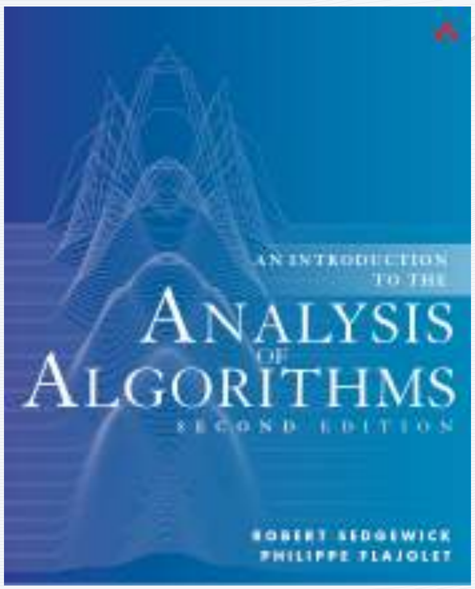
ANALYTIC COMBINATORICS

PART ONE



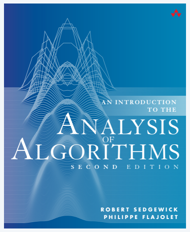
http://aofa.cs.princeton.edu

9. Words and Mappings

Orientation

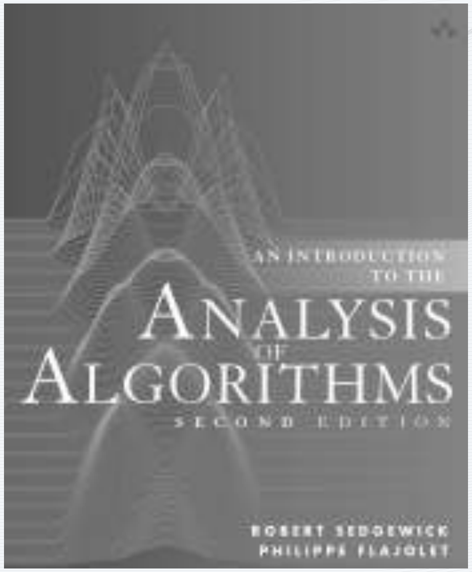
Second half of class

- Surveys fundamental combinatorial classes.
- Considers techniques from analytic combinatorics to study them .
- Includes applications to the analysis of algorithms.



chapter	*combinatorial classes*	*type of class*	*type of GF*
6	Trees	unlabeled	OGFs
7	Permutations	labeled	EGFs
8	Strings and Tries	unlabeled	OGFs
9	Words and Mappings	labeled	EGFs

Note: Many more examples in book than in lectures.



http://aofa.cs.princeton.edu

9. Words and Mappings

- Words
- Birthday problem
- Coupon collector problem
- Hash tables
- Mappings

Symbolic method for unlabelled objects (review)

Warmup: How many binary strings with N bits?

Class	B, the class of all binary strings
Size	$|b|$, the number of bits in b
OGF	$B(z) = \sum_{b \in B} z^{|b|} = \sum_{N \geq 0} B_N z^N$

Atoms

type	*class*	*size*	*GF*
0 bit	Z_0	1	z
1 bit	Z_1	1	z

Construction $$B = SEQ(Z_0 + Z_1)$$

"a binary string is a sequence of 0 bits and 1 bits"

OGF equation $$B(z) = \frac{1}{1-2z}$$

$$[z^N]B(z) = 2^N \quad ✓$$

Symbolic method for unlabelled objects (review)

How many strings drawn from an M-char alphabet with N chars?

Class	S, the class of all strings
Size	$\lvert s \rvert$, the number of chars in s
OGF	$S(z) = \sum_{s \in S} z^{\lvert s \rvert} = \sum_{N \geq 0} S_N z^N$

Atoms

type	*class*	*size*	*GF*
char 1	Z_1	1	z
char 2	Z_2	1	z
. . .			
char M	Z_M	1	z

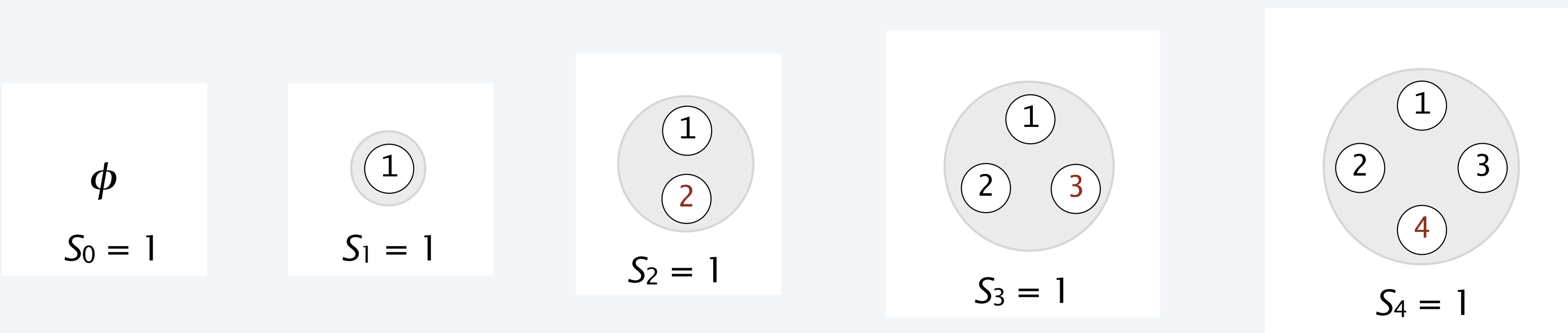
Construction $S = SEQ(Z_1 + Z_2 + \ldots + Z_M)$ "a string is a sequence of chars"

OGF equation $$S(z) = \frac{1}{1 - Mz}$$

Extract coefficients $[z^N]S(z) = M^N$ ✓

Symbolic method for labelled objects (review): sets

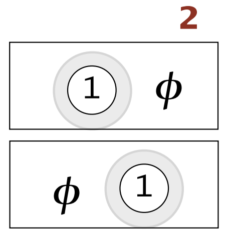
Q. How many labeled sets (urns) of size N?



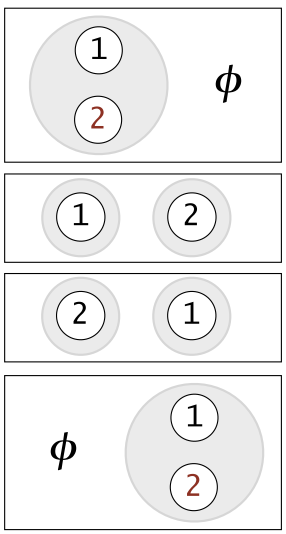
A. One.

Labelled objects review (continued): sets

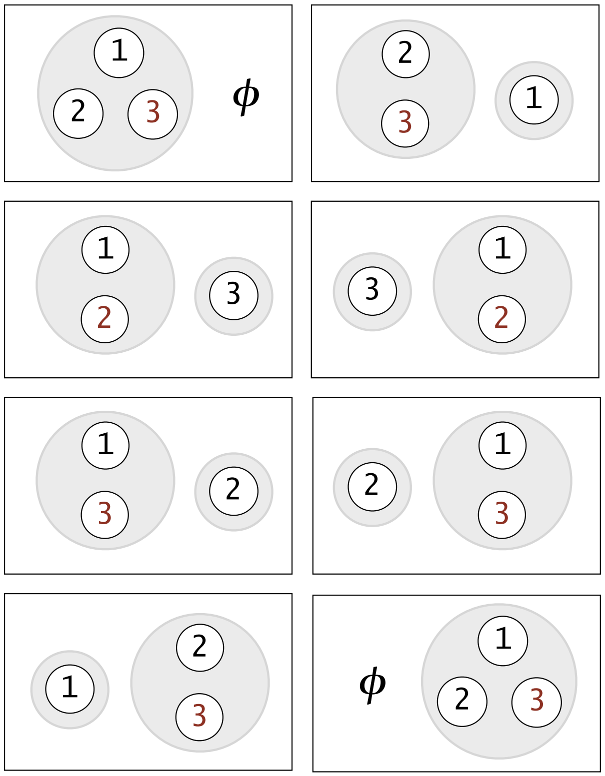
Q. How many ordered pairs of labelled sets of N objects?



$P_1 = 2$



$P_2 = 4$



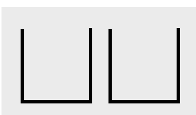
$P_3 = 8$

A. 2^N

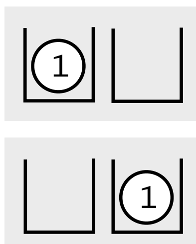
Q. How many sequences of length M of urns with N objects in total ?

Balls-and-urns viewpoint

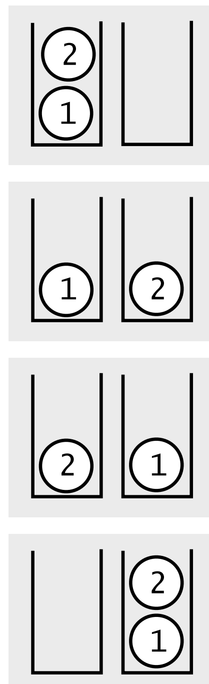
Q. How many different ways to throw N balls into 2 urns?



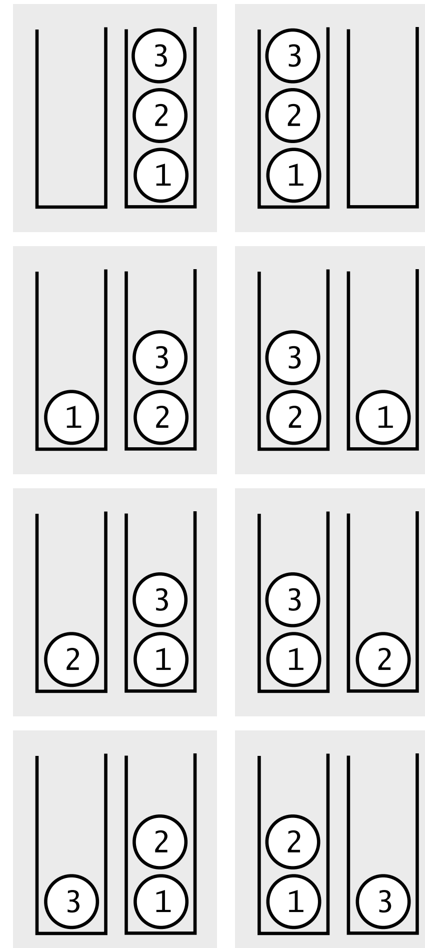
$W_0 = 1$



$W_1 = 2$



$W_2 = 4$



$W_3 = 8$

A. 2^N

The symbolic method for labelled classes (review)

Theorem. Let A and B be combinatorial classes of labelled objects with EGFs $A(z)$ and $B(z)$. Then

construction	*notation*	*semantics*	*EGF*
disjoint union	$A + B$	disjoint copies of objects from A and B	$A(z) + B(z)$
labelled product	$A \star B$	ordered pairs of copies of objects, one from A and one from B	$A(z)B(z)$
sequence	$SEQ_k(A)$	k- sequences of objects from A	$A(z)^k$
	$SEQ(A)$	sequences of objects from A	$\dfrac{1}{1 - A(z)}$
set	$SET_k(A)$	k-sets of objects from A	$A(z)^k/k!$
	$SET(A)$	sets of objects from A	$e^{A(z)}$
cycle	$CYC_k(A)$	k-cycles of objects from A	$A(z)^k/k$
	$CYC(A)$	cycles of objects from A	$\ln \dfrac{1}{1 - A(z)}$

Words

Def. A *word* is a sequence of M urns holding N objects in total.

"throw N balls into M urns"

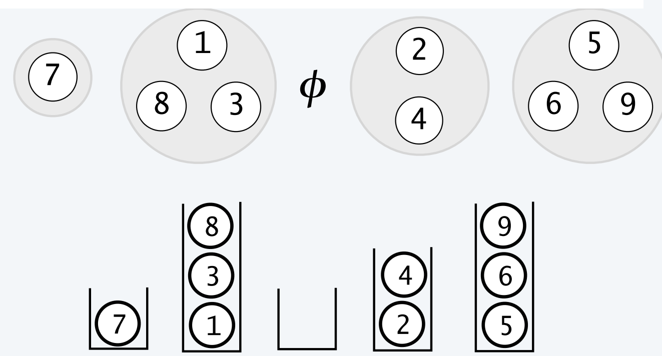
Q. How many words ?

Class	W_M, the class of M-sequences of urns
Size	$\|w\|$, the number of objects in w
EGF	$W_M(z) = \sum_{w \in W_M} \frac{z^{\|w\|}}{\|w\|!} = \sum_{N \geq 0} W_{MN} \frac{z^N}{N!}$

Atom

type	class	size	GF
labelled atom	Z	1	z

Example { 7 } { 1 8 3 } { } { 2 4 } { 5 6 9 }



Construction $\quad W_M = SEQ_M(SET(Z))$

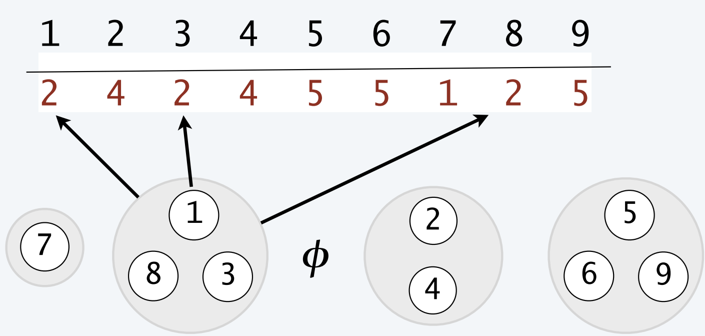
OGF equation $\quad W_M(z) = (e^z)^M = e^{Mz}$

Counting sequence $\quad N![z^N]W_M(z) = M^N$

A 1:1 correspondence

A string is a sequence of N characters (from an M-char alphabet). There are M^N strings.

A word is a sequence of M labelled sets (having N objects in total). There are M^N words.



Typical string

2 4 2 4 5 5 1 2 5

Typical word

{ 7 } { 1 8 3 } { } { 2 4 } { 5 6 9 }

Correspondence

- For each i in the kth set in the word set the ith char in the string to k.
- If the ith char in the string is k, put i into the kth set in the word.

Strings and Words

Familiar definition.
A string is a sequence of N characters (from an M-char alphabet).

Combinatorial definition.
A word is a sequence of M labeled sets (having N objects in total).

1-1 correspondence between words and strings

- Length of sequence in word: number of chars M in the alphabet.
- Number of objects in the set: length of string N.
- k th set in the sequence: indices where k appears in the string.

Q. What is the difference between strings and words?

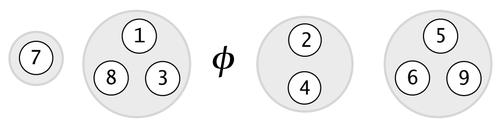
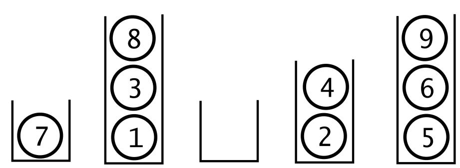
A. Only the point of view.

- With strings (last lecture) we study the sequence of characters.
- With words (this lecture) we study the sets of indices.

$N = 3$
$M = 2$

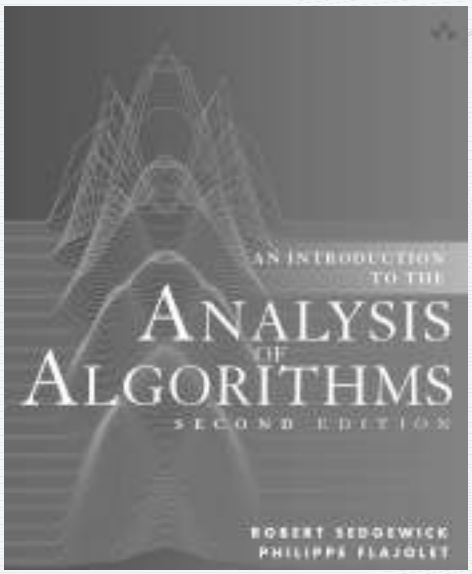
1 2 3	
0 0 0	{ 1 2 3 } { }
0 0 1	{ 1 2 } { 3 }
0 1 0	{ 1 3 } { 2 }
0 1 1	{ 1 } { 2 3 }
1 0 0	{ 2 3 } { 1 }
1 0 1	{ 2 } { 1 3 }
1 1 0	{ 3 } { 1 2 }
1 1 1	{ } { 1 2 3 }

Strings and Words (summary)

class	*type*	*GF type*	*typical*	*construction*	*GF*	*count*
STRING	unlabelled	OGF	2 4 2 4 5 5 1 2 5	$S = \text{SEQ}(Z_1 + \ldots + Z_M)$	$S(z) = \dfrac{1}{1 - Mz}$	M^N
WORD	labelled	EGF	 { 7 } { 1 8 3 } { } { 2 4 } { 5 6 9 } 2 4 2 4 5 5 1 2 5 	$W_M = \text{SEQ}_M(\text{SET}(Z))$	$W_M(z) = e^{Mz}$	M^N

ANALYTIC COMBINATORICS

PART ONE



http://aofa.cs.princeton.edu

9. Words and Mappings

- Words
- **Birthday problem**
- Coupon collector problem
- Hash tables
- Mappings

9b.Words.Birthday

Birthday problem

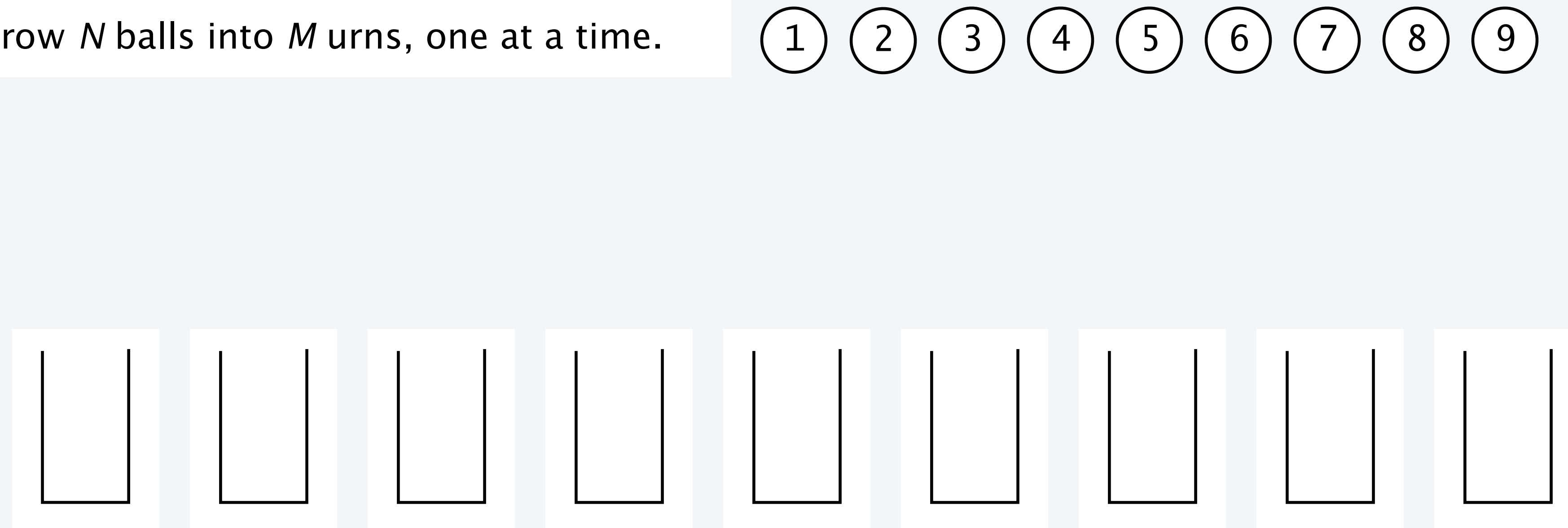
One at a time, ask each member of a group of people their birth date.

Q. How many people asked before finding two with the same birthday?

Quick answer: at most 365

Birthday problem

Throw N balls into M urns, one at a time.



Q. How long until some urn gets two balls (for $M = 365$) ?

Birthday sequences (words with no duplicates)

Def. A *birthday sequence* is a word where no set has more than one element.

a string with no duplicate letters

Q. How many birthday sequences?

Class	B_M, the class of birthday sequences
EGF	$B_M(z) = \sum_{w \in B_M} \frac{z^{\|w\|}}{\|w\|!} = \sum_{N \geq 0} B_{MN} \frac{z^N}{N!}$

Example

{ 3 } { } { 5 } { 1 } { } { } { 4 } { 2 } { }

4 8 1 7 3

Construction $B_M = SEQ_M(E + Z)$

OGF equation $B_M(z) = (1 + z)^M$

Counting sequence $N![z^N]B_M(z) = N!\binom{M}{N} = \frac{M!}{(M-N)!}$

$$= M(M-1)\ldots(M-N+1)$$

Birthday problem

Number of N-char M-words where no char is repeated

$$M(M-1)(M-2)\ldots(M-N+1) = \frac{M!}{(M-N)!}$$

Probability that no char is repeated in a random M-word of length N.

$$\frac{M!}{M^N(M-N)!}$$

Same as the probability that the first repeat position is > N.

Expected position of the first repeat

$$\sum_{0\le N\le M} \frac{M!}{M^N(M-N)!}$$

Laplace method to estimate Ramanujan Q-function (see Asymptotics lecture)

$$= 1 + Q(M) \sim \sqrt{\pi M/2}$$

Theorem. Expected position of the first repeated character in a random M-word is $\sim \sqrt{\pi M/2}$

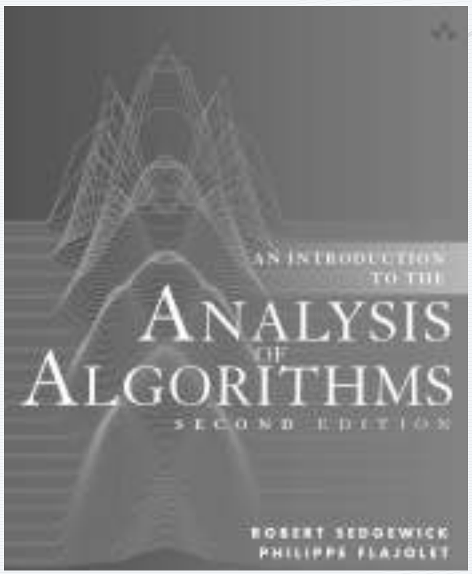
Birthday problem

Q. How many people asked before finding two with the same birthday?

A. About 24.

```
% bc
scale = 5
sqrt(3.14159*365/2)
23.94453
```

$$\sim \sqrt{\pi M/2}$$



http://aofa.cs.princeton.edu

9. Words and Mappings

- Words
- Birthday problem
- **Coupon collector problem**
- Hash tables
- Mappings

Coupon collector problem

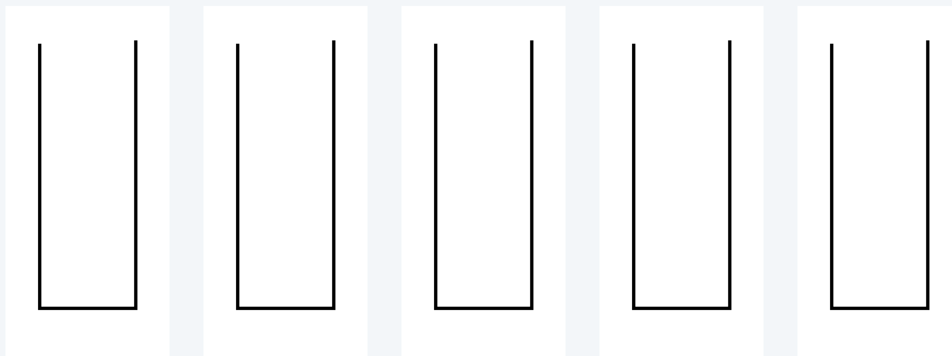
One at a time, ask each member of a group of people their birth date.

Q. How many people asked before finding *every day of the year*?

Quick answer: at *least* 365

Coupon collector problem

Throw N balls into M urns, one at a time.



Q. How long until each urn has at least one ball ?

Coupon collector problem

A collector buys coupons, each randomly chosen from *M* different types

Q. How many coupons collected before having every possible coupon?

Quick answer: at *least* 365

Coupon collector problem

Roll an *M*-sided die.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
9	12	19	3	5	20	10	17	16	20	13	8	2	13	9	2	15	17	3	9	11	7	18	2	10	1	20	12	10	8	14	5	5	9	4	5	6

first repeat

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Q. How many rolls until seeing all *M* values?

Coupon collector (classical analysis)

Probability that more than j rolls are needed to get the (k+1)st coupon

$$\left(\frac{k}{M}\right)^j$$

Expected number of rolls to get the (k+1)st coupon

$$\sum_{j\geq 0}\left(\frac{k}{M}\right)^j = \frac{1}{1-k/M} = \frac{M}{M-k}$$

Expected number of rolls to get all coupons

$$\sum_{0\leq k<M}\frac{M}{M-k} = MH_M \sim M\ln M$$

by linearity of expectation

Theorem. Expected number of coupons needed to complete a collection of size M is ~M ln M.

Motivation for studying in more detail:

- Discover variance and other properties of the distribution.
- Learn structure suitable for analyzing variants and extensions.

Coupon collector sequences (M-words with no empty sets)

Def. A *coupon collector sequence* is an M-word with no empty set.

a string that uses all the letters in the alphabet

Q. How many coupon collector sequences?

Example ($M = 26$)

```
the quick brown fox jumps over the lazy dog
```

Example ($M = 5$)

```
2 4 2 4 5 5 1 5 3
{ 7 } { 1 3 } { 9 } { 2 4 } { 5 6 8 }
```

Class	R_M, the class of coupon collector sequences
EGF	$R_M(z) = \sum_{w \in R_M} \frac{z^{\lvert w \rvert}}{\lvert w \rvert!} = \sum_{N \geq 0} R_{MN} \frac{z^N}{N!}$

Construction $\qquad R_M = SEQ_M(SET_{>0}(Z))$

EGF equation $\qquad R_M(z) = (e^z - 1)^M$

Counting sequence

$$N![z^N]R_M(z) = N![z^N] \sum_j \binom{M}{j} (-1)^j e^{(M-j)z}$$

$$= \sum_j \binom{M}{j} (-1)^j (M-j)^N \quad \sim M^N$$

Coupon collector sequences (EGF analysis, continued)

Probability that a random M-word of length N is a coupon collector sequence.

$$\frac{1}{M^N}\sum_j \binom{M}{j}(-1)^j(M-j)^N = \sum_j \binom{M}{j}(-1)^j\Big(1-\frac{j}{M}\Big)^N$$

Probability that collection in a random M-word completes in $>N$ chars.

$$1-\sum_j \binom{M}{j}(-1)^j\Big(1-\frac{j}{M}\Big)^N$$

Average number of chars to complete a collection in a random M-word.

$$\sum_{N\ge 0}\Big(1-\sum_j \binom{M}{j}(-1)^j\Big(1-\frac{j}{M}\Big)^N\Big)$$

$$= -M\sum_{j\ge 1}\binom{M}{j}\frac{(-1)^j}{j}$$

$$= MH_M$$

Knuth Exercise 1.2.7-13

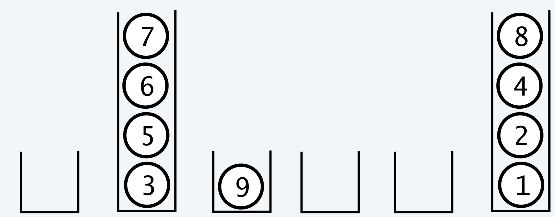
Coupon collector (OGF analysis)

Class	W_{Mk}, the class of M-words with k different letters and the last letter appearing only once
OGF	$W_{Mk}(z) = \sum_{w \in W_{Mk}} z^{\lvert w \rvert} = \sum_{N \geq 0} W_{MNk} z^N$
PGF	$W_{Mk}(z/M) = \sum_{N \geq 0} W_{MNk} \frac{z^N}{M^N}$
Mean wait time for k coupons	$w_{Mk} \equiv W'_{Mk}(z/M)\Big\rvert_{z=1} = \sum_{N \geq 0} N \frac{W_{MNk}}{M^N} z^N$

Example

6 6 2 6 2 2 2 6 3

{ } { 3 5 6 7 } { 9 } { } { } { 1 2 4 8 }



Coupon collector (OGF analysis, continued)

$W_{Mk} \equiv$ M-words with k different letters and the last letter appearing only once.

Construction $\quad W_{Mk} = (k-1)Z \times W_{Mk} + (M-k+1)Z \times W_{M(k-1)}$

OGF equation $\quad (1-(k-1)z)W_{Mk}(z) = (M-(k-1))zW_{M(k-1)}(z)$

OGF
$W_{Mk}(z)$

Evaluate at z/M

$$(M-(k-1)z)W_{Mk}(z/M) = (M-(k-1))zW_{M(k-1)}(z/M)$$

PGF
$W_{Mk}(z/M)$

Differentiate and evaluate at 1

$$(M-(k-1))w_{Mk} - (k-1) = (M-(k-1))(w_{M(k-1)}+1)$$

Wait time for k coupons
$$w_{Mk} \equiv W'_{Mk}(z/M)\Big|_{z=1}$$

Rearrange terms and telescope

$$w_{Mk} = w_{M(k-1)} + \frac{k-1}{M-(k-1)} + 1 = w_{M(k-1)} + \frac{M}{M-(k-1)}$$

$$= \sum_{0\le j<k} \frac{M}{M-j} = M(H_M - H_{M-k})$$

Wait time for full collection

$$w_{MM} = MH_M$$

Coupon collector problem

A collector buys coupons, each randomly chosen from M different types

Q. How many coupons collected before having every possible coupon?

A. $\sim M \ln M$.

Coupon collector problem

Roll an *M*-sided die.

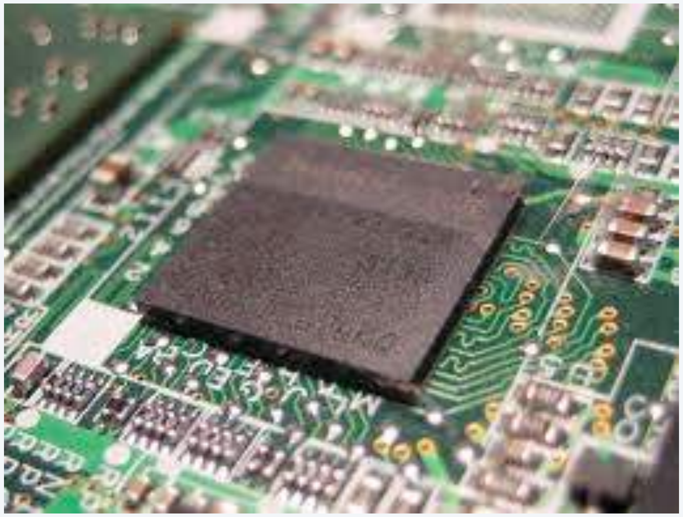
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
9	12	19	3	5	20	10	17	16	20	13	8	2	13	9	2	15	17	3	9	11	7	18	2	10	1	20	12	10	8	14	5	5	9	4	5	6

Q. How many rolls until seeing all *M* values?

A. $\sim M \ln M$. ← About 60 for a 20-sided die

Coupon collector problem: Sample application

A program randomly accesses an M-page memory.



Q. How many memory accesses before hitting every page, when $M = 2^{20}$?

A. About 14.5 million.

```
% bc -l
l(2)
.69314718055994530941
2^20*l(2^20)
14536349.96005650425534480384
```

Surjections

Def. An *M-surjection* is an *M*-word with no empty set.

Alt name for "coupon collector sequence"

Def. A *surjection* is a word that is an *M*-surjection for some *M*.

Q. How many surjections of length *N* ?

Class	R_M, the class of *M*-surjections

Construction

$$R_M = SEQ_M(SET_{>0}(Z))$$

EGF equation

$$R_M(z) = (e^z - 1)^M$$

Coefficients

$$R_{MN} \sim M^N$$

Class	R, the class of surjections

Construction

$$R = SEQ(SET_{>0}(Z))$$

EGF equation

$$R(z) = \frac{1}{1-(e^z-1)} = \frac{1}{2-e^z}$$

Coefficients

$$N![z^N]R(z) \sim \frac{N!}{2(\ln 2)^{N+1}}$$

Best handled with complex asymptotics (stay tuned for Part II)

1

$R_1 = 1$

1 1
1 2
2 1

$R_2 = 3$

1 1 1
1 1 2
1 2 1
1 2 2
1 2 3
1 3 2
2 1 1
2 1 2
2 1 3
2 2 1
2 3 1
3 1 2
3 2 1

$R_3 = 13$