

Math 55 - Spring 2004- Lecture notes #17 - March 18 (Thursday)

Reader (Lenstra): sections 1, 2, 3, 4

Recommended reader: CS70 notes, starting with lecture 15
inst.eecs.berkeley.edu/~cs70

Goals for today: Introduce discrete probability theory

Long term goal: we would like to make sense of statements like:

- 1) "The chance of getting a "flush" (all cards the same suit) in 5-card poker is about 2 in 1000."
- 2) "If you flip a fair coin 50 times, and each time it comes out heads, then the chance you get a head the 51st time is still 50%."
- 3) "If quicksort picks a random "pivot item" at each step, then it will sort n numbers, in $O(n \log n)$ time with high probability".
- 4) "With this algorithm for balancing the workload among computers the probability (or chance) that a user has to wait more than 1 minute is 2%."
- 5) "There is a 60% chance of the Big One (large earthquake) hitting Northern California in the next 30 years."
- 6) In an election, if B's reported vote total is 2,912,790, G's reported vote total is 2,912,253, and the chance of a vote having been mistakenly counted wrong is 1 out of 100,000, what is the probability that G actually won the election?

To understand these examples and others, we must specify a

- 1) an experiment, whose outcome is "random"
- 2) the set of possible outcomes (the "sample space")
- 3) probability of each possible outcome

EX 1): The experiment is shuffling a deck of cards and dealing 5 of them
The set of outcomes is all possible 5 card subsets of 52 cards

ASK&WAIT: How many ways are there?

The probability of each outcome is equal (assuming we shuffle well)

ASK&WAIT: What is the probability of any particular shuffle?

EX 2): The experiment is flipping a fair coin 51 times

The set of outcomes is all possible sequences is of 51 H's and T's

ASK&WAIT: How many outcomes are there?

The probability of each outcome is equal (assuming the coin is fair)

ASK&WAIT: What is the probability of any particular sequence?

EX 3 through 5 are trickier, especially 5) (ask your local geologist!)
We'll return to EX 6 later.

DEF: A sample space is a finite (or countable) set S
together with a function (called probability function, or just probability)
 $P: S \rightarrow [0,1]$
such that
 $\sum_{x \in S} P(x) = 1.$

S is the set of all possible outcomes of the experiment,
with $P(x)$ equal to the probability that the outcome is x .

The most notable case is when S is finite and $P(x)$ has the same value
for all x in S , i.e. when all events are equally likely. In this case,
we say S has a "uniform probability distribution."

What is $P(x)$ equal to in this case? We have

$$1 = \sum_{x \in S} P(x) = P(x) |S|$$

for all x in S , so, in the uniform distribution case, we have

$$P(x) = 1/|S|$$

for all x in S .

EX 1 and 2 above: $|S| = C(52,5)$ or $|S| = 2^{51}$

DEF: An event E is a subset of the sample space S , and the probability
 $P(E)$ of an event E is given by

$$P(E) = \sum_{x \in E} P(x).$$

Note: the empty set has probability 0. The whole sample space S has
probability 1.

In the case of a uniform distribution, we have

$$\begin{aligned} P(E) &= \sum_{x \in E} P(x) \\ &= \sum_{x \in E} 1/|S| \\ &= |E|/|S|. \end{aligned}$$

EX: One toss of a fair coin.

$$S = \{H,T\} \quad P(H) = 1/2, \quad P(T) = 1/2.$$

EX: 3 tosses of a fair coin

$$\begin{aligned} S &= \{HHH, HHT, \dots, TTT\}, \quad P(\text{any particular outcome}) = 1/8 \\ E &= \{2 \text{ heads and a tail}\} \end{aligned}$$

ASK&WAIT: What is $P(E)$?

EX: 3 tosses of a biased coin. This means $P(H)$ not equal to $P(T)$, i.e. not $1/2$
Suppose $P(H) = 1/3$

ASK&WAIT: What is $P(T)$?

ASK&WAIT: What is $P(HHH)$? $P(HTH)$?

ASK&WAIT: What is $P(E)$, E as above?

EX: A roll of a die.

$S = \{1,2,3,4,5,6\}$.

$P(x) = 1/6$ for all x in S .

Let $E =$ "the roll of the die is odd" $= \{1,3,5\}$

ASK&WAIT: What is $P(E)$?

EX: A roll of two dice, one red and one blue.

$S = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$,

i.e. all pairs $S = \{(i,j), 1 \leq i \leq 6, 1 \leq j \leq 6\}$

$P(x) = 1/36$ for all x in S , since $|S| = 36$

ASK&WAIT: What is $P(E)$, $E =$ "the first die is a 6"?

ASK&WAIT: What is $P(E)$, $E =$ "at least one die is a 6"?

ASK&WAIT: What is $P(E)$, $E =$ "the dice sum to 7"

ASK&WAIT: What is $P(E)$, $E =$ "the dice sum to 10"

EX: A roll of two indistinguishable dice (eg both blue)

indistinguishable means that, say $(1,6)$ and $(6,1)$ no longer different

$S = \{(1,1), (1,2), \dots, (1,6), (2,2), (2,3), \dots, (2,6), (3,3), \dots, (3,6), (4,4), \dots, (6,6)\}$

$S = \{(i,j), 1 \leq i \leq j \leq 6\}$

ASK&WAIT: What is $P(i,i)$? What is $P(i,j)$ for $i \neq j$?

ASK&WAIT: What is $P(E)$, $E =$ "dice sum to 10"

EX: A single poker hand, gotten by shuffling a deck of 52 card and taking 5.

S has $C(52,5) = 2,598,960$ elements, which I will not list here.

$P(x) = 1/2,598,960$ for all x in S .

ASK&WAIT: What is $P(E)$, $E =$ "royal flush" $=$ "A,K,Q,J,10" of the same suit?

ASK&WAIT: What is $P(E)$, $E = \text{"the hand has four of a kind"}$

ASK&WAIT: What is $P(E)$, $E = \text{"the hand contains a full house"}$

ASK&WAIT: What is $P(E)$, $E = \text{"the hand contain a flush"}$

EX: Balls and bins. Suppose you take 20 distinguishable balls (tasks) and throw them into 10 distinguishable bins (computers) so that each ball has an equal chance of landing in each bin. (This is a common way of distributing work among multiple computers, eg web requests coming into a company)

ASK&WAIT: What is S ? $|S|$? $P(\text{any particular outcome})$?

What is $P(E)$, $E = \{\text{each bin has at most 4 balls}\}$
 $= \{\text{each computer has at most 4 requests}\}$?

We will eventually answer this...

ASK&WAIT: What is $P(\{\text{each bin has exactly 2 balls}\})$?
i.e. that the balls are perfectly evenly distributed?

EX: Strategy for a TV game show, where you have to pick one of 3 doors, and you win whatever is behind the door.

First a prize is placed behind one of three doors, each with equal probability. You are then allowed first to choose one door.

Then, one of the other two doors is revealed (behind which, of course, no prize appears).

Finally, you are allowed the option of switching to another door.

You will win whatever is behind the door you select.

Should you switch to the third door, stay where you are, or does it not matter?

Answer: you should switch, always! Why?

Let's figure out the sample space describing the situation up to the moment you have to choose whether to switch:

$S = \{(i,j,k) \text{ where}$
 $i=1,2 \text{ or } 3 \text{ indicates the door where the prize is,}$
 $j=1,2 \text{ or } 3 \text{ indicates the door you originally choose, and}$
 $k=1,2 \text{ or } 3 \text{ indicates the door opened on the show}\}$

So i and j and take any values from $\{1,2,3\}$ with equal probability.

But k is restricted: if $i \neq j$, then k must be chosen not

to equal either i or j , so its value is determined. But if $i=j$, then k can equal either of the other 2 values with equal probability.

So here is the sample space with probabilities shown below each outcome in parentheses.

For example $i=2, j=1$ means $k=3$, and has probability $(1/3)*(1/3)=1/9$.

For example $i=2, j=2$ means $k=1$ or 3 , and $i=2, j=2, k=3$ has probability $(1/3)*(1/3)*(1/2)=1/18$

	i=1	i=2	i=3
j=1	k=2 or k=3 (1/18) (1/18)	k=3 (1/9)	k=2 (1/9)
j=2	k=3 (1/9)	k=1 or k=3 (1/18) (1/18)	k=1 (1/9)
j=3	k=2 (1/9)	k=1 (1/9)	k=1 or k=2 (1/18) (1/18)

Now suppose that your strategy is not to switch doors;

what is the probability of the event $E = \{\text{you win!}\}$?

ASK&WAIT? Can you indicate which parts of the sample space is in E ? What is $P(E)$?

Now suppose that your strategy is to switch doors;

what is the probability of the event $E = \{\text{you win!}\}$?

ASK&WAIT? Can you indicate which parts of the sample space is in E ? What is $P(E)$?

ASK&WAIT: What is the best strategy, switch or not?

Now we go on to techniques that make it easier to compute the probabilities of certain events.

THEOREM: Let E be an event in a sample space S . The probability of the event $S-E$, the complement of E in S , is given by $1-P(E)$.

PROOF: $1 = \sum_{\{x \text{ in } S\}} P(x)$
 $= \sum_{\{x \text{ in } E\}} P(x) + \sum_{\{x \text{ in } S-E\}} P(x)$
 $= P(E) + P(S-E).$

So $P(S-E) = 1 - P(E)$.

THEOREM: Let E and F be events in a sample space S . Then $P(E \text{ union } F) = P(E) + P(F) - P(E \text{ intersect } F)$.

PROOF: Similar to the proof of inclusion-exclusion, which may be stated as

$$\sum_{x \text{ in } E \text{ union } F} 1 = \sum_{x \text{ in } E} 1 + \sum_{x \text{ in } F} 1 - \sum_{x \text{ in } E \text{ intersect } F} 1.$$

Just replace 1 in the sums above by $P(x)$.

EX: What is the probability that a randomly chosen integer between 1 and 100 is divisible by 5 or 7?

THEOREM: Let E_1, E_2, \dots, E_n be pairwise disjoint event in a sample space S .

$$\text{Then } P(E_1 \cup E_2 \cup \dots \cup E_n) = P(E_1) + P(E_2) + \dots + P(E_n)$$

ASK&WAIT: What is the proof?