CS 5602 Lecture 12 Probability Homeworks

George Markowsky
Department of Computer Science
Missouri University of Science & Technology

1

The Birthday Problem

- How many people need to be in a room before you have a 50% chance that at least two of them have the same birthday?
- The surprising answer is that once you have 23 people in a room, you have at least a 50% chance that two of them have the same birthday
- How to see this?
- What is the sample space? the probability space?

2

The Sample Space

- We can use S = {(b_1,b_2,…,b_{23}) | b_i ∈<u>365</u> }
- What about February 29^{th} birthdays?
- What probability distribution do you want to use?
- How about uniform?
- Is this reasonable?
- How large is S?
- |S| = 365^{23}

3

The Birthday Problem

- S = {(b_1,b_2,…,b_{23}) | b_i ∈<u>365</u> } and we are using the uniform distribution
- Let A = { s ∈ S | at least two elements of s are equal}
- It is tricky to compute Pr(A) because you need to look at pairs being equal, triples, etc.
- Consider B = { s ∈ S | no two components in S are equal }

4

The Birthday Problem

- Let A = { s ∈ S | at least two elements of s are equal}
- Let B = { s ∈ S | no two components in S are equal }
- Note that A = S – B, so Pr(A) = 1 – Pr(B)
- |B| = P(365,23)
- Pr(B) = P(365,23)/365^{23} = (365/365)*(364/365)*(363/365)*…(343/365) ≈ .4927
- So Pr(A) ≈ .5073

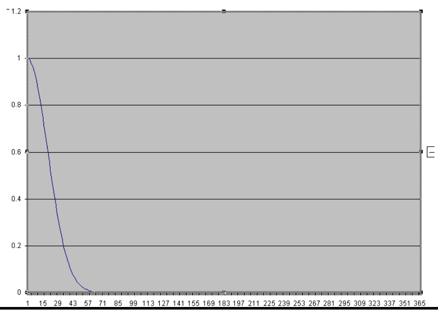
5

Some Calculations

N	Pr(B)	Pr(A)	23	0.492703	0.507297	343	3E-122	1
1	1	0	24	0.461656	0.538344	344	1.8E-123	1
2	0.99726	0.00274	25	0.4313	0.5687	345	1.1E-124	1
3	0.991796	0.008204	26	0.401759	0.598241	346	5.8E-126	1
4	0.983644	0.016356	27	0.373141	0.626859	347	3E-127	1
5	0.972864	0.027136	28	0.345539	0.654461	348	1.5E-128	1
6	0.959538	0.040462	29	0.319031	0.680969	349	6.9E-130	1
7	0.943764	0.056236	30	0.293684	0.706316	350	3E-131	1
8	0.925665	0.074335	31	0.269545	0.730455	351	1.2E-132	1
9	0.905376	0.094624	32	0.246652	0.753348	352	4.8E-134	1
10	0.883052	0.116948	33	0.225028	0.774972	353	1.7E-135	1
11	0.858859	0.141141	34	0.204683	0.795317	354	5.6E-137	1
12	0.832975	0.167025	35	0.185617	0.814383	355	1.7E-138	1
13	0.80559	0.19441	36	0.167818	0.832182	356	4.6E-140	1
14	0.776897	0.223103	37	0.151266	0.848734	357	1.1E-141	1
15	0.747099	0.252901	38	0.135932	0.864068	358	2.5E-143	1
16	0.716396	0.283604	39	0.12178	0.87822	359	4.8E-145	1
17	0.684992	0.315008	40	0.108768	0.891232	360	7.9E-147	1
18	0.653089	0.346911	41	0.096848	0.903152	361	1.1E-148	1
19	0.620881	0.379119	42	0.08597	0.91403	362	1.2E-150	1
20	0.588562	0.411438	43	0.076077	0.923923	363	9.7E-153	1
21	0.556312	0.443688	44	0.067115	0.932885	364	5.3E-155	1
22	0.524305	0.475695	45	0.059024	0.940976	365	1.5E-157	1
23	0.492703	0.507297	46	0.051747	0.948253	366	0	1

6

A Chart



7

A General Result

- If you have a set of n birthdays and k people in a room all of whose birthdays are from the set, the probability (assuming a uniform distribution) that at least two people have the same birthday is:

$$\frac{P(n,k)}{n^k} = \frac{n}{n} \times \frac{(n-1)}{n} \times \frac{(n-2)}{n} \times \ldots \times \frac{(n-(k-1))}{n}$$

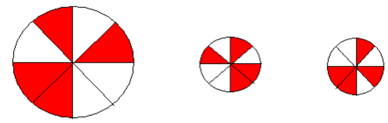
8

The Two Disks Problem

- Suppose you have two disks each with 128 pie-shaped sectors, 64 of which are painted red and 64 of which are painted white
- The two patterns can be completely arbitrary
- Can we place one over the other such that at least 64 sectors on each disk have the same color?

9

The Two Disks Problem



Can we place the smaller over the larger so that that are at least 4 matches?

10

How Can You Solve This Problem?

- Is this a probability problem?
- Why would you expect to solve it using probabilistic methods?
- Because probabilistic methods are really combinatorial in nature
- You need to start with the correct sample space

11

Sample Space

- What is the sample space?
- S = {1, 2, …, 128 } = <u>128</u>
- Why? What do the numbers mean?
- The number j means that slice 1 of the small disk is placed over slice j of the large disk
- Have R.V. Matches:S → **R** given by Matches(j) = number of matches when small disk's slice 1 is over large disk's slice j

12

Now What?

- Want to compute E(Matches), but how?
- Use **indicator variables**
- Let M_i:S → **R** be given by
- $M_i(j) = 1$ if slice i of the small disk has a match when the small disk is in position j
- $M_i(j) = 0$ otherwise
- Note that $E(\text{Matches}) = \sum_{i=1..128} E(M_i)$

13

Well, Now What?

- It is easy to see that $E(M_i) = ½$ because half the slices of the large disk are the same color as slice i of the small disk and half are the other color
- $E(M_i) = 1*(1/2) + 0*(1/2) = ½$
- Thus, E(Matches) = 64
- So what?
- By the Average Value Theorem, there exists a $j \in \underline{128}$ s.t. $\text{Matches}(j) \geq 64$

14

HW 01

- As I read through the HWs I found that there was a big disconnect between what I was getting and what I was expecting so I thought we should treat HW 01 as a learning exercise
- I will now review some of the issues that need to be addressed
- Will review orally and give you a more detailed document on Canvas by tomorrow

15

	A	B	C
1	C(n,k)	n	
2	k	23	24
3	0	1	1
4	1	23	24
5	2	253	276
6	3	1,771	2,024
7	4	8,855	10,626
8	5	33,649	42,504
9	6	100,947	134,596
10	7	245,157	346,104
11	8	490,314	735,471
12	9	817,190	1,307,504
13	10	1,144,066	1,961,256
14	11	1,352,078	2,496,144
15	12	1,352,078	2,704,156
16	13	1,144,066	2,496,144
17	14	817,190	1,961,256
18	15	490,314	1,307,504
19	16	245,157	735,471
20	17	100,947	346,104
21	18	33,649	134,596
22	19	8,855	42,504
23	20	1,771	10,626
24	21	253	2,024
25	22	23	276
26	23	1	24
27	24		1
28	Sum	8,388,608	16,777,216
29		8,388,608	16,777,216

16