Randomized Algorithms

Balls and Bins: Some basics

Joseph Chuang-Chieh Lin Dept. CSIE, Tamkang University

Outline

- The Birthday Paradox Revisited
- Balls into Bins
- Poisson Distribution

• By definition of e^x :

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots$$

$$\Rightarrow e = \sum_{k=0}^{\infty} \frac{1}{k!} = 1 + x + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} \cdots$$

• By the Binomial Theorem:

$$\left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^n \binom{n}{k} \left(\frac{1}{n}\right)^k$$

$$= \sum_{k=0}^n \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \cdot \frac{1}{n^k}$$

$$= \sum_{k=0}^n \left(\frac{n}{n}\right) \cdot \left(\frac{n-1}{n}\right) \cdots \left(\frac{n-k+1}{n}\right) \cdot \frac{1}{k!}.$$

By the Binomial Theorem:

$$\left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^n \binom{n}{k} \left(\frac{1}{n}\right)^k$$

$$= \sum_{k=0}^n \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \cdot \frac{1}{n^k}$$

$$= \sum_{k=0}^n \left(\frac{n}{n}\right) \cdot \left(\frac{n-1}{n}\right) \cdots \left(\frac{n-k+1}{n}\right) \cdot \frac{1}{k!}.$$

$$\therefore \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^\infty \frac{1}{k!}.$$

Useful approaches:

$$\left(1+\frac{1}{n}\right)^{-m} = \left(\left(1+\frac{1}{n}\right)^n\right)^{-m/n} \approx e^{-m/n}.$$

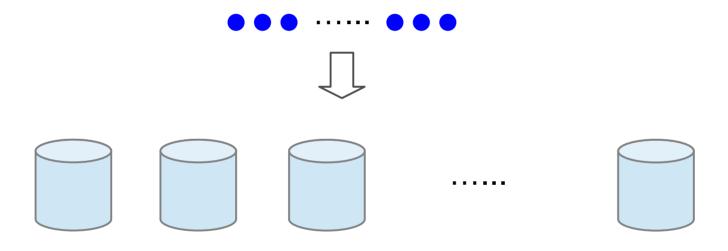
$$\frac{k^k}{k!} < \sum_{i=0}^{\infty} \frac{k^i}{i!} = e^k.$$

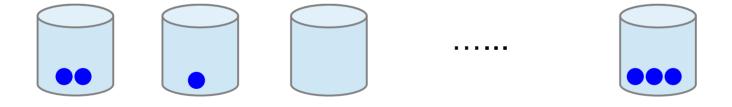
$$\therefore k! > \left(\frac{k}{e}\right)^k.$$

Assignment 04

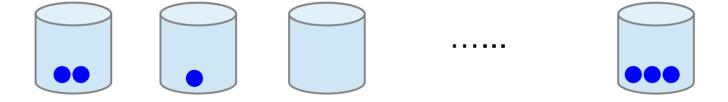
• Show that for $|x| \le 1$, $e^x(1-x^2) \le 1 + x \le e^x$.

• Let
$$\left(1 - \left(1 - \frac{1}{n}\right)^{km}\right)^k \approx (1 - e^{\square})^k$$
.
Find $\square = ?$

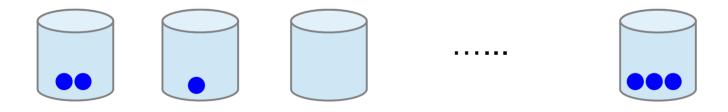




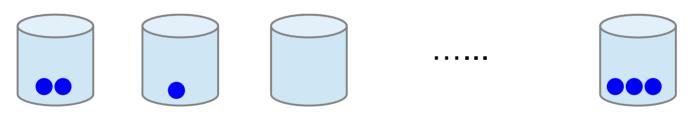
• Example. balls: people, bins: birthdays (mm/dd).



- Example. balls: people, bins: birthdays (mm/dd).
- Let's say there are 30 people in a room. What's the probability that no two people in the room share the same birthday?

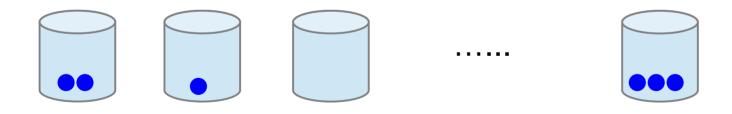


- Example. balls: people, bins: birthdays (mm/dd).
- Let's say there are 30 people in a room. What's the probability that no two people in the room share the same birthday?



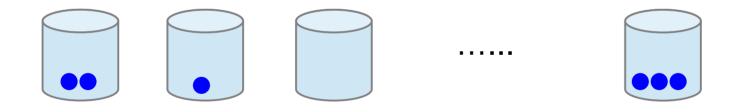
$$\frac{\binom{365}{30}30!}{365^{30}}.$$

- Example. balls: people, bins: birthdays (mm/dd).
- Let's say there are 30 people in a room. What's the probability that no two people in the room share the same birthday?



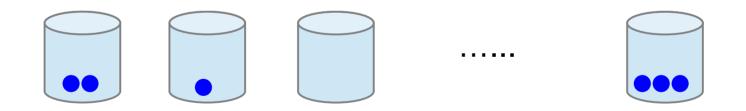
$$\left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{29}{365}\right).$$

- Example. balls: people, bins: birthdays (mm/dd).
- In general, for *m* people in a room and *n* possible birthdays, what's the probability that no two people in the room share the same birthday?



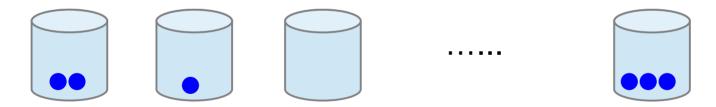
$$\left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right) = \prod_{j=1}^{m-1} \left(1 - \frac{j}{n}\right).$$

- Example. balls: people, bins: birthdays (mm/dd).
- In general, for *m* people in a room and *n* possible birthdays, what's the probability that no two people in the room share the same birthday?



$$\left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right) = \prod_{j=1}^{m-1} \left(1 - \frac{j}{n}\right) \approx \prod_{j=1}^{m-1} \frac{e^{-j/n}}{n} = \exp\left\{-\sum_{j=1}^{m-1} \frac{j}{n}\right\} = e^{-m(m-1)/2n} \approx e^{-m^2/2n}.$$

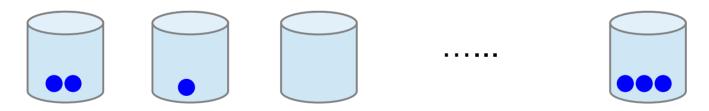
- Example. balls: people, bins: birthdays (mm/dd).
- In general, for *m* people in a room and *n* possible birthdays, what's the probability that no two people in the room share the same birthday?



• Set the probability threshold to be $\frac{1}{2}$:

$$\frac{m^2}{2n} = \ln 2 \Rightarrow m = \sqrt{2n \ln 2} \approx 22.49.$$

- Example. balls: people, bins: birthdays (mm/dd).
- In general, for *m* people in a room and *n* possible birthdays, what's the probability that no two people in the room share the same birthday?



• Set the probability threshold to be $\frac{1}{2}$:

$$\frac{m^2}{2n} = \ln 2 \Rightarrow m = \sqrt{2n \ln 2} \approx 22.49.$$
 Matching your observation?

- *n* balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than L?
 - Maximum number of balls in one bin.

- *n* balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than *L*?
 - Maximum number of balls in one bin.
 - The average load is n/n = 1?!

- *n* balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than L?
 - Maximum number of balls in one bin.

• The probability that bin 1 receives at least M balls is at most

$$\binom{n}{M} \left(\frac{1}{n}\right)^M$$

- *n* balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than L?
 - Maximum number of balls in one bin.

• The probability that bin 1 receives at least M balls is at most

$$\binom{n}{M} \left(\frac{1}{n}\right)^M$$
 union bound

- *n* balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than *L*?
 - Maximum number of balls in one bin.

• The probability that bin 1 receives at least M balls is at most

$$\binom{n}{M} \left(\frac{1}{n}\right)^M = \frac{n(n-1)\cdots(n-M+1)}{M!} \left(\frac{1}{n}\right)^M \le \frac{1}{M!} \le \left(\frac{e}{M}\right)^M.$$

• The probability that bin 1 receives at least M balls is at most

$$\binom{n}{M} \left(\frac{1}{n}\right)^M = \frac{n(n-1)\cdots(n-M+1)}{M!} \left(\frac{1}{n}\right)^M \le \frac{1}{M!} \le \left(\frac{e}{M}\right)^M.$$

Applying the union bound again:

• The probability that bin 1 receives at least *M* balls is at most

$$\binom{n}{M} \left(\frac{1}{n}\right)^M = \frac{n(n-1)\cdots(n-M+1)}{M!} \left(\frac{1}{n}\right)^M \le \frac{1}{M!} \le \left(\frac{e}{M}\right)^M.$$

Applying the union bound again:

$$n\left(\frac{e}{M}\right)^M \le n\left(\frac{e\ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n}$$

• The probability that bin 1 receives at least M balls is at most

$$\binom{n}{M} \left(\frac{1}{n}\right)^M = \frac{n(n-1)\cdots(n-M+1)}{M!} \left(\frac{1}{n}\right)^M \le \frac{1}{M!} \le \left(\frac{e}{M}\right)^M.$$

Applying the union bound again:

$$n\left(\frac{e}{M}\right)^{M} \leq n\left(\frac{e\ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n}$$

$$\leq n\left(\frac{\ln\ln n}{\ln n}\right)^{3\ln n/\ln\ln n} = e^{\ln n}\left(\frac{e^{\ln\ln\ln n}}{e^{\ln\ln n}}\right)^{3\ln n/\ln\ln n}$$

$$= e^{\ln n}(e^{\ln\ln\ln n-\ln\ln n})^{3\ln n/\ln\ln n}$$

$$= e^{-2\ln n+3(\ln n)(\ln\ln n)/\ln\ln n}$$

$$\leq \frac{1}{n}.$$

• A set of $n = 2^m$ numbers chosen uniformly at random in $[0, 2^k]$, $k \ge m$.

Bucket Sort:

- Stage 1: place the elements into *n* buckets.
 - j^{th} bucket: holds all elements whose first m binary digits corresponds to j.
 - = e.g., $n = 2^{10}$, bucket 3 contains all elements whose first 10 digits are 000000011.
- Stage 2: sort each bucket using any standard sorting algorithm.

• A set of $n = 2^m$ numbers chosen uniformly at random in $[0, 2^k]$, $k \ge m$.

Bucket Sort: Expected *O*(*n*) time?!

- Stage 1: place the elements into *n* buckets.
 - j^{th} bucket: holds all elements whose first m binary digits corresponds to j.
 - = e.g., $n = 2^{10}$, bucket 3 contains all elements whose first 10 digits are 000000011.
- Stage 2: sort each bucket using any standard sorting algorithm.

• A set of $n = 2^m$ numbers chosen uniformly at random in $[0, 2^k)$, $k \ge m$.

Bucket Sort:

- Stage 1: place the elements into *n* buckets.
 - j^{th} bucket: holds all elements whose first m binary digits corresponds to j.
 - = e.g., $n = 2^{10}$, bucket 3 contains all elements whose first 10 digits are 000000011.
 - The number of elements landing in a specific bucket: Binomial(n, 1/n).
- Stage 2: sort each bucket using any standard sorting algorithm.

- X_j : the number of elements landing in bucket j.
- The time to sort bucket $j: c(X_i)^2$, for some constant c.
- The expected time for sorting in Stage 2:

$$\mathbf{E}\left[\sum_{j=1}^{n} c(X_{j})^{2}\right] = c\sum_{j=1}^{n} \mathbf{E}[X_{j}^{2}] = cn\mathbf{E}[X_{1}^{2}].$$

- X_j : the number of elements landing in bucket j.
- The time to sort bucket $j: c(X_i)^2$, for some constant c.
- The expected time for sorting in Stage 2:

$$\mathbf{E}\left[\sum_{j=1}^n c(X_j)^2\right] = c\sum_{j=1}^n \mathbf{E}[X_j^2] = cn\mathbf{E}[X_1^2].$$
 The same for all buckets.

- *X*_i: the number of elements landing in bucket *j*.
- The time to sort bucket $j: c(X_i)^2$, for some constant c.
- The expected time for sorting in Stage 2:

The expected time for sorting in Stage 2:
$$\mathbf{E}\left[\sum_{j=1}^{n}c(X_{j})^{2}\right] = c\sum_{j=1}^{n}\mathbf{E}[X_{j}^{2}] = cn\mathbf{E}[X_{1}^{2}]. \quad \mathbf{E}[X_{1}^{2}] = \mathbf{Var}[X_{1}] + (\mathbf{E}[X_{1}])^{2}$$

$$= n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right) + \left(n \cdot \frac{1}{n}\right)^{2}$$

$$= 2 - \frac{1}{n}$$

$$< 2.$$

Assignment 05

- Suppose that *n* balls are thrown independently and uniformly at random into *n* bins.
 - Find the conditional probability that bin 1 has one ball given that exactly one ball fell into the first three bins.
 - Find the conditional expectation of the number of balls in bin 1 under the condition that bin 2 received no balls.

Questions

- What is the probability that **a given bin is empty**?
- What is the expected number of **empty bins**?

• The probability that the *i*th bin remains empty is

$$\left(1 - \frac{1}{n}\right)^m \approx e^{-m/n}.$$

• The probability that the *i*th bin remains empty is

$$\left(1 - \frac{1}{n}\right)^m \approx e^{-m/n}.$$

The expected number of empty bins:

$$n \cdot \left(1 - \frac{1}{n}\right)^m \approx ne^{-m/n}.$$

• The probability that the *i*th bin remains empty is

$$\left(1 - \frac{1}{n}\right)^m \approx e^{-m/n}.$$

The expected number of empty bins:

$$n \cdot \left(1 - \frac{1}{n}\right)^m \approx n e^{-m/n}.$$

The expected "fraction"

• Generalization:

The probability that a given bin has r balls is

Generalization:

The probability that a given bin has r balls is

$$p_r = {m \choose r} \left(\frac{1}{n}\right)^r \left(1 - \frac{1}{n}\right)^{m-r}$$

$$= \frac{1}{r!} \frac{m(m-1)\cdots(m-r+1)}{n^r} \left(1 - \frac{1}{n}\right)^{m-r}.$$

$$\approx \frac{e^{-m/n}(m/n)^r}{r!}.$$
Expected number of bins with

Expected number of bins with exactly r balls $\approx np_r$.



Siméon Poisson (1781–1840) Wikipedia

• A discrete Poisson random variable X with parameter μ is given by the following probability distribution on j = 0, 1, 2, ...:

$$\Pr[X=j] = \frac{e^{-\mu}\mu^{j}}{j!}.$$



Siméon Poisson (1781–1840) Wikipedia

• A discrete Poisson random variable X with parameter μ is given by the following probability distribution on j = 0, 1, 2, ...:

$$\Pr[X=j] = \frac{e^{-\mu}\mu^{j}}{j!}.$$



$$\frac{e^{-m/n}(m/n)^r}{r!}.$$



Siméon Poisson (1781–1840) Wikipedia

• Try to verify if it's proper:

$$\sum_{j=0}^{\infty} \Pr[X=j] = \sum_{j=0}^{\infty} \frac{e^{-\mu} \mu^j}{j!}$$

$$= e^{-\mu} \sum_{j=0}^{\infty} \frac{\mu^j}{j!}$$

$$= 1.$$



Siméon Poisson (1781–1840) Wikipedia

• The expectation:

$$\mathbf{E}[X] = \sum_{j=0}^{\infty} j \Pr[X = j] = \sum_{j=1}^{\infty} j \frac{e^{-\mu} \mu^j}{j!}$$

$$= \mu \cdot \sum_{j=1}^{\infty} \frac{e^{-\mu} \mu^{j-1}}{(j-1)!}$$

$$= \mu \cdot \sum_{j=0}^{\infty} \frac{e^{-\mu} \mu^j}{j!}$$

$$= \mu.$$



Siméon Poisson (1781 - 1840)Wikipedia

$$\Pr[X + Y = j]$$

$$\Pr[X + Y = j] = \sum_{i=1}^{n} \Pr[(X = k) \cap (Y = j - k)]$$

- The sum of finite number of independent Poisson random variables is a Poisson random variable.
- $= \sum_{k=0}^{J} \frac{e^{-\mu_1} \mu_1^k}{k!} \frac{e^{-\mu_2} \mu_2^{(j-k)}}{(j-k)!}$

$$= \frac{e^{-(\mu_1 + \mu_2)}}{j!} \sum_{k=0}^{j} \frac{j!}{k!(j-k)!} \mu_1^k \mu_2^{(j-k)}$$

$$= \frac{e^{-(\mu_1 + \mu_2)}}{j!} \sum_{k=0}^{j} {j \choose k} \mu_1^k \mu_2^{(j-k)}$$
$$= \frac{e^{-(\mu_1 + \mu_2)} (\mu_1 + \mu_2)^j}{j!}.$$

$$= \frac{e^{-(\mu_1 + \mu_2)}(\mu_1 + \mu_2)^j)}{j!}.$$



Siméon Poisson (1781–1840) Wikipedia

• The moment generating function of a Poisson random variable with parameter μ is

$$M_X(t) = e^{\mu(e^t - 1)}.$$



Siméon Poisson (1781–1840) Wikipedia

• The moment generating function of a Poisson random variable with parameter μ is

$$M_X(t) = e^{\mu(e^t - 1)}.$$

For any t,

$$\mathbf{E}[e^{tX}] = \sum_{k=0}^{\infty} \frac{e^{-\mu} \mu^k}{k!} e^{tk} = e^{\mu(e^t - 1)} \sum_{k=0}^{\infty} \frac{e^{-\mu e^t} (\mu e^t)^k}{k!}$$

$$= e^{\mu(e^t-1)}.$$



Siméon Poisson (1781–1840) Wikipedia

• The moment generating function of a Poisson random variable with parameter μ is

$$M_X(t) = e^{\mu(e^t - 1)}.$$

For any t,

$$\mathbf{E}[e^{tX}] = \sum_{k=0}^{\infty} \frac{e^{-\mu} \mu^k}{k!} e^{tk} = e^{\mu(e^t - 1)} \sum_{k=0}^{\infty} \frac{e^{-\mu e^t} (\mu e^t)^k}{k!}$$

$$= e^{\mu(e^t - 1)}.$$

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{(\mu_1 + \mu_2)(e^t - 1)}$$

Limit of the Binomial Distribution

- Theorem. Let $X_n \sim \text{Binomial}(n, p)$ be a binomial random variable
 - *p*: a function of *n*
 - − $\lim_{n\to\infty} np = \lambda$ is a constant, independent of *n*.

Then for any fixed k,

$$\lim_{n \to \infty} \Pr[X_n = k] = \frac{e^{-\lambda} \lambda^k}{k!}.$$

Scenario

- *m* balls into *n* bins.
 - m = f(n);
 - $\lim_{m\to\infty} m/n = \lambda$;
 - X_m : the number of balls in a specific bin.
 - Binomial(m, 1/n).
- From the theorem:

$$\lim_{m \to \infty} \Pr[X_m = r] = \frac{e^{-m/n} (m/n)^r}{r!}.$$