RECAP

OCCUPANCY / BALLS-INTO BINS

PROBLEM: $m \geq n$ identical balls thrown randomly and independently into n bins (boxes)

1) Maximum number of balls into any bin

Let X_b = "# balls into bin b", $1 \leq b \leq n$

$$X_b \sim \text{Binomial}\left(m, \frac{1}{n}\right)$$

$$\Pr(X_b = i) = \binom{m}{i} \frac{1}{n^i} \left(1 - \frac{1}{n}\right)^{m-i}$$

$C_b(k)$ = "bin b has $\geq k$ balls"

$$\Pr(C_b(k)) = \sum_{i \geq k} \Pr(X_b = i) \leq 2 \left(\frac{e\,m}{k\,n}\right)^k$$

$\boxed{m = n}$ $\Pr(C_b(\bar{k})) \leq \frac{1}{n^2}$ for $\bar{k} = \Theta(\ln n / \ln\ln n)$

$\Rightarrow \Pr(\exists b : C_b(\bar{k}) \text{ holds}) \leq \frac{1}{n}$

(union bound)

$\Rightarrow$ Insertion/search in hash table with chaining ($m = n$ slots/keys): $O(\log n / \log\log n)$ w.h.p.

$\boxed{\text{Case 2: } m = n \ln n}$

Set $k^* = e^2 \ln n \; \left(> 2\frac{e m}{n}\right)$

$$\Pr(C_b(k^*)) \leq 2\left(\frac{e m}{k^* n}\right)^{k^*}$$

$$< 2\left(\frac{\cancel{e}\, \cancel{n} \cancel{\ln n}}{e^{\cancel{2}}\, \cancel{n} \cancel{\ln n}}\right)^{e^2 \ln n} \leq \frac{2}{n^{e^2}} < \frac{1}{n^2}$$

Thus

$$\Pr(\exists \text{ bin with } \geq k^* = e^2 \ln n \text{ balls}) =$$

$$\Pr\left(\bigcup_{b=1}^{n} C_b(k^*)\right) \leq \sum_{b=1}^{n} \Pr(C_b(k^*)) < \frac{n}{n^2} = \frac{1}{n}$$

Therefore, the maximum number of balls into a bin, when $m = n \ln n$, is $O(\log n) = O\left(\frac{m}{n}\right)$ w.h.p.

This means that balls are roughly balanced among the bins!

This phenomenon also holds for $m = \Omega(n \log n)$! (exercise)

FURTHER RESULTS

$m = n$ $\exists$ bin b with $\Theta(\log n / \log\log n)$ balls w.h.p. (no proof)

$m = n \ln n$:

$Pr(\exists$ bin b with $\Theta(\log n)$ balls$) = 1$

(PIGEON HOLE)

APPLICATION: Scheduling $m \gg n$ $(\Omega(n \log n))$ work units among n workers. A random allocation guarantees weak balancing: $\Theta\left(\frac{m}{n}\right)$ units/worker whp!

2. Number of empty bins ($m=n$)

For a fixed bin b:

$$\Pr(X_b=0) = \left(1-\frac{1}{n}\right)^n$$

Consider an indicator variable $Y_b = 1 \Leftrightarrow X_b=0$. $\Pr(Y_b=1) = \left(1-\frac{1}{n}\right)^n$

Then $Y = \sum_{b=1}^{n} Y_b = \#$ empty bins

The Y_b's are not independent, but

$$E[Y] = \sum_{b=1}^{n} E[Y_b] = n\left(1-\frac{1}{n}\right)^n$$

For $n \geq 2$: $\frac{1}{4} \leq \left(1-\frac{1}{n}\right)^n < e^{-1}$

(for large n: $\left(1-\frac{1}{n}\right)^n \simeq e^{-1}$)

On average there are

$\forall n \quad \frac{n}{4} \leq E[Y] \leq \frac{n}{e}$

empty bins!

MORALE: When $m=n$, randomly distributing work units results into a constant fraction of idle workers! $\Rightarrow$ MANY CONFLICTS

Case $m \ll n$: BIRTHDAY PARADOX

- Class of $m = 30$ students
- What is the probability that two students have the same birthday?

(assume fully random birthdays)

e.g. no leap years
no twins
etc...

Experiment: throwing $m = 30$ balls (students) into $n = 365$ bins (birthdays):

evaluate $P = \Pr(\text{no bin with} > 1 \text{ balls})$

There are $\binom{365}{30}$ $\binom{n}{m}$ choices of picking distinct birthdays, and $30!$ ways to distribute the distinct birthdays to the students.

The number of all possible birthday configurations is 365^{30}

Therefore

$$P = \frac{\binom{365}{30} 30!}{365^{30}} = \frac{365 \cdot 364 \cdot \; \cdots \; \cdot (365-29)}{365^{30}}$$

n, n, $n-(n-1)$

$$= 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdot \quad \cdot \left(1 - \frac{29}{365}\right)$$

< 0.3 !

In $> 70\%$ of the cases, there will be two students with the same birthday!

GENERALIZING:

$$P_{m,n} = \prod_{j=1}^{m-1}(1 - j/n) \simeq \prod_{j=1}^{m-1} e^{-j/n} = e^{-\sum_{j=1}^{m-1} j/n}$$

$$= e^{-(m-1)m/2n} \simeq e^{-m^2/2n}$$

The probability stays $\geq \frac{1}{2}$ for m up to $\sim \sqrt{2\ln 2\, n} = \Theta(\sqrt{n})$.

It becomes vanishing $\left(\xrightarrow[n\to+\infty]{} 0\right)$ for $m = \Omega\left(\sqrt{n\log n}\right)$