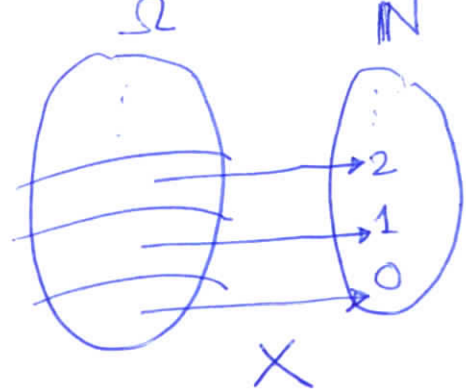
Random Variables and Expectation.

Def A random variable on space $(\Omega, \Pr)$ is a map $X : \Omega \to \mathbb{R}$.

Typically $X : \Omega \to \{0, 1, 2, \ldots\}$.

Def $\Pr[X = j] = \Pr[X^{-1}(j)]$.



Def Expectation of a r.v.

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} \Pr[\omega] \cdot X(\omega)$$

$$= \sum_{j=0,1,2,\ldots} j \cdot \Pr[X = j].$$

Note If $(\Omega, \Pr)$ is uniform, then $\mathbb{E}[X] = \frac{1}{|\Omega|} \cdot \sum_{\omega \in \Omega} X(\omega)$ is just the average.

Examples

① Ω = Students.

X = Height. $\qquad \mathbb{E}[X]$ = Average height of student in class.

② $\Omega = \{H, T\}^n$.

$X(\omega)$ = # heads in the seq. of tosses ω.

$$\mathbb{E}[X] = \sum_{j=0}^{n} j \cdot \Pr[X = j]$$

$$= \sum_{j=0}^{n} j \cdot \frac{\binom{n}{j}}{2^n}$$

$$= \frac{n}{2} \qquad !$$

Linearity of Expectation

Fact If X, Y are r.v.'s on $(\Omega, \Pr)$ then so is $X+Y$.

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Proof $\mathbb{E}[X+Y] = \sum_{\omega \in \Omega} (X+Y)(\omega)$

$$= \sum_{\omega \in \Omega} X(\omega) + Y(\omega)$$

$$= \sum_{\omega \in \Omega} X(\omega) + \sum_{\omega \in \Omega} Y(\omega)$$

$$= \mathbb{E}[X] + \mathbb{E}[Y].$$

$\blacksquare$

Example $\Omega = \{H, T\}^n$. $\quad X = \#\text{heads}.$

Let X_i be a random variable ($1 \le i \le n$, "indicator")

s.t. $X_i(\omega) = \begin{cases} 1 & \text{if } i^{th} \text{ toss is Heads} \\ 0 & \text{otherwise.} \end{cases}$

$\therefore \quad X = X_1 + X_2 + \cdots + X_n$

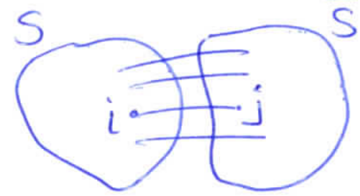
$\therefore \quad \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \cdot \frac{1}{2}.$

$\mathbb{E}[X_i] = 1 \cdot \Pr[X_i = 1] + 0 \cdot \Pr[X_i = 0] = \frac{1}{2}.$

$\blacksquare$

Finding Cuts in Graphs (MAX-CUT)

Def In an undirected graph $G(V,E)$, $V=\{1,\cdots,n\}$, a cut $(S,\bar{S})$ is a partition of V into disjoint sets S and $\bar{S}$.



Def "Edges cut".

$$e(S,\bar{S}) = \{(i,j) \in E \mid i \in S,\ j \in \bar{S}\}.$$

Problem Given $G(V,E)$, find a cut so as to maximize $|e(S,\bar{S})|$. NP-complete.

Theorem In a graph $G(V,E)$, $|E| = m$, there always exists a cut $(S,\bar{S})$ s.t.

$$|e(S,\bar{S})| \geq \frac{m}{2}.$$

Such a cut can be found in polynomial time by "a randomized algorithm."

Idea 1 We show that for a "random cut", expected #edges cut is $\frac{m}{2}$.

"Random cut" = Each vertex placed on S or $\bar{S}$ side w.p. $\frac{1}{2}$, independently

② Then we show that the expectation is achieved with reasonable prob. ($\geq \frac{1}{n^2}$).

③ With (say) n^3 repeated trials, one can get such a cut (i.e. of size $\geq \frac{m}{2}$) with probability $1 - 2^{-n}$.

<u>Algorithm / Analysis</u>

Construct a cut $(S, \bar{S})$ as follows.

Place each $i \in V$ into S or $\bar{S}$ w.p. $1/2$ each independently.

Let $X_{(i,j)}$ be indicator random var.

$$X_{(i,j)} = \begin{cases} 1 & \text{if edge } (i,j) \text{ is cut} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}[X_{(i,j)}] = \Pr[X_{(i,j)} = 1] = \frac{1}{2}.$$

Now we look at the r.v. $|e(S,\bar{S})|$.

$$|e(S,\bar{S})| = \sum_{(i,j)\in E} X_{(i,j)}$$

$$\therefore \mathbb{E}[|e(S,\bar{S})|] = \sum_{(i,j)\in E} \mathbb{E}[X_{(i,j)}]$$

$$= \frac{m}{2}. \qquad |E| = m.$$

<u>Claim</u> With probability at least $\geq \frac{1}{n^2}$, the cut $(S,\bar{S})$ output has $|e(S,\bar{S})| \geq \frac{m}{2}$.

<u>Proof</u> Let p = prob that $|e(S,\bar{S})| \geq \frac{m}{2}$.

$\therefore$ With prob $1-p$, $|e(S,\bar{S})| \leq \frac{m}{2} - \frac{1}{2}$.

With prob p, $\boxed{|e(S,\bar{S})| \leq m}$

$$\therefore \frac{m}{2} = \mathbb{E}[|e(S,\bar{S})|]$$

$$\leq (1-p)\left(\frac{m}{2} - \frac{1}{2}\right) + p \cdot m$$

$$\therefore \cancel{\frac{m}{2}} \leq \cancel{\frac{m}{2}} - \frac{1}{2}pm - \frac{1}{2} + \frac{p}{2} + pm$$

$$\therefore \frac{1}{2} - \frac{p}{2} \leq \frac{1}{2}pm$$

$\therefore \quad \frac{1}{1+m} \leq p \qquad\qquad m \leq \binom{n}{2} = \frac{n(n-1)}{2}.$

$\therefore \quad \frac{1}{n^2} \leq p.$

$\blacksquare$

Repeated trials

- Repeat the algorithm n^3 times.
- "Prob of success" in each trial is $\geq \frac{1}{n^2}$.
- $\therefore$ Prob that it fails in all trials is at most $\left(1 - \frac{1}{n^2}\right)^{n^3} \leq 2^{-n}$.

$\blacksquare$

Randomized Algorithms

Type 1

- Outputs correct answer w.p. $\geq 99\%$

  May err.
- Runs in polytime.

Type 2

- Outputs correct answer w.p. $\geq 99\%$ <u>or does not answer.</u>
- Runs in <u>expected</u> polytime.

Randomized Quicksort (RQ)

- Given $(a_1, a_2, \ldots, a_n)$
- Pick "pivot" a_i at random, $1 \le i \le n$.
- $S_1 = \{ a_j \mid a_j < a_i \}$
  $S_2 = \{ a_j \mid a_j > a_i \}$.
- Output $RQ(S_1) \cdot a_i \cdot RQ(S_2)$.

Theorem Let $T(n)$ be expected running time. Then $T(n)$ is $O(n \log n)$.

Proof We show, by induction, that

$T(n) \le D \cdot n \log n$ for some constant D.

Note.

$$T(n) = Cn + \frac{1}{n} \sum_{i=1}^{n} T(i-1) + T(n-i)$$

$$= Cn + \frac{2}{n} \sum_{i=1}^{n-1} T(i). \qquad T(0) = 0.$$

$$\therefore\ T(n) \leq Cn + \frac{2}{n}\left(\sum_{i=1}^{\frac{n}{2}} T(i) + \sum_{i=\frac{n}{2}+1}^{n-1} T(i)\right)$$

$$\leq Cn + \frac{2}{n}\cdot D\cdot\left(\sum_{i=1}^{\frac{n}{2}} i\,(\log n - 1) + \sum_{i=\frac{n}{2}+1}^{n-1} i \log n\right)$$

$$= Cn + \frac{2D}{n}\left(\sum_{i=1}^{n-1} i\right)\log n - \frac{2D}{n}\left(\sum_{i=1}^{\frac{n}{2}} i\right)$$

$$= Cn + \frac{\cancel{2}D}{\cancel{n}}\cdot\frac{\cancel{n}(n-1)}{\cancel{2}}\cdot\log n - \frac{\cancel{2}D}{\cancel{n}}\cdot\frac{\frac{\cancel{n}}{\cancel{2}}\left(\frac{n}{2}+1\right)}{2}$$

$$\leq Cn + Dn\log n - \frac{Dn}{4}$$

$$\leq Dn\log n \qquad \left(\text{provided } \frac{D}{4} \geq C\right).$$

$\blacksquare$