2.2–2.3 Introduction to Probability and Sample Spaces

Prof. Tesler
Math 186
Winter 2016
Course overview

Probability: Determine likelihood of events
- Roll a die. The probability of rolling 1 is 1/6.

Descriptive statistics: Summarize data
- Mean, median, standard deviation, ...

Inferential statistics: Infer a conclusion/prediction from data
- Test a drug to see if it is safe and effective, and at what dose.
- Poll to predict the outcome of an election.
- Repeatedly flip a coin or roll a die to determine if it is fair.

Bioinformatics
- We’ll apply these to biological data, such as DNA sequences and microarrays.
Related courses

- **Math 183**: Usually uses the same textbook and chapters as Math 186. Focuses on the examples in the book. The mathematical content is the same, but Math 186 has extra material for bioinformatics.

- **Math 180ABC plus 181ABC**: More in-depth: a year of probability and a year of statistics.

- **CSE 103, Econ 120A, ECE 109**: One quarter intro to probability and statistics, specialized for other areas.

- **Math 283**: Graduate version of this course. Review of basic probability and statistics, with a lot more applications in bioinformatics.
2.2 Sample spaces

- Flip a coin 3 times. The possible outcomes are
  
  $\text{HHH HHT HTH HTT THH THT TTH TTT}$

- The sample space is the set of all possible outcomes:
  
  $S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}$

- The size of the sample space is
  
  $N(S) = 8$  Our book's notation
  
  $|S| = 8$  A more common notation in other books

- We could count this by making a $2 \times 2 \times 2$ table:
  
  2 choices for the first flip
  
  $\times$ 2 choices for the second flip
  
  $\times$ 2 choices for the third flip
  
  $= 2^3 = 8$

- The number of strings $x_1 x_2 \ldots x_k$ or sequences $(x_1, x_2, \ldots, x_k)$ of length $k$ with $r$ choices for each entry is $r^k$. 
Rolling two dice

- Roll two six-sided dice, one red, one green:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(1,1)</td>
<td>(1,2)</td>
<td>(1,3)</td>
<td>(1,4)</td>
<td>(1,5)</td>
<td>(1,6)</td>
</tr>
<tr>
<td>2</td>
<td>(2,1)</td>
<td>(2,2)</td>
<td>(2,3)</td>
<td>(2,4)</td>
<td>(2,5)</td>
<td>(2,6)</td>
</tr>
<tr>
<td>3</td>
<td>(3,1)</td>
<td>(3,2)</td>
<td>(3,3)</td>
<td>(3,4)</td>
<td>(3,5)</td>
<td>(3,6)</td>
</tr>
<tr>
<td>4</td>
<td>(4,1)</td>
<td>(4,2)</td>
<td>(4,3)</td>
<td>(4,4)</td>
<td>(4,5)</td>
<td>(4,6)</td>
</tr>
<tr>
<td>5</td>
<td>(5,1)</td>
<td>(5,2)</td>
<td>(5,3)</td>
<td>(5,4)</td>
<td>(5,5)</td>
<td>(5,6)</td>
</tr>
<tr>
<td>6</td>
<td>(6,1)</td>
<td>(6,2)</td>
<td>(6,3)</td>
<td>(6,4)</td>
<td>(6,5)</td>
<td>(6,6)</td>
</tr>
</tbody>
</table>

- The sample space is

\[
S = \{(1,1), (1,2), \ldots, (6,6)\}
= \{(i,j) \in \mathbb{Z}^2 : 1 \leq i \leq 6, \ 1 \leq j \leq 6\}
\]

where \( \mathbb{Z} = \text{integers} \quad \mathbb{Z}^2 = \text{ordered pairs of integers} \)

- \( N(S) = 6^2 = 36 \)
A codon is a DNA sequence of length 3, in the alphabet of nucleotides \{ A, C, G, T \}:

\[ S = \{ \text{AAA, AAC, AAG, AAT, \ldots, TTT} \} \]

How many codons are there?

\[ N(S) = 4^3 = 64 \]
A continuous sample space

Consider this disk (filled-in circle):

\[ S = \{ (x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 2^2 \} \]

Complications

- The sample space is infinite and continuous.
- The choices of \( x \) and \( y \) are dependent. E.g.:
  - at \( x = 0 \), we have \(-2 \leq y \leq 2\);
  - at \( x = 2 \), we have \( y = 0 \).
Flip a coin 3 times. The sample space is

\[ S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \]

An *event* is a subset of the sample space \((A \subset S)\):

- \(A = \text{“First flip is heads”} = \{HHH, HHT, HTH, HTT\}\)
- \(B = \text{“Two flips are heads”} = \{HHT, HTH, THH\}\)
- \(C = \text{“Four flips are heads”} = \emptyset\) (*empty set or null set*)
Using set operations to form new events

\[ A = \text{“First flip is heads”} = \{HHH, HHT, HTH, HTT\} \]

\[ B = \text{“Two flips are heads”} = \{HHT, HTH, THH\} \]

**Union:** All elements that are in \( A \) or in \( B \)

\[ A \cup B = \{HHH, HHT, HTH, HTT, THH\} \]

“\( A \) or \( B \)”: “The first flip is heads or two flips are heads”

This is *inclusive or*: one or both conditions are true.

**Intersection:** All elements that are in both \( A \) and in \( B \)

\[ A \cap B = \{HHT, HTH\} \]

“A and B”: “The first flip is heads and two flips are heads”

**Complement:** All elements of the sample space not in \( A \)

\[ A^c = \{THT, TTH, TTT, THH\} \]

“Not \( A \)”: “The first flip is not heads”
Venn diagram and set sizes

\[ A = \{HHH, HHT, HTH, HTT\} \]
\[ B = \{HHT, HTH, THH\} \]
\[ A \cup B = \{HHH, HHT, HTH, HTT, THH\} \]
\[ A \cap B = \{HHT, HTH\} \]

Relation between sizes of union and intersection

- Notice that \( N(A \cup B) = N(A) + N(B) - N(A \cap B) \)
  \[
  5 = 4 + 3 - 2
  \]
- \( N(A) + N(B) \) counts everything in the union, but elements in the intersection are counted twice. Subtract \( N(A \cap B) \) to compensate.

Size of complement

\[ N(B^c) = N(S) - N(B) \]
\[
5 = 8 - 3
\]
Algebraic rules for set theory

**Commutative laws**
\[ A \cup B = B \cup A \]
\[ A \cap B = B \cap A \]

**Associative laws**
\[ (A \cup B) \cup C = A \cup (B \cup C) \]
\[ (A \cap B) \cap C = A \cap (B \cap C) \]

One may omit parentheses in \( A \cap B \cap C \) or \( A \cup B \cup C \). But don’t do that with a mix of \( \cup \) and \( \cap \).

**Distributive laws**
\[ A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \]
\[ A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \]
These are like \( a(b + c) = ab + ac \)

**Complements**
\[ A \cup A^c = S \]
\[ A \cap A^c = \emptyset \]

**De Morgan’s laws**
\[ (A \cup B)^c = A^c \cap B^c \]
\[ (A \cap B)^c = A^c \cup B^c \]
Distributive laws

Visualizing identities using Venn diagrams: \( A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \)
Two events are **mutually exclusive** if their intersection is $\emptyset$.

$A =$ “First flip is heads” $= \{HHT, HTH, HTT\}$

$B =$ “Two flips are heads” $= \{HHT, HTH, THH\}$

$C =$ “One flip is heads” $= \{HTT, THT, TTH\}$

$A$ and $B$ are not mutually exclusive, since $A \cap B = \{HHT, HTH\} \neq \emptyset$.

$B$ and $C$ are mutually exclusive, since $B \cap C = \emptyset$.

For mutually exclusive events, since $N(B \cap C) = 0$, we get:

$N(B \cup C) = N(B) + N(C)$

Events $A_1, A_2, \ldots$ are **pairwise mutually exclusive** when

$A_i \cap A_j = \emptyset$ for $i \neq j$.
2.3 Probability functions

Historically, there have been several ways of defining probabilities. We’ll start with *Classical Probability*:

### Classical probability

- Suppose the sample space has $n$ outcomes ($N(S) = n$) and all of them are equally likely.

- Each outcome has a probability $1/n$ of occurring:
  \[ P(s) = \frac{1}{n} \text{ for each outcome } s \in S \]

- An event $A \subseteq S$ with $m$ outcomes has probability $m/n$ of occurring:
  \[ P(A) = \frac{m}{n} = \frac{N(A)}{N(S)} \]

### Example: Rolling a pair of dice

- $N(S) = n = 36$
- $P(\text{first die is } 3) = P(\{(3, 1), (3, 2), \ldots, (3, 6)\}) = \frac{6}{36}$
- $P(\text{the sum is } 8) = P(\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}) = \frac{5}{36}$
Drawbacks

- What if outcomes are not equally likely?
- What if there are infinitely many outcomes?
Empirical probability

Use long-term frequencies of different outcomes to estimate their probabilities.

- Flip a coin a lot of times. Use the fraction of times it comes up heads to estimate the probability of heads.
- 520 heads out of 1000 flips leads to estimating $P(\text{heads}) = 0.520$.

This estimate is only approximate because
- Due to random variation, the numerator will fluctuate.
- Precision is limited by the denominator. 1000 flips can only estimate it to three decimals.

More on this later in the course in Chapter 5.3.
Empirical probability

- *E. coli* has been sequenced: AGCTTTTTTCAT...

On the forwards strand:

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Count</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1,142,136</td>
<td>$P(A) = \frac{1,142,136}{4,639,221} \approx 0.2461913326$</td>
</tr>
<tr>
<td>C</td>
<td>1,179,433</td>
<td>$P(C) \approx 0.2536578878$</td>
</tr>
<tr>
<td>G</td>
<td>1,176,775</td>
<td>$P(G) \approx 0.2542308288$</td>
</tr>
<tr>
<td>T</td>
<td>1,140,877</td>
<td>$P(T) \approx 0.2459199508$</td>
</tr>
<tr>
<td>Total</td>
<td>4,639,221</td>
<td>1</td>
</tr>
</tbody>
</table>

Sample space: set of positions $S = \{1, 2, \ldots, 4639221\}$

Event $A$ is the set of positions with nucleotide A (similar for $C$, $G$, $T$).

- $A = \{1, 9, \ldots\}$  
- $C = \{3, 8, \ldots\}$  
- $G = \{2, \ldots\}$  
- $T = \{4, 5, 6, 7, 10\}$

Simplistic model: the sequence is generated from a biased 4-sided die with faces A, C, G, T.
Axiomatic probability

A definition of a probability function \( P \) based on events and the following axioms is the most useful. Each event \( A \subset S \) is assigned a probability that obeys these axioms:

**Axioms for a finite sample space**

- For any event \( A \subset S \): \( P(A) \geq 0 \)
- The total sample space has probability 1: \( P(S) = 1 \)
- For mutually exclusive events \( A \) and \( B \): \( P(A \cup B) = P(A) + P(B) \)

**Additional axiom for an infinite sample space**

- If \( A_1, A_2, \ldots \) (infinitely many) are pairwise mutually exclusive, then
  \[
  P \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i)
  \]
- \( \bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots \) is like \( \sum \) notation, but for unions.
Axiomatic probability — additional properties

Additional properties of the probability function follow from the axioms:

- \( P(A^c) = 1 - P(A) \)
- \( P(\emptyset) = 0 \)
- If \( A \subset B \) then \( P(A) \leq P(B) \)
- \( P(A) \leq 1 \)
- **Additive Law:** If \( A_1, A_2, \ldots, A_n \) are pairwise mutually exclusive \((A_i \cap A_j = \emptyset \text{ for all } i \neq j)\) then

\[
P \left( \bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} P(A_i)
\]

(The first three axioms only lead to a proof of this for finite \( n \), so we have to introduce the infinite case as an additional axiom.)

- **De Moivre’s Law:** \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)
Axiomatic probability — additional properties

- \( P(A^c) = 1 - P(A) \)

**Example:**
- \( P(\text{die roll} = 1) = \frac{1}{6} \)
- \( P(\text{die roll} \neq 1) = 1 - \frac{1}{6} = \frac{5}{6} \)

**Proof.**

\( A \) and \( A^c \) are mutually exclusive, so
\[ P(\overline{A \cup A^c}) = P(A) + P(A^c). \]

Also, \( P(A \cup A^c) = P(S) = 1. \)

Thus, \( P(A^c) = 1 - P(A). \)

- \( P(\emptyset) = 0 \)

**Proof.**

\[ P(\emptyset) = P(S^c) = 1 - P(S) = 1 - 1 = 0 \]
Axiomatic probability — additional properties

- If \( A \subset B \) then \( P(A) \leq P(B) \)

**Proof.**

Write \( B = A \cup (A^c \cap B) \).

\( A \) and \( A^c \cap B \) are mutually exclusive, so

\[
P(B) = P(A) + P(A^c \cap B).
\]

The first axiom gives \( P(A^c \cap B) \geq 0 \), so

\[
P(B) \geq P(A).
\]

- \( P(A) \leq 1 \)

**Proof.**

\( A \subset S \) so \( P(A) \leq P(S) = 1 \).
Axiomatic probability — additional properties

- **Additive Law:** If $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive ($A_i \cap A_j = \emptyset$ for all $i \neq j$) then

$$P \left( \bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} P(A_i)$$

**Proof.**

Use induction to prove this, starting with $n = 1$ (trivial) and $n = 2$ (given as an axiom).

Note that induction only proves it for finite $n \geq 1$, so we have to introduce an additional axiom for the infinite case.

□
**De Moivre’s Law:** \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \)

**Proof.**

\[
P(A) = P(A \cap B) + P(A \cap B^c) \\
P(B) = P(A \cap B) + P(A^c \cap B)
\]

\[
P(A) + P(B) = P(A \cap B) + (P(A \cap B) + P(A \cap B^c) + P(A^c \cap B)) \\
= P(A \cap B) + P(A \cup B)
\]
Finite discrete sample space

- Each outcome has a probability between 0 and 1 and the probabilities add up to 1:
  \[ 0 \leq P(s) \leq 1 \text{ for each } s \in S \]
  \[ \sum_{s \in S} P(s) = 1 \]
- For an event \( A \subset S \), define \( P(A) = \sum_{s \in A} P(s) \).
- E.g., \( P(\{HHT, HTH, THH\}) = P(HHT) + P(HTH) + P(THH) \).

Biased die

- A die has \( n \) faces, 1, 2, \ldots, \( n \).
- The probability of face \( i \) is \( q_i \), with \( 0 \leq q_i \leq 1 \), and \( q_1 + \cdots + q_n = 1 \).
- \( P(\text{even number}) = q_2 + q_4 + \cdots \)

DNA

- Generate a DNA sequence by rolling a biased ACGT-die.
- Of course, it’s not random: there is structure in genes, codons, repeats, etc. Violations of what’s expected from random rolls of a dice are used to detect structure in the DNA sequence.