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*

Materials

Textbook:	Probability and Statistical Inference, Hogg, Tanis, & Zimmerman Pearson, 9^{th} Edition, 2015 ISBN: 9780321923271
Exams:	Midterm: Wednesday, June 4^{th}, 2:00 - 4:30 pm Final: Thursday, June $19^t h$ 2:00 - 4:30 pm
Holidays:	Monday, May 26^{th} - Memorial Day

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.

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

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
$\{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}$	$\{x \mid x \text{ is prime}\}$

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	$\{x \mid x \in A \text{ or } B\}$
$A \cap B$, AB	is the intersection of A and B	$\{x \mid x \in A \text{ and } B\}$
$A \subseteq B$	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	$A' = \{x \mid x \notin A\}$
$\lvert A \rvert$	is the cardinality of A	

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 disjuction (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:

1. $A \cup B \cap C = (A \cup B) \cap (A \cup C)$
2. $(A \cup C) \cap A \cup (A \cap C) \cup C = A \cup C$
3. $B \cup A \cap (B \cup C) \cup (B \cap C) = B \cup (A \cap C)$

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:

Set Theory	Logic
$A \cup B$	p or q
$A \cap B$	p and q
$A = B$	$p \leftrightarrow q$
$A \subseteq B$	$p \rightarrow q$
$(A \cup B)' = A' \cap B'$	$(p \text{ or } q)' \leftrightarrow p' \text{ and } q'$
$(A \cap B)' = A' \cup B'$	$(p \text{ and } q)' \leftrightarrow p' \text{ or } q'$

More Correspondence

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

$$A \cup B \longrightarrow x \in (A \cup B) \leftrightarrow x \in A \text{ or } x \in B$$

Example - let A be the set of possible weather for today,

$$\{sun, clouds, rain, snow\}.$$

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,
$$P(E) = p$$

Terminology

Outcome space (Ω) - set of all possible outcomes (ω).

Examples:	3 coin tosses	{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
	One die roll	{1,2,3,4,5,6}
	Sum of two rolls	{2,3,...,11,12}
	Seconds waiting for bus	$[0, \infty)$

Event (E) - subset of Ω ($E \subseteq \Omega$) that might happen, or might not

Examples:	2 heads	{HHT, HTH, THH}
	Roll an even number	{2,4,6}
	Wait $<$ 2 minutes	$[0, 120)$

Random Variable (X) - a value that depends somehow on chance

Examples:	# of heads	{3, 2, 2, 1, 2, 1, 1, 0}
	# flips until heads	{3, 2, 1, 1, 0, 0, 0, 0}
	2ˆdie	{2, 4, 8, 16, 32, 64}

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:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) \text{ if } E_i E_j = \emptyset \text{ for } i \neq j$$

(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:

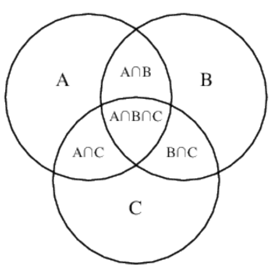
$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Generalized Inclusion-Exclusion

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i \leq n} P(E_i) - \sum_{i<j\leq n} P(E_i E_j) + \sum_{i<j<k\leq n} P(E_i E_j E_k) - \ldots + (-1)^{n+1} P(E_1 \ldots E_n)$$

For the case of $n = 3$:

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



Equally Likely Outcomes

$$P(E) = \frac{\#(E)}{\#(\Omega)} = \sum_i \frac{1_{\omega_i \in E}}{\#(\Omega)}$$

Notation:

Cardinality - $\#(S)$ = number of elements in set S

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

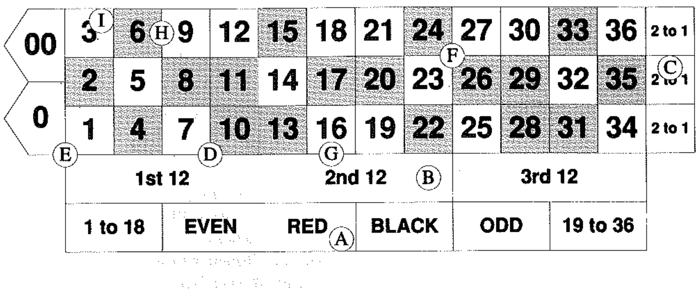
Probability of rolling an even number with a six sided die?

$$E = \{2, 4, 6\} \text{ and } \Omega = \{1, 2, 3, 4, 5, 6\}$$

$$P(E) = 3/6 = 1/2$$

Roulette

FIGURE 1. Layout of a Nevada roulette table. Key to colors: 0 and 00 = Green, *unshaded numbers* = Red, *shaded numbers* = Black.



Roulette, cont.

Play	Set of winning numbers	Payoff odds
A. Even money play	Group of 18 numbers as marked in the box	1 to 1
B. Dozen play	12 numbers marked in the box	2 to 1
C. Column play	12 numbers in column (shown here as a row)	2 to 1
D. Line play	Six numbers above	5 to 1
E. House special	0, 00, 1, 2, 3	6 to 1
F. Quarter play	Four numbers in square	8 to 1
G. Street play	Three numbers above	11 to 1
H. Split play	Two adjoining numbers	17 to 1
I. Straight play	Single number	35 to 1

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

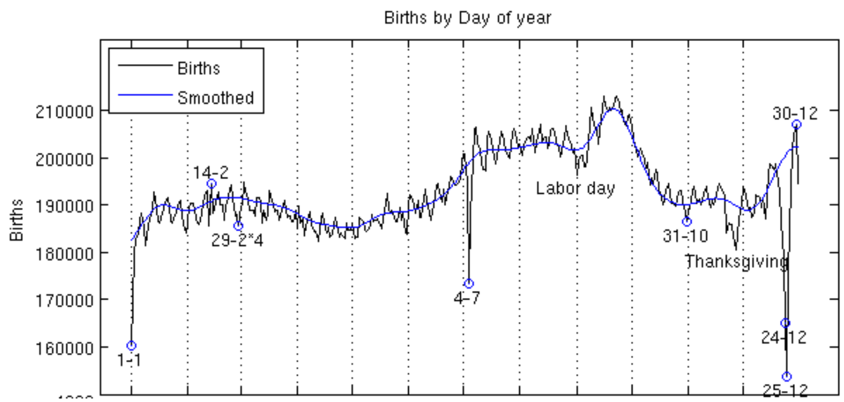
Two common options + one extra for completeness:

- 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?

Birthday Problem, cont.



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)