LECTURE 1

The Mathematical Theory of Probability

1 Introduction

Today we will do §2.1 and §2.2. We all have an intuitive notion of probability. What is the probability of tossing two heads in a row with a fair coin?

Method 1
List *all* possible outcomes

$$\left\{\boxed{\text{HH}}, HT, TH, TT\right\}$$

so $p =?$.

Question
What did we just assume to arrive at that answer?

Another Way

$$P\left(\underbrace{H}_{1^{st}\text{ toss}} \quad \underbrace{H}_{2^{nd}\text{ toss}}\right) = \frac{1}{2} \cdot \frac{1}{2}$$

However, it is important to put probability into a formal mathematical framework for many reasons.

1. Even "elementary" problems become too hard unless we can break them down into simpler problems using the rules of *set theory*.

Examples
Let's see how you can deal with these now and later (there is another reason which we will run into later – we often have infinite sets and need calculus, e.g. financial math.)

Problems

1. What is the probability of getting exactly *one* head in one hundred tosses of a fair coin?

2. What is the probability of getting exactly 27 heads in one hundred tosses of a fair coin?

2 Transition from the naive theory to the formal mathematical theory

To make the transition we introduce the word "experiment" which will be taken to mean "any action or process whose outcome is subject to uncertainty" (text – pg. 47).

Examples

- Tossing a fair coin 100 times.
- Dealing 5 cards from a 52 card deck – a poker hand.
- Dealing 13 cards from a 52 card deck – a bridge hand.

Definition
The set of all possible outcomes of an experiment will be called the *sample space* of that experiment and denote $\mathcal{S}$.

Experiment
3 tosses of a fair coin

$$\mathcal{S} = \left\{ \begin{array}{l} HHH, HHT, HTH, HTT, \\ THH, THT, TTH, TTT \end{array} \right\}$$

Definition A subset A of $\mathcal{S}$ is called an event.

Problem
Find

$$P\left(\begin{array}{l} \text{at least one head in} \\ \text{3 tosses of a fair coin} \end{array} \right).$$

We are looking for $P(A)$ where A is a subset of the previous $\mathcal{S}$.

$$A = \left\{ \begin{array}{l} HHH, HHT, HTH, HTT, \\ THH, THT, TTH \end{array} \right\}.$$

We will call this "our favorite sample space" from now on.

3 The Formal Mathematical Theory

Let $\mathcal{S}$ be a set (the sample space). A *probability measure* P on $\mathcal{S}$ is a rule (function) which assigns a real number $P(A)$ to any subset A of $\mathcal{S}$ (i.e., to any event) such that the following axioms are satisfied:

- For any event $A \subset \mathcal{S}$ we have $P(A) \geq 0$.

- $P(\mathcal{S}) = 1$.
- If $A_1, A_2, , \cdots, A_n, \cdots$ is a possibly infinite collection of pairwise disjoint (mutually exclusive) events then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots) \\ = \underbrace{\sum_{n=1}^{\infty} P(A_n).}_{\substack{\text{sum of an infinite series} \\ \text{not just ordinary sum}}}$$

  Here, mutually exclusive means $A_i \cap A_j = \emptyset$ for any pair i, j with $i \neq j$.

Special Cases

1. Two mutually-exclusive events A_1 and A_2 (so $A_1 \cap A_2 = \emptyset$).

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

2. n mutually-exclusive events $A_1, A_2, \cdots, A_n$

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

A Class of Examples
Let $\mathcal{S}$ be a set with n elements. Let $A \subset \mathcal{S}$ be any subset. Define

$$P(A) = \frac{\#(A)}{\#(\mathcal{S})} = \frac{\#(A)}{n}.$$

Then P satisfies the axioms 1., 2. and 3.

Here $\#(A)$ means the number elements in A. This is called the "equally likely probability measure".

An Example in the Above Class
Take our favorite sample space

$$\mathcal{S} = \left\{ \begin{array}{c} HHH, HHT, HTH, HTT, \\ THH, THT, TTH, TTT \end{array} \right\}.$$

Let A be the subset (event) of outcomes with at least one head and one tail. All the outcomes are equally likely (because the coin is fair) so

$$P(A) = \frac{\#(A)}{\#(\mathcal{S})} = \frac{?}{8}.$$

A Continuous Example
Consider the unit square $\mathcal{S}$ in the plane. Let $A \subset \mathcal{S}$ be any subset. Define

$$P(A) = \text{Area of } A.$$

Then P satisfies the axioms 1., 2. and 3.

Let A be the subset of points in the square below the diagonal. What is $P(A)$?

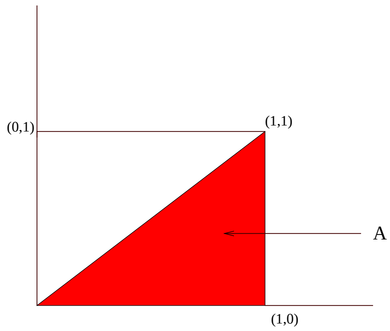

Figure 1: Red Half Square.

Can you find A so that

$$P(A) = \frac{1}{\pi}?$$

4 A Quick Trip Through Set-Theory (pgs. 49-50)

Let $\mathcal{S}$ be a set and A and B be subsets. Then we have $A \cup B$ (union), $A \cap B$ (intersection) and A' complement).

$$A \cup B = \text{"everything in } \mathcal{S} \text{ that is in either } A \textit{ or } B\text{"}$$

$$A \cap B = \text{"everything in } \mathcal{S} \text{ that is in } A \textit{ and } B\text{"}$$

Here are the Venn diagrams for complement, union and intersection.

First the diagram for complement.

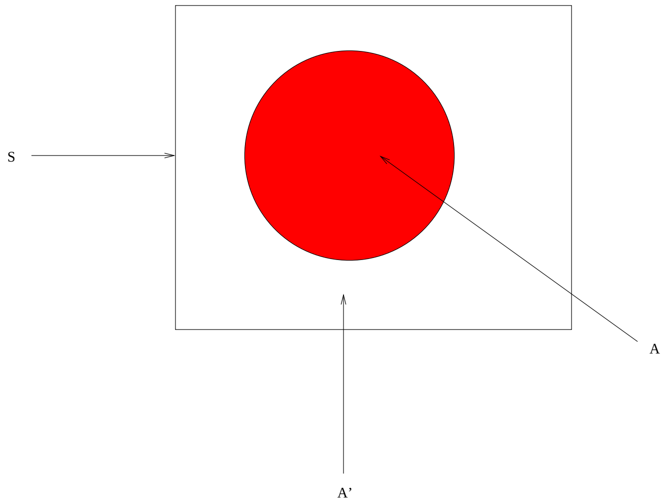

Figure 2: The Complement of A.

Now the diagram for union.

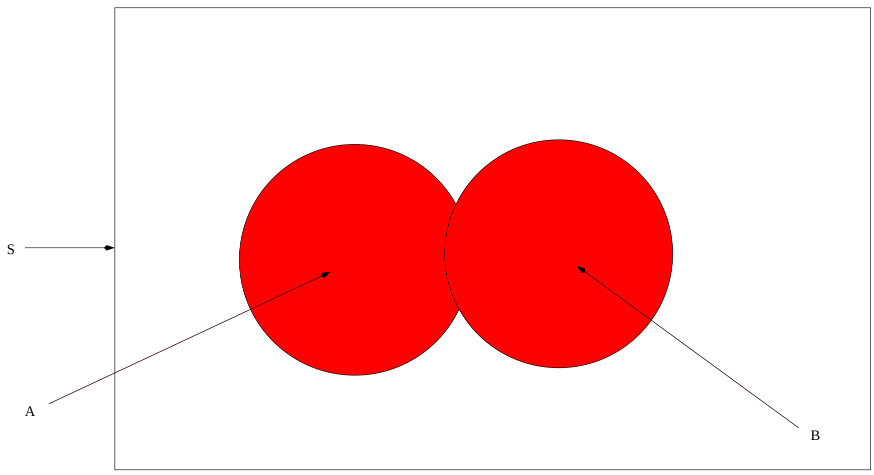

Figure 3: The Union of A and B.

Lastly the diagram for intersection.

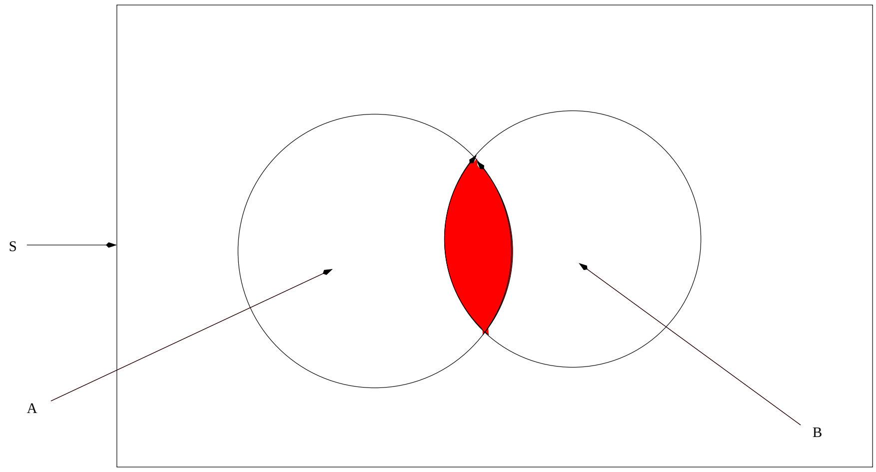

Figure 4: The Intersection of A and B.

The formulas that relate $\cup, \cap$ and $'$

To help you remember the formulas that follow, use the analogy

$$\begin{array}{ccl} \mathcal{S} & \longleftrightarrow & \text{set of numbers} \\ \cup & \longleftrightarrow & + \\ \cap & \longleftrightarrow & \cdot \end{array}$$

The Commutative Laws

$$\begin{aligned} A \cup B &= B \cup A \begin{pmatrix} \text{analogue} \\ a + b = b + a \end{pmatrix} \\ A \cap B &= B \cap A \begin{pmatrix} \text{analogue} \\ a \cdot b = b \cdot a \end{pmatrix} \end{aligned}$$

The Associative Laws

$$\begin{aligned} (A \cup B) \cup C &= A \cup (B \cup C) \begin{pmatrix} \text{analogue} \\ (a + b) + c = a + (b + c) \end{pmatrix} \\ (A \cap B) \cap C &= A \cap (B \cap C) \begin{pmatrix} \text{analogue} \\ (a \cdot b) \cdot c = a - (b - c) \end{pmatrix} \end{aligned}$$

Now we have laws that relate two or more of $\cup, \cap$ and $'$.

The Distributive Laws

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \quad \left(\begin{array}{c} \text{analogue} \\ a \cdot (b+c) \equiv (a \cdot b) + (a \cdot c) \end{array} \right) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \quad \text{no analogue} \end{aligned}$$

Problem
What would the analogue of the second distributive law say? It isn't true.

De Morgan's Laws
(no analogy with $+, \cdot$)

$$\begin{array}{ccc} (A \cup B)' & = & A' \cap B' \\ (A \cap B)' & = & A' \cup B' \\ C \subset D & \underset{\text{if and only if}}{\Longleftrightarrow} & C' \supset D' \end{array}$$

(so complement reverses $\cup, \cap$ and $\subset$).

5 Consequences of the Axioms of Probability Theory (pgs.54-56)

We will prove two propositions which will be *extremely useful* to you.

Proposition 1 (Complement Law)

$$P(A') = 1 - P(A).$$

Proof

$$A \cup A' = \mathcal{S}$$

so

$$P(A \cup A') = P(\mathcal{S}) = 1 \qquad \text{(axiom 2)}. \tag{\#}$$

But $A \cap A' = \emptyset$ so by axiom 3, special case 1

$$P(A \cup A') = P(A) + P(A'). \tag{\#\#}$$

Putting (#) and (##) together we get

$$1 = P(A) + P(A').$$

$\square$

Corollary 1

$$P(\emptyset) = 0.$$

Proof

$$\emptyset = \mathcal{S}'$$

so

$$P(\emptyset) = 1 - P(\mathcal{S}) = 1 - 1 = 0.$$

$\square$

Remark
$\emptyset$ is not the Greek letter phi, it is a Norwegian letter. The symbol was chosen by André Weil.

Corollary 2

$$P(A) \leq 1.$$

Proof

$$P(A) = 1 - P(A') \leq 1.$$

because $P(A') \geq 0$. $\square$

Bottom line

$$0 \leq P(A) \leq 1.$$

To illustrate the use of Proposition 1, let us go back to computing

$$P(\text{at least one head in three tosses}).$$

Put

$$\begin{aligned}
\mathcal{S} &= \text{our favorite sample space} \\
A &= \text{at least one head} \quad \text{SO} \\
A' &= \text{no heads} = \text{all tails} = TTT
\end{aligned}$$

so

$$P(A) = 1 - P(TTT) = 1 - \frac{1}{8} = \frac{7}{8}.$$

Recall that two events A and B are mutually exclusive if $A \cap B = \emptyset$ and axiom 3 says in this case

$$P(A \cup B) = P(A) + P(B). \tag{\#}$$

The following proposition is absolutely critical for computations.

Proposition 2(Additive Law)

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

Note that this is consistent with (#) above because if $A \cap B = \emptyset$ then

$$P(A \cap B) = P(\emptyset) = 0.$$

Proof

The proof is hard. It depends on the following Venn diagram.

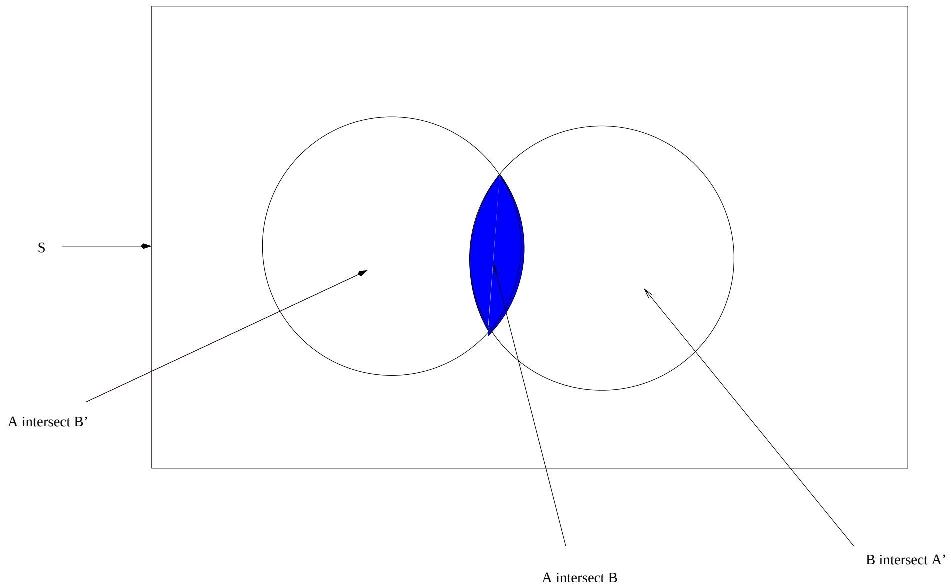

Figure 5: The Decomposition of $A \cup B$ into Three *Disjoint* Sets.

We see that $A \cup B$ is the union of three *disjoint* sets

$$A \cup B = (A \cap B') \cup (A \cap B) \cup (B \cap A')$$

so by axiom 3 with $n = 3$,

$$P(A \cup B) = P(A \cap B') + P(A \cap B) + P(B \cap A'). \tag{\#\#}$$

How do we compute the first and third terms?

We have a disjoint union (i.e., union of mutually exclusive sets)

$$A = (A \cap B') \cup (A \cap B) \text{ the two left-most sets in the figure above}$$

so by axiom 3

$$P(A) = P(A \cap B) + P(A \cap B')$$

whence

$$P(A \cap B') = P(A) - P(A \cap B). \tag{1}$$

Similarly,

$$P(B \cap A') = P(B) - P(A \cap B). \tag{3}$$

Plug (1) and (3) into (##). □

What about the intersection of three terms?

Proposition 3

$$\begin{gathered} P(A \cup B \cup C) = P(A) + P(B) + P(C) \\ -P(A \cap B) - P(A \cap C) - P(B \cap C) \\ +P(A \cap B \cap C). \end{gathered}$$

This is (more or less) “the principle of exclusion and inclusion”.

1. Include the singletons A, B, C.

2. Exclude the pairs $A \cap B$, $A \cap C$, $B \cap C$

3. Include the triple $A \cap B \cap C$.