Randomized algorithms 2/3/2022

L6: Birthday paradox, occupancy and collector

Lemma $\forall x \in \mathbb{R}$

(i.) $1+x \le e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$ and $1-x \le e^{-x}$ $\forall x \in \mathbb{R}$.

(ii.) $(1+1/n)^n \le e$ $\quad e \le (1+\frac{1}{n})^{n+1}$ $\quad (1+\frac{1}{n})^n \xrightarrow[n\to\infty]{} e$

(iii.) $(1-\frac{1}{n})^n \le \frac{1}{e} \le (1-\frac{1}{n})^{n-1}$

(iv.) $n! \ge (n/e)^n$

(v.) $\forall 1 \le k \le n \quad \left(\frac{n}{k}\right)^k \le \binom{n}{k} \le \left(\frac{ne}{k}\right)^k$.

Birthday Paradox

n people $\quad m = 365$

Lemma

$X_1, X_2, \dots, X_n$ picked randomly, uniformly and independently from $[m] = \{1, 2, \dots, m\}$.

Then the expected number of collisions is $\binom{n}{2}\frac{1}{m}$.

Proof $Y_{i,j} = 1 \iff X_i = X_j$

$P[Y_{i,j}=1] = P[X_i = X_j] = \frac{1}{m}$

$E\left[\sum_{i<j} Y_{i,j}\right] = \sum_{i<j} E[Y_{i,j}] = \sum_{i<j} \frac{1}{m} = \binom{n}{2}\frac{1}{m}$ ∎

$n = 28$

$\binom{n}{2} > 364 \implies E[\#\text{ of collisions}] \ge 1$.

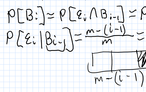
$\mathcal{E}_i = X_i$ is different than all values in $X_1, X_2 \dots, X_{i-1}$.

B_i = There is no collision in $X_1, \dots, X_i$:

$B_i = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \dots \cap \mathcal{E}_i$

Balls and bins

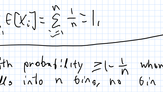
m bins



$B_i = \mathcal{E}_i \cap \mathcal{E}_2 \cap \dots \cap \mathcal{E}_i = B_{i-1} \cap \mathcal{E}_i$

$P[B_i] = P[\mathcal{E}_i \cap B_{i-1}] = P[\mathcal{E}_i | B_{i-1}] \cdot P[B_{i-1}]$

$P[\mathcal{E}_i | B_{i-1}] = \frac{m-(i-1)}{m} = 1 - \frac{i-1}{m} \le \exp(-\frac{i-1}{m})$.



$P[B_i] = P[\mathcal{E}_i | B_i] P[B_{i-1}] \le \exp(-\frac{i-1}{m}) P[B_{i-1}]$

$\le \exp(-\frac{i-1}{m}) \exp(-\frac{i-2}{m}) \cdots$

$= \prod_{k=1}^{i} \exp(-\frac{k-1}{m})$

$= \exp\left(-\sum_{k=1}^{i} \frac{k-1}{m}\right) = \exp\left(-\frac{1}{m}\binom{i}{2}\right)$

$\sum_{k=1}^{i} (k-1) = 1+2+\dots+(i-1)$

$= \frac{i(i-1)}{m}$

Claim For $X_1, \dots, X_n$ independently picked from $[m]$, the prob they are all distinct

$\le \exp\left(-\binom{n}{2}\frac{1}{m}\right)$

$n = 28 \quad m = 365 \quad \le \frac{1}{2}$

Occupancy problem

n balls into n bins.

$E[\#\text{ of balls in the first bin}] = 1$

$X_i = 1$ if ith ball fall into first bin

$P[X_i = 1] = \frac{1}{n}$

$E[\sum X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{n} = 1$.

Theorem with probability $\ge 1 - \frac{1}{n}$ when we throw n balls into n bins, no bin has more than

$\left\lceil \frac{3 \ln n}{\ln \ln n} \right\rceil$.

Proof

prob that a specific bin gets exactly i balls?

$p = \frac{1}{n}$ $\quad \binom{n}{i} p^i (1-p)^{n-i} \le \binom{n}{i} \frac{1}{n^i} \le \left(\frac{ne}{i}\right)^i \frac{1}{n^i} = \left(\frac{e}{i}\right)^i$

$C(k) \equiv$ event that the first bin has k or more balls

$P[C(k)] = \sum_{i=k}^{n} \binom{n}{i} \frac{1}{n^i} \le 2\left(\frac{e}{k}\right)^k$ $\quad k \ge 2e$

$k = c\frac{\ln n}{\ln \ln n}$

$\le 2\left(\frac{e}{k}\right)^k \le 2\exp\left(k - k\ln k\right)$

$\le 2\exp\left(-\frac{k \ln k}{2}\right)$ $\quad \ln \ln n$

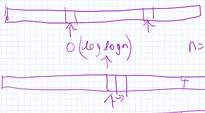
$\le 2\exp\left(-\frac{c \ln n}{2 \ln \ln n} \ln\left(\frac{c \ln n}{\ln \ln n}\right)\right)$

$\le 2\exp(-c \ln n) \le \frac{1}{n^2}$ $\quad \iff c \ge 6$

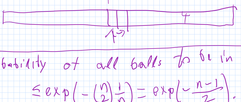
$\implies n \cdot P[C(k)] = n \cdot \frac{1}{n^2} \le \frac{1}{n}$

$P[\text{any bin contains more than } \frac{3\ln n}{\ln \ln n} \text{ balls}]$ ∎

The power of two choices.



$n = 2^{20} \approx 10^6$



Probability of all balls to be in their own bins

$\le \exp\left(-\binom{n}{2}\frac{1}{n}\right) = \exp\left(-\frac{n-1}{2}\right)$.

$\frac{n!}{n^n}$

Coupon collector problem

n coupons.

collect all n coupons?

$X_i \equiv$ # of coupons you buy till you get the ith distinct coupon.

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

$E[X_i]$ $\quad p_i = \frac{n-(i-1)}{n}$

$X_i \sim \text{Geom}\left(\frac{n-(i-1)}{n}\right)$

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

$E[X] = E\left[\sum X_i\right]$

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

$= n + \sum_{i=1}^{n-1} \frac{n-i}{i}$

$\le n + n\sum_{i=1}^{n-1} \frac{1}{i} = n + n\ln n$.

$X_i \sim \text{Geom}(p_i)$

$V[X_i] = \frac{1-p_i}{p_i^2} \le \frac{1}{p_i^2}$ $\quad p_i \ge 1 - \frac{i-1}{n}$

$V[X] = \sum_{i=1}^{n} V[X_i] \le \sum_{i=1}^{n} \frac{1-p_i}{p_i^2}$

$V[X] = V\left[\sum X_i\right]$

$= \sum V[X_i]$ $\quad$ indep

$= n^2 \sum_{i=1}^{n} \frac{1}{i^2} \sim nH_n \approx \frac{\pi^2}{6} n^2$

$\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$

$\sigma_X = \sqrt{V[X]} = \sqrt{\frac{\pi^2}{6} n^2} \approx \frac{\pi}{\sqrt{6}} n$

$P[X \ge n\ln n + \frac{n}{4} + t\, n\frac{\pi}{\sqrt{6}}] \le P[|X - E[X]| \ge t\sigma_X] \le \frac{1}{t^2}$ ∎