Probability I

Sumanta Basu

Reading

Probability Concepts

- Reading:Textbook Section 2.1
- Recommended exercise: 2.1, 2.5, 2.7, 2.11

`for`, `if` statements in R, and Monte Carlo Simulation

- Practice Lab (Tuesday: Problems 1, 2; Thursday: Problems 3, 4; in lab 4: brief review + Problem 5)
- Openintro Lab on Probability (under Blackboard ‘Labs/Lab4’)
- Recommended: Chapter 2 Supplement (under Blackboard ‘Reading’) Section 2.2
- Simple Random Sampling demo (under Blackboard ‘Labs/Lab2’)

Probability

A bit of context

So far:

- Sampling
- Graphical methods
- Descriptive statistics

Helpful for summarizing data collected

How can we quantify uncertainty in the conclusions we draw from data?

Quantifying uncertainty

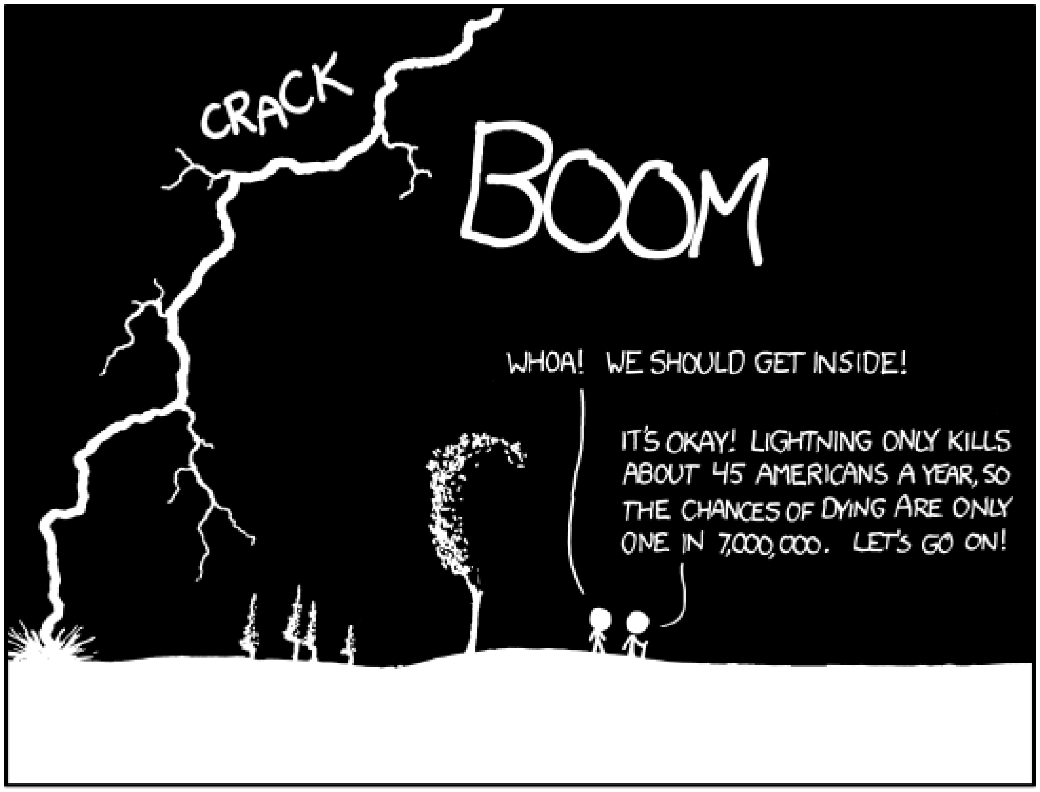
Statements about population based on a sample are uncertain.

Next up: Tools (and language) for quantifying our level of uncertainty.

Will develop **models** that are **probabilistic** in nature - describe “long run behavior” imagining if we repeated the sampling process over and over

We must know more about **probability**!

Relying too much on probability?



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

or …

> “When a statistician passes the airport security check, they discover a bomb in his bag. He explains.”Statistics shows that the probability of a bomb being on an airplane is 1/1000. However, the chance that there are two bombs at one plane is 1/1000000. So, I am much safer…"

source: http://www.math.utah.edu/~cherk/mathjokes.html

Our favorite examples

Flip a fair coin

- random guessing on True/False test
- sex of your child

Flip a coin with P(Heads)=0.80

- some vaccine working
- some factory producing a nondefective device

Roll a die or **Pick a card**

- random guessing on multiple choice
- which door you choose

Flip a coin

Possible **Outcomes:** Head or Tail (H or T)

A **fair** coin:

$$P(H) = \frac{1}{2} \quad \text{and} \quad P(T) = \frac{1}{2}$$

Flip it 10 times:

```
sample(c("H","T"), 10, replace=TRUE)
```

```
##  [1] "H" "H" "T" "H" "H" "H" "T" "T" "T" "T"
```

Flip a coin

```r
flips = sample(c("H","T"), 10, replace=TRUE)
sum(flips=="H")
```

```
## [1] 4
```

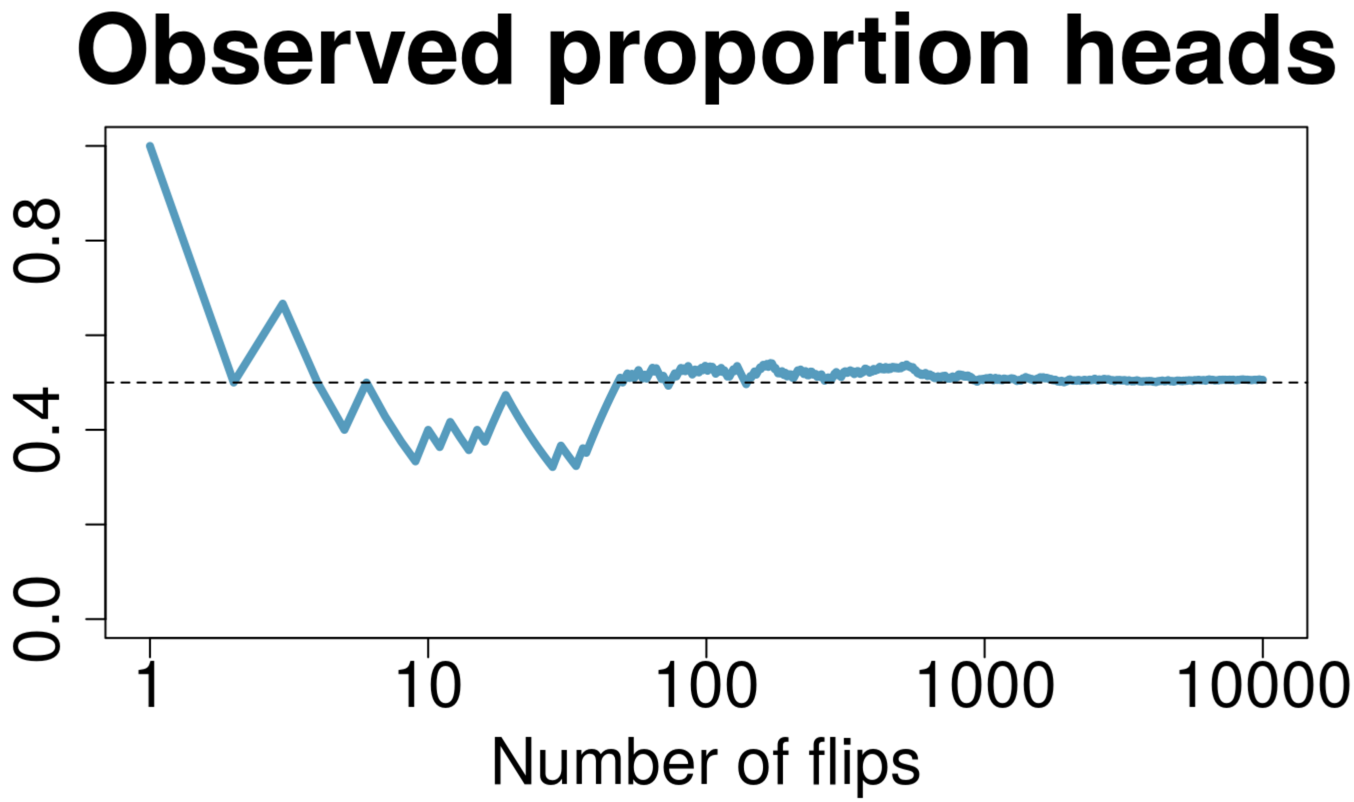
```r
flips = sample(c("H","T"), 100, repl=TRUE)
sum(flips=="H")
```

```
## [1] 59
```

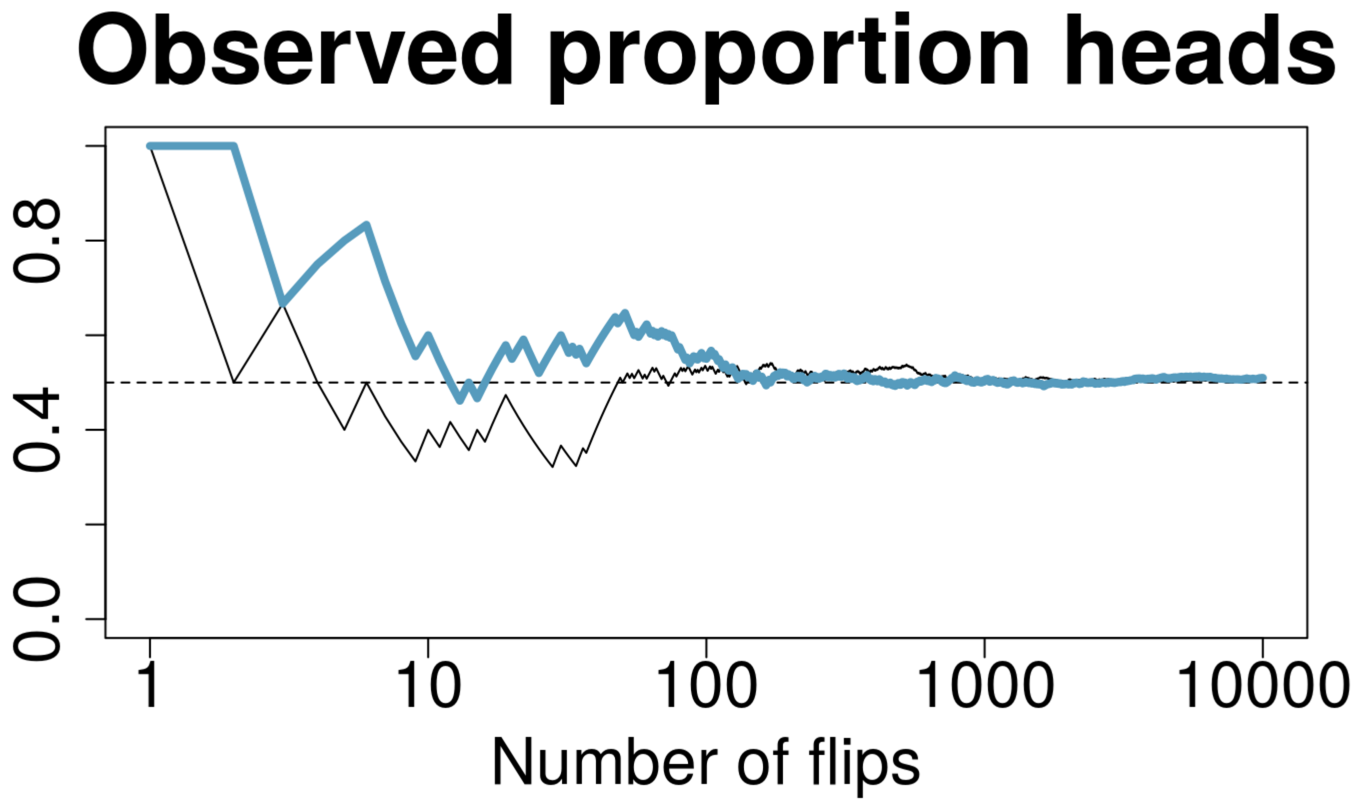
```r
flips = sample(c("H","T"), 10000, repl=TRUE)
sum(flips=="H")
```

```
## [1] 4950
```

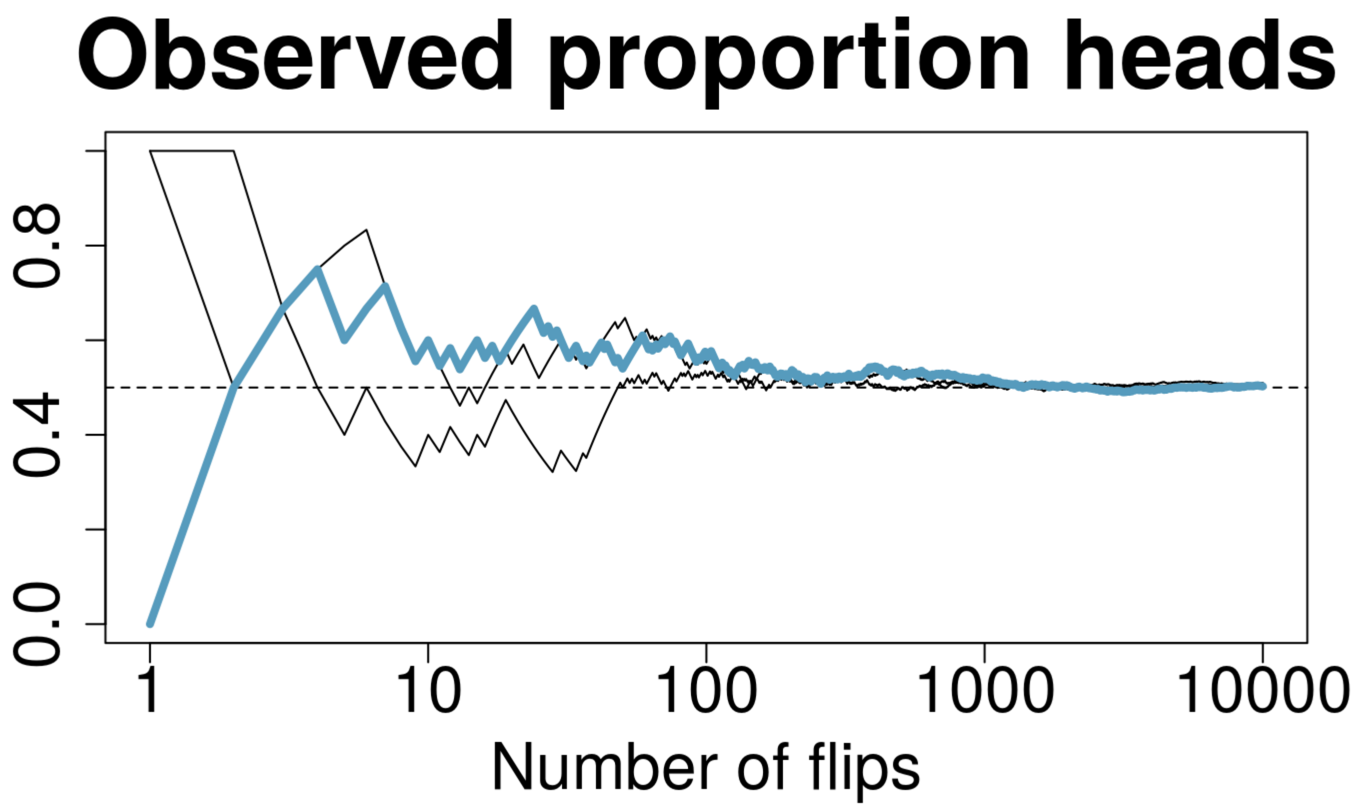
Probability as long-term frequency



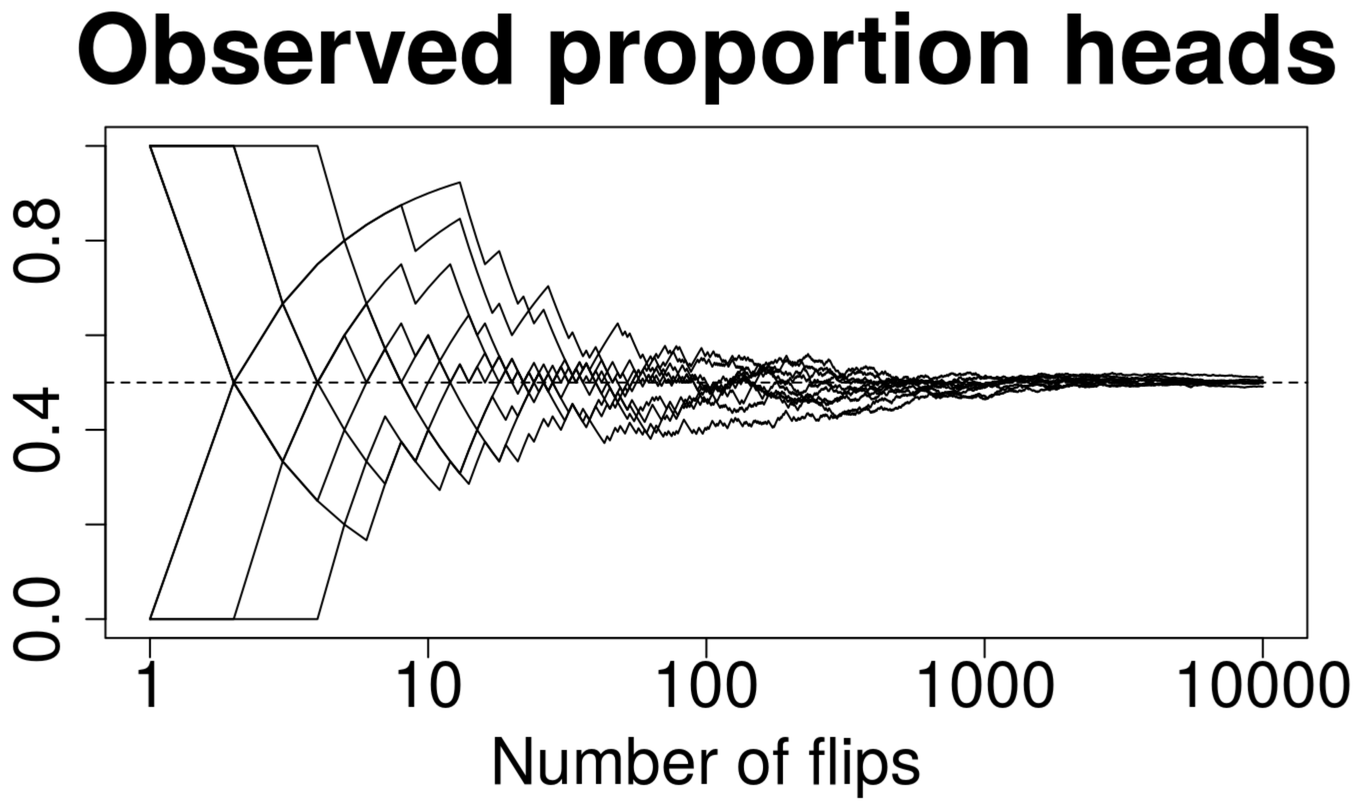
Probability as long-term frequency



Probability as long-term frequency



Probability as long-term frequency



Probability as long-term frequency

- We can't know whether one flip will be heads.
- But we can talk about the **probability** of heads.

> **Probability** - proportion of times something would occur if we could repeat random process an infinite number of times.

$P(H) = \frac{1}{2}$ means we expect about 5000 Heads in 10000 flips.

How to describe a process that is random

An **experiment** (or **trial**) - generates a realization of a random outcome - flip a coin; roll a die

Sample space - set of all possible outcomes - flip a coin: $\Omega = \{H, T\}$ - roll a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Event - a set of outcomes we can describe - e.g., $A = \{\text{Roll an even number}\} = \{2, 4, 6\}$ - e.g., $B = \{\text{Roll higher than a } 5\} = \{6\}$

Experiment: Roll two dice

Outcome?

(first die, second die)

Sample space?

$$\Omega = \{(1,1), (1,2), (1,3), \ldots, (6,6)\}$$

Events?

- $A = \{\text{get double 6s}\} = \{(6,6)\}$
- $B = \{\text{sum is 9}\} = \{(3,6), (4,5), (5,4), (6,3)\}$

Experiment: Roll two dice

Two special events

- $\{\text{sum is } 15\} = \emptyset$ - empty set or null event
- $\{\text{sum is not } 15\} = \Omega$ - sample space

When all outcomes are equally likely…

then event A has probability

$$P(A) = \frac{\text{number of outcomes in } A}{\text{total number of possible outcomes}}$$

Experiment: Roll two dice

Are all outcomes equally likely?

$$P(\{\text{roll doubles}\}) = ?$$

What is the total number of possible outcomes?

How many outcomes are in this event?

The complement

Given an event A, its **complement** is the set of all outcomes **not** in A.

- written A^C
- draw picture

$$P(A^C) = 1 - P(A)$$

Experiment: Roll two dice

$$P(\{\text{don't roll doubles}\}) = ?$$

The union

Given events A and B, the **union** is the set of outcomes in A **or** B.

- written $A \cup B$ or *"A or B"*
- draw picture

$$A \cup A^C = ?$$

Experiment: Roll two dice

$$A = \{\text{roll doubles}\} \qquad B = \{\text{sum is } 3\}$$

$$A \cup B = ?$$

$$P(A \cup B) = ?$$

(What is the total number of possible outcomes?)

(How many outcomes are in this event?)

The intersection

Given events A and B, the **intersection** is the set of outcomes in A **and** B.

- written $A \cap B$ or *"A and B"*
- draw picture

Experiment: Roll two dice

$$A = \{\text{roll doubles}\} \qquad B = \{\text{sum is less than } 4\}$$

$$A \cap B = ?$$

$$P(A \cap B) = ?$$

(What is the total number of possible outcomes?)

(How many outcomes are in this event?)

Disjoint or mutually exclusive

A and B are **disjoint** if they have **no outcomes in common**.

- $A \cap B = \emptyset$
- aka **mutually exclusive**
- draw picture

Axioms of probability

1. For any A, > $0 \leq P(A) \leq 1$.

2. $P(\Omega) = 1$

(**some** outcome occurs).

1. For any disjoint events A_1 and A_2

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

General addition rule

For any two events A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- don't double count outcomes in intersection
- picture proof
- extreme case: $A = B$.

Experiment: Roll two dice

$$A = \{\text{roll doubles}\} \qquad B = \{\text{sum is less than 4}\}$$

$$P(A \cup B) = ?$$