

Probability

Independence

Last time:

Two events are indpt if knowing that one did or did not happen tells you nothing about whether the other will or will not. It doesn't change the probability.

Example: coinflips, dice rolls, sex of child in family.

Useful for Multiplication Rule:

When two events, A and B, are independent,  $P(A \text{ and } B) = P(A) P(B)$ .

Example: Flip a coin two times. What's probability both are heads?

A = first flip is heads

B = second flip is heads

$$P(A \text{ and } B) = P(A) P(B) = .5 * .5 = .25$$

Or:

Sample Space = HH, HT, TH, TT

A = {HH, HT}

B = {HH, TH}

# of outcomes in both A and B = 1.

# of events in S = 4

$$P(A \text{ and } B) = 1/4$$

The first method is often much easier. For example: Flip a coin 10 times. What's probability all heads? Writing all outcomes down is very time consuming. But easy enough this way.

Example: What's probability of rolling two "aces"?

$$1/6 * 1/6 = 1/36$$

NOTE: 3 ways of computing  $P(A \text{ and } B)$ :

1)  $\#(\text{in both A and B}) / \#(S)$

2) If A and B are mutually exclusive,  $P(A \text{ and } B) = 0$

3) If A and B are independent,  $P(A \text{ and } B) = P(A) * P(B)$

Example:

China had a one-child rule, that was ammended in some cities and rural areas to a "one boy" rule. You may have as many children as it takes, as long as only one boy. Assume that under these conditions, families will have babies until their first boy. What's the

probability a family will have 1 child? 2? 3? (Assume sex of successive children is indpt, and that the probability of a boy is equal to a girl.)

Answer: The probability that the first child is a boy is  $1/2$ . So  $P(\text{one child}) = 1/2$ .  
What about the probability that they have 2? That means the first child is a girl and the second is a boy. It's just like flipping a coin twice and asking  $P(\text{TH})$ ? Since we assume these are indtp.  $P(\text{first is girl AND second is boy}) = P(\text{first is girl}) P(\text{second is boy}) = 1/2 * 1/2 = 1/4$   
3? Means GGB  $P(\text{GGB}) = P(\text{G}) P(\text{G}) P(\text{B}) = 1/8$ .

Do you see a pattern?  
 $P(\text{number of children is } x) = (1/2)^x$

Could you simulate by tossing a coin? (Say  $P(\text{get two children})$ )  
One simulation: toss until first head; count tosses  
Event of interest: takes two tosses  
Repetitions: 1000? 100?  
Experimental probability: number of times it took 2 tosses/1000.

Note that the sample space consists of things like:

GB, B, GGB, GGGGB, etc.

But we're only concerned with counting the number of kids in the family. When you take a sample space and reduce your interest to a number, you are creating a random variable:

## Random Variables:

Another way of looking at outcomes and events is the random variable. A random variable is a number associated with a random event. Think of it like this:

It is the outcome of an experiment:

It is variable, meaning the outcome is not always the same.

It is random, because we can't predict with certainty what the value of the variable will be.

Usual notation: X, Y. For example:

X = number of children born in a family.

Y = number of heads in 10 tosses of a coin.

Z = number of inches of rain in LA next year

X = a randomly selected person's height

Random Variables come in two flavors: continuous and discrete

Discrete: the values are "discrete" numbers, like the whole numbers: 1,2,3, or -2, -1, 0, 1, ....

the important features is that there is "space" between values.

Example: X = number of kittens in a litter.

There can be 0, 1, 2, 3, .....  
kittens in the litter. But there is space between these because you can't have .5 kittens.

Example:

A lottery has small prizes of \$1, \$1.5, \$2, and \$100.

$X$  = amount you win

is discrete because there are spaces between the values. For example, the number 1.75 is between 1.5 and 2, and yet is not a value.

Continuous: There are no spaces between values. The possible values occur over an interval including all numbers.

Example:  $X$  = time spent waiting in line at the Food Court. Yes, possible values are 1 minute and 2 minutes, but 1.5 is another possible value between 1 and 2. And 1.25 is a possible value between 1 and 1.5, and 1.125 is between 1 and 1.25, and 1.005, etc.

## Probability Distribution Functions (PDFs)

Once you've got a random variable, it's usually quite useful to be able to write down the probabilities associated with ALL of the outcomes. For example

One-Boy Rule:  $X$  = number of children in a family under the one-boy policy.  $X$  is discrete, and possible values are 1,2,3,4,.....

We found that  $P(X = x) = (1/2)^x$

Or we could put (some) of this information into a table:

$x$	$P(X = x)$
1	1/2
2	1/4
3	1/8
4	1/16
etc	etc.

Or we could make a graph:

Another example:

Toss a coin 3 times. Let  $X$  = number of heads. What's the probability distribution function (PDF) for  $X$ ?

Values for  $X$  are 0,1,2,3

Write down all outcomes:

HHH

HTT

THT

TTH

HTH

HHT  
THH  
TTT

Each is equally likely. So

$$P(X = 0) = P(\text{no heads}) = 1/8$$

$$P(X = 1) = P(\text{HTT OR THT OR TTH}) = 1/8 + 1/8 + 1/8 = 3/8$$

$$P(X = 2) = 3/8$$

$$P(X = 3) = 1/8$$

x	p
0	1/8
1	3/8
2	3/8
3	1/8

Q:  $P(X =$

Note:

$$\text{Sum } P(X) = 1$$

Notation:  $P(X = x) = p(x)$  So we might write  $P(X = 3)$  or  $p(3)$ .

Simulation:

Let's return to simulating the one-boy rule. I did it with coin flips, and here's what I found: (H is a boy, T a girl)

50 trials:

TH

TTH

H

H

H

TH

H

H

TH

TH

H

H

TTH

H

H

TTTH

H

TTH

TTH

TTTTTH

H  
H  
TTH etc.

Putting this together, I got:

x	frequency	P-hat	p
1	28	.56	.5
2	9	.18	.25
3	9	.18	.125
4	0	0	0.0625
5	0	0	0.03125
6	1	0.02	0.015625

You can make a plot of the empirical probabilities and also of the theoretical probabilities to help compare.

The plot of experimental probabilities is very much like a histogram. We can pretend that we sampled a population and this is what we got. The theoretical probabilities is a plot of the pdf. We can think of it as a picture of the population from which we sampled.

In toher words, another way of doing this simulationN:

Put the tickets 1,2,3,4,...  
in a box in the proportions shown in the theoretical probability. Then pull out 50 (with replacement). You'll get a result very much like (but not exactly the same as), our experiment.