Balls & Bins

Anil Maheshwari

School of Computer Science
Carleton University
Canada

Outline

Model

Collisions

Size of Bins

Model

Balls & Bins

Model

We have m Balls and n Bins. We throw each ball in a bin uniformly at random.

What is the probability of following events:

1. Balls i and j are in the same bin.
2. Expected number of balls in a bin.
3. Expected number of empty bins.
4. Bin $\#i$ receives (a) 0 balls, (b) k balls, and (c) $\geq k$ balls.
5. All bins have $\leq \frac{c \ln n}{\ln \ln n}$ balls.

Applications: Birthday Paradox, Load Balancing, Perfect Hashing

Collisions

Probability of ball j is in the bin containing ball i

Number of Balls = m
Number of Bins = n.

$Pr[\text{Balls } i \text{ and } j \text{ in same bin}] = \frac{1}{n}$.

Define r.v. X_{ij} ($1 \leq i \leq m-1$, $i+1 \leq j \leq m$) as follows:

$$X_{ij} = \begin{cases} 1 & \text{if ball } j \text{ is Bin of ball } i \\ 0 & \text{Otherwise} \end{cases}$$

$Pr(X_{ij} = 1) = \frac{1}{n}$ and $E[X_{ij}] = \frac{1}{n}$

Expected number of balls in a bin

Number of Balls = m
Number of Bins = n.

$$
\begin{aligned}
E[\#\ \text{Balls in Bin } i] &= E[\sum_{j=1}^{n} X_{ij}] \\
&= \sum_{j=1}^{m} E[X_{ij}] \\
&= \sum_{j=1}^{m} P(\text{Ball } j \text{ is in Bin } i) \\
&= \sum_{j=1}^{m} 1/n \\
&= m/n
\end{aligned}
$$

Expected number of collisions

Number of Balls = m
Number of Bins = n.

Define $X = \sum_{i,j} X_{ij}$ = Total # of collisions

$E[X] = E[\sum_{i,j} X_{ij}]$

By Linearity of Expectation: $E[\sum_{i,j} X_{ij}] = \sum_{i,j} E[X_{ij}]$

Thus $E[X] = \frac{1}{n}\binom{m}{2}$

Birthday Paradox

Number of Balls = m = Number of Students
Number of Bins = n = Number of days in a Year.

For two students to have same Birthday:
What value of m will result in $E[X] = \frac{1}{n}\binom{m}{2} \geq 1$

For $m \geq 28$, $E[X] = \frac{1}{365}\frac{m(m-1)}{2} \geq 1$

Birthday Paradox Contd.

What is minimum value of m so that the probability that two students share the same birthday is $\geq \frac{1}{2}$?

Probability that all m students have distinct birthday's is given by:
$(1-\frac{1}{n})(1-\frac{2}{n})(1-\frac{3}{n})\ldots(1-\frac{m-1}{n})$

Let us use the inequality $1-x \leq e^{-x}$.

We want:

$$\begin{aligned} e^{-\frac{1}{n}}e^{-\frac{2}{n}}e^{-\frac{3}{n}}\ldots e^{-\frac{m-1}{n}} &\leq \frac{1}{2} \\ e^{-\frac{m(m-1)}{2n}} &\leq \frac{1}{2} \end{aligned}$$

Now using $n = 365$, we have $e^{-\frac{m(m-1)}{2*365}} \leq \frac{1}{2}$ or $m \geq 23$.

Intuition: There are $m^2 \approx n$ pairs where a collision can occur.

Size of Bins

Number of Balls in Bin i

Number of Balls = m; Number of Bins = n.

Problem I

What is the probability that Bin i receives no balls?

$$\left(1 - \frac{1}{n}\right)^m \le e^{-\frac{m}{n}}$$

If $n = m$, $(1 - \frac{1}{n})^n \le e^{-1} = 0.37$.

Problem II

What is the probability that Bin i receives exactly k balls?

$$\binom{m}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{m-k}$$

Number of Balls in Bin i contd.

Number of Balls = m; Number of Bins = n.

Problem III

What is the probability that Bin i receives $\geq k$ balls?

$$\leq \binom{m}{k}\left(\frac{1}{n}\right)^k$$

Fix a set S of k balls.

Pr(Bin i receives all balls in S)$= (\frac{1}{n})^k$.

The remaining balls can go anywhere.

Take the union bound over all possible subset of k balls.

Useful Bounds: $(\frac{n}{k})^k < \binom{n}{k} < \frac{n^k}{k!} < (\frac{en}{k})^k$.
If $n = m$, we have $\binom{n}{k}\left(\frac{1}{n}\right)^k \leq \left(\frac{e}{k}\right)^k$

Expected Number of Balls in a Bin

Number of Balls = m; Number of Bins = n.

Problem IV

What is Expected # of Balls in a Bin?

$$= \frac{m}{n}$$

You can use random variable B_{ij} to compute this probability.

Expected Number Empty Bins

Number of Balls = m; Number of Bins = n.

Problem V

What is Expected # of Empty Bins?

Define a r.v. X_i such that

$$X_i = \begin{cases} 1 & \text{if Bin } i \text{ is empty} \\ 0 & \text{Otherwise} \end{cases}$$

From Problem I, $Pr(X_i = 1) \leq e^{-\frac{m}{n}}$ and $E[X_i] \leq e^{-\frac{m}{n}}$

Thus, $E[\text{\# of Empty Bins}] = \sum_{i=1}^{n} E[X_i] \leq ne^{-\frac{m}{n}}$

When $n = m$, $E[\text{\# of Empty Bins}] \leq \frac{n}{e}$

Max # Balls in Bins

Number of Balls = Number of Bins = n.

Max # of Balls in Bins

With probability $\geq 1 - \frac{1}{n}$ all bins receive fewer than $3\frac{\ln n}{\ln \ln n}$ balls.

Pr(Bin i has more that k balls) $\leq \binom{m}{k}\left(\frac{1}{n}\right)^k = \binom{n}{k}\left(\frac{1}{n}\right)^k \leq \left(\frac{e}{k}\right)^k$

Substitute $k = 3\frac{\ln n}{\ln \ln n}$ and show that

Pr(Bin i has more that k balls) $\leq \frac{1}{n^2}$

Thus by Union Bound,

Pr(Any bin has more that k balls) $\leq \frac{1}{n}$

Max # Balls in Bins Contd.

Claim

For $k = 3\frac{\ln n}{\ln\ln n}$, $\left(\frac{e}{k}\right)^k \leq \frac{1}{n^2}$

Proof.

$$
\begin{aligned}
\left(\frac{e}{k}\right)^k &= \left[e\frac{\ln\ln n}{3\ln n}\right]^{\frac{3\ln n}{\ln\ln n}} \\
&= \left[e^1 e^{\ln\frac{\ln\ln n}{3\ln n}}\right]^{\frac{3\ln n}{\ln\ln n}} \\
&= e^{\frac{3\ln n}{\ln\ln n}[1+\ln\ln\ln n - \ln(3\ln n)]} \\
&= e^{\frac{3\ln n}{\ln\ln n}[1+\ln\ln\ln n - \ln 3 - \ln\ln n]} \\
&\leq e^{\frac{3\ln n}{\ln\ln n}[\ln\ln\ln n - \ln\ln n]} \\
&= e^{\left[-3\ln n + \frac{3\ln n\ln\ln\ln n}{\ln\ln n}\right]}
\end{aligned}
$$

For large values of n, $-3\ln n + \frac{3\ln n\ln\ln\ln n}{\ln\ln n} \leq -2\ln n$.
Thus, $\left(\frac{e}{k}\right)^k \leq e^{-2\ln n} = \frac{1}{n^2}$ $\square$

References

1. Probability and Computing by Mitzenmacher and Upfal, Cambridge Univ. Press 2005.
2. Introduction to Probability by Blitzstein and Hwang, CRC Press 2015.
3. Courses Notes of COMP 2804 by Michiel Smid.
4. My Notes on Algorithm Design.