2. "w/o replacement", i.e.

for $i = 1$ to r do
pick rand elem $z_i \in S - \{z_1, \ldots, z_{i-1}\}$ indep.
put z_i in R

$(|R| = r)$

(each subset chosen w. prob $\frac{1}{\binom{n}{r}}$)

3. flip coins, i.e.

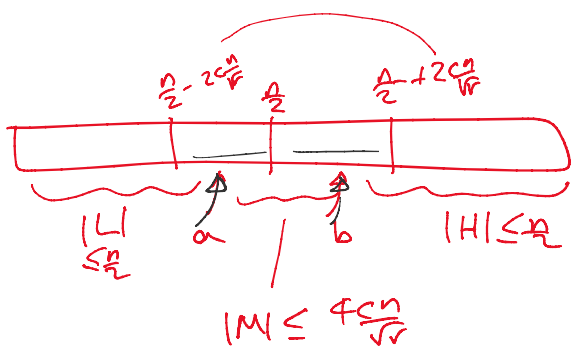
for $z \in S$
indep'ly decide to put z in R w. prob r/n

$(E[|R|] = r)$

here, with replacement .. (option 1)

Error Analysis:

$$\Pr[\text{error}] = \Pr(\underline{(*)} \text{ false})$$



$$\le \Pr\Big[\ \text{rank}_S(a) < \tfrac{n}{2} - 2c\tfrac{n}{\sqrt{r}}$$
$$\text{or } \text{rank}_S(a) > \tfrac{n}{2}$$
$$\text{or } \text{rank}_S(b) < \tfrac{n}{2}$$
$$\text{or } \text{rank}_S(b) > \tfrac{n}{2} + 2c\tfrac{n}{\sqrt{r}}\Big]$$

$$\Pr\Big[\text{rank}_S(a) < \tfrac{n}{2} - 2c\tfrac{n}{\sqrt{r}}\Big]$$

$a = (\frac{r}{2} - c\sqrt{r})$-th smallest of R

$$= \Pr\Big[\Big(\#\text{ elems of } R \text{ that have rank (in } S) \le \tfrac{n}{2} - 2c\tfrac{n}{\sqrt{r}}\Big) > \tfrac{r}{2} - c\sqrt{r}\Big]$$

first $\frac{n}{2} - 2c\frac{n}{\sqrt{r}}$ smallest elems in S

X = # elems of R that rank in S $\leq \frac{n}{2} - 2c\frac{n}{\sqrt{r}}$

$$= \sum_{i=1}^{r} X_i$$

where $X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ elem chosen in } R \text{ has rank} \leq \frac{n}{2} - 2c\frac{n}{\sqrt{r}} \\ 0 & \text{else} \end{cases}$

$$E[X_i] = \frac{n}{2} - 2c\frac{n}{\sqrt{r}} = \frac{1}{2} - \frac{2c}{\sqrt{r}}$$

$$\mu := E[X] = \sum_{i=1}^{r} E[X_i] = \frac{r}{2} - 2c\sqrt{r}$$

Fact If $X = \sum_i X_i$ and X_i's are pairwise indep,

then $Var(X) = \sum_i Var(X_i)$

Pf: $Var(X) = E[(X-\mu)^2]$ where $\mu = E[X]$

let $\mu_i = E[X_i]$

$\sum_i \mu_i = \mu$

let $Y_i = X_i - \mu_i$

$E[Y_i] = 0$.

$$= E\left[\left(\sum_i Y_i\right)^2\right]$$

$$= E\left(\sum_i Y_i^2 + 2\sum_{i<j} Y_i Y_j\right)$$

$$= \sum_i E(Y_i^2) + 2\sum_{i<j} E(Y_i Y_j)$$

$$= \sum_i Var(X_i) + 2\sum_{i<j} \underbrace{E(Y_i)E(Y_j)}_{0}$$

$\cancel{\sum_{i<j}}$ by indep. $\square$

In our appl'n,

$$\mathrm{Var}(X_i) = E(X_i^2) - E(X_i)^2$$
$$\leq E(X_i) \leq \frac{1}{2}.$$

$\Rightarrow$ by Fact, $\sigma^2 := \mathrm{Var}(X) \leq \frac{r}{2}$

$\Rightarrow \quad \sigma \leq \sqrt{r}.$

$$\Pr\left(\mathrm{rank}_S(a) \leq \frac{n}{2} - 2c\frac{n}{\sqrt{r}}\right)$$
$$\leq \Pr\left(X \geq \frac{r}{2} - c\sqrt{r}\right)$$
$$= \Pr\left(X - \mu \geq c\sqrt{r}\right)$$
$$\leq \Pr\left(X - \mu \geq c\sigma\right)$$
$$\leq \frac{1}{c^2}$$

("2nd Moment Method")

by Chebyshev's ineq

Other parts similar

$\Rightarrow \quad \Pr(\text{error}) \leq \frac{4}{c^2}.$

Rmk: convertible to Las Vegas

$$\text{expected \# comp} \leq 1.5n\left(1 - \frac{4}{c^2}\right) + \frac{4}{c^2} \cdot O(n \log n) + o(n)$$

Set $c = n^{0.1}$

$$= 1.5n + o(n)$$

Rmk. with stronger "Chernoff bound",

$\Pr[\text{error}] < \frac{1}{\exp(c)}$

Rmk. With [illegible]

$$\Pr[\text{error}] \le \frac{1}{2^{\Theta(c)}}.$$

Rmk weaker bd by 2nd moment only requires pairwise indep.

$\Rightarrow$ requires less random bits

Fact ("2-Point Sampling")

Fix prime p.

Pick random $a, b \in \{0, \dots, p-1\}$.

Let $Z_i := (ai + b) \bmod p$.

1. $\Pr[Z_i = s] = \frac{1}{p}$. $\forall s \in \{0, \dots, p-1\}$
2. $Z_0, \dots, Z_{p-1}$ are pairwise indep.

Pf: Fix $s, t \in \{0, \dots, p-1\}$.

1. $\forall i$,

$$\Pr(Z_i = s) = \Pr[ai + b \equiv s \pmod p]$$
$$= \Pr[b \equiv s - ai \pmod p]$$
$$= \frac{1}{p}.$$

2. $\forall i, j$, $i \neq j$,

$$\Pr[Z_i = s \wedge Z_j = t]$$
$$= \Pr_{a,b}\left[\begin{array}{l} ai + b \equiv s \\ aj + b \equiv t \end{array} \pmod p\right]$$
$$= \Pr\left[a \equiv \frac{s-t}{j-i} \wedge b \equiv s - \left(\frac{s-t}{j-i}\right) i \pmod p\right]$$
$$= \frac{1}{p^2}. \qquad \square$$

for prime $p \in [n, 2n]$.

Let $R = \{ S[z_1], S[z_2], \ldots, S[z_r] \}$.

$\Rightarrow O(\log n)$ rand. bits suffice.

Rmk- generalizes to "k-point sampling" for k-wise indep.

(pick rand $a_0, \ldots, a_{k-1}$, let $z_i = (a_{k-1} i^{k-1} + \ldots + a_0) \bmod p$).

What about Question 2 on time/space?

[Munro-Raman'92 / C.'09]

idea - iterate in M

at iteration i, maintain interval $[a_i, b_i]$ containing answer

let S_i = elems in $[a_i, b_i]$, $n_i = |S_i|$.

pick rand sample R_i of S of size $\frac{r_i}{n_i} \cdot n$.

$$E(|R_i \cap S_i|) = r_i$$

$$\Rightarrow n_{i+1} \leq \frac{4 c_i n_i}{\sqrt{r_i}} \ll n_i^{0.8}$$

Set $r_i = n_i^{2/3}$, $c_i = n_i^{0.1}$

after ℓ iterations, $n_\ell \ll n^{0.8^\ell} = s$

$$0.8^\ell = \frac{\log s}{\log n}$$

$$\Rightarrow \ell = \log_{0.8}\left(\frac{\log n}{\log s}\right)$$

iteration i takes expected time

$$O(n) + O(|R_i| \log n_i)$$

$$O(n) + O(|R_i| \log n_i)$$

quickselect with $O(\log n_i)$ expected # of passes

$$= O\left(\frac{r_i\, n}{n_i} \log n_i\right)$$

$r_i = n_i^{2/3}$

$$= O\left(\frac{n}{n_i^{1/3}} \log n_i\right)$$

$$= O(n).$$

$\Rightarrow$ total expected time

$$\boxed{O\left(n \log\left(\frac{\log n}{\log s}\right)\right)}$$

(det. $O\left(\frac{n \log n}{\log s} + n \log^* n\right)$).

(e.g. $s = n^{0.1} \Rightarrow O(n)$ time

$s = \log n \Rightarrow O(n \log\log n)$ time)

Issue - can't store sample R_i!

use 2-point sampling

$O(\log n)$ extra bits of space

Rmk: rand. lower bd $\Omega\left(n \log\left(\frac{\log n}{\log s}\right)\right)$

for comp-based alg'ms [C.'09]