

## Lecture 10: October 5, 2009

CS 330 Discrete Structures  
Fall Semester, 2009

### 1 Probability

When the weather-woman says “there is a 30% chance of rain”, what does she mean? Does she mean that:

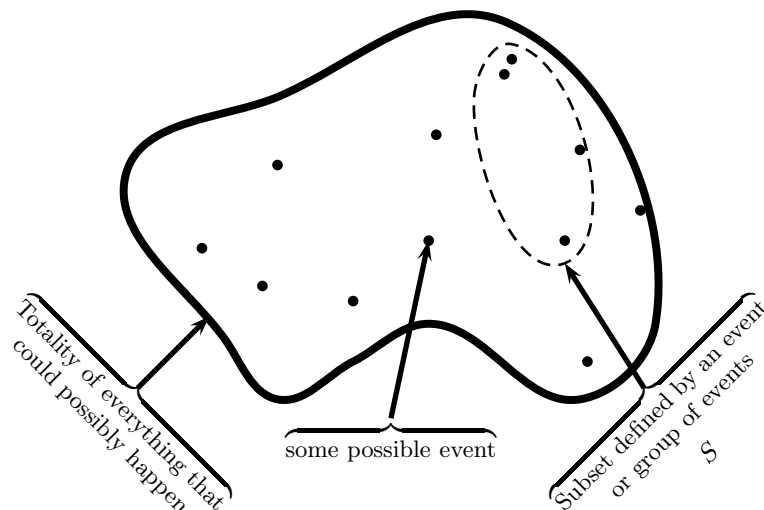
- rain will fall on 30% of the viewing area?
- in the last 100 years, it rained 30 times on this date?
- under present conditions, recorded history for comparable conditions shows it rained 30% of the time?

Similarly, what does it mean for an algorithm to be correct 99% of the time? Can we trust it to guard our nuclear warheads?

In order to answer these questions, we first need to know a little about **probability**.

### 2 Basics of Probability

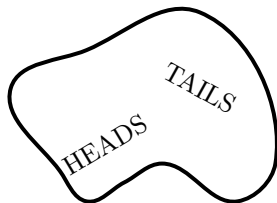
Consider the following amoeba-like graph:



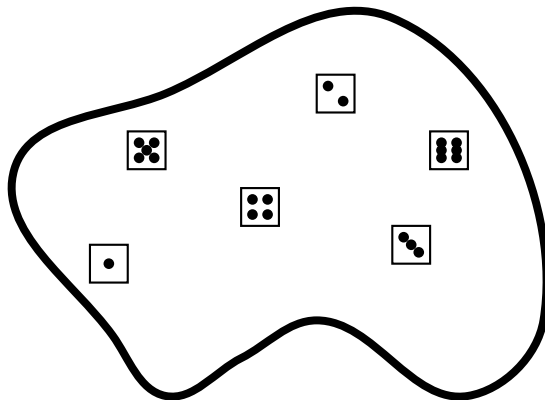
We wish to measure the likelihood of event  $S$  occurring. We call this “Probability of  $S$ ” and we write it as

$$\Pr(S) \text{ or } P[S].$$

Consider the following experiment: We will flip a “fair” coin. We have the following “universe” of outcomes:



Our intuition tells us that  $\Pr(TAILS)$  is 50% or  $\frac{1}{2}$  because TAILS happens about half the time. Consider this experiment: We will roll a “fair” die. We have the following “universe” of outcomes:



Our intuition tells us that  $\Pr(\text{Rolling a one})$  is  $\frac{1}{6}$ , because rolling a “one” occurs about one sixth of the time. From these we can see a simple definition of Probability as:

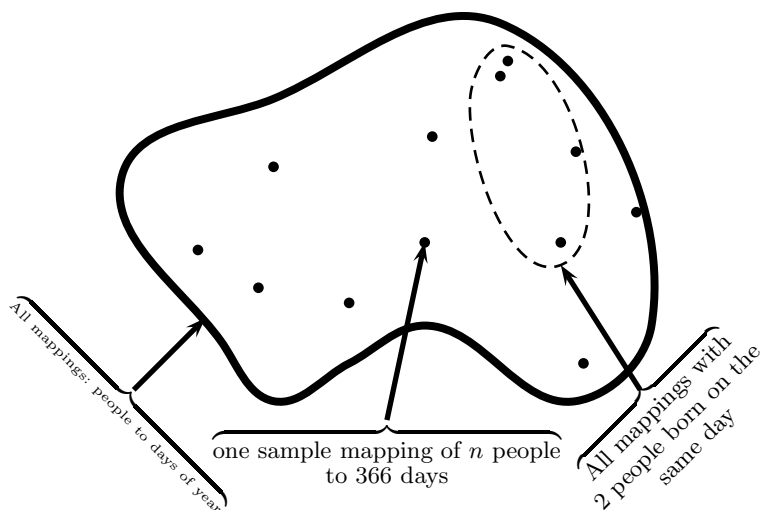
$$\Pr(S) = \frac{\text{number of events in } S}{\text{number of events altogether}}$$

This definition will be ample for our purposes.

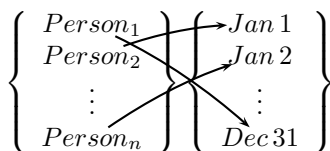
### 3 The Birthday Problem

An illustration of the power of probability is the Birthday problem: If we have  $n$  people in a room, what is the probability that two people celebrate their birthday on exactly the same day?

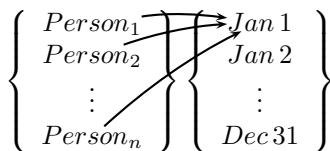
After a careful examination, you will notice that the set we have defined is:



A mapping simply relates people to a day of the year corresponding to their birthday. For instance



and



are mappings.

There are  $366^n$  ways to map  $n$  people to birthdays. There are  $\frac{366!}{(366-n)!}$  ways to pick the birthdays so that no two people share the same birthday. Using our definition of Probability:

$$\Pr(\text{No two birthdays on the same day}) = \frac{366!}{366^n} = 1 \left(1 - \frac{1}{366}\right) \left(1 - \frac{2}{366}\right) \cdots \left(1 - \frac{n-1}{366}\right)$$

Here we can see how this probability decreases when  $n$  grows:

$n$	1	2	3	4	5	6	7	8	9	10
$\frac{366!}{(366-n)!}$	1	0.9973	0.9918	0.9837	0.9729	0.9596	0.9439	0.9259	0.9056	0.8834
$\frac{366!}{366^n}$	11	12	13	14	15	16	17	18	19	20
$\frac{366!}{(366-n)!}$	0.8592	0.8334	0.8061	0.7774	0.7477	0.7171	0.6857	0.6539	0.6217	0.5894
$n$	21	22	23	24	25	26	27	28	29	30
$\frac{366!}{(366-n)!}$	0.5572	0.5252	0.4937	0.4627	0.4323	0.4028	0.3742	0.3466	0.3201	0.2947

It turns out that when there are 23 people in the room the  $\Pr(\text{No two birthdays on the same day}) \approx 0.5$ . This result has direct relation to computer science and hash tables, because it says that a hash table can be very empty (23 people compared to 366 days of the year, in our analogy), but it is still very likely that there will be a hashing conflict (that you will have two people with the same birthday, in our analogy).

Food for thought: How many people are needed to have the probability of the same birth *month* be at least 0.5?

## 4 An Application—GUIDs

According to Wikipedia,

A globally unique identifier or GUID is a special type of identifier used in software applications to provide a reference number which is unique in any context (hence, “globally”), for example, in defining the internal reference for a type of access point in a software application, or for creating unique keys in a database. While each generated GUID is not guaranteed to be unique, the total number of unique keys ( $2^{128} \approx 3.4 \times 10^{38}$ ) is so large that the probability of the same number being generated twice is extremely small.

But how small? Specifically, suppose that a million GUIDs are generated every hour of every day for 100 years; what is the probability of a duplicate GUID?

The total number of GUIDs will be less than  $10^6 \times 24 \times 366 \times 100 < 10^{12}$ . Reasoning as in the birthday problem, the probability of a duplication is

$$\frac{2^{128!}}{(2^{128})^{10^6} (2^{128} - 10^6)!} < 10^{-38}$$

by Stirling’s approximation.

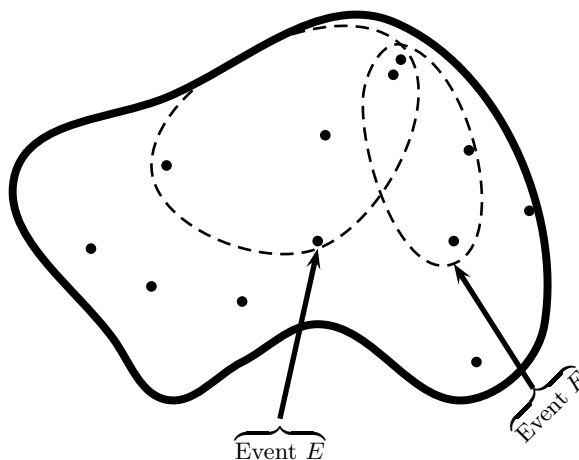
## 5 Conditional Probability

Consider now the following experiment: You are given three identical urns. Urn 1 contains 2 black balls. Urn 2 contains 2 gray balls. Urn 3 contains 1 black and 1 gray ball. You pick urn at random and take out a gray ball. How likely was it that you picked Urn 1? Urn 2? Urn 3?

For Urn 1, the result is easy since there are no gray balls. There are 2 gray balls in urn 2, so it has a  $\frac{2}{3}$  probability while urn 3 has a  $\frac{1}{3}$  probability.

Now lets replace the gray ball back into the urn and pick a ball again from the *same* urn. Again we get a gray ball; what is the probability that we we pick a gray ball? To answer this question we need to study **Conditional Probability**.

Conditional probability is saying “What is the likelihood of Event  $E$  happening if we already know that Event  $F$  happened” and is written as:  $\Pr(E|F)$ . The situation is as follows:

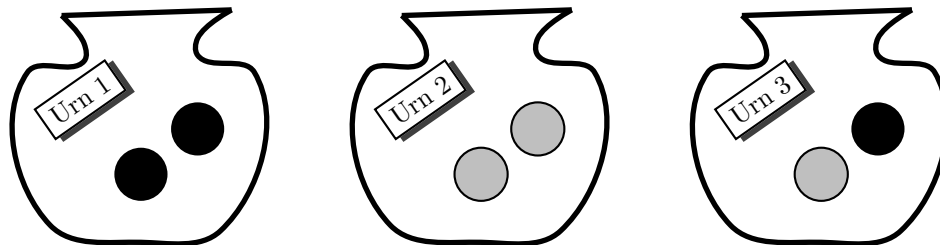


If we know that Event  $F$  happened, we know that the only part of Event  $E$  that can happen is in the intersection of  $E$  and  $F$ . From this we have the result:

$$\begin{aligned}\Pr(E|F) &= \frac{\Pr(E \wedge F)}{\Pr(F)} \\ \Pr(F|E) &= \frac{\Pr(F \wedge E)}{\Pr(E)} \\ \Rightarrow \Pr(E \wedge F) &= \Pr(E)\Pr(F|E) \\ \Rightarrow \Pr(E|F) &= \frac{\Pr(E)\Pr(F|E)}{\Pr(F)}\end{aligned}\tag{1}$$

The identity in Equation 1 is quite important, and is known as **Bayes' Theorem**. Note also that there is nothing special about the letters “E” and “F”; we can interchange what events “E” and “F” represent and still have the same identities.

Let's go back to the experiment with urns and balls. We have three urns, the Urn 1 contains two black balls, Urn 2 contains two gray balls and Urn 3 contains a black and a gray ball.



Having selected an urn at random, we pull out a gray ball. What is the probability, for each urn, that we had selected that urn. In other words, if someone came to you and told you that they had picked a gray ball, and asked you to guess to which urn the ball belonged; what would be the probability of you being right if you guessed a particular urn? In order to answer this question, let us first examine the probability of pulling out a gray ball from each of the three urns.

The first urn has  $\Pr(\text{picked gray}) = \frac{0}{2} = 0$  since there are no gray balls to pick.



The second urn has  $\Pr(\text{picked gray}) = \frac{2}{2} = 1$  since there are only gray balls to pick.



The third urn has  $\Pr(\text{picked gray}) = \frac{1}{2}$  since there is one gray and one black ball to pick.



We can now apply Conditional Probability to the problem: If we choose an urn at random and get a gray ball, what was the probability of having picked a particular urn? We can write three separate equations (and apply the definition of conditional probability for the first and Bayes' Theorem for the next two):

$$\begin{aligned}\Pr(\text{Urn 1}|\text{Gray}) &= \frac{\Pr(\text{Urn 1}) \times \Pr(\text{Gray}|\text{Urn 1})}{\Pr(\text{Gray})} = \frac{0}{\frac{1}{2}} = 0 \\ \Pr(\text{Urn 2}|\text{Gray}) &= \frac{\Pr(\text{Urn 2}) \times \Pr(\text{Gray}|\text{Urn 2})}{\Pr(\text{Gray})} = \frac{\frac{1}{3} \times 1}{\frac{1}{2}} = \frac{2}{3} \\ \Pr(\text{Urn 3}|\text{Gray}) &= \frac{\Pr(\text{Urn 3}) \times \Pr(\text{Gray}|\text{Urn 3})}{\Pr(\text{Gray})} = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{2}} = \frac{1}{3}\end{aligned}$$

Let us consider a related problem: If we choose an urn and pick one ball, return it to the urn and pick another ball from the same urn, and get two gray balls, what was the probability of having picked a particular urn? Applying Bayes' Theorem:

$$\Pr(\text{Urn 1}|2 \text{ Gray Balls}) = \frac{\Pr(\text{Urn 1}) \times \Pr(2 \text{ Gray Balls}|\text{Urn 1})}{\Pr(2 \text{ Gray Balls})} = \frac{\frac{1}{3} \times 0}{\frac{1}{3} \times 0 + \frac{1}{3} \times 1 + \frac{1}{3} \times \frac{1}{4}} = 0 \quad (2)$$

$$\Pr(\text{Urn 2}|2 \text{ Gray Balls}) = \frac{\Pr(\text{Urn 2}) \times \Pr(2 \text{ Gray Balls}|\text{Urn 2})}{\Pr(2 \text{ Gray Balls})} = \frac{\frac{1}{3} \times 1}{\frac{1}{3} \times 0 + \frac{1}{3} \times 1 + \frac{1}{3} \times \frac{1}{4}} = \frac{4}{5} \quad (3)$$

$$\Pr(\text{Urn 3}|2 \text{ Gray Balls}) = \frac{\Pr(\text{Urn 3}) \times \Pr(2 \text{ Gray Balls}|\text{Urn 3})}{\Pr(2 \text{ Gray Balls})} = \frac{\frac{1}{3} \times \frac{1}{4}}{\frac{1}{3} \times 0 + \frac{1}{3} \times 1 + \frac{1}{3} \times \frac{1}{4}} = \frac{1}{5} \quad (4)$$