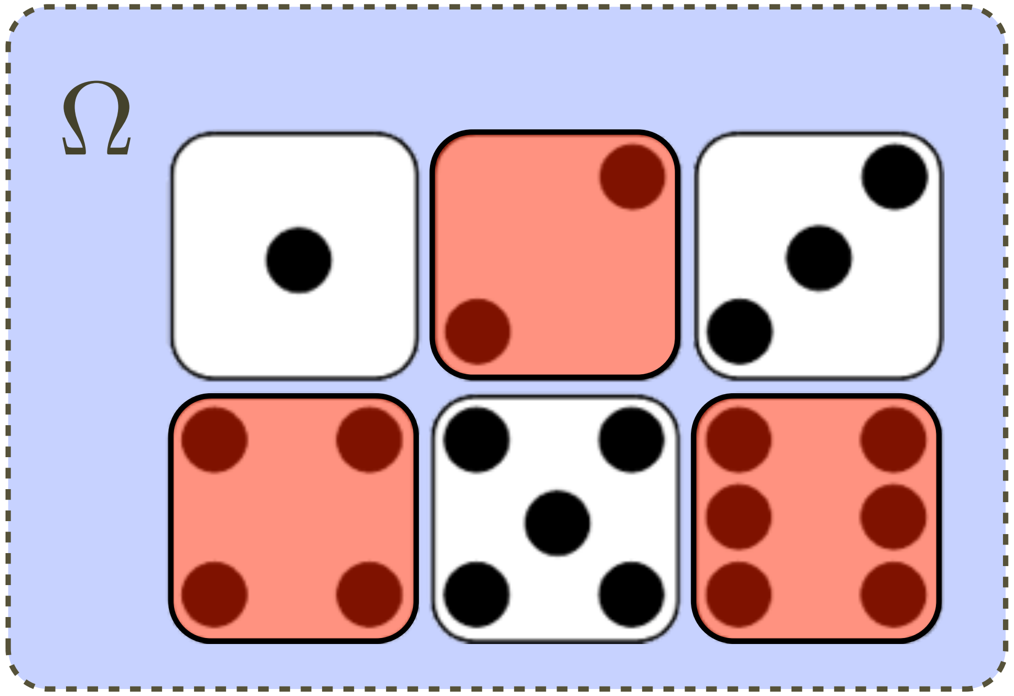
Randomized Algorithms

南京大学

尹一通

Probability Space



Sample space: Ω

Probability measure:

$$\Pr : \Omega \to [0,1]$$

s.t. $$\sum_{e \in \Omega} \Pr(e) = 1$$

event $A \subseteq \Omega$

probability $$\Pr(A) = \sum_{e \in A} \Pr(e)$$

Probability Space

Kolmogorov (1933)

Sample space Ω: set of all elementary events (samples)

Set of events Σ: each event is a subset of Ω

(K1) $\emptyset, \Omega \in \Sigma.$ impossible event, certain event

(K2) Σ is closed under $\cup, \cap, \setminus.$ σ-algebra

Probability measure $\Pr : \Sigma \to [0,1]$

(K3) $\Pr(\Omega) = 1$

(K4) $A \cap B = \emptyset \Rightarrow \Pr(A \cup B) = \Pr(A) + \Pr(B)$

(K5*) for $A_1 \supset \cdots \supset A_n \supset \cdots$ with $\bigcap_n A_n = \emptyset$

$$\lim_{n \to \infty} \Pr(A_n) = 0$$

(K1) $\emptyset, \Omega \in \Sigma$.

(K2) Σ is closed under $\cup, \cap, \setminus$.

(K3) $\Pr(\Omega) = 1$

(K4) $A \cap B = \emptyset \Rightarrow \Pr(A \cup B) = \Pr(A) + \Pr(B)$

$$\Pr(\Omega \setminus A) = 1 - \Pr(A)$$

If $A \subseteq B$, then $\Pr(A) \leq \Pr(B)$.

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

The Union bound

Works for arbitrary dependency!

Union bound (Boole's inequality):

$$\Pr\left(\bigcup_i A_i\right) \le \sum_i \Pr(A_i)$$

Inclusion-Exclusion:

$$\Pr\left(\bigcup_{i\in[n]} A_i\right) = \sum_{k=1}^{n}(-1)^{k-1}\sum_{S\in\binom{[n]}{k}}\Pr\left(\bigcap_{i\in S} A_i\right)$$

Boole-Bonferroni:

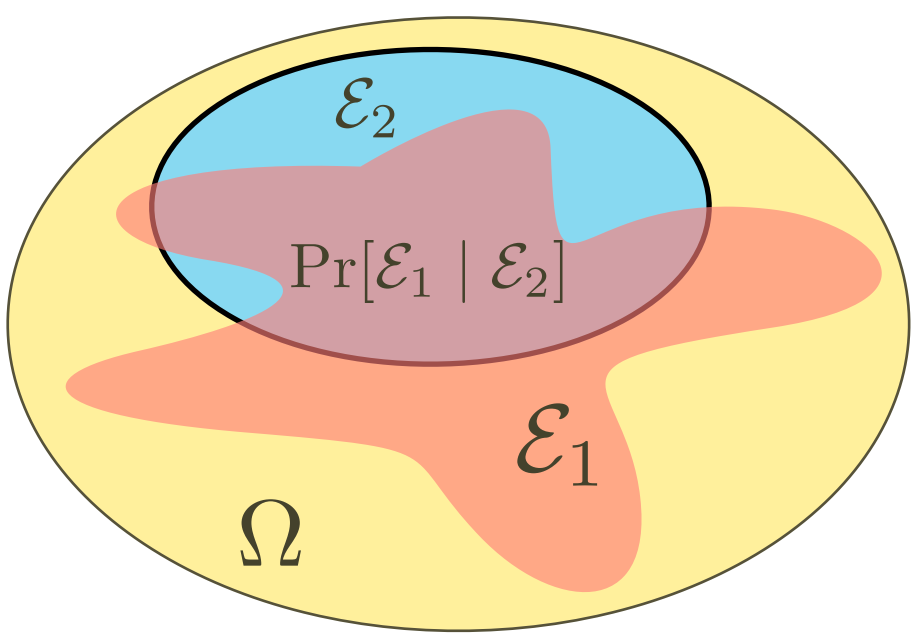
$$\sum_{k=1}^{2\ell}(-1)^{k-1}\sum_{S\in\binom{[n]}{k}}\Pr\left(\bigcap_{i\in S} A_i\right) \le \Pr\left(\bigcup_{i\in[n]} A_i\right) \le \sum_{k=1}^{2\ell+1}(-1)^{k-1}\sum_{S\in\binom{[n]}{k}}\Pr\left(\bigcap_{i\in S} A_i\right)$$

Conditional Probability

Definition:

The **conditional probability** that event $\mathcal{E}_1$ occurs given that event $\mathcal{E}_2$ occurs is

$$\Pr[\mathcal{E}_1 \mid \mathcal{E}_2] = \frac{\Pr[\mathcal{E}_1 \wedge \mathcal{E}_2]}{\Pr[\mathcal{E}_2]}.$$



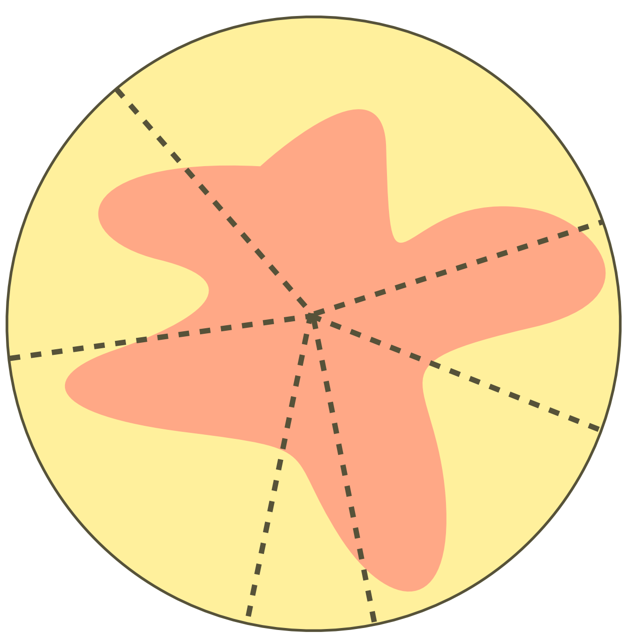
For independent $\mathcal{E}_1, \mathcal{E}_2$,

$$\Pr[\mathcal{E}_1 \mid \mathcal{E}_2] = \frac{\Pr[\mathcal{E}_1 \wedge \mathcal{E}_2]}{\Pr[\mathcal{E}_2]} = \frac{\Pr[\mathcal{E}_1] \cdot \Pr[\mathcal{E}_2]}{\Pr[\mathcal{E}_2]} = \Pr[\mathcal{E}_1]$$

Law of Total Probability

Law of total probability: For disjoint $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ that $\bigcup_i^n \mathcal{E}_i = \Omega$,

$$\Pr[\mathcal{E}] = \sum_{i=1}^{n} \Pr[\mathcal{E} \wedge \mathcal{E}_i] = \sum_{i=1}^{n} \Pr[\mathcal{E} \mid \mathcal{E}_i] \cdot \Pr[\mathcal{E}_i].$$



Analyze the probability by cases!

Law of Successive Conditioning

(chain rule)

Theorem

For any $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$,

$$\Pr\left[\bigwedge_{i=1}^{n} \mathcal{E}_i\right] = \prod_{k=1}^{n} \Pr\left[\mathcal{E}_k \mid \bigwedge_{i<k} \mathcal{E}_i\right].$$

Proof:

$$\Pr\left[\mathcal{E}_n \,\middle|\, \bigwedge_{i=1}^{n-1} \mathcal{E}_i\right] = \frac{\Pr\left[\bigwedge_{i=1}^{n} \mathcal{E}_i\right]}{\Pr\left[\bigwedge_{i=1}^{n-1} \mathcal{E}_i\right]}$$

recursion!

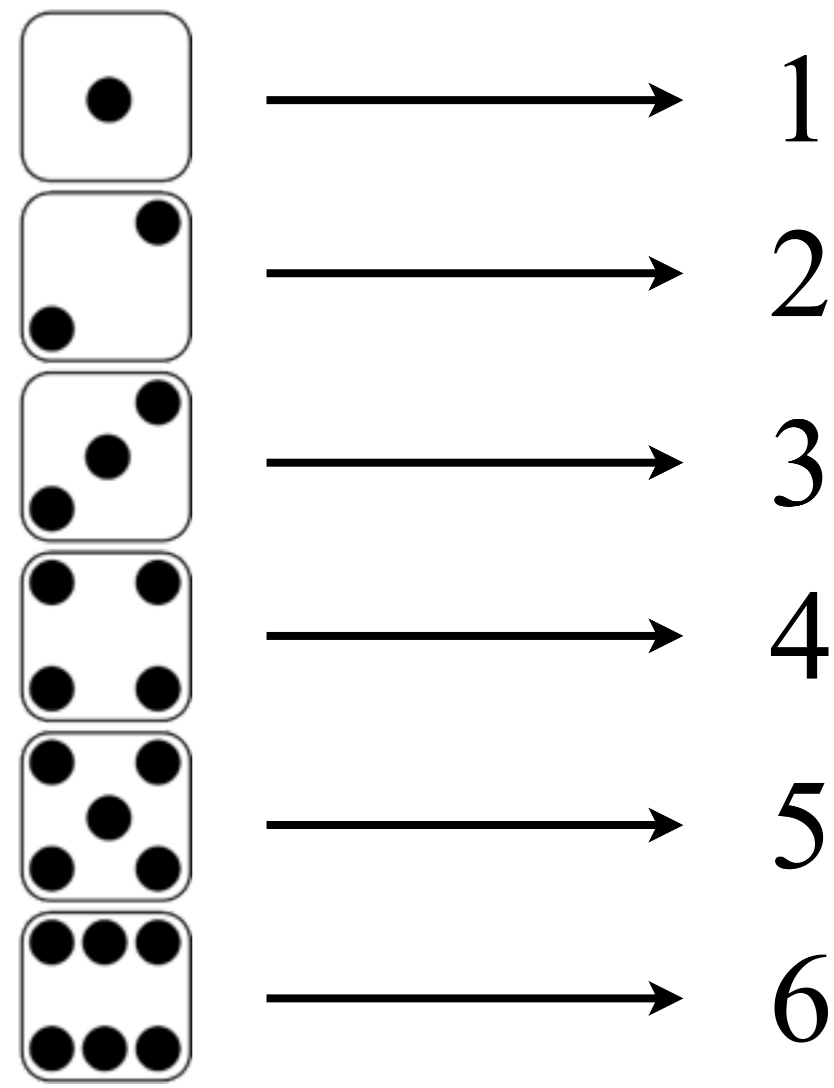
Random Variables

probability space:

$$(\Omega, \Sigma, \Pr)$$

random variable X

X is the outcome



Random Variables

probability space:

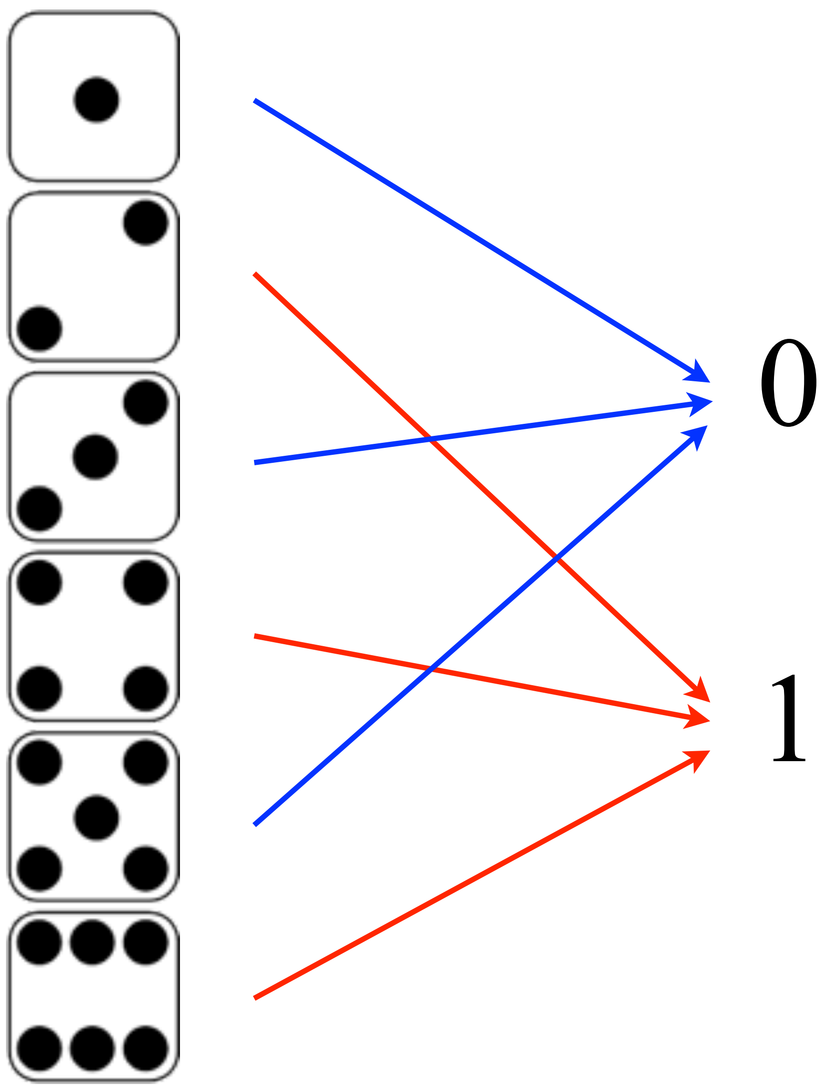
$$(\Omega, \Sigma, \Pr)$$

random variable X

a function defined over the sample space

$$X : \Omega \to \mathbb{R}$$

X indicates the evenness



Random Variables

random variable X

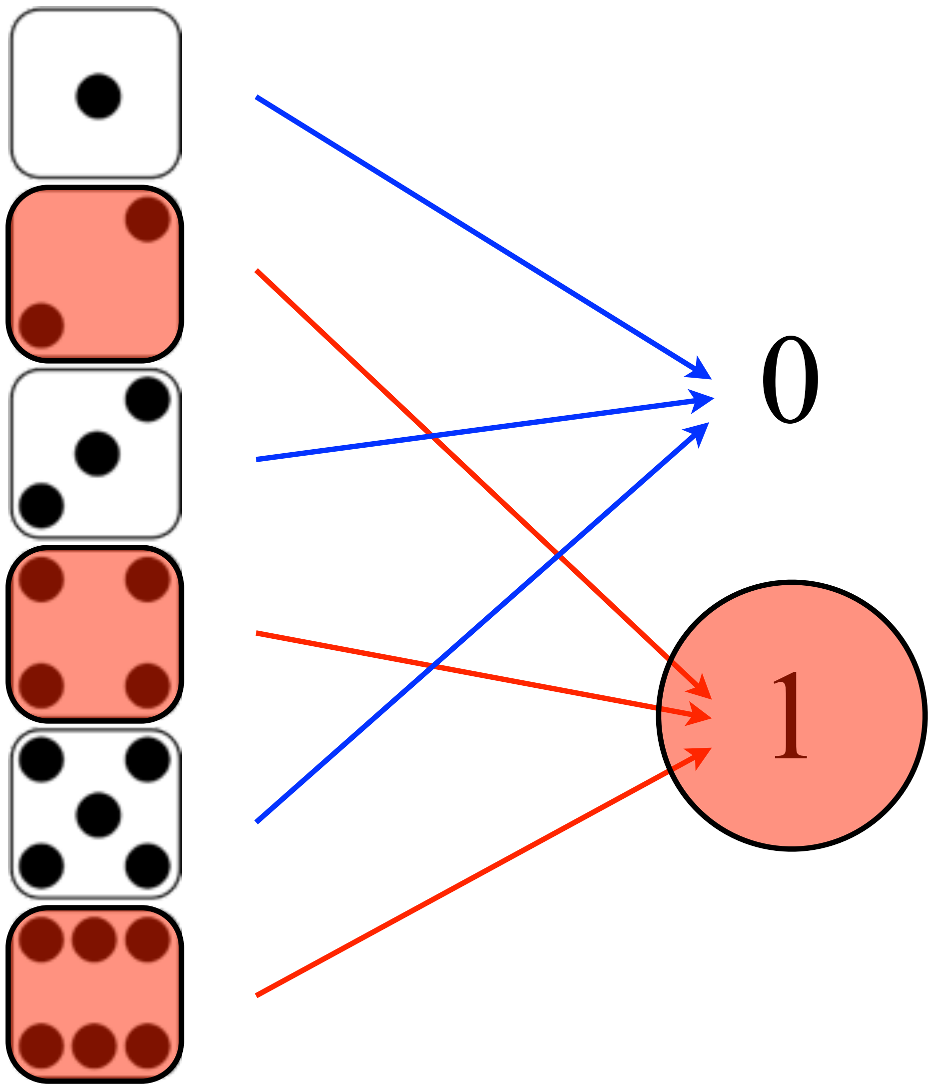
a function defined over the sample space

$$X : \Omega \to \mathbb{R}$$

event "$X=x$"

$$\Pr[X = x]$$
$$= \Pr\left(\{s \in \Omega \mid X(s) = x\}\right)$$

X indicates the evenness



Expectation

Definition:

The **expectation** of a discrete random variable X is

$$\mathbf{E}[X] = \sum_x x \cdot \Pr[X = x]$$

where the sum is over all values x in the range of X.

Linearity of expectations:

$$\mathbf{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \cdot \mathbf{E}[X_i].$$

Works for arbitrary dependency!

Linearity of Expectations

A monkey randomly types in 1 billion letters.
Expected number of “proof”s.

X_i indicates a “proof” started at position i

linearity + indicator ⇒ counter

$$\mathbf{E}\left[\sum_{i=1}^{10^9-4} X_i\right] = \sum_{i=1}^{10^9-4} \mathbf{E}[X_i] = (10^9-4)\Pr(\text{“proof”}) = \frac{10^9-4}{26^5} \approx 84$$

Coin Flipping

flip a *biased* coin:

- distribution of one flipping
  Bernoulli
- # of flips until HEADs occurs
  geometric
- # of HEADs in *n* flips
  binomial

Geometric distribution

(hitting time)

# of coin flips until a HEAD occurs.

- Run i.i.d. Bernoulli trials until succeeded.

  (Independently and Identically Distributed)

- X is the number of trials / coin flips.

$$\Pr[X = k] = (1-p)^{k-1}p$$

X follows the geometric distribution with parameter p.

Geometric distribution

Geometric X:

$$\Pr[X=k]=(1-p)^{k-1}p$$

brutal force:

$$\mathbf{E}[X]=\sum_{k=1}^{\infty}k\Pr[X=k]$$

$$=\sum_{k=1}^{\infty}k(1-p)^{k-1}p$$

$$\cdots \quad \cdots$$

$$=\frac{1}{p}$$

indicators:

$$Y_k=\begin{cases}1 & \text{the first } k \text{ trials fail}\\ 0 & \text{otherwise}\end{cases}$$

$$\Pr[Y_k=1]=(1-p)^k$$

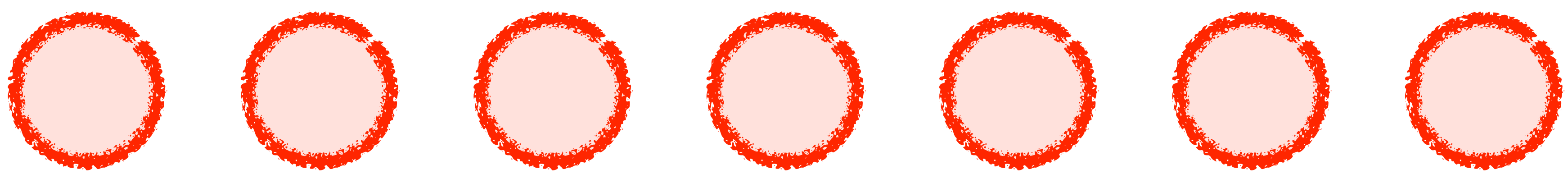
$$X=\sum_{k=0}^{\infty}Y_k$$

$$\mathbf{E}[X]=\sum_{k=0}^{\infty}\mathbf{E}[Y_k]$$ linearity of expectation

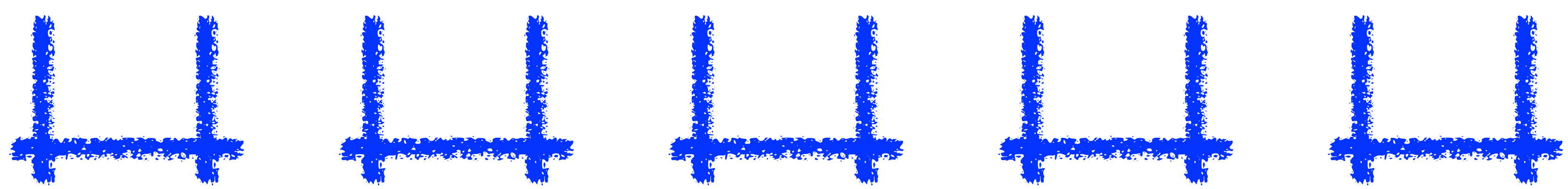
geometric $$=\sum_{k=0}^{\infty}(1-p)^k=\frac{1}{p}$$

Balls and Bins

m balls



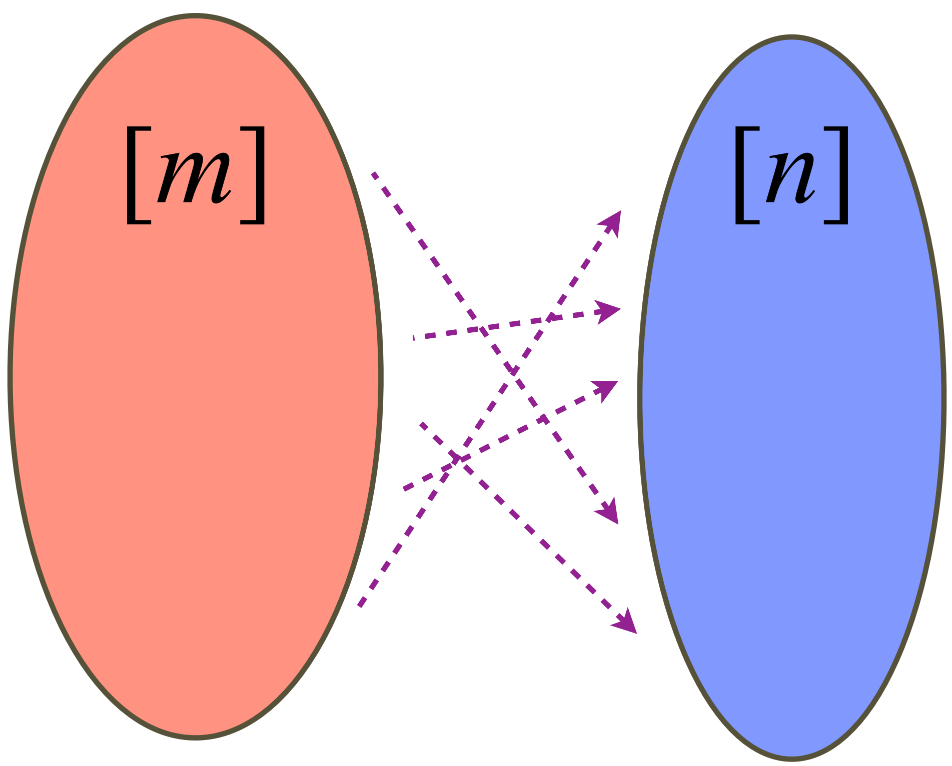
uniformly & independently



n bins

birthday problem, coupon collector problem,
occupancy problem, ...

Random function



uniformly random function

balls-into-bins:

$$\Pr[\text{assignment}] = \underbrace{\frac{1}{n} \cdot \frac{1}{n} \cdots \frac{1}{n}}_{m} = \frac{1}{n^m}$$

random function:

$$\Pr[\text{assignment}] = \frac{1}{|[m] \to [n]|} = \frac{1}{n^m}$$

1-1	birthday problem
on-to	coupon collector
pre-images	occupancy problem

Birthday Paradox

Paradox:

(i) a statement that leads to a contradiction;

(ii) a situation which defies intuition.

birthday paradox:

Assumption: birthdays are uniformly & independently distributed.

In a class of m>57 students, with >99% probability, there are two students with the same birthday.

m-balls-into-n-bins:

$\mathcal{E}$: there is no bin with > 1 balls.

Birthday Paradox

m-balls-into-n-bins:
$\mathcal{E}$: there is no bin with > 1 balls.

uniformly random $f : [m] \to [n]$,
$\mathcal{E}$: f is one-one.

$$\Pr[\mathcal{E}] = \frac{\left|[m] \xrightarrow{\text{1-1}} [n]\right|}{\left|[m] \to [n]\right|} = \frac{n \cdot (n-1) \cdots (n-m+1)}{n^m}$$

$$= \prod_{k=0}^{m-1} \left(1 - \frac{k}{n}\right)$$

Birthday Paradox

m-balls-into-n-bins:
$\mathcal{E}$: there is no bin with > 1 balls.

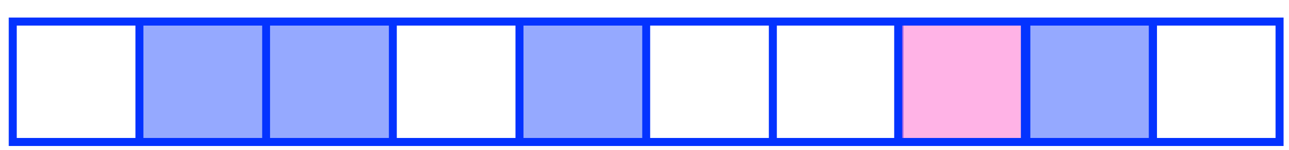
$$\Pr[\mathcal{E}] = \prod_{k=0}^{m-1}\left(1-\frac{k}{n}\right)$$

suppose balls are thrown one-by-one:

$$\Pr[\mathcal{E}] = \Pr[\text{no collision for all } m \text{ balls}]$$

$$= \prod_{k=0}^{m-1} \Pr[\text{no collision for the } (k+1)\text{th ball} \mid \text{no collision for the first } k \text{ balls}]$$

chain rule



Birthday Paradox

m-balls-into-n-bins:
$\mathcal{E}$: there is no bin with >1 balls.

$$\Pr[\mathcal{E}] = \prod_{k=0}^{m-1}\left(1-\frac{k}{n}\right)$$

Taylor's expansion: $e^{-k/n} \approx 1 - k/n$

$$\prod_{k=1}^{m-1}\left(1-\frac{k}{n}\right) \approx \prod_{k=1}^{m-1} e^{-\frac{k}{n}}$$

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

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

$$\approx e^{-m^2/2n}$$

Birthday Paradox

m-balls-into-n-bins:
$\mathcal{E}$: there is no bin with > 1 balls.

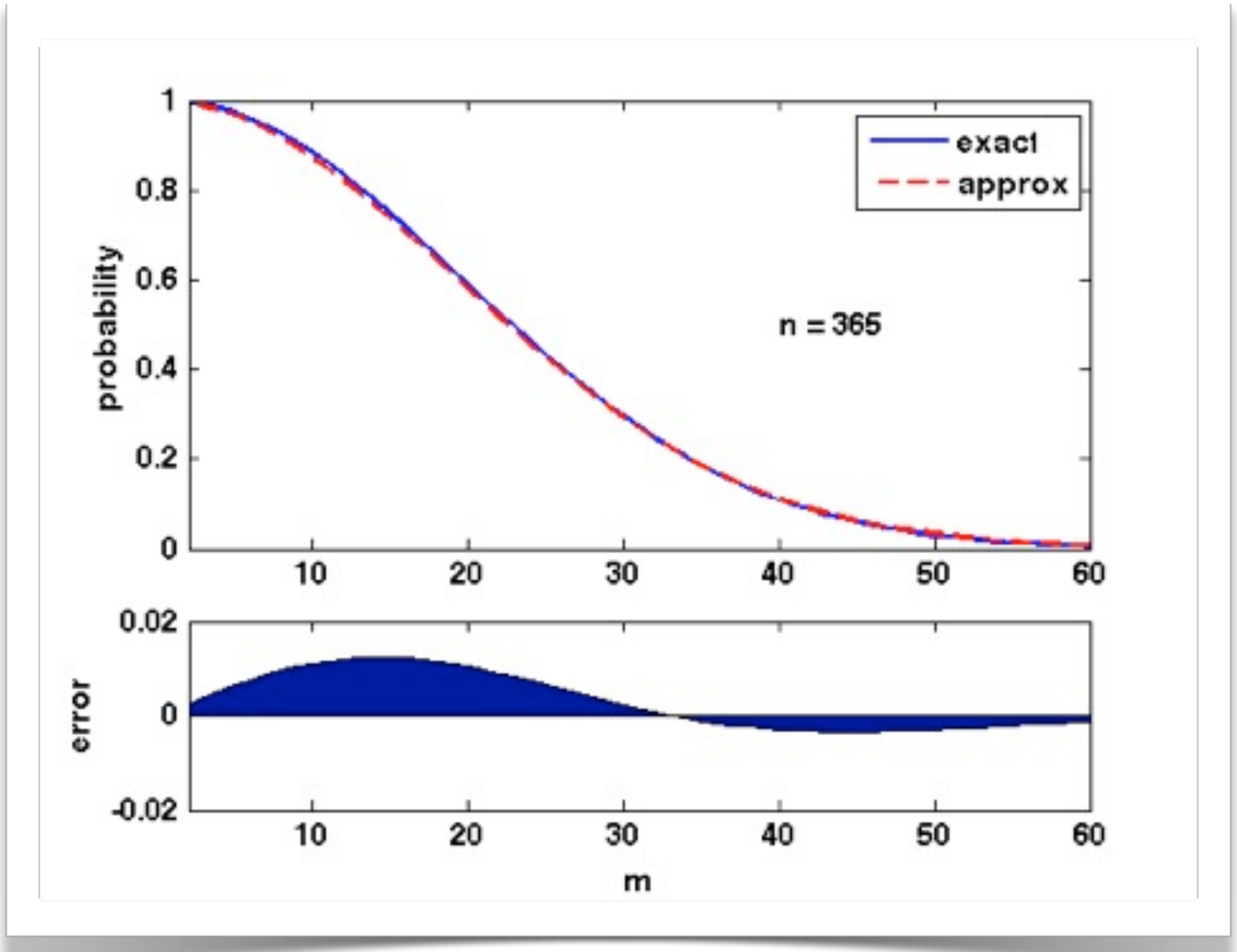
$$\Pr[\mathcal{E}] = \prod_{k=0}^{m-1}\left(1 - \frac{k}{n}\right)$$

$$\prod_{k=1}^{m-1}\left(1 - \frac{k}{n}\right) \approx e^{-m^2/2n}$$

$$\text{for } m = \sqrt{2n \ln \frac{1}{\epsilon}},$$

$$\Pr[\mathcal{E}] \approx \epsilon$$

$m = \theta(\sqrt{n})$ for constant ϵ



Perfect Hashing

$S = \{ a, b, c, d, e, f \}$

uniform random h $[N] \to [M]$ Pr[*perfect*] > 1/2

Table T:

e	b		d		f		c	a	

$M = O(n^2)$ birthday!

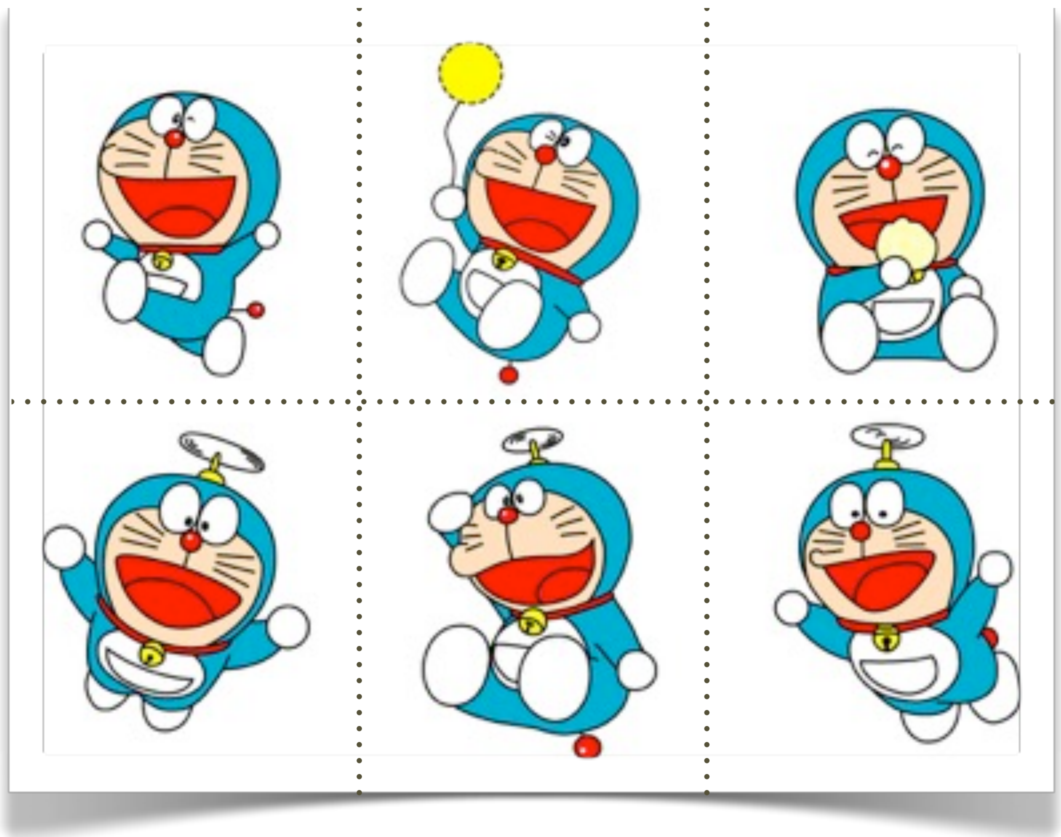
UHA: Uniform Hash Assumption

search(x): retrieve h;
check whether $T[h(x)] = x$;

Coupon Collector

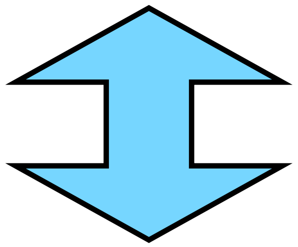
(cover time)

coupons in cookie box



each box comes with a uniformly random coupon

number of boxes bought to collect all n coupons

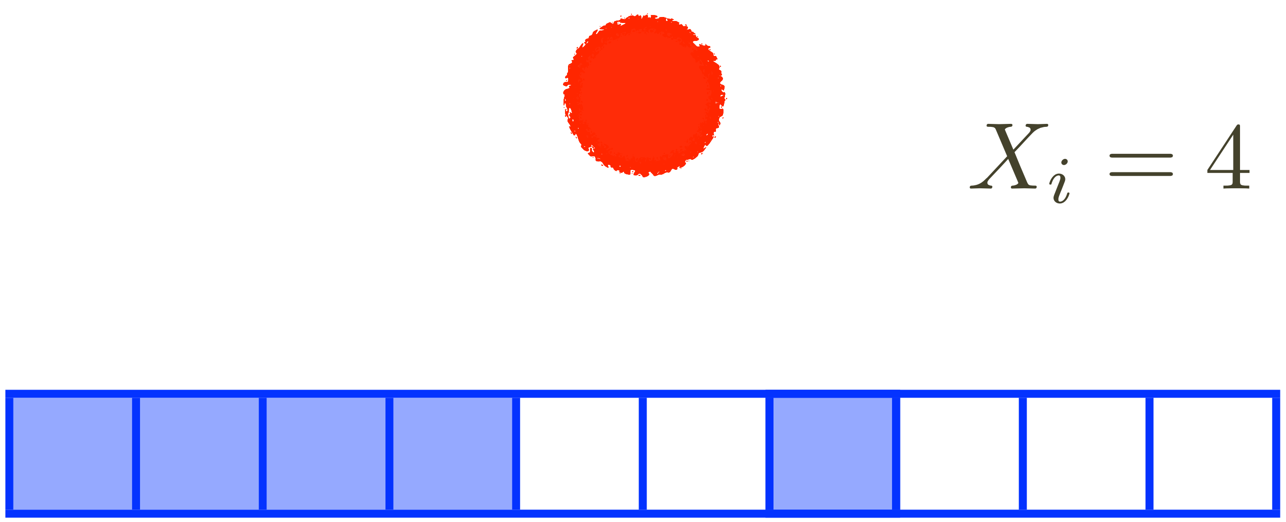


number of balls thrown to cover all n bins

Coupon Collector

$X:$ number of balls thrown to make all the n bins nonempty

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



X_i is geometric!

with $p_i = 1 - \frac{i-1}{n}$

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

Coupon Collector

X :	number of balls thrown to make all the n bins nonempty
X_i :	number of balls thrown while there are exactly $(i$-$1)$ nonempty bins

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

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

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i]$$

linearity of expectations

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

Expected $n \ln n + O(n)$ balls!

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

$$= nH(n)$$

Harmonic number

Coupon Collector

X : number of balls thrown to make all the n bins nonempty

Theorem: For $c > 0$

$$\Pr[X \geq n \ln n + cn] < e^{-c}$$

Proof: For one bin, it misses all balls with probability

$$\left(1 - \frac{1}{n}\right)^{n \ln n + cn} = \left(1 - \frac{1}{n}\right)^{n(\ln n + c)}$$

$$< e^{-(\ln n + c)}$$

$$= \frac{1}{ne^c}$$

Coupon Collector

X : number of balls thrown to make all the n bins nonempty

Theorem: For $c > 0$

$$\Pr[X \geq n \ln n + cn] < e^{-c}$$

Proof: For one bin, it misses all balls with probability

$$< \frac{1}{ne^c}$$

For all n bins, union bound!

$$\Pr[\exists \text{ a bin misses all balls}] \leq n \cdot \Pr[\text{one bin misses all balls}]$$

$$< n \cdot \frac{1}{ne^c} = e^{-c}$$

Coupon Collector

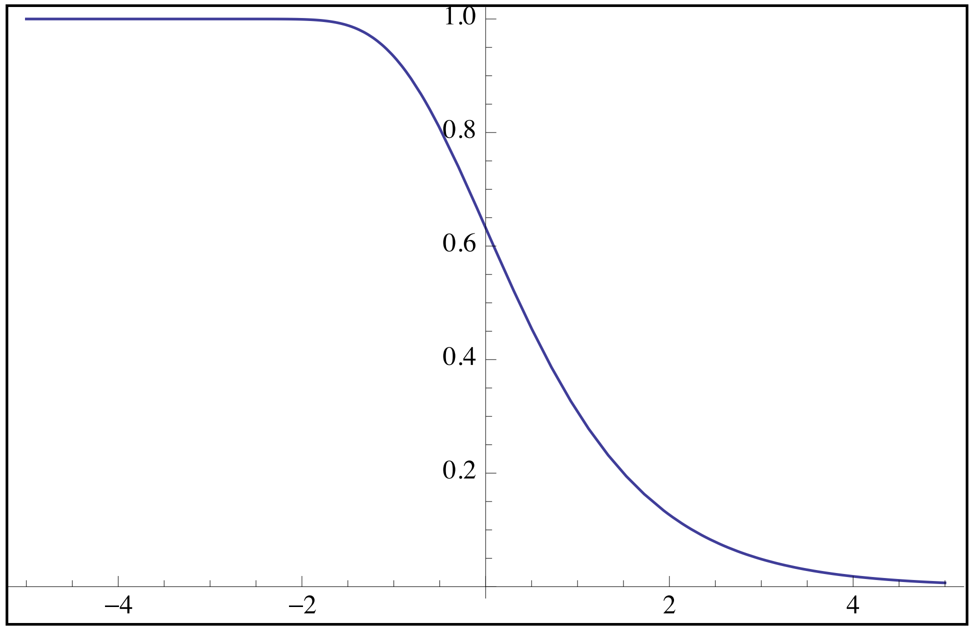
X : number of balls thrown to make all the n bins nonempty

Theorem: For $c > 0$

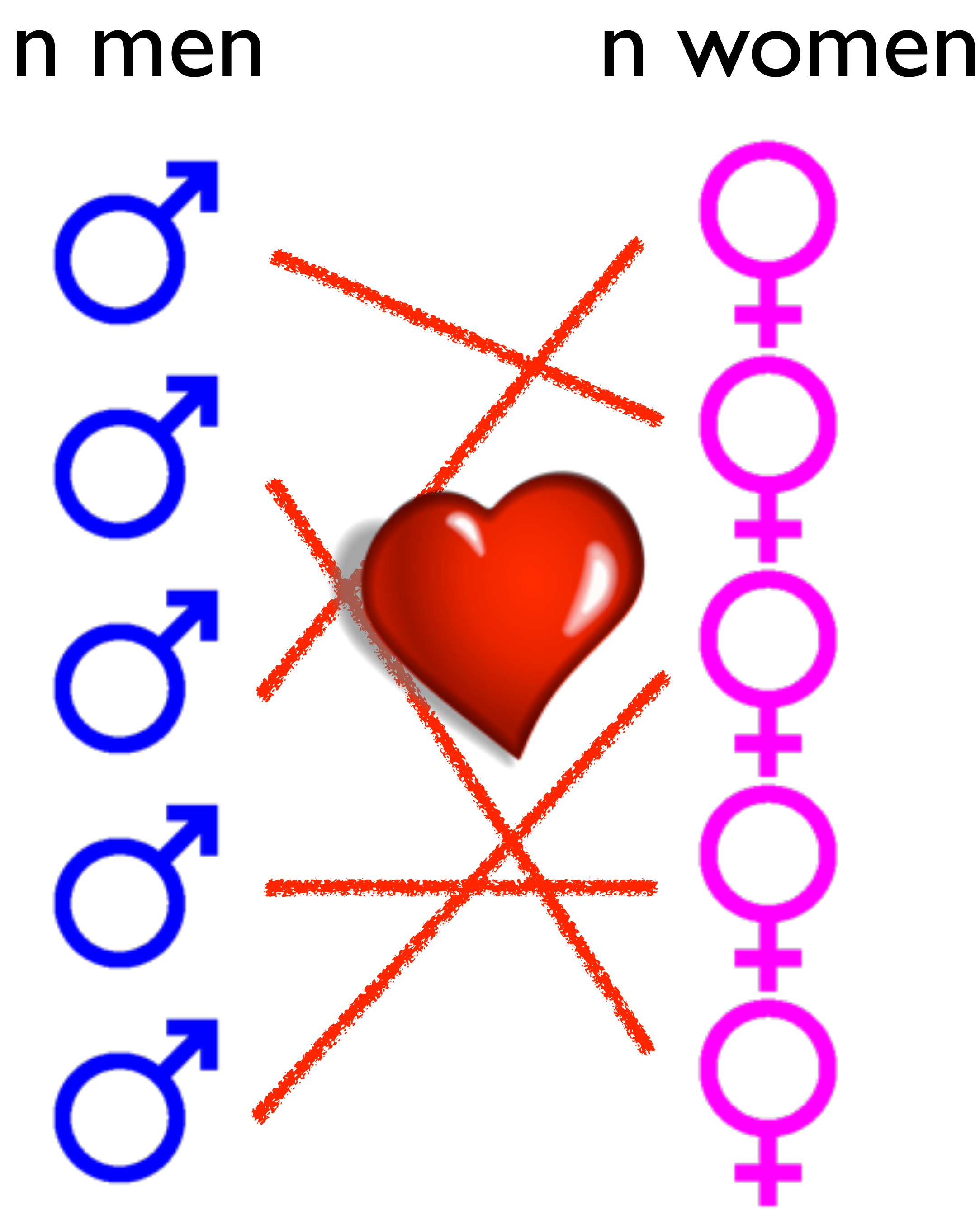
$$\Pr[X \geq n \ln n + cn] < e^{-c}$$

a sharp threshold:

$$\lim_{n\to\infty} \Pr[X \geq n \ln n + cn] = 1 - \mathrm{e}^{-\mathrm{e}^{-c}}$$

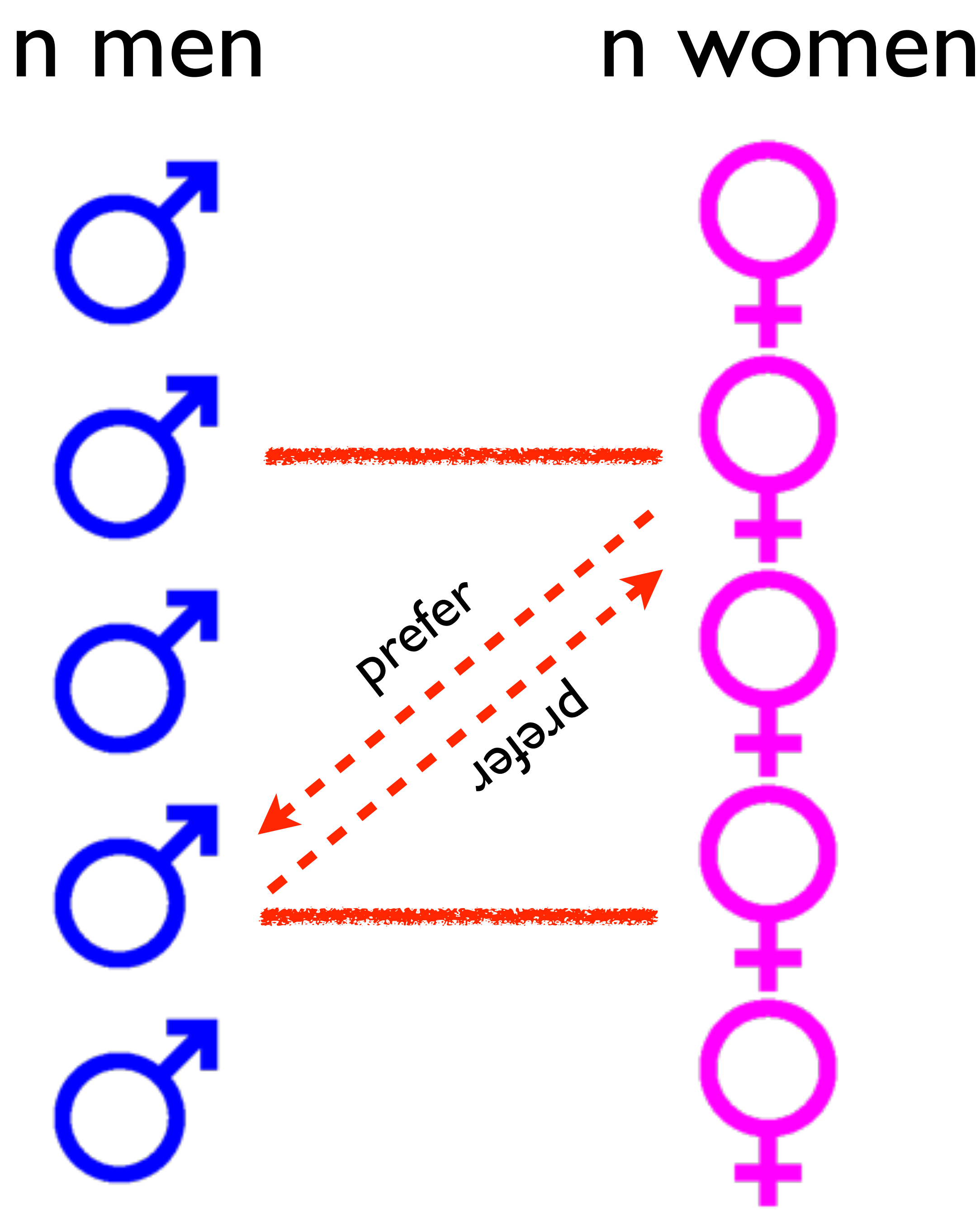


Stable Marriage



- each man has a preference order of the *n* women;
- each woman has a preference order of the *n* men;
- solution: *n* couples
- Marriages are stable!

Stable Marriage



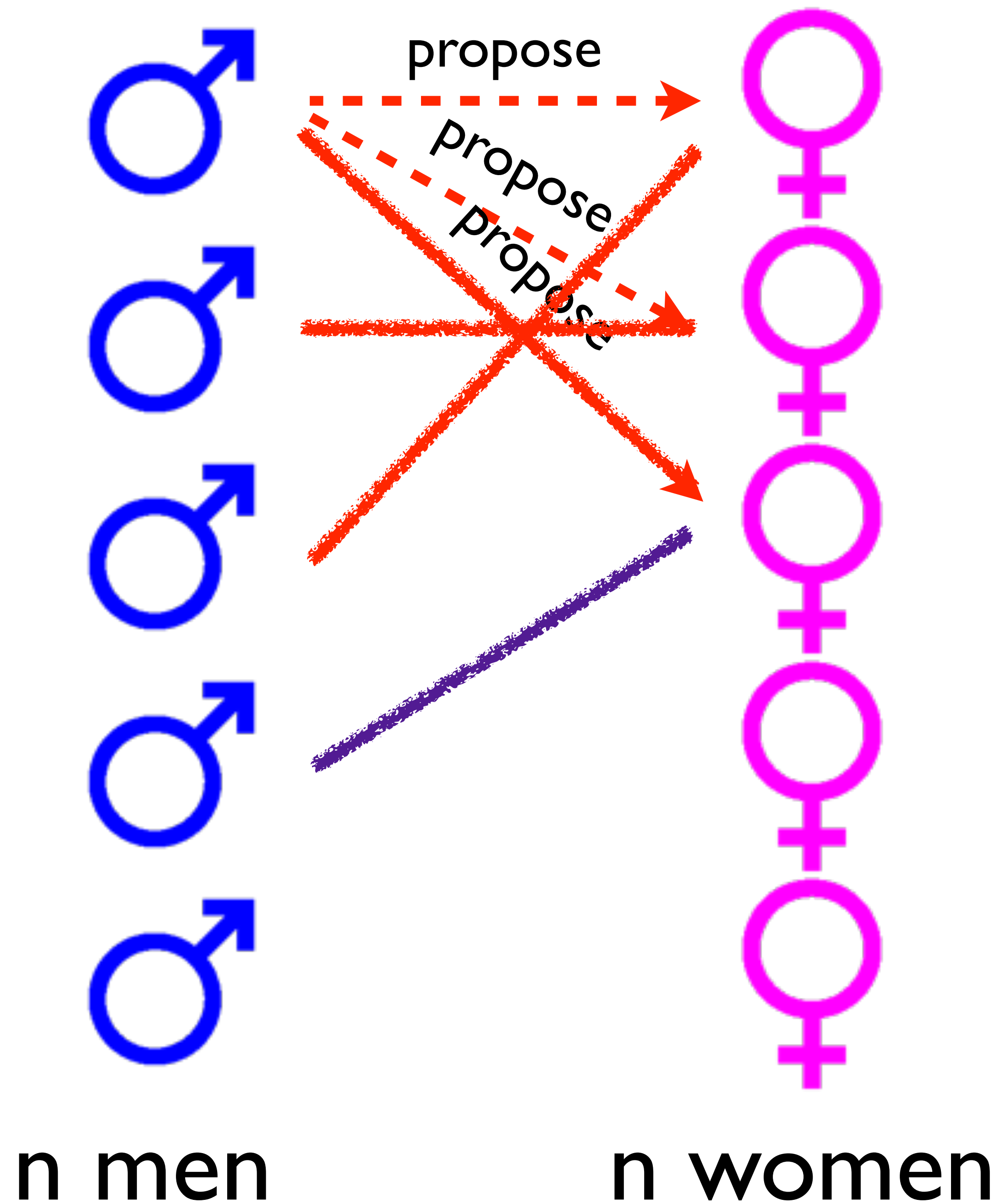
unstable (blocking pair):
a man and a woman, who prefer each other to their current partners

stable: no blocking pairs

local optimum
fixed point
equilibrium
deadlock

Proposal Algorithm

(Gale-Shapley 1962)



Single man:

propose to the most preferable women who has not rejected him

Woman:

upon received a proposal: accept if she's single or married to a less preferable man (divorce!)

Proposal Algorithm

- woman: once got married always married

(will only switch to better men!)

- man: will only get worse ...

- once all women are married, the algorithm terminates, and the marriages are stable

- total number of proposals:

$$\leq n^2$$

Single man:
propose to the most preferable women who has not rejected him

Woman:
upon received a proposal: ... single or ... a less ... man ... !)

if "A" and "b" prefer each other than their current partners "a" and "B", then "A" would have proposed to "b" before to "a", and "b" should have accepted

this proves the existence of stable matching by construction

Average-case

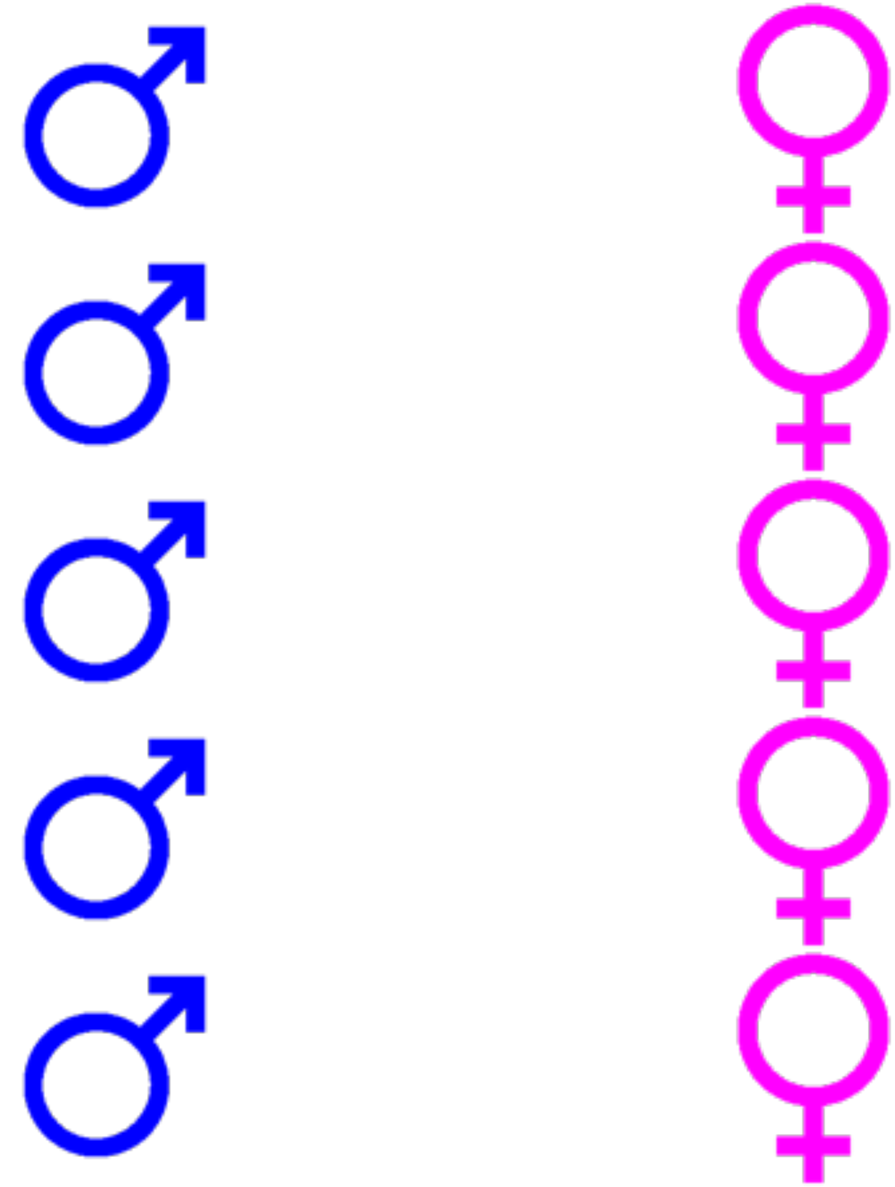
- every man/woman has a uniform random permutation as preference list

- total number of proposals?

Looks very complicated!

men propose

women change minds



Principle of Deferred Decisions

Principle of deferred decision

The decision of random choice in the random input is deferred to the running time of the algorithm.

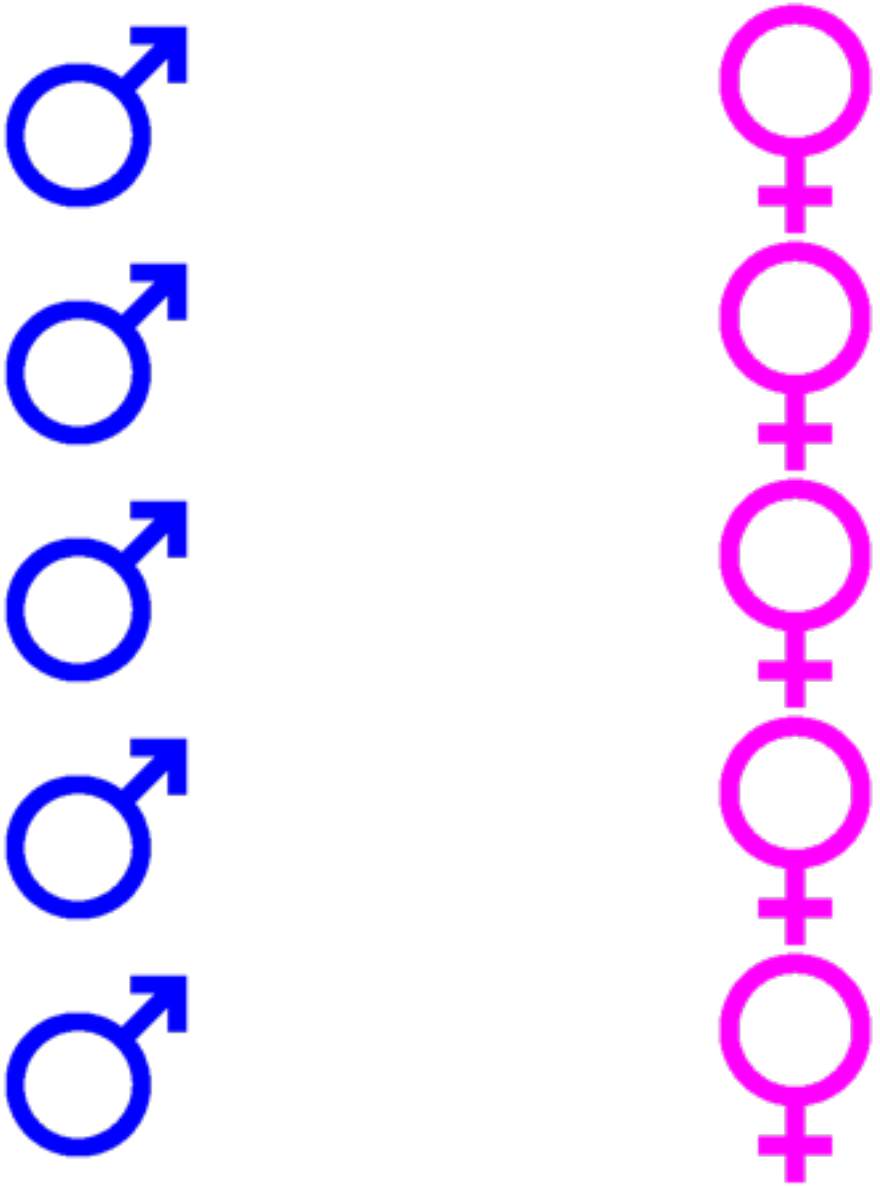
Principle of Deferred Decisions

proposing in the order of a uniformly random permutation

⇕

at each time, proposing to a uniformly random woman who has not rejected him

decisions of the inputs are deferred to the time when Alg accesses them



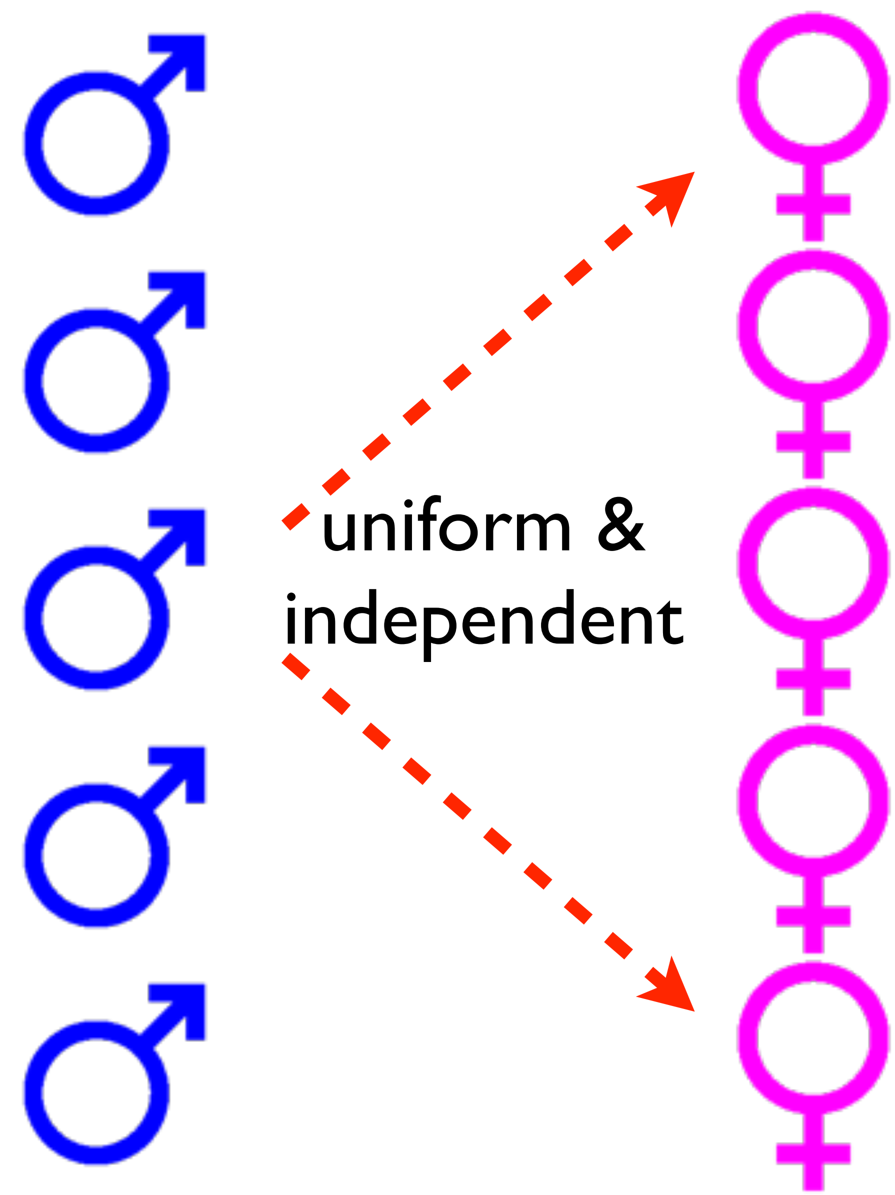
Coupling

at each time, proposing to a uniformly random woman who has not rejected him

$\leq$

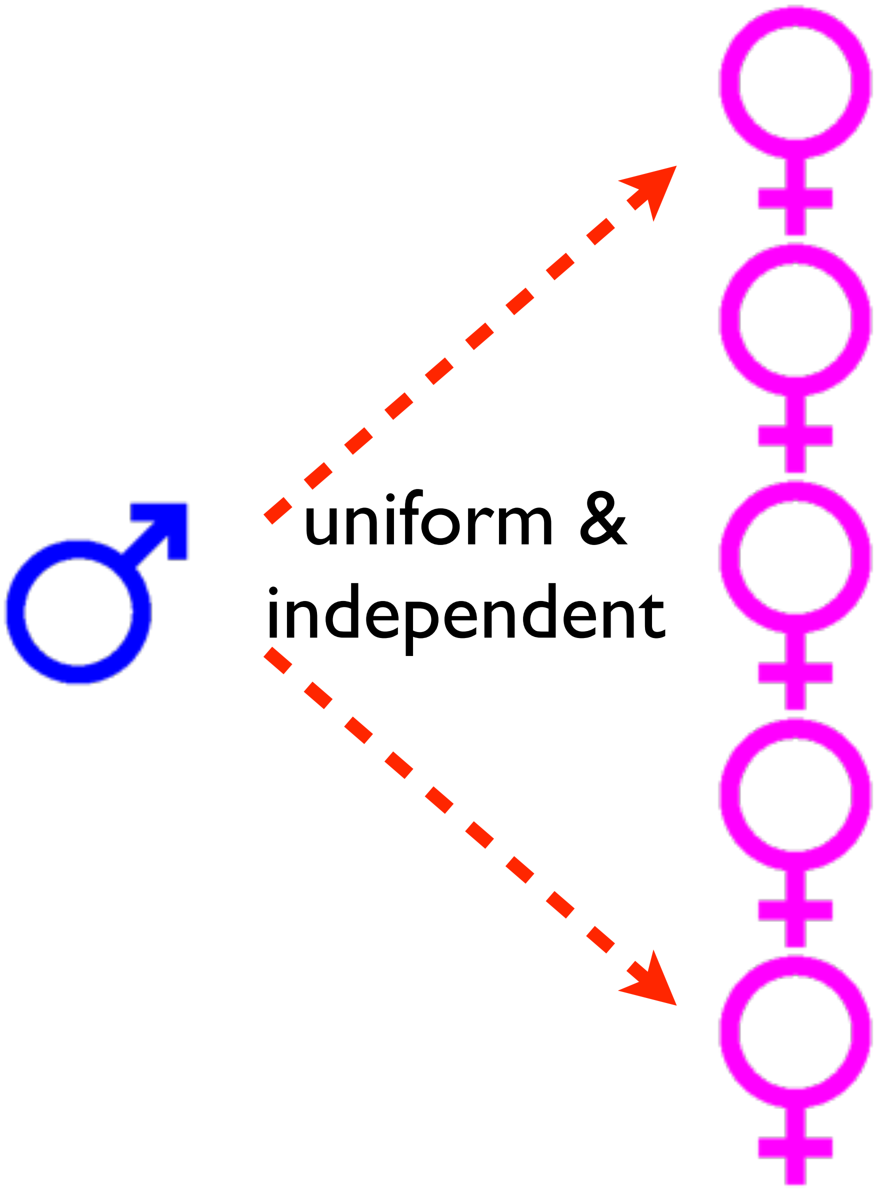
at each time, proposing to a uniformly & independently random woman

the man forgot who had rejected him (!)



Average-case

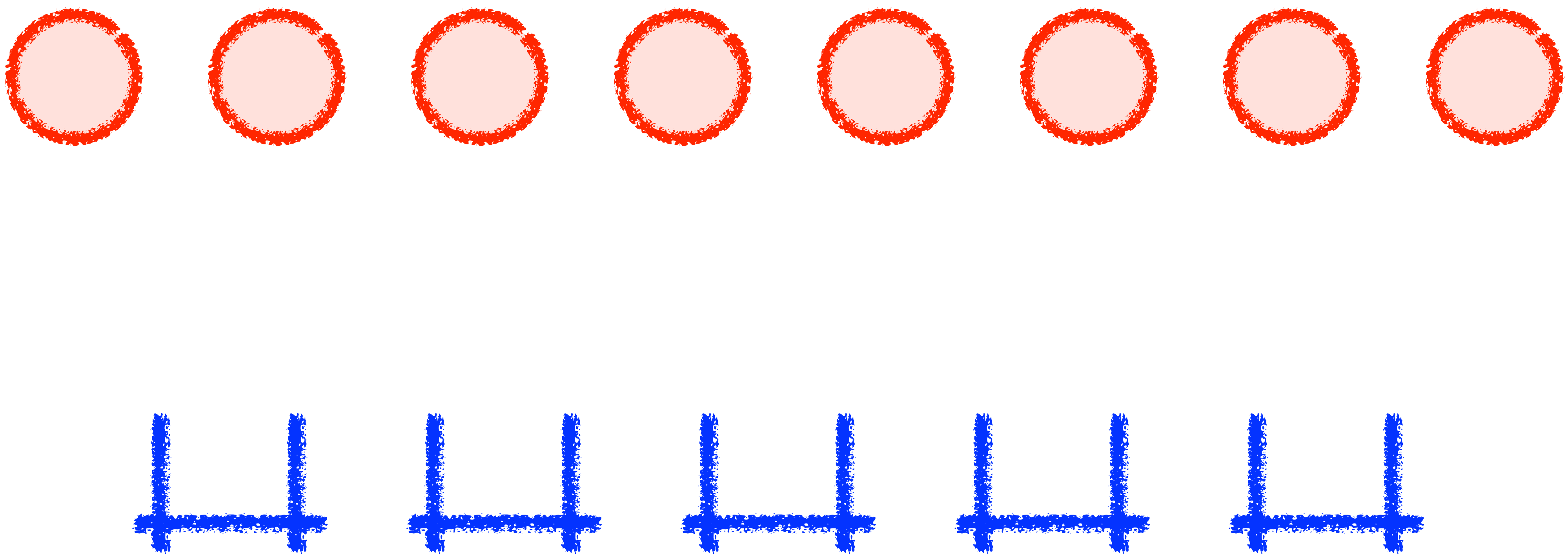
- uniformly and independently proposing to *n* women
- Alg stops once all women got proposed.
- Coupon collector!
- Expected O(*n* ln *n*) proposals.



Occupancy Problem

(load balancing)

m balls



n bins

$X_1, X_2, \ldots, X_n$

loads of bins

maximum load?

Occupancy Problem

m balls

n bins

$X_1, X_2, \ldots, X_n$

loads of bins

$$\max_{1 \le i \le n} \mathbf{E}[X_i] = ?$$

$$\sum_{i=1}^{n} X_i = m \Rightarrow \sum_{i=1}^{n} \mathbf{E}[X_i] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = m$$

Symmetry! $\Rightarrow$ All $\mathbf{E}[X_i]$ are equal.

$$\max_{1 \le i \le n} \mathbf{E}[X_i] = \frac{m}{n}$$

Occupancy Problem

$$\max_{1 \le i \le n} \mathbf{E}[X_i] = \frac{m}{n}$$

Theorem:

If $m = n$, the max load is $O\left(\frac{\ln n}{\ln \ln n}\right)$ with high probability.

w.h.p.: $\Pr = 1 - O(\frac{1}{n^c})$ or $\Pr = 1 - o(1)$

n balls into n bins:

$$\Pr[\text{ bin-1 has } \geq t \text{ balls }]$$

$$\leq \Pr[\exists \text{ a set } S \text{ of } t \text{ balls s.t. all balls in } S \text{ are in bin-1 }]$$

$$\binom{n}{t} \qquad \frac{1}{n^t}$$

union bound

$$\leq \sum_{\text{set } S \text{ of } t \text{ balls}} \Pr[\text{all balls in } S \text{ are in bin-1}]$$

$$\leq \frac{1}{n^t}\binom{n}{t} = \frac{n(n-1)(n-2)\cdots(n-t+1)}{t!n^t} \leq \frac{1}{t!} \leq \left(\frac{e}{t}\right)^t$$

Stirling approximation

n balls into n bins:

$$\Pr[\text{ bin-1 has } \geq t \text{ balls }] \leq \left(\frac{\mathrm{e}}{t}\right)^t$$

$$\Pr[\text{ max load } \geq t] = \Pr[\exists \text{ bin-}i \text{ has } \geq t \text{ balls}]$$

$$\leq n\Pr[\text{ bin-1 has } \geq t \text{ balls }]$$ union bound

$$\leq n\left(\frac{\mathrm{e}}{t}\right)^t$$ **choose** $t = \dfrac{3\ln n}{\ln\ln n}$

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

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

$$\leq ne^{-3\ln n + 3(\ln\ln\ln n)(\ln n)/\ln\ln n}$$

$$\leq ne^{-2\ln n} = \frac{1}{n}$$

Occupancy Problem

Theorem: m balls into n bins:

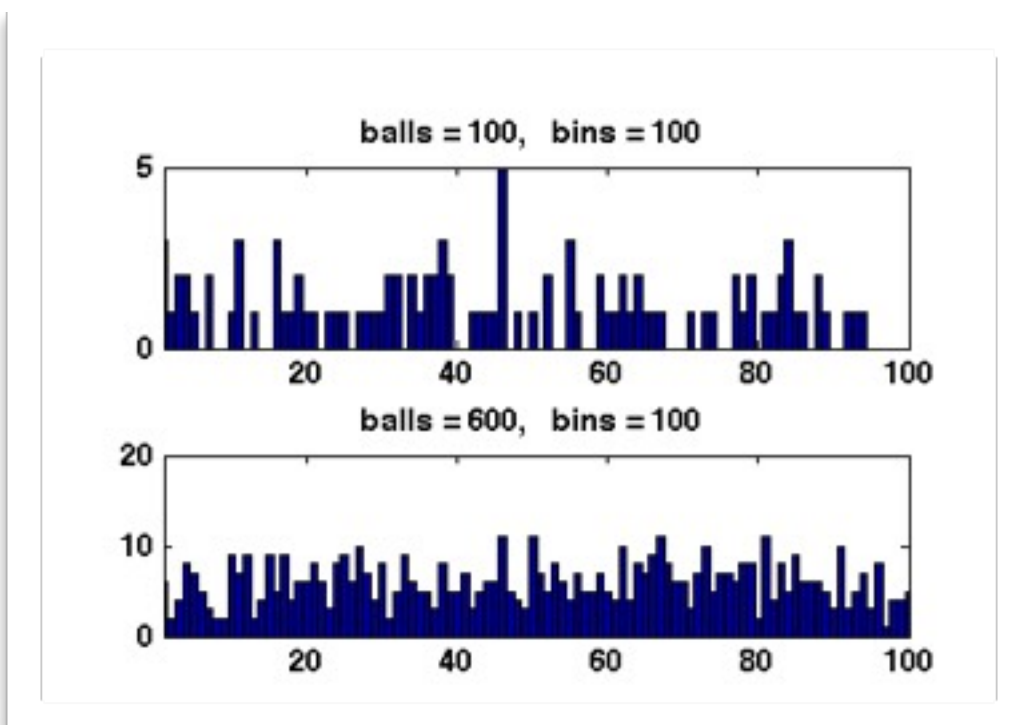
If $m = n$, the max load is $O\left(\frac{\ln n}{\ln \ln n}\right)$ with high probability.

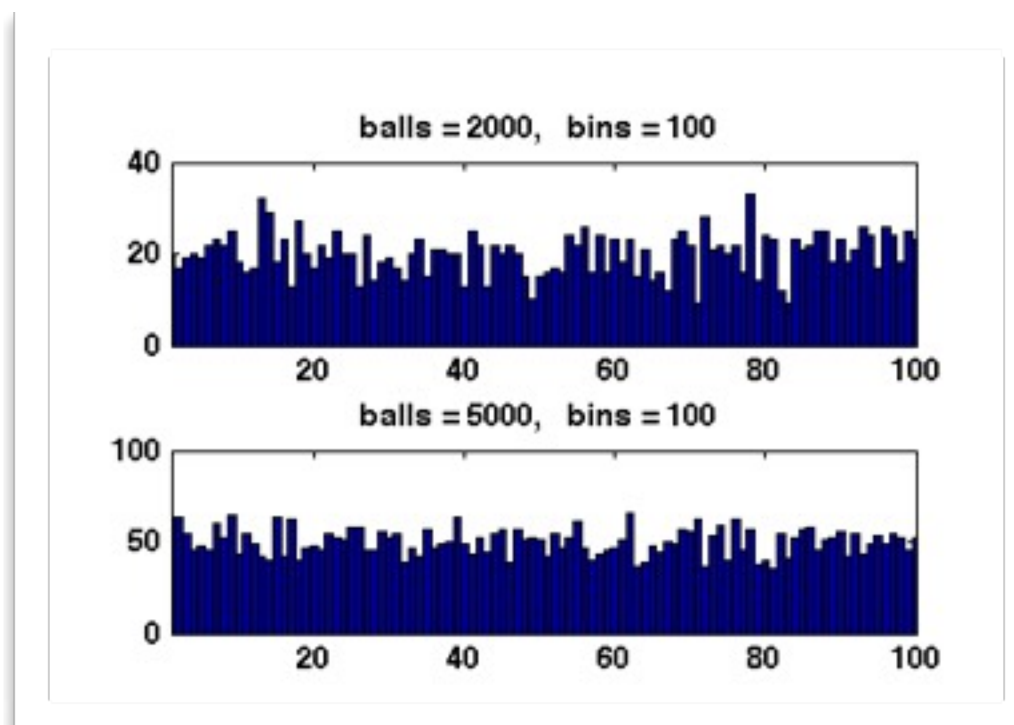
Occupancy Problem

Theorem: m balls into n bins:

If $m = n$, the max load is $O\left(\frac{\ln n}{\ln \ln n}\right)$ with high probability.

When $m = \Omega(n \log n)$, the max load is $O(\frac{m}{n})$ with high probability





Balls-into-bins model

throw *m* balls into *n* bins
uniformly and independently

uniform random function
$$f : [m] \to [n]$$

1-1	birthday problem
on-to	coupon collector
pre-images	occupancy problem

- The threshold for being 1-1 is $m = \Theta(\sqrt{n})$.
- The threshold for being on-to is $m = n \ln n + O(n)$.
- The maximum load is

$$\begin{cases} O\left(\frac{\ln n}{\ln \ln n}\right) & \text{for } m = \Theta(n), \\ O\left(\frac{m}{n}\right) & \text{for } m = \Omega(n \ln n). \end{cases}$$