + \cdots + X_n$ *such that each* X_i *assume value* 1 *or* -1, *with probability* $\frac{1}{2}$, *independently. Then*

$$\Pr(X \leq -a) \leq e^{\frac{-a^2}{2n}}, \qquad \text{for any } a > 0.$$

Corollary: *Let r.v.* $X = X_1 + \cdots + X_n$ *such that each* X_i *assume value* 1 *or* -1, *with probability* $\frac{1}{2}$, *independently. Then*

$$\Pr(|X| \geq a) \leq 2e^{\frac{-a^2}{2n}}, \qquad \text{for any } a > 0.$$

Corollary: *Let r.v.* $Y = Y_1 + \cdots + Y_n$ *be a sum of independent r.v. such that each* Y_i *assume value* 0 *or* 1, *with probability* $\frac{1}{2}$. *Then*

1) For any $a > 0$ *we have* $\Pr(Y \geq \mu + a) \leq e^{\frac{-2a^2}{n}}$.

2) For any $\delta > 0$ *we have* $\Pr(Y \geq (1+\delta)\mu) \leq e^{-\delta^2 \mu}$.

Proof. Proof of item 1).

Let X_i be such that $Y_i = \frac{X_i+1}{2}$ and $X = \sum_{i=1}^{n} X_i$ (i.e. $X_i \in \{-1, 1\}$).

$$\text{So, } Y = \sum_{i=1}^{n} Y_i = \frac{1}{2}\sum_{i=1}^{n} X_i + \frac{n}{2} = \frac{X}{2} + \mu$$

Therefore

$$\begin{aligned} \Pr(Y \geq \mu + a) &= \Pr(\frac{X}{2} + \mu \geq \mu + a) \\ &= \Pr(X \geq 2a) \\ &\leq e^{\frac{-4a^2}{2n}} \end{aligned}$$

Proof of item 2)

2) For any $\delta > 0$ we have $\Pr(Y \geq (1+\delta)\mu) \leq e^{-\delta^2\mu}$.

This follows from setting $a = \delta\mu = \delta\frac{n}{2}$ in the bound of item 1). That is,

$$\begin{aligned}
\Pr(Y \geq (1+\delta)\mu) &= \Pr(\frac{X}{2} + \mu \geq (1+\delta)\mu) \\
&= \Pr(X \geq 2\delta\mu) \\
&\leq e^{\frac{-(2\delta\mu)^2}{2n}} \\
&= e^{\frac{-4\delta^2\mu\frac{n}{2}}{2n}} \\
&= e^{-\delta^2\mu}
\end{aligned}$$

Corollary: *Let $Y_1, \ldots, Y_n$ independent r.v. with* $\Pr(Y_i = 0) = \Pr(Y_i = 1) = \frac{1}{2}$, $Y = \sum_{i=1}^{n} Y_i$ *and* $\mu = E[Y] = \frac{n}{2}$. *Then*

1) For any $0 < a < \mu$ *we have* $\Pr(Y \leq \mu - a) \leq e^{\frac{-2a^2}{n}}$.

2) For any $0 < \delta < 1$ *we have* $\Pr(Y \leq (1-\delta)\mu) \leq e^{-\delta^2 \mu}$.

Proof. Exercise (similar to the previous corollary). □

Application: Set Balancing

Given a matrix A with inputs in $\{0, 1\}$ find vector b such that $b_i \in \{-1, 1\}$

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \longleftarrow \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

and $\|A \cdot b\|_\infty = \max_{i=1,\ldots,m} |c_i|$ is minimized.

Application: Consider a matrix A where the lines are features and the columns are individuals. Each element represents the presence or absence of the feature in the individual. The vector b divides the set of individuals in two balanced parts, considering each feature.

Theorem: *For any n, there exists matrix $A \in \mathbb{R}^{n \times n}$ of the Set Balancing problem such that $\|A \cdot b\|_\infty$ is $\Omega(\sqrt{n})$.*

Algorithm: *SetBalancing(A),*

1. *For $i \leftarrow 1$ to n do*
2. *Let $b_i \in \{-1, 1\}$ be chosen with probability $\frac{1}{2}$.*
3. *Return b*

Theorem: *If b is obtained by algorithm SetBalancing then*

$$\Pr(\|A \cdot b\|_\infty \geq \sqrt{4m \ln n}) \leq \frac{2}{n}$$

Theorem: *If b is obtained by algorithm SetBalancing then*

$$\Pr(\|A \cdot b\|_\infty \geq \sqrt{4m \ln n}) \leq \frac{2}{n}$$

Proof.

Let A_{i*} the i-th line of A, $c_i \leftarrow A_{i*} \cdot b$ and k the number of 1's in A_{i*}.

If $k \leq \sqrt{4m \ln n}$ then $|c_i| \leq \sqrt{4m \ln n}$.

If $k > \sqrt{4m \ln n}$ then the k non-null terms in the sum $c_i \leftarrow \sum_{i=1}^{m} A_{ij} b_j$ are like independent variables with probability $\frac{1}{2}$ to have value -1 or 1.

By Chernoff, we have

$$\begin{aligned}
\Pr(|c_i| > \sqrt{4m \ln n}) &\leq 2e^{-\frac{(\sqrt{4m \ln n})^2}{2k}} \\
&= 2\left(e^{\ln n}\right)^{-4m/2k} \\
&= 2n^{-2m/k} \\
&\leq \frac{2}{n^2}. \qquad \text{because } m \geq k
\end{aligned}$$

By the union bound, the probability that exists i such that $|c_i| > \sqrt{4m \ln n}$ is

$$\Pr(\text{exists } i \text{ s.t. } |c_i| > \sqrt{4m \ln n})) \leq n \cdot \frac{2}{n^2} = \frac{2}{n}.$$

Application: Packet Routing in Sparse Networks

Communication problem in Parallel Computing:

- Each node (or processor) is a routing switch.
- An edge is a communication channel.
- We consider a synchronous model in which:
  - Each edge can transmit at most one packet in one unit of time.
  - A packet can traverse no more than one edge per unit of time.
- Each node has a destination package to some other node.

Based in the Motwani and Raghavan book.

- A route is a sequence of edges a packet traverse to go from an origin to its destination.
- A packet can wait in a node until an edge become free. Each node has a buffer/queue to store packets waiting to be transmitted.
- A routing algorithm have to specify a queuing policy to solve conflicts between packets that want to follow to the same edge from a node.
- In one step, several packets can traverse distinct edges in parallel.

Def.: *A routing algorithm is said to be oblivious if the packet routing consider only its origin and destination nodes and the network topology (but not the other packets).*

Theorem: *For any deterministic oblivious algorithm in a networks of N nodes, each node with output degree d, there is a routing distance that require $\Omega(\sqrt{N/d})$ steps.*

Proof. Not in the scope of the course. □

We will see a probabilistic algorithm for the hypercube that transmit packets in $O(\log_2 N)$ steps, with high probability.

We consider that:

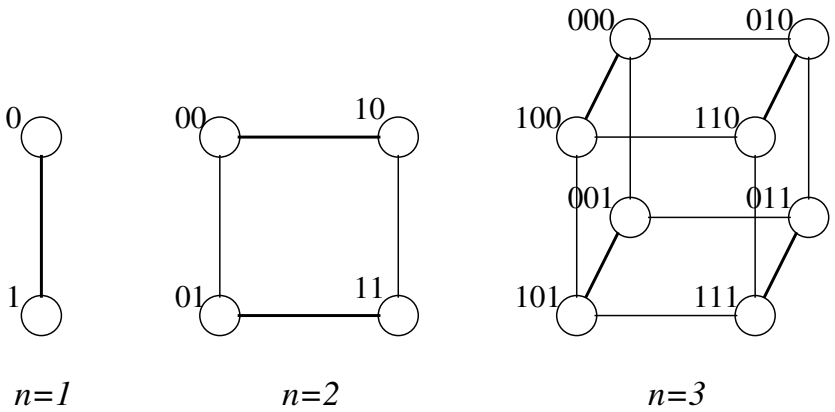
- A (sparse) hypercube network with $N = 2^n$ nodes has $O(N \log_2 N)$ edges.
- The nodes cannot compare the origins and destinations of the packets in the queue.
- The set of N nodes is given by $\mathcal{N} = \{i : 0 \leq i \leq N-1\}$.
- Permutation routing: Each node is the origin of exactly one packet and the destination of exactly one packet.

Def.: *Given $x \in \mathcal{N}$, denote by $\bar{x}$ the binary representation of x using exactly n bits.*

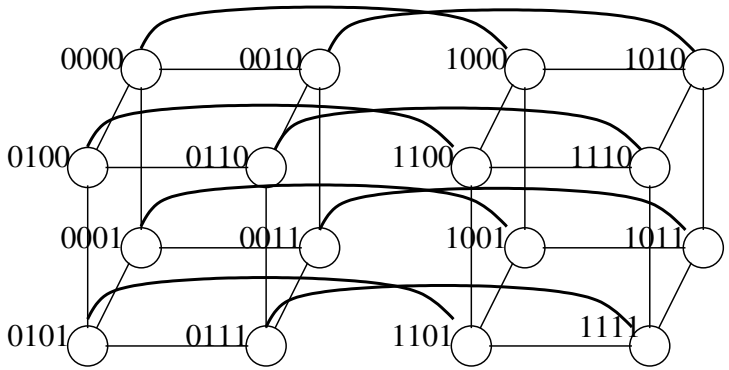
Def.: *A n-dimensional hypercube is a network with $N = 2^n$ nodes, and two nodes are connected by an edge $\{x, y\}$ if and only if $\bar{x}$ and $\bar{y}$ differ by exactly one bit.*

HYPERCUBE

Example of hypercubes:



n=1 *n=2* *n=3*



n=4

Packet route, given origin and destination in the hypercube:

Algorithm: *BitFixing*(a, b)

where a and b are the origin and destination nodes.

1. *Consider* $\overline{a} = (a_1, \ldots, a_n)$ *and* $\overline{b} = (b_1, \ldots, b_n)$.
2. *For* $i \leftarrow 1$ *to* n *do*
3. *If* $a_i \neq b_i$ *traverse the edge*
4. $(b_1, \ldots, b_{i-1}, a_i, a_{i+1} \ldots, a_n) \rightarrow (b_1, \ldots, b_{i-1}, b_i, a_{i+1}, \ldots, a_n)$

Proposition: *The BitFixing algorithm can take* $\Omega(\sqrt{N})$ *steps.*

Proof. Exercise 4.21. □

Probabilistic algorithm:

Algorithm: *TwoPhaseBitFixing*(a,b)

where a and b are the origin and destination nodes.

1. *Let r be a randomly chosen node in* $\mathcal{N}$
2. *Phase 1: Execute BitFixing*(a,r)
3. *Phase 2: Execute BitFixing*(r,b)

Theorem: *With probability* $1-\frac{1}{N}$, *TwoPhaseBitFixing do* $O(\log_2 N)$ *steps.*

We analyse Phase 1. The analysis of Phase 2 is analogous.

Given origen node i, denote by

- r_i the random destination node in Phase 1
- ρ_i the route generated from i to r_i
- i (same name of the origin node) the packet that goes from i to r_i
- d_i the number of delays of packet i
- $|\rho_i|$ the number of edges in ρ_i

Fact: *The number of time steps needed by packet i in Phase 1 is $|\rho_i| + d_i$.*

Fact: *Two routes intersects in at most one node.*

Proof. Note that bits are corrected from the left to the right. So, when two routes separate, the separation bit will not be changed again and the routes will never intersect again. □

Lemma: *Consider route* $\rho_i = (e_1, \ldots, e_k)$ *from* i *to* r_i. *Let* S *be the set of packets, including* i, *that use at least one edge of* ρ_i. *Then,* $d_i \leq |S|$.

Proof. Exercise. □

Let

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

Let $H = \sum_{j=1}^{N} H_{ij}$. Then

$$d_i \leq \sum_{j=1}^{N} H_{ij} = H \quad \text{for any } i \in \mathcal{N}$$

Fact: *Given $i \in \mathcal{N}$, the r.v.'s H_{ij} are Poisson Trials (independent binary r.v.).*

Fact: *The size of any route ρ_i is at most n and $E[|\rho_i|] = \frac{n}{2}$.*

Lemma: *For any $i \in \mathcal{N}$ we have*

$$E\left[\sum_{j=1}^{N} H_{ij}\right] \leq n.$$

Proof. Compute $\Pr(H_{ij} = 1)$ is hard. We use another bound.
Let $T(e)$ (or T_e) the number of routes that traverse edge e and let $\rho_i = (e_1, \ldots, e_k)$.
Note that

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

and therefore

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

Fact: *If e and f are any two edges, then*

$$E[T(e)] = E[T(f)].$$

Proof. Exercise (follows from symmetry). □

Fact: *If e is an edge, then $E[T_e] \leq 1$.*

Proof. As $E[|\rho_i|] = \frac{n}{2}$, we have

$$E\left[\sum_{\text{edge } f} T_f\right] = E\left[\sum_{i=1}^{N} |\rho_i|\right] \leq N \cdot \frac{n}{2}.$$

As we have $N \cdot \frac{n}{2}$ edges in the hypercube, and $E[T_e] = E[T_f]$ for edges e and f, then

$$E[T_e] \leq 1 \quad \text{for any edge } e$$

□

To complete the proof of the lemma (that $E\left[\sum_{j=1}^{N} H_{ij}\right] \leq n$), we have

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

□

Fact: *If* $H = \sum_{j=1}^{N} H_{ij}$ *then*

$$\Pr(H \geq 6n) \leq \frac{1}{2^{6n}}.$$

Proof. From the previous lemma, we have that

$$E\left[H\right] \leq n.$$

As H is the sum of a sequence of Poisson Trials, we can use Chernoff bound:

$$\Pr(H \geq 6n) \quad \leq \quad \frac{1}{2^{6n}} \quad \text{because } 6n \geq 6\,E[H]$$

□

Lemma: *The probability that there exists route $i \in \mathcal{N}$ such that $|\rho_i| + d_i > 7n$ is at most $\frac{1}{2^{5n}}$.*

Proof. Given node i, we know that

- $|\rho_i| \leq n$
- $d_i \leq H$, onde $H = \sum_{j=1}^{N} H_{ij}$

And therefore

$$\begin{aligned} \Pr(|\rho_i| + d_i > 7n) &\leq \Pr(d_i > 6n) \\ &\leq \Pr(H > 6n) \\ &\leq \frac{1}{2^{6n}} \end{aligned}$$

So, the probability that there exists i such that $|\rho_i| + d_i > 7n$ is (bounded by the probability of sum) is at most

$$N \cdot \frac{1}{2^{6n}} = 2^n \frac{1}{2^{6n}} = \frac{1}{2^{5n}}.$$

□

So, we conclude that

Lemma: *The probability that there exists $i \in \mathcal{N}$ such that packet i uses more than $7n$ steps to go from i to r_i in Phase 1 is at most $\frac{1}{2^{5n}}$.*

We suppose that the algorithm wait the Phase 1 finish for all packets, before to start Phase 2.

We can prove similar result for Phase 2.

Lemma: *The probability that there exist $i \in \mathcal{N}$ such that packet i uses more than $7n$ steps to go from r_i to its destination in Phase 2 is at most $\frac{1}{2^{5n}}$.*

Theorem: *The probability that there exists $i \in \mathcal{N}$ such that packet i use more than $14 \log N$ steps is at most $\frac{1}{N}$.*

Proof. The probability that there exists a packet i that uses more than $7n$ steps in Phase 1 or Phase 2 is at most

$$2\frac{1}{2^{5n}} \leq \frac{1}{2^{4n}} = \frac{1}{N^4} \leq \frac{1}{N}.$$

□

See also the packet routing in the *ButterFly* topology, in the book of Upfal and Mitzenmacher.

Balls, Bins and Random Graphs

In several problems we have situations like balls randomly thrown into bins.

Some questions:

- How many of the bins are empty ?
- How many bins have at least two balls ?
- How many balls are in the fullest bin ?
- etc.

EXAMPLE: THE BIRTHDAY PARADOX

Let E_{30} the event that two people in 30 share the same birthday.

What is $\Pr(E_{30})$?

Let us consider that

- It is not a leap year
- We ignore the possibility of twins
- A person's birthday is equally likely to be any day of the year

Let us calculate $\Pr(\overline{E_{30}})$.

$$\Pr(\overline{E_{30}}) = \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdot \ldots \cdot \left(1 - \frac{29}{365}\right) = 0.2937$$

So, $\Pr(E_{30}) \approx 70.63\%$.

Doing the same calculation for 22 people, we have
$\Pr(E_{22}) \approx 47.57\%$

I.e., the probability that 22 people born in distinct days of the same year is approximately 52.43%.

GENERALIZING FOR m PEOPLE (BALLS) AND n DAYS (BINS)

Consider that m is small, compared to n.

Let E be the event that m balls are thrown randomly in distinct bins

Using the fact that $e^{-j/n} = 1 - (j/n) + \frac{(j/n)^2}{2!} - \frac{(j/n)^3}{3!} + \cdots$ we can approximate $\Pr(E)$ as

$$\begin{aligned}
\Pr(E) &= \Pi_{j=1}^{m-1}\left(1-\frac{j}{n}\right) \approx \Pi_{j=1}^{m-1} e^{-j/n} \\
&= \exp\left(-\sum_{j=1}^{m-1}\frac{j}{n}\right) \\
&= e^{\frac{m(m-1)}{2n}} \approx e^{\frac{-m^2}{2n}}
\end{aligned}$$

Fact: *Consider m balls are thrown into n bins, with m small compared to n. To have probability p that m balls fall in distinct bins, we must have $m \approx \sqrt{-2n \ln p}$.*

Example: *In the Birthday Paradox, to have probability 0.5 that all people born in distinct days, we need to approximately have*

$$m = \sqrt{-2 \cdot 365 \ln(0.5)} = 22.49 \text{ people.}$$

That is consistent with the previous probability obtained for 22 people.

Corollary: *If $m = \sqrt{2n \ln n}$ the probability that all m balls are thrown in distinct bins is approximately $\frac{1}{n}$ (small probability).*

Example: *We will do another analysis to bound m and have probability at least 0.5 that all birthdays are distinct.*
Let E_i the event that the i-th person do not match the same birthday with the previous $i-1$ people.

$$\begin{aligned}
\Pr(\text{have two birthdays in } k \text{ people}) &= \Pr(\overline{E_1} \cup \overline{E_2} \cup \ldots \cup \overline{E_k}) \\
&\leq \sum_{i=1}^{k} \Pr(\overline{E_i}) \\
&\leq \sum_{i=1}^{k} \frac{i-1}{n}) \\
&= \frac{k(k-1)}{2n}
\end{aligned}$$

I.e., If $k < \sqrt{n}$ then $\Pr(\text{have two birthdays}) < 0.5$.

Suppose we have $\lceil \sqrt{n} \rceil$ distinct birthdays. The probability that each person match its birthday with one of the previous is $\geq \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$.

So,

$$\begin{aligned} \Pr(\text{next } \lceil \sqrt{n} \rceil \text{ do not match birthdays}) &\leq \left(1 - \frac{1}{\sqrt{n}}\right)^{\lceil \sqrt{n} \rceil} \\ &\leq \frac{1}{e} \end{aligned}$$

Fact: *If we have $m = (\ln(n) + 1)\lceil \sqrt{n} \rceil$ the probability that all birthdays are distinct (all balls are thrown into distinct bins) is $\leq (1/e)^{\ln n} = \frac{1}{n}$ (i.e., with small probability)*

BALLS INTO BINS

Now we bound (with high probability) the maximum number of balls in a bin

Lemma: *When n balls are thrown uniformly at random into n bins, the probability that there exists a bin with at least M balls ($M \leq n$) is at most $n \cdot \left(\frac{e}{M}\right)^M$.*

Proof.

$$\begin{aligned}
\Pr(\text{Bin } i \text{ receives} \geq M \text{ balls}) &\leq \binom{n}{M}\left(\frac{1}{n}\right)^M \\
&= \frac{n \cdot (n-1) \cdot \ldots \cdot (n-M+1)}{M!} \cdot \frac{1}{n^M} \\
&\leq \frac{1}{M!}
\end{aligned}$$

Using the inequality $\frac{k^k}{k!} < \sum_{i=0}^{\infty} \frac{k^i}{i!} = e^k$

we have that $\frac{1}{k!} < \frac{e^k}{k^k}$. So,

$$\Pr(\text{Bin } i \text{ receives} \geq M \text{ balls}) \leq \frac{1}{M!} < \left(\frac{e}{M}\right)^M$$

Using the union bound, we have

$$\begin{aligned} \Pr(\text{there exists a bin with} \geq M \text{ balls}) &\leq \sum_{i=1}^{n} \Pr(\text{bin } i \text{ receives} \geq M \text{ balls}) \\ &< n \cdot \left(\frac{e}{M}\right)^M \end{aligned}$$

Lemma: *When n balls are thrown uniformly at random into n bins, the probability that there exists bin with at least $3\ln n/\ln\ln n$ balls is at most $\frac{1}{n}$.*

Proof. If $M \geq 3\ln n/\ln\ln n$, the probability to have a bin with at least M balls is bounded by

$$\begin{aligned}
n\cdot\left(\frac{e}{M}\right)^M &\leq n\left(\frac{e\ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n}\\
&\leq n\left(\frac{\ln\ln n}{\ln n}\right)^{3\ln n/\ln\ln n}\\
&= e^{\ln n}\cdot\left(e^{\ln\ln\ln n-\ln\ln n}\right)^{3\ln n/\ln\ln n}\\
&= e^{-2\ln n+3(\ln n)(\ln\ln\ln n)/\ln\ln n}\\
&\leq \frac{1}{n} \qquad \text{(for sufficiently large } n\text{)}
\end{aligned}$$

□

APPLICATION: BUCKETSORT

Algorithm: *BucketSort*(S)
where S *has* $n = 2^m$ *integers in* $\{0, \ldots, 2^k - 1\}$, *with* $k \geq m$

1. *Consider all integers in* S *represented with* k *bits*
2. *For each* $t \in \{0, \ldots, 2^m - 1\}$ *do* $B[t] \leftarrow \emptyset$
3. *For each* $x \in S$ *do*
4. *Let* $\bar{x}$ *the integer obtained with the* m *more significant bits of de* x
5. *Put* x *in the bucket* $B[\bar{x}]$
6. *For each* $t \in \{0, \ldots, 2^m - 1\}$ *sort* $B[t]$ *with InsertionSort*
7. *Return* $(B[0] \,\|\, B[1] \,\|\, \ldots \,\|\, B[2^m - 1])$

Lemma: *The BucketSort algorithm sort* S *correctly.*

Proof. Exercise. □

Lemma: *If S contains $n = 2^m$ integers uniformly distributed in $\{0, \ldots, 2^k - 1\}$, with $k \geq m$, then the expected time of algorithm BucketSort is $O(n)$.*

Proof. Clearly, all steps, with the exception the sorting step (step 6) can be performed in time $O(n)$.

Let X_j the number of elements in bucket $B[j]$. Each bucket $B[j]$ is sorted in at most $c \cdot (X_j)^2$ (time of InsertionSort).

$$
\begin{aligned}
E\left[\sum_{j=0}^{2^m-1} c \cdot (X_j)^2\right] &= c \cdot n \cdot E[(X_1)^2] \\
&\qquad \text{(as } X_j \text{ is a binomial r.v. } B(n, p = \tfrac{1}{n})\text{)} \\
&= c \cdot n \cdot (n \cdot (n-1) \cdot p^2 + n \cdot p) \\
&= c \cdot n \cdot \left(\frac{n \cdot (n-1)}{n^2} + 1\right) < 2 \cdot c \cdot n
\end{aligned}
$$

□

The Poisson Distribution

Lemma: *If m balls are thrown into n bins, then the number of expected empty bins is $\approx n \cdot e^{-m/n}$.*

Proof. Let V_i the event that bin i is empty.
For a bin i, we have

$$\begin{aligned} \Pr(V_i) &= \left(1 - \frac{1}{n}\right)^m = \left[\left(1 - \frac{1}{n}\right)^n\right]^{\frac{m}{n}} \\ &\approx (e^{-1})^{m/n} \end{aligned}$$

Let X_i a binary r.v. with value 1 if and only if V_i happens.

Let $X = \sum_{i=1}^{n} X_i$. Then $E[X] = \sum_{i=1}^{n} E[X_i] = n \cdot e^{-m/n}$ □

Lemma: *Consider m balls randomly thrown into n bins. The number of bins with exactly r balls is approximately $n\frac{e^{-m/n}(m/n)^r}{r!}$, when r is small compared to m and n.*

Proof.

Let E_r the event that bin j has exactly r balls.

$$\begin{aligned}
\Pr(E_r) &= \binom{m}{r} \cdot \left(\frac{1}{n}\right)^r \cdot \left(1-\frac{1}{n}\right)^{m-r} \\
&= \frac{1}{r!} \cdot \frac{m \cdot (m-1) \cdot \ldots \cdot (m-r+1)}{n^r} \cdot \left(1-\frac{1}{n}\right)^{m-r} \\
&\quad \text{(when } r \text{ is small compared to } m \text{ and } n\text{)} \\
&\approx \frac{1}{r!} \cdot \frac{m^r}{n^r} \cdot \left(1-\frac{1}{n}\right)^{m} \\
&\approx \frac{e^{-m/n}(m/n)^r}{r!}
\end{aligned}$$

Let X_j a binary r.v. with value 1 if and only if bin j has exactly r balls.

Let $X = \sum_{j=1}^{n} X_j$. Then

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} E[X_j] \\ &= \sum_{i=1}^{n} \Pr(E_r) \\ &\approx n\frac{e^{-m/n}(m/n)^r}{r!}. \end{aligned}$$

Note that the number of expected balls in a bin is $\mu = m/n$.

The probability $\Pr(E_r)$ leads to the following distribution:

Def.: *A discrete Poisson r.v. with parameter* μ *is given by*

$$\Pr(X = j) = \frac{e^{-\mu} \cdot \mu^j}{j!}.$$

Exercise: *Use the Taylor expansion* $e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!}$ *to show that the above function is in fact a probability function.*

Fact: *If X is a discrete Poisson r.v., then $E[X] = \mu$.*

Proof.

$$\begin{aligned}
E[X] &= \sum_{j=1}^{\infty} j \cdot \Pr(X = j) \\
&= \sum_{j=1}^{\infty} j \cdot \frac{e^{-\mu} \cdot \mu^j}{j!} \\
&= \mu \cdot e^{-\mu} \sum_{j=1}^{\infty} \cdot \frac{\mu^{j-1}}{(j-1)!} \\
&= \mu \cdot e^{-\mu} \sum_{j=0}^{\infty} \cdot \frac{\mu^j}{j!} \\
&= \mu \cdot e^{-\mu} e^{\mu} = \mu
\end{aligned}$$

□

Lemma: *The sum of independent discrete Poisson r.v. is also a discrete Poisson r.v.*

Proof. Proof for two variables. Let X and Y discrete Poisson r.v.

$$\begin{aligned}
\Pr(X+Y=j) &= \sum_{k=0}^{j} \Pr((X=k)\cap(Y=j-k)) \\
&= \sum_{k=0}^{j} \frac{e^{-\mu_1}\cdot\mu_1^k}{k!}\cdot\frac{e^{-\mu_2}\cdot\mu_2^{j-k}}{(j-k)!} \\
&= \frac{e^{-(\mu_1+\mu_2)}}{j!}\sum_{k=0}^{j}\frac{j!}{k!(j-k)!}\cdot\mu_1^k\cdot\mu_2^{(j-k)} \\
&= \frac{e^{-(\mu_1+\mu_2)}}{j!}\sum_{k=0}^{j}\binom{j}{k}\cdot\mu_1^k\cdot\mu_2^{(j-k)} \\
&= \frac{e^{-(\mu_1+\mu_2)}(\mu_1+\mu_2)^j}{j!}
\end{aligned}$$

□

Same result using moment generating function (m.g.f.)

Lemma: *The m.g.f. of a Poisson r.v. with parameter* μ *is*

$$M_x(t) = e^{\mu(e^t-1)}.$$

Proof. For any t we have

$$\begin{aligned}
E[e^{tX}] &= \sum_{k=0}^{\infty} e^{tk} \cdot \Pr(X=k) = \sum_{k=0}^{\infty} e^{tk} \cdot \frac{e^{-\mu} \cdot \mu^k}{k!} \\
&= \sum_{k=0}^{\infty} e^{tk} \cdot \frac{e^{-\mu} \cdot \mu^k}{k!} \cdot e^{\mu e^t} \cdot e^{-\mu e^t} \\
&= e^{\mu(e^t-1)} \sum_{k=0}^{\infty} \frac{e^{-\mu e^t} \cdot (\mu \cdot e^t)^k}{k!} = e^{\mu(e^t-1)} \sum_{k=0}^{\infty} \frac{e^{-\mu_Y} \cdot (\mu_Y)^k}{k!} \\
&= e^{\mu(e^t-1)} \sum_{k=0}^{\infty} \Pr(Y=k) \quad \text{where } Y \text{ is discrete Poisson r.v.} \\
&= e^{\mu(e^t-1)}
\end{aligned}$$

□

Corollary: *The sum of independent discrete Poisson r.v. is discrete Poisson r.v.*

Proof. For two independent r.v. X and Y with expectations μ_1 and μ_2, resp. Note that $E[X+Y] = \mu_1 + \mu_2$.
If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Then

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= e^{\mu_1(e^t-1)} \cdot e^{\mu_2(e^t-1)} \\ &= e^{(\mu_1+\mu_2)(e^t-1)} \end{aligned}$$

As the moment of a variable (in the case, $X+Y$) define its distribution in a unique way, we have that $X+Y$ is a Poisson r.v. with expectation $\mu_1 + \mu_2$. □

Chernoff bounds for Poisson r.v.

Theorem: *Let X be a Poisson r.v. with parameter μ.*

1. *If $x > \mu$ then*
$$\Pr(X \geq x) \leq \frac{e^{-\mu}(e\cdot\mu)^x}{x^x}$$
2. *If $x < \mu$ then*
$$\Pr(X \leq x) \leq \frac{e^{-\mu}(e\cdot\mu)^x}{x^x}$$

Proof. For any $t > 0$ and $x > \mu$ then

$$\begin{aligned}
\Pr(X \geq x) &= \Pr(e^{tX} \geq e^{tx}) \\
&\leq \frac{E[e^{tX}]}{e^{tx}} \\
&= \frac{e^{\mu(e^t-1)}}{e^{tx}} \qquad \forall t > 0.
\end{aligned}$$

So, $\Pr(X \geq x) \leq \dfrac{e^{\mu(e^t-1)}}{e^{tx}} \quad \forall t > 0.$

Let $t = \ln(x/\mu) > 0$. Then

$$\begin{aligned}
\Pr(X \geq x) &\leq \frac{e^{\mu(e^{\ln(x/\mu)}-1)}}{e^{\ln(x/\mu)\cdot x}} \\
&= \frac{e^{\mu(x/\mu-1)}}{(x/\mu)^x} \\
&= \frac{e^x \cdot e^{-\mu}}{x^x} \cdot \mu^x \\
&= \frac{e^{-\mu} \cdot (e \cdot \mu)^x}{x^x}
\end{aligned}$$

The proof of item 2 is similar (exercise).

Limit of the binomial distribution

In general, the Poisson distribution is bounded by the binomial distribution with parameter n and p, when n is large and p small.

Theorem: *Let X_n be a binomial r.v. with parameters n and p, where p is function of n and $\lim_{n\to\infty} n \cdot p = \lambda$ is a constant independent of n. Then, for each fixed k,*

$$\lim_{n\to\infty} \Pr(X_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

First, let's see an application.

Example: *Consider m balls and n bins where m is a function of n and $\lim_{n\to\infty} \frac{m}{n} = \lambda$. Then, if X_n is the number of balls in a specific bin, then*

$$\lim_{n\to\infty} \Pr(X_n = r) = \frac{e^{-m/n}(m/n)^r}{r!},$$

that is the approximation obtained before.

To see this, consider the bin 1 (w.l.o.g.).

Each of the m balls enters into the bin 1 with probability $p = \frac{1}{n}$. As X_n is a binomial r.v. with parameters n and p, the number of balls in the bin 1 tends to

$$\lim_{n\to\infty} m \cdot p = \lim_{n\to\infty} m \cdot \frac{1}{n} = \lambda$$

and therefore, from the previous theorem

$$\lim_{n\to\infty} \Pr(X_n = r) = \lim_{n\to\infty} \frac{e^{-\lambda}\lambda^r}{r!} = \frac{e^{-m/n}(m/n)^r}{r!}.$$

Example: *Gramatical errors in a text: Each word of a text can be typed wrongly. The number of errors is a binomial r.v. with n large and p small that can be consider a Poisson r.v.*

Proof. (of the Theorem): We can write

$$\Pr(X_n = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

For $|x| \leq 1$ and $k \geq 0$ it is valid that (exercise: use Taylor expansion)

$$e^x(1-x^2) \leq 1+x \leq e^x \qquad \text{e} \qquad (1-p)^k \geq 1-p\cdot k$$

Then,

$$\begin{aligned}
\Pr(X_n = k) &= \frac{n!}{k!(n-k)!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\
&\leq \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k!} \cdot p^k \cdot \frac{(e^{-p})^n}{(1-p)^k} \\
&\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(e^{-pn})}{1-pk} \\
&= \frac{e^{-pn}}{k!} \cdot \frac{(np)^k}{1-pk} \qquad (*)
\end{aligned}$$

On the other hand, we have:

$$\begin{aligned}
\Pr(X_n = k) &= \frac{n!}{k!(n-k)!} \cdot p^k \cdot (1-p)^{n-k} \\
&\geq \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} \cdot p^k \cdot (1-p)^n \\
&\geq \frac{(n-k+1)^k}{k!} \cdot p^k \cdot (e^{-p}(1-p^2))^n \\
&\geq \frac{e^{-pn}((n-k+1) \cdot p)^k}{k!} \cdot (1-p^2 n) \qquad (**)
\end{aligned}$$

Combining $(*)$ with $(**)$:

$$\Pr(X_n = k) \leq \frac{e^{-pn}}{k!} \cdot \frac{(np)^k}{1-pk} \qquad (*)$$

$$\Pr(X_n = k) \geq \frac{e^{-pn}((n-k+1)\cdot p)^k}{k!} \cdot (1-p^2 n) \qquad (**)$$

we have

$$\frac{e^{-pn}}{k!} \cdot \frac{(np)^k}{1-pk} \geq \Pr(X_n = k) \geq \frac{e^{-pn}((n-k+1)\cdot p)^k}{k!} \cdot (1-p^2 n)$$

Dividing both parts by $\frac{e^{-pn}(pn)^k}{k!}$, we have

$$\frac{1}{1-pk} \geq \frac{\Pr(X_n = k)}{\frac{e^{-pn}(pn)^k}{k!}} \geq \frac{(n-k+1)^k}{n^k} \cdot (1-p^2 n) \qquad (***)$$

$$\frac{1}{1-pk} \geq \frac{\Pr(X_n = k)}{\frac{e^{-pn}(pn)^k}{k!}} \geq \frac{(n-k+1)^k}{n^k} \cdot (1-p^2 n) \quad (***)$$

When $n \to \infty$, $p \to 0$ and $pn \to \lambda$ we have

$$\frac{1}{1-pk} \to 1, \qquad 1-p^2 n \to 1 \qquad \text{and} \qquad \frac{(n-k+1)^k}{n^k} \to 1$$

and therefore

$$1 \geq \lim_{n\to\infty} \frac{\Pr(X_n = k)}{\frac{e^{-pn}(pn)^k}{k!}} \geq 1$$

Poisson Approximation

One of the difficulties to analyse problems with balls and bins is to consider dependencies.

- If bin 1 is empty, it is less probably that bin 2 is empty.
- If the number of balls in the first $n-1$ bins is known, then the number of balls in the n-th bin is totally determined.

Suppose that m balls are thrown uniformly at random into n bins.

- Let $X_i^{(m)}$ a r.v. that gives the number of balls in the i-th bin (note that the r.vs. $X_1^{(m)}, \ldots, X_n^{(m)}$ are not independent).
- We can approximate these variables by independent Poisson r.vs. $Y_i^{(m)}$ with average m/n. I.e., each bin receives independently a random quantity of balls.
- Note that with variables $Y_i^{(m)}$ we can obtain a total number of balls that is different from m.

Theorem: *The distribution* $(Y_1^{(m)}, \ldots, Y_n^{(m)})$ *conditioned to* $\sum_i Y_i^{(m)} = k$ *is the same as* $(X_1^{(k)}, \ldots, X_n^{(k)})$*, independently from* m*.*

Proof. We will show that

$$\Pr\left((Y_1^{(m)}, \ldots, Y_n^{(m)}) = (k_1, \ldots, k_n) | \sum_i Y_i^{(m)} = k \right)$$

and

$$\Pr\left((X_1^{(k)}, \ldots, X_n^{(k)}) = (k_1, \ldots, k_n) \right)$$

attain the same value.

- When we throw k balls into n bins, the probability that $(X_1^{(k)}, \ldots, X_n^{(k)}) = (k_1, \ldots, k_n)$ is given by

$$\Pr\left((X_1^{(m)}, \ldots, X_n^{(m)}) = (k_1, \ldots, k_n) = k \right)$$

$$\begin{aligned} &= \frac{\binom{k}{k_1;k_2;\ldots;k_n}}{n^k} \\ &= \frac{k!}{k_1! \cdot k_2! \cdot \ldots \cdot k_n! \cdot n^k} \end{aligned}$$

- Let $k_1, \ldots, k_n$ such that $\sum_i k_i = k$.

$$\Pr\left((Y_1^{(m)}, \ldots, Y_n^{(m)}) = (k_1, \ldots, k_n) | \sum_i Y_i^{(m)} = k \right)$$

$$= \frac{\Pr(Y_1^{(m)} = k_1) \cap \Pr(Y_2^{(m)} = k_2) \cap \ldots \Pr(Y_n^{(m)} = k_n)}{\Pr(\sum_i Y_i^{(m)} = k)}$$

$$= \frac{\prod_{i=1}^n \frac{e^{-\frac{m}{n}} (\frac{m}{n})^{k_i}}{k_i!}}{\frac{e^{-m} m^k}{k!}}$$

$$= \frac{(e^{-\frac{m}{n}})^n \cdot (\frac{m}{n})^{k_1 + \cdots + k_n}}{k_1! \cdot \ldots \cdot k_n!} \cdot \frac{k!}{e^{-m} \cdot m^k}$$

$$= \frac{k!}{k_1! \cdot \ldots \cdot k_n! \cdot n^k}$$

Using the previous result, we can prove stronger results for any non-negative function on the number of ball into bins.

Theorem: *Let $f(x_1, \ldots, x_n)$ a non-negative function. Then,*

$$E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \leq e\sqrt{m} \cdot E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})]$$

Proof.

$$\begin{aligned}
& E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})] \\
&= \sum_{k=0}^{\infty} E[f(Y_1^{(m)}, \ldots, Y_n^{(m)}) | \sum_i Y_i^{(m)} = k] \cdot \Pr(\sum_i Y_i^{(m)} = k) \\
&\geq E[f(Y_1^{(m)}, \ldots, Y_n^{(m)}) | \sum_i Y_i^{(m)} = m] \cdot \Pr(\sum_i Y_i^{(m)} = m) \\
&= E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \cdot \frac{e^{-m} \cdot m^n}{m!} \qquad (**)
\end{aligned}$$

$$E[f(Y_1^{(m)},\ldots,Y_n^{(m)})]$$

$$\begin{aligned}
&\geq E[f(X_1^{(m)},\ldots,X_n^{(m)})]\cdot\frac{e^{-m}\cdot m^n}{m!} \qquad (**)\\
&\quad \text{(using the inequality } m! < e\sqrt{m}\cdot\left(\tfrac{m}{e}\right)^m \text{ (exercise))}\\
&\geq E[f(X_1^{(m)},\ldots,X_n^{(m)})]\cdot\frac{e^{-m}\cdot m^n}{e\sqrt{m}\cdot m^n\cdot e^{-m}}\\
&= \frac{E[f(X_1^{(m)},\ldots,X_n^{(m)})]}{e\sqrt{m}}
\end{aligned}$$

□

The previous theorem is valid for any non-negative function, inclusively for indicator functions (0 or 1) of the occurrence of an event.

We call the scenario
- with m balls and n bins as Exact Case
- approximated by independent r.vs. with average $\frac{m}{n}$ as Poisson Case

Corollary: *Any event that occur with probability p in the Poisson case, occur with probability at most $p \cdot e \cdot \sqrt{m}$ in the exact case.*

Proof. Let f an event indicator function. So, $E[f(\cdot)]$ is exactly the probability of the event occur. The result follows from the previous theorem. □

The corollary above is very useful, mainly when p is small.

A particular case with better approximation:

Theorem: *If $f(x_1, \ldots, x_n)$ is a non-negative function such that $E[f(X_1^{(m)}, \ldots, X_n^{(m)})]$ is monotonically increasing, or monotonically decreasing in m, then*

$$E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \leq 2 \cdot E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})]$$

Proof. Exercise. □

The proof of the next corollary is straightforward.

Corollary: *Let $\mathcal{E}$ an event with probability that is monotonically increasing or monotonically decreasing in the number of balls. If $\mathcal{E}$ has probability p in the Poisson case, then $\mathcal{E}$ has probability at most $2 \cdot p$ in the exact case.*

Example: *Maximum number of balls in a bin: Consider n balls uniformly thrown into n bins and $X_{\max}$ the maximum number of balls in a bin.*
We prove that $\Pr(X_{\max} \geq 3\ln n/\ln\ln n) \leq \frac{1}{n}$. *Now, we use the previous result and prove that* $\Pr(X_{\max} \leq \ln n/\ln\ln n) \leq \frac{1}{n}$. *I.e., with high probability, the maximum number of balls is* $\Theta(\ln n/\ln\ln n)$.

Lemma: *When n balls are thrown into n bins we have* $\Pr(X_{\max} \geq \ln n/\ln\ln n) \geq 1 - \frac{1}{n}$ *for n sufficiently large.*

Proof. Let $M = \frac{\ln n}{\ln\ln n}$ and consider the Poisson case:

$$\Pr(\text{Bin } i \text{ has} \geq M \text{ balls in the Poisson case})$$

$$\begin{aligned} &\geq \Pr(\text{Bin } i \text{ has exactly } M \text{ balls in the Poisson case}) \\ &= \frac{e^{-n/n}(n/n)^M}{M!} = \frac{1}{e \cdot M!} \end{aligned}$$

I.e.,

$$\Pr(\text{Bin } i \text{ do not have } M \text{ balls in the Poisson case}) \leq 1 - \frac{1}{e \cdot M!}$$

and therfore

$$\begin{aligned}
\Pr(\text{No bin have } M \text{ balls in the Poisson case}) &\leq \left(1 - \frac{1}{e \cdot M!}\right)^n \\
&\leq \left(e^{-\frac{1}{eM!}}\right)^n \\
&= e^{-\frac{n}{eM!}}
\end{aligned}$$

Now, we show that $e^{-\frac{n}{eM!}} \leq \frac{1}{n^2}$. Because, proving this inequality, we can use the Poisson approximation to bound the probability in the exact case.

$$
\begin{aligned}
&\Pr(\text{No bin have } M \text{ balls in the exact case}) \\
&\quad \leq \quad e\sqrt{n}\Pr(\text{No bin have } M \text{ balls in the Poisson case}) \\
&\quad \leq \quad e\sqrt{n}\frac{1}{n^2} \\
&\quad \leq \quad \frac{1}{n}, \qquad \text{for sufficiently large } n.
\end{aligned}
$$

So, we show that $e^{-\frac{n}{eM!}} \leq n^{-2}$:

Applying $\ln(\cdot)$ in both sides, it is equivalent to prove that

$$\frac{n}{eM!} \geq 2 \ln n$$

Isolating $M!$, we have to prove that

$$M! \leq \frac{n}{2 \cdot e \cdot \ln n}$$

Using the fact that $M! \leq e\sqrt{M}\left(\frac{M}{e}\right)^M \leq M\left(\frac{M}{e}\right)^M$, it is sufficient to prove that

$$M\left(\frac{M}{e}\right)^M \leq \frac{n}{2 \cdot e \cdot \ln n} \qquad (*)$$

$$M\left(\frac{M}{e}\right)^M \leq \frac{n}{2\cdot e\cdot \ln n} \qquad (*)$$

Applying $\ln(\cdot)$ in both sides, it is sufficient to prove that

$$\ln M + M\ln M - M \leq \ln n - \ln\ln n - \ln(2e)$$

Replacing $M = \frac{\ln n}{\ln\ln n}$ and developing only the left side:

$$\begin{aligned}
&\ln M + M\ln M - M \\
&= (\ln\ln n - \ln\ln\ln n) + \frac{\ln n}{\ln\ln n}(\ln\ln n - \ln\ln\ln n) - \frac{\ln n}{\ln\ln n} \\
&\leq \ln n - \frac{\ln n}{\ln\ln n} \\
&\leq \ln n - \ln\ln n - \ln(2e) \quad \text{(for sufficiently large } n\text{)}
\end{aligned}$$

Application: Hashing

Problem: *Password Verification: Given the set of words in a Dictionary (unacceptable passwords) P, verify if a given word x belongs to P.*

Suppose we have a hashing function $h : U \to \{0, \ldots, n-1\}$ such that

- $h(x) = j$ with probability $\frac{1}{n} \quad \forall j \in \{0, \ldots, n-1\}$
- $h(x)$ is independent for different values of x.

We consider that words are stored in a Chain Hashing.

Chain Hashing of n position:

- Each position of the hashing stores words with linked list.
- Insert each $x \in P$ in the list of index $h[x]$.

Observations:

- Expected number of words in each position: $\frac{m}{n}$
- Expected number of comparisons to search $x \notin P$ is $\leq \frac{m}{n}$
- Expected number of comparisons to search $x \in P$ is $\leq 1 + \frac{m-1}{n}$
- If $n = m$, maximum number of comparisons is $\Theta(\frac{\ln n}{\ln \ln n})$ with high probability
- Several positions may stay empty

Hashing Bit Strings: Consider the application of unacceptable passwords

- Suppose that $h(x) \rightarrow \{0,1\}^b$, where b is a number of bits.
- The function $h(x)$ works as a fingerprint of x.
- Instead of storing the word x, we only store the b bits of $h(x)$.
- Let $H = \{h(x) : x \in P\}$.
- To verify if $x \in P$, we ask if $h(x) \in H$.
- Although we can save space, we may have false positives (word $y \notin P$ such that $h(y) \in H$).
- As b increases, the chance to obtain false positives decreases.

How large does b have to be so that $\Pr(\text{word is false positive}) \leq \frac{1}{m}$?

How large does b have to be so that $\Pr(\text{word is false positive}) \leq \frac{1}{m}$?
Let P be a set of m unaccepted passwords, $x \notin P$.

Then

$$\Pr(h(x) = h(y)) = \frac{1}{2^b} \quad \text{for any word } y \in P$$

$$\Rightarrow \quad \Pr(h(x) \neq h(y)) = 1 - \frac{1}{2^b} \quad \text{for any word } y \in P$$

$$\Rightarrow \quad \Pr(h(x) \neq h(z) : \forall z \in P) = \left(1 - \frac{1}{2^b}\right)^m$$

$$\Rightarrow \quad \Pr(h(x) = h(z) : \text{ for some } z \in P) = 1 - \left(1 - \frac{1}{2^b}\right)^m$$

$$\Rightarrow \quad \Pr(x \text{ is false positive}) = 1 - \left(1 - \frac{1}{2^b}\right)^m$$

$$\Rightarrow \quad \Pr(x \text{ is false positive}) \approx 1 - e^{-\frac{m}{2^b}}$$

$\Pr(x \text{ is false positive}) \approx 1 - e^{-\frac{m}{2^b}}$

Given constant probability p, to have $\Pr(x \text{ is false positive}) \leq p$ we must have $1 - e^{-\frac{m}{2^b}} \leq p$

$$\begin{aligned} &\Rightarrow \quad e^{-\frac{m}{2^b}} \geq 1 - p \\ &\Rightarrow \quad b \geq \lg_2\left(\frac{m}{\ln(1/(1-p))}\right) \end{aligned}$$

I.e., $b = \Omega(\ln(m))$.

On the other hand, if we use $b = 2\lg_2(m)$ and $x \notin P$, we have

$$\begin{aligned}
\Pr(x \text{ is false positive}) &= 1 - \left(1 - \frac{1}{2^b}\right)^m \\
&= 1 - \left(1 - \frac{1}{2^{2\lg_2 m}}\right)^m \\
&= 1 - \left(1 - \frac{1}{m^2}\right)^m \\
&\leq \frac{1}{m} \quad \text{(exercise)}
\end{aligned}$$

So, if we have 65000 words,

$$\Pr(x \text{ is false positive}) \leq \frac{1}{65000}$$

Bloom Filters

- is a generalization of the previous model.
- Using k hashing functions $h_i : U \to \{0, \ldots, n-1\}$, for $i = 1, \ldots, k$.

Initialization of bit vector:

- For $i \leftarrow 0$ to $n-1$ do $A[i] \leftarrow 0$
- For each $x \in P$ do

  For $i \leftarrow 1$ to k do

  $A[h_i(x)] \leftarrow 1$

Query of a word y:

- Return $(A[h_1(y)] \bigwedge A[h_2(y)]) \bigwedge \ldots \bigwedge A[h_k(y)])$

Can also leads to false positives.

Let $y \notin P$. What is $\Pr(y \text{ is false positive})$?

Let $i \in \{1, \ldots, k\}$.

$\Pr(A[h_i(y)] = 0) = \left(1 - \frac{1}{n}\right)^{km}$

$$\Rightarrow \quad \Pr(A[h_i(y)] = 1) = 1 - \left(1 - \frac{1}{n}\right)^{km}$$

$$\Rightarrow \quad \Pr(y \text{ is false positive}) = \Pr(1 = A[h_1(y)] = \ldots = A[h_k(y)])$$

$$\approx \left(1 - \left(1 - \frac{1}{n}\right)^{km}\right)^k \approx \left(1 - e^{-km/n}\right)^k$$

Note that

- If k is large the chance to have false positives is small, as more bits are tested.
- On the other hand, if k is large, more bits of A will be set to 1.

Given m and n, what is the best value for k ?

We know that if $y \notin P$,

$$\Pr(y \text{ is false positive}) \approx \left(1 - e^{-km/n}\right)^k.$$

Let $f = \left(1 - e^{-km/n}\right)^k$ and $g = k \ln(1 - e^{-km/n})$.
The minimum value of f can pode obtained deriving f:

$$\begin{aligned}
\frac{df}{dk} &= e^g \frac{dg}{dk} \\
&= e^g \left[(\ln(1 - e^{-\frac{km}{n}})) + \frac{k \cdot m}{n} \left(\frac{e^{-\frac{km}{n}}}{1 - e^{-\frac{km}{n}}} \right) \right] \\
&= 0, \quad \text{when } k = (\ln 2) \cdot \frac{n}{m}.
\end{aligned}$$

So, the smallest value for

$$\Pr(y \text{ is false positive})$$

is obtained when $k = (\ln 2) \cdot \frac{n}{m}$. I.e.,

$$\begin{aligned}
\Pr(y \text{ is false positive when } k = (\ln 2) \cdot \tfrac{n}{m}) &\approx \left(1 - e^{-\ln 2 \cdot \frac{m}{n} \cdot \frac{n}{m}}\right)^{(\ln 2) \cdot \frac{n}{m}} \\
&= \left(1 - \frac{1}{2}\right)^{(\ln 2) \cdot \frac{n}{m}} \\
&\approx 0.6185^{n/m}
\end{aligned}$$

Example: *In many applications, as the password example, it is sufficient to have* $\Pr(y \text{ is false positive}) \approx 2\%$.
To have this, we set $n \leftarrow 8 \cdot m$ *and* $k \leftarrow 6$*:*

$$f = \left(1 - e^{-km/n}\right)^k = \left(1 - e^{-6 \cdot \frac{1}{8}}\right)^6 = 0.0215\ldots$$

Breaking Symmetry

We can also consider the hashing functions $h(\cdot)$ in the set S, of n elements to

- Choose a leader S
  We may choose the largest $h(x)$ for $x \in S$.
- Choose a permutation of S
  We may sort S using $h(x)$

The problem are the collisions/ties!

What is the probability of collisions/ties occur ?

Given $x \in S$, what is the probability that $h(x) = h(y)$ for some other $y \in S$?

$$\Pr(h(x) = h(y), \text{ for fixed } y \in S,\ x \neq y) = \frac{1}{2^b}$$

$$\Rightarrow \quad \Pr(h(x) \neq h(y), \text{ for fixed } y \in S,\ x \neq y) = 1 - \frac{1}{2^b}$$

$$\Rightarrow \quad \Pr(h(x) \neq h(y), \text{ for any } y \in S,\ x \neq y) = \left(1 - \frac{1}{2^b}\right)^{n-1}$$

$$\Rightarrow \quad \Pr(h(x) = h(y), \text{ for some } y \in S,\ x \neq y) = 1 - \left(1 - \frac{1}{2^b}\right)^{n-1}$$

$$\leq \frac{n-1}{2^b} \text{ (exercise)}$$

What is the probability to exists $x, y \in S$ such that $h(x) = h(y)$?

Let $S = \{x_1, \ldots, x_n\}$.

$$\begin{aligned}
&\Pr(\exists x, y \in S : h(x) = h(y)) \\
&\quad = \Pr((\exists y \in S : h(x_1) = h(y)) \cup \ldots \cup (\exists y \in S : h(x_n) = h(y)) \\
&\quad \leq n\frac{n-1}{2^b} \\
&\quad \leq \frac{n^2}{2^b}
\end{aligned}$$

So, setting $b = 3\log_2 n$ we have

$$\begin{aligned}
\Pr(\exists x, y \in S : h(x) = h(y)) &\leq \frac{n^2}{2^{3\log_2 n}} \\
&= \frac{n^2}{n^3} = \frac{1}{n}
\end{aligned}$$

Random Graphs

We will consider the random graph models $G_{n,p}$ e $G_{n,N}$.

Notation:

- V set of vertices
- n number of vertices in V.
- K_n set of edges in a complete graph of n vertices

Some problems can be

- modeled/investigated using random graphs
- and have good properties with high probability

Model $G_{n,p}$: Graph with n vertices and each pair $\{i,j\}$ for $i,j \in V$ is an edge with probability p.

- Each graph with m edges has probability $p^m(1-p)^{\binom{n}{2}-m}$ to be in the graph.
- Expected number of edges is $\binom{n}{2} \cdot p$.
- Expected degree of a node is $(n-1) \cdot p$.

Model $G_{n,N}$: Graph with n vertices and exactly N edges.

To generate a graph in $G_{n,N}$:

1. $E \leftarrow \emptyset$
2. Repeat N times
3. Choose $e \in K_n \setminus E$ uniformly at random
4. $E \leftarrow E + e$
5. Return $G(V, E)$.

- There is $\binom{\binom{n}{2}}{N}$ distinct graphs, each one equally probable.

There are similarities with the Balls and Bins model:
Throwing edges into the graph (to generate $G_{n,N}$) is like throwing balls into bins. Each edge can be seen as two balls in two distinct bins.

Theorem: *Let $N = \frac{1}{2}(n \ln n + cn)$. Then, the probability that there exists an isolated vertex (degree 0) in $G_{n,N}$ converges to $e^{-e^{-c}}$.*

Proof.

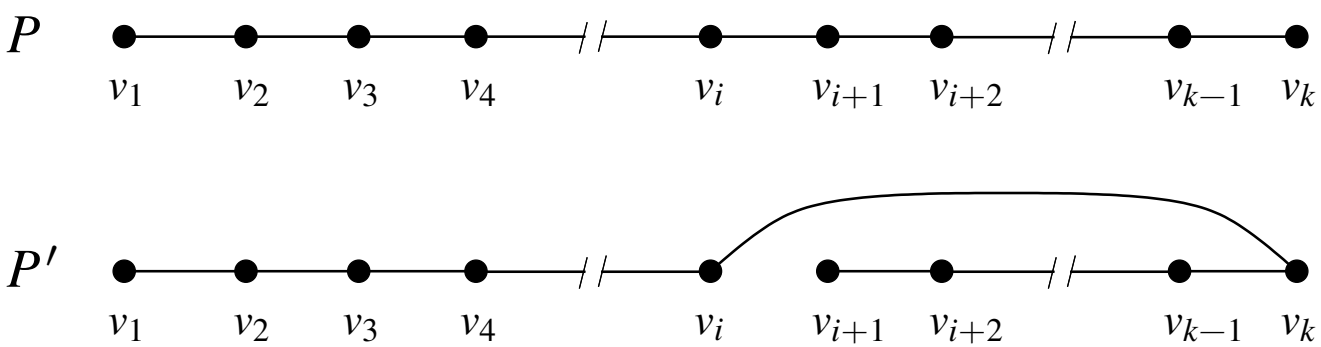
Exercise. Idea: In the coupon collector problem, after $n \ln n + cn$ balls, the probability to have empty bins converges to $e^{-e^{-c}}$. □

Hamiltonian Circle in Random Graphs

Find a Hamiltonian Circle (H.C.) in a graph is an NP-hard problem.
We describe an algorithm to find a H.C. in a random graph with high probability.
The algorithm do the following:

Rotation(P, i, k)

1. Let $P = (v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_{k-1}, v_k)$
2. Return $P' = (v_1, \ldots, v_{i-1}, v_i, v_k, v_{k-1}, \ldots, v_{i+1})$



Notation:

- $Head(P)$: last node in P, also called head of P.
  Ex.: If $P = (v_1, \ldots, v_k)$ then $Head(P) = v_k$.

- $Used(v)$: incident edges to v already considered by the algorithm. Initially empty.

- $NonUsed(v)$: non-used edges of v. Initially is the set of edges incidents to v.

We will see

- An Algorithm 1, that is natural, but hard to analyse.
- An Algorithm 2, that is less efficient but find H.C. with high probability for certain graphs (used/non-used edges model).
- An Algorithm 3, that uses Algorithm 2 and obatin H.C. with high probability in a graph $G_{n,p}$, for $p \geq \frac{40 \ln n}{n}$.

Algorithm 1 $(G = (V, E))$**, where** $|V| = n$**.**

1. $NonUsed(v) \leftarrow \{e \in E : \ e \text{ is edge incident to } v\}$
2. choose $v \in V$ uniformly at random
3. let $P \leftarrow (v)$ and v the head of P
4. while $NonUsed(Head(P)) \neq \emptyset$
5. let $P = (v_1, \ldots, v_k)$ where v_k is the head
6. let $\{v_k, u\}$ the first edge in $NonUsed(Head(P))$
7. remove (v_k, u) from $NonUsed(v_k)$ and from $NonUsed(u)$.
8. if $u \notin P$ then $P \leftarrow P \| u$ (and u becomes head)
9. else
10. let $u = v_i$, where $v_i \in P$
11. if $i = 1$ and $k = n$ then return the obtained H.C.
12. else let $P \leftarrow Rotation(P, i, k)$ (and v_{i+1} becomes the head)
13. return *fail*.

It is difficult to analyse Algorithm 1, as each step is dependent with the previous.

About Algorithm 2

- At each iteration, the current path is either increased, inverted or rotated.
- The algorithm consider that there are lists of *Used* and *NonUsed* edges for each head of a path.
- Initially the set $Used(v)$ is empty for each vertex v.
- Initially the set $NonUsed(v)$ is obtained choosing independently each one of the $n-1$ possible edges with probability q, for each vertex v. I.e., if $e = \{u, v\}$ it can also occur that $e \in NonUsed(v)$ and $e \notin NonUsed(u)$.

About Algorithm 2

- When the edge $e = \{u, v\}$ is removed from $NonUsed(v)$ do not remove the edge from $NonUsed(u)$ (in case it belongs to this set).

- Just after the list $NonUsed(v)$ is generated, it is shuffled, for each v.

Algorithm 2$(G = (V, E))$

where $|V| = n$, $NonUsed(v)$ and $Used(v)$ are part of the input, under the previous conditions.

1. choose v uniformly at random from V
2. let $P \leftarrow (v)$ and the head of P as vertex v

3. while $NonUsed(Head(P)) \neq \emptyset$ and has no H.C.
4. let $P = (v_1, \ldots, v_k)$ where v_k is head
5. execute *i)* or *ii)* or *iii)* (exactly one of them) with probability
6. $\frac{1}{n}$, $\frac{|Used(v_k)|}{n}$ and $1 - \frac{1}{n} - \frac{|Used(v_k)|}{n}$, respectively
7. *i)* revert the path P making v_1 as head
8. *ii)* choose a random edge $\{v_k, v_i\} \in Used(v_k)$
9. rotate and make v_{i+1} the new head
10. *iii)* let $\{v_k, u\}$ the first edge of $NonUsed(v_k)$.
11. if $u \notin P$ then let $u = v_{k+1}$ and v_{k+1} the new head
12. else if $u = v_i$, rotate with $\{v_k, v_i\}$ and make v_{i+1} head
13. (the loop finish if edge $\{v_n, v_1\}$ is chosen).
14. update $Used(v_k)$ and $NonUsed(v_k)$
15. return H.C. if found, or *fail* otherwise.

Lemma: *Consider Algorithm 2. Let x_t the head after the t-th step. Then, for all nodes u, whenever there exists a non-used edge in the head vertex,*

$$\Pr(x_{t+1} = u | x_t = u_t, x_{t-1} = u_{t-1}, \ldots, x_0 = u_0) = \frac{1}{n}$$

I.e., We can consider that the head is chosen randomly between all vertices in each step, independently from the previous steps.

Proof. Let $P = (v_1, \ldots, v_k)$. We can consider the possible cases, for the vertices in P and outside P.

- The unique manner that v_1 becomes head is reverting the list. In this case, the applied probability is $\frac{1}{n}$.
- If u is a vertex of the path and $\{v_k, u\}$ is edge in $Used(v_k)$ then the probability that $x_{t+1} = u$ is

$$\frac{|Used(v_k)|}{n} \cdot \frac{1}{|Used(v_k)|} = \frac{1}{n}$$

The first fraction is the probability to enter the case. The second fraction is the probability to choose a specific vertex by one a used edge.

- Now, suppose that we pick the edge $e \in NonUsed(v_k)$, where $e = (v_k, u)$

Although the edges have been chosen/generated with probability q, the probability that e is incident to one of the $n - 1 - |Used(v_k)|$ possible vertices is the same (given that we have e). So, the probability that $x_{t+1} = u$ is

$$\left(1 - \frac{1}{n} - \frac{|Used(v_k)|}{n}\right) \cdot \left(\frac{1}{n - 1 - |Used(v_k)|}\right) = \frac{1}{n}.$$

The first term is the probability to enter this case. The second is the probability to chose a specific vertex using a non-used edge. □

Generate a random graph in the model of Used/NonUsed edges

Let H a graph, in the model of *Used* and *NonUsed* edges, where each (possible) edge of a vertex v is inserted in $NonUsed(v)$ with probability $q = \frac{20 \ln n}{n}$.

We show that Algorithm 2 obtain a H.C. in H

- in $O(n \ln n)$ steps,
- with high probability: $1 - O(\frac{1}{n})$.

Lemma: *If Algorithm 2 make at least* $3n \ln n$ *iterations of step 3, then the probability to not find a H.C. is at most* $\frac{2}{n}$.

Proof.

Note that in this case, we suppose that the algorithm do not fail before this number of iterations.

First, we prove that with $2n \ln n$ iterations of step 3, we obtain all vertices of H with probability $\frac{1}{n}$.

The proof is basically the same of the Coupon Collector problem, where we want n different vertices (cupons) in the path.

In each iteration, the chance to take a vertex as head is the same: $\frac{1}{n}$.

The probability that a specific vertex is not chosed after $2n \ln n$ iterations is:

$$\left(1 - \frac{1}{n}\right)^{2n \ln n} \leq e^{-2 \ln n} = \frac{1}{n^2}.$$

So, the probability that some vertex is not chosed after $2n \ln n$ iterations is (bounded by the probability of the sum) $n \cdot \frac{1}{n^2} = \frac{1}{n}$

Now, we can show that with more $n \ln n$ iterations, the last vertex close with the first vertex. The probability to not close is:

$$\left(1 - \frac{1}{n}\right)^{n \ln n} \leq e^{-\ln n} = \frac{1}{n}.$$

Therefore, the probability that Algorithm 2 do not find a H.C. is at most $2 \cdot \frac{1}{n}$.

□

Lemma: *The probability that there exists a vertex v of H with less than $10 \ln n$ edges in $NonUsed(v)$ is at most $\frac{1}{n}$.*

Proof.

Let Y_v^e a binary variable to indicate if an edge e is chosen in $NonUsed(v)$ and $Y_v = \sum_e Y_v^e$ the number of edges in $NonUsed(v)$.

We have that $E[Y_v] = (n-1) \cdot q = (n-1) \cdot \frac{20 \ln n}{n} \geq 19 \ln n$, for n sufficiently large. So,

$$\begin{aligned} \Pr(Y \leq 10 \ln n) &= \Pr(Y \leq (1 - \frac{9}{19}) 19 \ln n) \\ &\leq e^{-(19 \ln n)(9/19)^2/2} \\ &\leq \frac{1}{n^2} \end{aligned}$$

Therefore, the probability that one of the vertices do not have $10 \ln n$ non used edges (by the prob. of the sum) is at most $n \cdot \frac{1}{n^2}$. □

Lemma: *The probability that Algorithm 2 remove at least* $9 \ln n$ *edges from* $NonUsed(v)$ *for some vertex* v *of* H *in* $3n \ln n$ *iterations is at most* $\frac{1}{n}$.

Proof.

Let $v \in V$. The number of edges removed from $NonUsed(v)$ is bounded by the number of times v was head.

The probability that v becomes head in an iteration is $\frac{1}{n}$. Let X the number of times vertex v was head.

$$E[X] = (3n \ln n) \cdot \frac{1}{n} = 3 \ln n.$$

So,

$$\begin{aligned}
\Pr(X \geq 9 \ln n) &= \Pr(X \geq (1+2)3 \ln n) \\
&< \left(\frac{e^2}{(1+2)^{(1+2)}} \right)^{3 \ln n} \\
&< \left(\frac{e^2}{e^3} \right)^{3 \ln n} \\
&\leq \frac{1}{n^2}
\end{aligned}$$

Therefore, the probability that one of the vertices have at least $9 \ln n$ removed non-used edges (by the prob. of the sum) is at most $n \cdot \frac{1}{n^2}$. □

Theorem: *Algorithm 2 obtain a H.C. of H in $3n \ln n$ iterations with probability $1 - O(\frac{1}{n})$ (in the used/non-used model).*

Proof.

- The probability that there exists a vertex v in H with $|NonUsed(v)| \leq 10 \ln n$ is at most $\frac{1}{n}$.
- The probability the algorithm remove from a vertex v at least $9 \ln n$ edges from $NonUsed(v)$ in $3n \ln n$ iterations, is at most $\frac{1}{n}$.

• So, the probability that algorithm fail in $3n \ln n$ iterations for having set $NonUsed(v)$ empty is (by the sum) at most $\frac{2}{n}$.

• The probability the algorithm do not find a H.C. of H in $3n \ln n$ iterations is at most $\frac{2}{n}$.

Therefore, the probability of fail in $3n \ln n$ iterations is at most $\frac{4}{n}$. □

Algorithm 3($G_{n,p}$), where $p \geq \frac{40 \ln n}{n}$.

1. Construct a graph H as follows:
2. Let $q \in [0,1]$ be such that $p = 2q - q^2$.
3. For each edge $e = \{u, v\}$ in $G_{n,p}$ execute
4. *i)* or *ii)* or *iii)* (exactly one of them) with probability
5. $p_u = \frac{q(1-q)}{2q-q^2}$, $p_v = \frac{q(1-q)}{2q-q^2}$ and $p_{uv} = \frac{q^2}{2q-q^2}$, respectively
6. *i)* Put e in the set of non-used edges of u.
7. *ii)* Put e in the set of non-used edges of v.
8. *iii)* Put e in the set of non-used edges of u and v
9. Execute the Algorithm 2 in H.
10. If obtained H.C. then return the circuit.
11. Else return *fail*.

First, note that the probabilities of step 5 are well defined, as $p_u + p_v + p_{uv} = 1$.

Fact: *If q is obtained from p, as made in Algorithm 3, then $q \geq \frac{20 \ln n}{n}$.*

Proof.

$$\begin{aligned} q &\geq q - \frac{q^2}{2} \\ &= \frac{2q - q^2}{2} = \frac{p}{2} \\ &\geq \frac{40 \ln n}{2n} = \frac{20 \ln n}{n} \end{aligned}$$

□

Lemma: *An edge $e = \{u, v\}$ is inserted in $NonUsed(u)$ with probability q and $NonUsed(v)$ with probability q. Furthermore, both insertions are made in an independent way.*

Proof.

The edge e will belong to $NonUsed(u)$ if e belongs to $G_{n,p}$ and if e is inserted in *i)* or *iii)*. So,

$$\begin{aligned}
\Pr(e \in NonUsed(u)) &= p \cdot (p_u + p_{uv}) \\
&= (2q - q^2) \cdot \left(\frac{q(1-q)}{2q-q^2} + \frac{q^2}{2q-q^2} \right) \\
&= q
\end{aligned}$$

The proof that $\Pr(e \in NonUsed(v)) = q$ is analogous.

Now, we prove that the insertions of $e = \{u, v\}$ in $NonUsed(u))$ and in $NonUsed(v))$ are independent.

$$
\begin{aligned}
&\Pr(e \in NonUsed(u) \ \cap \ e \in NonUsed(v)) \\
&\qquad = \quad p \cdot p_{uv} \\
&\qquad = \quad (2q - q^2) \cdot \frac{q^2}{2q - q^2} \\
&\qquad = \quad q^2 \\
&\qquad = \quad \Pr(e \in NonUsed(u)) \cdot \Pr(e \in NonUsed(v))
\end{aligned}
$$

As the graph H generated by Algorithm 3 satisfy all required conditions by Algorithm 2, this obtain a H.C. from H, and consequently from G, with high probability.

Theorem: *The Algorithm 3 obtain a H.C. for the graph $G_{n,p}$, where $p \geq \frac{40 \ln n}{n}$, with probability $1 - O(\frac{1}{n})$.*

Corollary: *A graph in $G_{n,p}$, for $p \geq \frac{40 \ln n}{n}$, has a H.C. with high probability.*

The Probabilistic Method

Method to prove existence of an object.

Idea: To prove the existence of an object with certain properties, it is sufficient to prove that the probability of its existence is positive.

The Basic Counting Argument

Theorem: *If* $\binom{n}{k}2^{-\binom{k}{2}+1} < 1$ *then it is possible to color the edges of a graph* K_n *with two colors such that there is no subgraph monochromatic* K_k.

Proof.

- Consider a random coloring, each edge of K_n colored with probability $\frac{1}{2}$.
- Consider all the $\binom{n}{k}$ different cliques with k vertices in an order $i = 1, \ldots, \binom{n}{k}$.
- Let A_i the event that clique i is monochromatic. Then,

$$\Pr(A_i) = \frac{2}{2^{\binom{k}{2}}} = 2^{-\binom{k}{2}+1}$$

The probability that one subgraph K_k is monochromatic is less than 1:

$$\begin{aligned}
\Pr\left(\cup_{i=1}^{\binom{n}{k}} A_i\right) &\leq \sum_{i=1}^{\binom{n}{k}} \Pr(A_i) \\
&= \binom{n}{k} 2^{-\binom{k}{2}+1} \\
&< 1
\end{aligned}$$

So, the probability that there is no monochromatic K_k is

$$\Pr\left(\cap_{i=1}^{\binom{n}{k}} \overline{A_i}\right) = 1 - \Pr\left(\cup_{i=1}^{\binom{n}{k}} A_i\right) > 0$$

As such probability is not null, there must exists such coloring. □

Example: *It is possible to color a K_{1000} without monochromatic K_{20}.*
For $n \leq 2^{k/2}$ and $k \geq 3$, we have

$$\begin{aligned}
\binom{n}{k} 2^{-\binom{k}{2}+1} &\leq \frac{n^k}{k!} \cdot 2^{-\frac{k(k-1)}{2}+1} \\
&\leq \frac{(2^{\frac{k}{2}})^k \cdot 2^{-\frac{k(k-1)}{2}+1}}{k!} \\
&= \frac{2^{\frac{k}{2}+1}}{k!} \\
&< 1
\end{aligned}$$

From the previous theorem, there exists a coloring of K_{1000} without monochromatic K_{20}.

How do we can find such a coloring ?

Now, we will find a coloring of K_{1000} without monochromatic K_{20}.

From the previous theorem, the probability to have a random coloring with monochromatic K_k is at most

$$\binom{n}{k} 2^{-\binom{k}{2}+1} \leq \frac{2^{\frac{k}{2}+1}}{k!} = \frac{2^{\frac{20}{2}+1}}{20!} < 8.5 \times 10^{-16}$$

Then, a random coloring already leads to a coloring without monochromatic K_{20} with good probability (Monte Carlo Algorithm).

If we need a Las Vegas algorithm, for constant k, it is sufficient to repeat the Monte Carlo algorithm and stop when we obtain a coloring without monochromatic K_k.

The Expectation Argument

This argument is based in the following lemma:

Lemma: *Suppose that we have a sample space $\mathcal{S}$ and r.v. X defined in $\mathcal{S}$ such that $E[X] = \mu$. Then*

$$\Pr(X \geq \mu) > 0 \quad \textit{and} \quad \Pr(X \leq \mu) > 0.$$

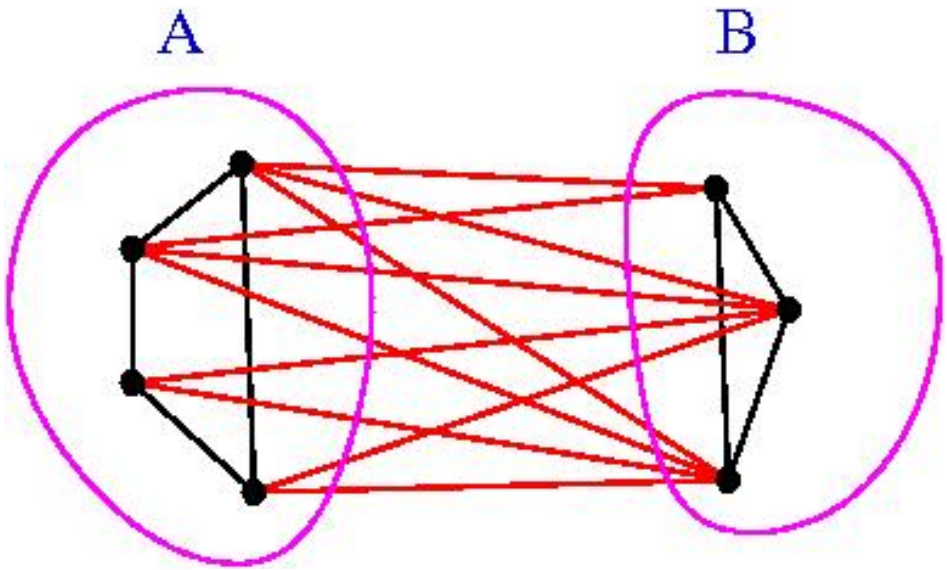
Proof. Suppose that $\Pr(X \geq \mu) = 0$. Then,

$$\begin{aligned} \mu &= \sum_x x \cdot \Pr(X = x) = \sum_{x<\mu} x \cdot \Pr(X = x) \\ &< \sum_{x<\mu} \mu \cdot \Pr(X = x) \\ &= \mu \quad \text{Contradiction!} \end{aligned}$$

The proof that $\Pr(X \leq \mu) = 0$ is analogous. □

Application: MaxCut

Problem: *MaxCut: Given a non-oriented graph $G = (V, E)$, find a partition (A, B) of V such that the number of edges in (A, B) is maximum.*



Theorem: *The MaxCut problem is NP-hard.*

Theorem: *Given a non-oriented graph $G = (V, E)$ with n vertices and m edges, there exists a partition of V into sets A and B such that there exists at least $\frac{m}{2}$ edges in the cut (A, B) (set of edges connecting vertices from A and B).*

Proof. Start A and B as empty sets and add each vertex of V to A or B with probability $\frac{1}{2}$.

Let $E = \{e_1, \ldots, e_m\}$. Let X_i a binary variable with value 1 if and only if $e_i \in (A, B)$ and $X = \sum_{i=1}^{m} X_i$ (X is the size of the cut).

Clearly, $E[X] = \frac{m}{2}$. Therefore, from the previous lemma, there must exists a cut with at least $\frac{m}{2}$ edges. □

Now, consider obtaining a Las Vegas algorithm to find a cut with at least $\frac{m}{2}$ edges.

First, obtain a bound to $p = \Pr(|C| \geq \frac{m}{2})$

$$
\begin{aligned}
\frac{m}{2} &= E[Y] \\
&= \sum_{i \leq \frac{m}{2}-1} i \cdot \Pr(Y = i) + \sum_{i \geq \frac{m}{2}} i \cdot \Pr(Y = i) \\
&\leq \sum_{i \leq \frac{m}{2}-1} \left(\frac{m}{2} - 1\right) \cdot \Pr(Y = i) + \sum_{i \geq \frac{m}{2}} m \cdot \Pr(Y = i) \\
&= \left(\frac{m}{2} - 1\right) \cdot \Pr(Y < \frac{m}{2}) + m \cdot \Pr(Y \geq \frac{m}{2}) \\
&= \left(\frac{m}{2} - 1\right) \cdot (1 - p) + m \cdot p.
\end{aligned}
$$

I.e.,

$$\frac{m}{2} \leq \left(\frac{m}{2} - 1\right) \cdot (1-p) + m \cdot p.$$

Isolating p, we have

$$p \geq \frac{1}{\frac{m}{2}+1}.$$

Therefore, the expected number of steps of the Las Vegas algorithm to obtain a cut of size $\frac{m}{2}$ is at most $\frac{m}{2} + 1$.

Application: MaxSat

Def.: *A formula is in Conjunctive Normal Form (CNF) if it is a conjunction of clauses (by logic $\wedge$'s), each one is a disjunction of literals (by logic $\vee$'s).*

Problem: *MaxSat: Given a boolean formula in CNF, find an assignment to the variables in such a way to maximize the number of true clauses.*

Example: *Does the formula below is satisfiable ?*

$$\phi = (x_1 \vee \overline{x_2} \vee \overline{x_3}) \wedge (\overline{x_1} \vee \overline{x_3}) \wedge (x_1 \vee x_2 \vee x_4) \wedge (x_4 \vee \overline{x_3}) \wedge (x_4 \vee \overline{x_1})$$

Theorem: *The MaxSat problem is NP-hard.*

Theorem: *Given a set of clauses, $\mathcal{C} = \{C_1, \ldots, C_m\}$, let k_i the number of literals in the clause C_i and $k = \min_{i \in [m]} k_i$. Then, there exists an assignment that satisfy at least*

$$\sum_{i=1}^{m} \left(1 - \frac{1}{2^{k_i}}\right) \geq m \left(1 - \frac{1}{2^k}\right) \quad \textit{clauses.}$$

Proof.

Assign each variable x_j with true or false with probability $\frac{1}{2}$.

Let $Y = \sum_{i=1}^{m} X_i$ the number of satisfied clauses and let X_i a r.v. that indicates if clause C_i is satisfied.

$$E[Y] = \sum_{i=1}^{m} E[X_i] = \sum_{i=1}^{m} \left(1 - \frac{1}{2^{k_i}}\right) \geq m \left(1 - \frac{1}{2^k}\right)$$

The result follows from the previous lemma. □

Derandomization Using Conditional Expectations

If we can compute conditional expectations in the choices made by the algorithm efficiently, we can derandomize the algorithm, obtaining a deterministic algorithm.

Consider the MaxCut problem. The algorithm considered assign each vertex to one of the parts with probability $\frac{1}{2}$ and finds a cut with expected size $E[Y] \geq \frac{m}{2}$, where Y is the expected number of edges in the cut and m is the number of edges in the graph.

The derandomized algorithm is such that

- it considers an arbitrary order of the vertices $v_1, \ldots, v_n$ and a deterministic assignment of them
- to chose an assignment of a vertex, consider the calculation of conditional expectations

Idea: let x_i the indication of assignment of v_i to A or B in the i-th iteration of the algorithm, then the algorithm chose x_i such that

$$\frac{m}{2} \leq E[Y] \leq E[Y|x_1] \leq E[Y|x_1, x_2] \leq \ldots \leq E[Y|x_1, x_2, \ldots, x_n]$$

The decision x_i is based in the calculation of conditional expectations.

Suppose that $x_1, \ldots, x_k$ were already decided such that

$$\frac{m}{2} \leq E[Y] \leq E[Y|x_1] \leq \ldots \leq E[Y|x_1, \ldots, x_k]$$

Let Y_{k+1} r.v. that indicate the assignment of v_{k+1} to one of the parts with probability $\frac{1}{2}$:

$$\begin{aligned} E[Y|x_1, \ldots, x_k] \quad = \quad & \frac{1}{2} E[Y|x_1, \ldots, x_k, Y_{k+1} = A] + \\ & \frac{1}{2} E[Y|x_1, \ldots, x_k, Y_{k+1} = B] \end{aligned}$$

Once we know how to calculate conditional expectations efficiently we verify which one is larger

$$E[Y|x_1, \ldots, x_k, Y_{k+1} = A] \quad \text{or} \quad E[Y|x_1, \ldots, x_k, Y_{k+1} = B]$$

If we have

$$E[Y|x_1,\ldots,x_k,Y_{k+1}=A] \leq E[Y|x_1,\ldots,x_k,Y_{k+1}=B]$$

then we make x_{k+1} as the assignment of v_{k+1} to B, else to A.

$$\begin{aligned}
\text{So,}\quad E[Y|x_1,\ldots,x_k] &= \frac{1}{2}E[Y|x_1,\ldots,x_k,Y_{k+1}=A] + \\
&\quad \frac{1}{2}E[Y|x_1,\ldots,x_k,Y_{k+1}=B] \\
&\leq \frac{1}{2}E[Y|x_1,\ldots,x_k,x_{k+1}] + \\
&\quad \frac{1}{2}E[Y|x_1,\ldots,x_k,x_{k+1}] \\
&= E[Y|x_1,\ldots,x_k,x_{k+1}]
\end{aligned}$$

Repeating this process, we obtain a deterministic algorithm that guarantee at least $\frac{m}{2}$ edges.

Simplification: Consider an assignment of v_{k+1}.

- Let (A', B') the cut defined by vertices $v_1, \ldots, v_k$.
- Each edge that connect v_{k+1} to the vertices of $A' \cup B'$ have probability $\frac{1}{2}$ to enter in the cut.
- Let N'_A and N'_B the number of edges that connect v_{k+1} to A' and B', resp.
- In the probabilistic algorithm, the contribution of v_{k+1} give us an expected number of $\frac{N'_A + N'_B}{2}$ more edges in the cut.
- The conditional expectation indicate that the set will give us more number of edges that will be added to the cut.

Therefore, the derandomized algorithm is basically a greedy algorithm that assign the vertex to the part which leads to a larger number of edges in the cut.

Theorem: *The MaxCut greedy algorithm guarantee at least $\frac{m}{2}$ edges in the cut.*

Sample and Modify

Idea: Break the argument in two stages.

- In the first stage, build a random structure which may not have the desired properties.
- In the second stage, modify the structure to obtain the desired propertity.

Application: Independent Set

Def.: *Given a graph $G = (V, E)$, a set $S \subseteq V$ is said independent if there are no edges in E connecting two edges in S.*

Problem: *Maximum Independent Set: Given a graph G, find an independent set of G of maximum size.*

Theorem: *The Independent Set Problem is NP-Hard.*

Theorem: *Let $G = (V, E)$ a graph with n vertices and m edges. Then, G has an independent set with at least $\frac{n^2}{4m}$ vertices.*

Proof.

Let $d = \frac{2m}{n}$ the average degree of the vertices in G.

Algorithm: *Algorithm $Ind(G)$*

1. *Remove each vertex of G and its edge with probability $1 - \frac{1}{d}$.*
2. *For each remaining edge erase one of the extremes.*
3. *Return the obtained independent set.*

Let X and Y the number of remaining vertices and edges, after step 1, resp. Each vertex survive with probability $\frac{1}{d}$, so

$$E[X] = \frac{n}{d}.$$

As we have m edges, we have

$$E[Y] = m \cdot \left(\frac{1}{d}\right)^2 = \frac{nd}{2} \cdot \frac{1}{d^2} = \frac{n}{2d}$$

As each edge of Y is removed in step 2, removing one of the extremes, the number of vertices Z of the independent set is such that

$$E[Z] = E[X - Y] = \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d} = \frac{n^2}{4m}$$

Application: Graphs with large girth

Def.: *Given a graph G, its girth is the size of the smallest cycle.*

Theorem: *For each $k \geq 3$, there is a graph with n vertices and at least $\frac{n^{1+\frac{1}{k}}}{4}$ edges and girth at least k.*

Proof.

Consider a random graph $G_{n,p}$ with $p = n^{\frac{1}{k}-1}$.

Let X be the number of edges in the graph. Then,

$$E[X] = p \cdot \binom{n}{2} = n^{\frac{1}{k}-1} \cdot \frac{n(n-1)}{2} = \left(1 - \frac{1}{n}\right) \frac{n^{\frac{1}{k}+1}}{2}$$

Let Y the number of cycles of size smaller than or equal to $k-1$.
Each cycle of size i, for $3 \leq i \leq k-1$ has probability p^i to occur.
Furthermore, there are N_i possible cycle of size i,

$$N_i = \binom{n}{i} \frac{(i-1)!}{2}$$

So,

$$\begin{aligned} E[Y] &= \sum_{i=3}^{k-1} N_i \cdot p^i = \sum_{i=3}^{k-1} \binom{n}{i} \frac{(i-1)!}{2} \cdot p^i \\ &\leq \sum_{i=3}^{k-1} \frac{n^i}{i!} \frac{(i-1)!}{2} \cdot p^i \\ &\leq \sum_{i=3}^{k-1} n^i \cdot p^i \end{aligned}$$

$$
\begin{aligned}
E[Y] &\leq \sum_{i=3}^{k-1} n^i \cdot p^i = \sum_{i=3}^{k-1} n^i \cdot n^{(\frac{1}{k}-1)i} \\
&= \sum_{i=3}^{k-1} n^{i/k} \\
&< k \cdot n^{(k-1)/k}
\end{aligned}
$$

Now, remove one edge from each of these small cycles. Now, the resulting graph has girth at least k.

Let Z the number of remaining edges:

$$\begin{aligned} E[Z] &\geq E[X-Y] \\ &\geq \left(1-\frac{1}{n}\right)\frac{n^{\frac{1}{k}+1}}{2} - k\cdot n^{(k-1)/k} \\ &= \frac{n^{\frac{1}{k}+1}}{2} - \frac{n^{\frac{1}{k}}}{2} - k\cdot n^{1-\frac{1}{k}} \\ &\geq \frac{n^{\frac{1}{k}+1}}{4}, \quad \text{for sufficiently large } n. \end{aligned}$$

METHOD OF THE SECOND MOMENT

Method based in the following theorem:

Theorem: *If X is an integer r.v. then*

$$\Pr(X = 0) \leq \frac{\text{var}[X]}{(E[X])^2}.$$

Proof.

$$\Pr(X = 0) \leq \Pr(|X - E[X]| \geq E[X]) \leq \frac{\text{var}[X]}{(E[X])^2}$$

□

Application: Threshold in Random Graphs

Consider the $G_{n,p}$ model and the property Π. It is frequent to have a function $f(n)$ such that

- When p is a bit smaller than $f(n)$ then almost all graphs do not satisfy Π.
- When p is a bit larger than $f(n)$ then almost all graphs satisfy Π.

Given property Π over certain structure, we denote by $\Pi(S)$ the event that S has the property Π.

Theorem: *Consider the graph $G_{n,p}$, with $p = f(n)$ and property Π that indicate if a graph has a clique of size 4 or more.*

a) *If $f(n) = o\left(\frac{1}{n^{2/3}}\right)$ then for $\varepsilon > 0$ and n sufficiently large,* $\Pr(\Pi(G)) \leq \varepsilon$.

b) *If $f(n) = \omega\left(\frac{1}{n^{2/3}}\right)$ then for $\varepsilon > 0$ and n sufficiently large,* $\Pr(\overline{\Pi(G)}) \leq \varepsilon$.

Proof.

Let $C_1, \ldots, C_{\binom{n}{4}}$ all sets with exactly 4 vertices.

Let X_i r.v. that indicates if C_i is a clique in G and $X = \sum_i X_i$.

Consider the case a) and $p = f(n) = o\left(\frac{1}{n^{2/3}}\right)$.

So,

$$\begin{aligned} E[X] &= \binom{n}{4} p^6 \\ &\leq \frac{n^4}{4!} p^6 \\ &= o(1) \end{aligned}$$

So, for sufficiently large n, $E[X] < \varepsilon$.

Therefore, as X is non-negative r.v.,

$$\begin{aligned}
\Pr(X \geq 1) &= \Pr(X = 1) + \Pr(X = 2) + \cdots \\
&\leq 1 \cdot \Pr(X = 1) + 2 \cdot \Pr(X = 2) + \cdots \\
&= E[X] \\
&< \varepsilon
\end{aligned}$$

So, the probability that the graph has Π is smaller than ε.

Consider the case b) and $p = f(n) = \omega\left(\frac{1}{n^{2/3}}\right)$.

Analogously, we show that

$$E[X] \to \infty \quad \text{when } n \to \infty.$$

This is not sufficient to prove that $G_{n,p}$ has K_4 with high probability.

We will use the theorem of the second moment. It is sufficient to prove that $\text{var}[X] = o((E[X])^2)$. With this, we prove that

$$\Pr(X = 0) \leq \frac{\text{var}[X]}{(E[X])^2} = o(1).$$

Lemma: *Let Y_i, for $i = 1, \ldots, m$, binary r.v.s and $Y = \sum_i Y_i$. Then* $\text{var}[Y] \leq E[Y] + \sum_{i \neq j} \text{cov}(Y_i, Y_j)$.

Proof.

For any sequence of r.v.s $Y_1, \ldots, Y_m$ we have

$$\text{var}[Y] = \text{var}[\sum_i Y_i] = \sum_i \text{var}[Y_i] + \sum_{i \neq j} \text{cov}(Y_i, Y_j).$$

When Y_i is binary r.v. we have $E[Y_i^2] = E[Y_i]$.

Therefore

$$\text{var}[Y_i] = E[Y_i^2] - (E[Y_i])^2 \leq E[Y_i^2] = E[Y_i]$$

So,

$$\sum_i \text{var}[Y_i] \leq \sum_i E[Y_i] = E[Y].$$

□

Now, we calculate the $\text{cov}(X_i, X_j)$.

Note that

$\text{cov}(X_i, X_j) = E[(X_i - E[X_i]) \cdot (X_j - E[X_j])] = E[X_i \cdot X_j] - E[X_i] \cdot E[X_j]$.

- If $|C_i \cap C_j| = 0$ or $|C_i \cap C_j| = 1$ are disjoint cliques then X_i and X_j are independent and $\text{cov}(X_i, X_j) = 0$.
- If $|C_i \cap C_j| = 2$

$$\text{cov}(X_i, X_j) = E[X_i \cdot X_j] - E[X_i] \cdot E[X_j] \leq E[X_i \cdot X_j] = p^{11},$$

and there are $\binom{n}{6}$ sets fo 6 vertices, each one of these with $\binom{6}{2;2;2}$ possibilities.

- If $|C_i \cap C_j| = 3$, in the same manner, we have

$$\text{cov}(X_i, X_j) \leq p^9,$$

and there are $\binom{n}{5}$ sets of 5 vertices and each one of these with $\binom{5}{3;1;1}$ possibilities.

$$\begin{aligned}
\text{var}[X] &\leq E[X] + \sum_{i \neq j} \text{cov}(X_i, X_j) \\
&\leq \binom{n}{4} \cdot p^6 + \binom{n}{6}\binom{6}{2;2;2} p^{11} + \binom{n}{5}\binom{5}{3;1;1} p^9 + 0 \\
&= o(n^8 p^{12})
\end{aligned}$$

As $(E[X])^2 = (\binom{n}{4} \cdot p^6)^2 = \Theta(n^8 p^{12})$, we have that

$$\Pr(X = 0) \leq \frac{\text{var}[X]}{(E[X])^2} = o(1).$$

□

The Conditional Expectation Inequality

- For sum of Bernoulli r.v.
- Easier to apply than the Second Moment Method.

Theorem: *Let $X = \sum_{i=1}^{n} X_i$, where X_i is binary r.v. Then*

$$\Pr(X > 0) \geq \sum_{i=1}^{n} \frac{\Pr(X_i = 1)}{E[X|X_i = 1]}.$$

Obs.: r.v.s X_i's do not need to be independent.

Proof.

Let $Y = \frac{1}{X}$ if $X > 0$ and $Y = 0$ otherwise. Then

$$\Pr(X > 0) = E[X \cdot Y] = \sum_{i=1}^{n} E[X_i \cdot Y]$$

$$= \sum_{i=1}^{n} (E[X_i \cdot Y|X_i = 1] \cdot \Pr(X_i = 1) + E[X_i \cdot Y|X_i = 0] \cdot \Pr(X_i = 0))$$

$$= \sum_{i=1}^{n} E[Y|X_i = 1] \cdot \Pr(X_i = 1)$$

So,

$$\begin{aligned}
\Pr(X > 0) &= \sum_{i=1}^{n} E[Y|X_i = 1] \cdot \Pr(X_i = 1) \\
&= \sum_{i=1}^{n} E[\frac{1}{X}|X_i = 1] \cdot \Pr(X_i = 1) \\
&\geq \sum_{i=1}^{n} \frac{1}{E[X|X_i = 1]} \cdot \Pr(X_i = 1)
\end{aligned}$$

where the last inequality follows from the Jensen Inequality and the fact that $f(x) = \frac{1}{x}$ is convex. □

An alternative proof for part *b)* of the theorem:

Theorem: *Consider a graph $G_{n,p}$, with $p = f(n)$ and a property Π that indicates if the graph has a clique of size at least 4.*

a) *If $f(n) = o\left(\frac{1}{n^{2/3}}\right)$ then for $\varepsilon > 0$ and sufficiently large n, $\Pr(\Pi(G)) \leq \varepsilon$.*

b) *If $f(n) = \omega\left(\frac{1}{n^{2/3}}\right)$ then for $\varepsilon > 0$ and sufficiently large n, $\Pr(\overline{\Pi(G)}) \leq \varepsilon$.*

- Let $C_1, \ldots, C_{\binom{n}{4}}$ all set of 4 vertices.
- Let X_i r.v. that indicates if C_i is a clique in G and let $X = \sum_i X_i$.
- In case *b)*, we have $p = f(n) = \omega\left(\frac{1}{n^{2/3}}\right)$.
- To use the previous theorem, we calculate $\Pr(X_i = 1)$ and $E[X|X_i = 1]$.
- For a set C_i, we have

$$\Pr(X_i = 1) = p^6$$

We calculate $E[X|X_j = 1]$. To this end, we consider each $\Pr(X_i = 1|X_j = 1)$, when C_i intersects or not set C_j.

- There are $\binom{n-4}{4}$ sets C_i that do not intersect C_j. Each one with probability p^6.
- There are $\binom{4}{1}\binom{n-3}{3}$ sets C_i that intersect C_j in exactly 1 vertex. Each one with probability p^6.
- There are $\binom{4}{2}\binom{n-4}{2}$ sets C_i that intersect C_j in exactly 2 vertices. Each one with probability p^5.
- There are $\binom{4}{3}\binom{n-4}{1}$ sets C_i that intersect C_j in exactly 3 vertices. Each one with probability p^3.

Summing up, we have

$$E[X|X_j = 1] = \sum_{i=1}^{\binom{n}{4}} E[X_i|X_j = 1]$$
$$= 1 + \binom{n-4}{4}p^6 + 4\binom{n-4}{3}p^6 + 6\binom{n-4}{2}p^5 + 4\binom{n-4}{1}p^3.$$

Applying the theorem, we have

$$\Pr(X > 0) \geq \frac{\binom{n}{4}p^6}{1 + \binom{n-4}{4}p^6 + 4\binom{n-4}{3}p^6 + 6\binom{n-4}{2}p^5 + 4\binom{n-4}{1}p^3}.$$

That tends to 1 as $n \to \infty$