Lecture 1: Introduction, Set Theory, and Boolean Algebra

Sta 111
Colin Rundel
May 13, 2014

General Info

Classroom: Perkins 2-072
Time: Mon - Fri, 2:00 - 3:15 pm
             Wed, 3:30 - 4:30 pm

Professor: Colin Rundel
Office: Old Chemistry 223E
Email: colin.rundel@stat.duke.edu

Office hours: Monday 12:30 - 1:30 pm
             Monday 3:30 - 5:00 pm
             or by appointment.

Course Website: http://stat.duke.edu/~cr173/Sta111_Su14

Grading

Homework 25%
Labs 15%
Midterm 30%
Final 30%

- Grades will be curved at the end of the course after overall averages have been calculated.
  - Average of 90-100 guaranteed A-.
  - Average of 80-90 guaranteed B-.
  - Average of 70-80 guaranteed C-.

- The more evidence there is that the class has mastered the material, the more generous the curve will be.

Materials

Textbook: Probability and Statistical Inference, Hogg, Tanis, & Zimmerman
          ISBN: 9780321923271

Exams: Midterm: Wednesday, June 4th, 2:00 - 4:30 pm
       Final: Thursday, June 19th, 2:00 - 4:30 pm

Holidays: Monday, May 26th - Memorial Day
Labs

Objective: Give you hands on experience with data analysis using statistical software.

http://beta.rstudio.org

- A gmail account is needed to register your Rstudio account, add your gmail address using the quiz on Sakai.
- 10 labs are planned.
- Write ups due the following Tuesday - most can be completed in class.

Policies

- There will not be make-ups for any of the homework, labs, or exams.
- All regrade requests on homework assignments and exams should be submitted promptly. There will be no grade changes after the final exam.
- Academic Integrity & Duke Community Standard
- Excused absences
- Late work policy - for homework and lab write ups:
  - late but during class: -10%
  - after class on due date: -20%
  - next day or later: no credit

Set Theory (Briefly)

What is a set?

Georg Cantor

A set is a gathering together into a whole of definite, distinct objects of our perception or of our thought - which are called elements of the set.

A set is a grouping of distinct objects. We will denote sets using capital letters (A,B) and the elements of the set using curly braces ({}).

To define a set we either enumerate all elements or use set-builder notation:

The set of card suits

\{♥, ♦, ♣, ♠\}

The set of all prime numbers

\{x|x \text{ is prime}\}

Set notation and operations

- \(x \in A\) \(x\) is an element of \(A\)
- \(A \cup B, A + B\) is the union of \(A\) and \(B\) \(\{x|x \in A \text{ or } B\}\)
- \(A \cap B, AB\) is the intersection of \(A\) and \(B\) \(\{x|x \in A \text{ and } B\}\)
- \(A \subseteq B\) means \(A\) is a subset of \(B\) \(x \in A \Rightarrow x \in B\)
- \(A \subset B\) means \(A\) is a proper subset of \(B\) \(A \subseteq B\) and \(A \neq B\)
- \(U\) is the universal set
- \(\emptyset, \{\}\) is the empty set
- \(A', \overline{A}, A^c\) is the complement of \(A\) \(A' = \{x|x \notin A\}\)
- \(|A|\) is the cardinality of \(A\)
Set Theory (Briefly)

Set operation properties

**Commutativity**
- \( A \cup B = B \cup A \)
- \( A \cap B = B \cap A \)

**Associativity**
- \((A \cup B) \cup C = A \cup (B \cup C)\)
- \((A \cap B) \cap C = A \cap (B \cap C)\)

**Distribution**
- \( A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \)
- \( A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \)

**Idempotence**
- \( A \cup A = A \)
- \( A \cap A = A \)

**Absorption**
- \( A \cup (A \cap B) = A \)
- \( A \cap (A \cup B) = A \)

- \( A \cup A' = U \)
- \( A \cap A' = \emptyset \)

De Morgan’s Laws

Useful rules that allow us to relate conjunction (union / or) and disjunction (intersection / and) using only negation.

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

Some examples

Show the following relations are true:

- \( A \cup B \cap C = (A \cup B) \cap (A \cup C) \)
- \((A \cup C) \cap A \cup (A \cap C) \cup C = A \cup C \)
- \( B \cup A \cap (B \cup C) \cup (B \cap C) = B \cup (A \cap C) \)

Logic / Boolean Algebra (Briefly)

A little Logic / Boolean Algebra

Logic statements are statements that must be either true or false. In general we indicate logic statements using lower case letters (e.g. \( p, q \)).

There is a natural correspondence between set theory and logic operators:

<table>
<thead>
<tr>
<th>Set Theory</th>
<th>Logic</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A \cup B )</td>
<td>( p \lor q )</td>
</tr>
<tr>
<td>( A \cap B )</td>
<td>( p \land q )</td>
</tr>
<tr>
<td>( A = B )</td>
<td>( p \iff q )</td>
</tr>
<tr>
<td>( A \subseteq B )</td>
<td>( p \rightarrow q )</td>
</tr>
<tr>
<td>((A \cup B)’ = A’ \cap B’)</td>
<td>((p \lor q)’ \iff p’ \land q’)</td>
</tr>
<tr>
<td>((A \cap B)’ = A’ \cup B’)</td>
<td>((p \land q)’ \iff p’ \lor q’)</td>
</tr>
</tbody>
</table>
More Correspondence

Additionally, we can construct logic statements from sets by asking if an element belongs to a set, e.g.

\[ A \cup B \rightarrow x \in (A \cup B) \leftrightarrow x \in A \text{ or } x \in B \]

Example - let \( A \) be the set of possible weather for today, \( \{\text{sun, clouds, rain, snow}\} \).

What does it mean to say that:

- The probability of rolling snake eyes is \( P(S) = 1/36? \)
- The probability of flipping a coin and getting heads is \( P(H) = 1/2? \)
- The probability Apple’s stock price goes up today is \( P(+) = 3/4? \)

Interpretations:
- Symmetry: If there are \( k \) equally-likely outcomes, each has \( P(E) = 1/k \)
- Frequency: If you can repeat an experiment indefinitely,
  \[ P(E) = \lim_{n \to \infty} \frac{\#E}{n} \]
- Belief: If you are indifferent between winning \( \$1 \) if \( E \) occurs or winning \( \$1 \) if you draw a blue chip from a box with \( 100 \times p \) blue chips, rest red,
  \[ P(E) = p \]

Event space is **big**

Even for relatively small outcome spaces there are a lot of possible events we can define.

Let \( \Omega = \{\omega_1, \omega_2, \omega_3\} \) then

\[ E = \{\emptyset, \omega_1, \omega_2, \omega_3, \omega_1 \cup \omega_2, \omega_1 \cup \omega_3, \omega_2 \cup \omega_3, \omega_1 \cup \omega_2 \cup \omega_3\} \]

\[ |E| = 8 = 2^3 \]

We can show that for an outcome space \( \Omega = \{\omega_1, \omega_2, \ldots, \omega_n\} \) the total number of possible events is \( 2^n \).

\[ E_1 = \emptyset, \ E_2 = \omega_1, \ E_3 = \omega_2, \ldots, \ E_{n+1} = \omega_n, \ E_{n+2} = \omega_1 \cup \omega_2, \ldots, \ E_{2^n} = \ \bigcup_{i=1}^{n} \omega_i \]

Why can we get away with exclusively using unions here?
Rules of Probability - Kolmogorov’s axioms

(1) Non-negative:
\[ P(E) \geq 0 \]

(2) Addition:
\[ P(E \cup F) = P(E) + P(F) \text{ if } E \cap F = \emptyset \]

(2)' Countable Addition:
\[ P\left( \bigcup_{i=1}^{\infty} E_i \right) = \sum_{i=1}^{\infty} P(E_i) \text{ if } E_i \cap E_j = \emptyset \text{ for } i \neq j \]

(3) Total one:
\[ P(\Omega) = 1 \]

Useful Identities

Complement Rule:
\[ P(\text{not } A) = P(A^c) = 1 - P(A) \]

Difference Rule:
\[ P(B \text{ and not } A) = P(B \cap A^c) = P(B) - P(A) \text{ if } A \subseteq B \]

Inclusion-Exclusion:
\[ P(A \cup B) = P(A) + P(B) - P(AB) \]

Generalized Inclusion-Exclusion

\[ P\left( \bigcup_{i=1}^{n} E_i \right) = \sum_{i \leq n} P(E_i) - \sum_{i < j \leq n} P(E_iE_j) + \sum_{i < j < k \leq n} P(E_iE_jE_k) - \ldots + (-1)^{n+1}P(E_1 \ldots E_n) \]

For the case of \( n = 3 \):
\[ P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \]

Equally Likely Outcomes

\[ P(E) = \frac{\#(E)}{\#(\Omega)} = \sum_{i} \frac{1_{\omega \in E}}{\#(\Omega)} \]

Notation:
Cardinality - \( \#(S) = \text{number of elements in set } S \)

Indicator function - \( 1_{x \in S} = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases} \)

Probability of rolling an even number with a six sided die?
\[ E = \{2, 4, 6\} \text{ and } \Omega = \{1, 2, 3, 4, 5, 6\} \]
\[ P(E) = \frac{3}{6} = \frac{1}{2} \]
Sampling

Imagine an urn filled with white and black marbles
...or a deck of cards
...or a bingo cage
...or a hat full of raffle tickets

Two common options + one extra for completeness:
- Sampling without replacement
- Sampling with replacement
- Pólya urn model

Birthday Problem

Ignoring leap years, and assuming birthdays are equally likely to be any day of the year, what is the chance of a tie in birthdays among the students in this class?
Something a little more complicated . . .

What is the probably that if you deal yourself 5 cards you will have a royal flush? (Ace, king, queen, jack, and 10 of the same suit)