STA 611: Introduction to Statistical Methods

Instructor: Jenný Brynjarsdóttir

TueThu 3:05 - 4:20 pm, Physics Building 130

August 28, 2012

Introduction

About me

From Iceland.
BS and MS from University of Iceland and PhD from The Ohio State University.
Postdoc at Samsi and Duke.

- Course website: `http://stat.duke.edu/courses/Fall12/sta611`
- Homework (30%), Midterm Oct 25 (30%), Final Dec 11 (40%).
- First two homework assignments are posted on the website.
- Office hour: Tuesdays 1:00-2:00 pm
- TA: Jianyu Wang. Office hours: TBD

Make sure you have enough calculus background!

You need to be familiar with the following

- Function, limit, continuity
- Differentiation e.g. product rule, quotient rule and chain rule:

$$
\begin{aligned}
(f(x)g(x))' &= f'(x)g(x) + f(x)g'(x) \\
\left(\frac{f(x)}{g(x)}\right)' &= \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2} \\
(f(g(x))) &= g'(x)f'(g(x))
\end{aligned}
$$

- Integration, integration techniques such as integration by parts and by substitution,

$$
\begin{aligned}
\int f(x)g'(x)dx &= f(x)g(x) - \int f'(x)g(x)dx \\
\int_{g(a)}^{g(b)} f(x)dx &= \int_a^b f(g(t))g'(t)dt
\end{aligned}
$$

First assignment

- Function of two variables, double integrals, partial derivatives, infinite series, maxima and minima of functions.
  - E.g: $\int\int f(x,y)dxdy$ over an area like $0 < x < y < 1$.

First Homework: Take the math quiz

- Before Thursday, take the math quiz on the website to help you determine if you have the necessary mathematical background.
- Solutions are also on the website.
- The quiz will not be graded.

Course overview

- Basic probability theory, Random variables, Expectations, Probability distribution functions, Convergence.
- Principles of statistical inference, Likelihood based, Frequentist and Bayesian paradigms.
- Estimation, sampling distributions and hypothesis testing.

We will cover most of the book “Probability and Statistics” By DeGroot and Schervish (Fourth edition).

- I will use a combination of slides and whiteboard. Slides will be posted on the course website.

Statistics

Statistics according to dictionary.com:

The science that deals with the collection, classification, analysis, and interpretation of numerical facts or data, and that, by use of mathematical theories of probability, imposes order and regularity on aggregates of more or less disparate elements.

- Collection: Design of Experiments

Statistical Analysis

- Use **probability models** to describe uncertainty
- **Statistical inference**: What does the data tell us?

Statistics and Probability

Example

- **Probability model**: If we toss a fair coin 10 times, what is the probability that we will get 5 tails?
- **Statistical inference**: If we toss a coin 10 times and get 5 tails, what does that tell us about fairness of the coin?

- Probability is the foundation of statistics

Probability

- Is there such thing as true *randomness* or *stochasticity*?
  - Is the toss of a coin "random" or completely determined (by physics)?
- Probability models are very useful for modeling uncertainty, even though the phenomenon under study is not "truly random".
- Different interpretations of probability - no agreement on which is the "right one"
- But: There is complete agreement on the mathematical rules that probability must follow

Interpretation of Probability

Frequentist interpretation of Probability

Probability of an event is the *relative frequency* in which that event happens it the process where repeated many times under similar conditions.

- Problems with this view: How many times is needed? How similar do the conditions need to be? What about events that cannot be repeated?
- This view is sometimes called “classical statistics” and was the main stream view until recently.
- Often contributed to Jerzy Neyman and Egon Pearson and had a fierce advocate in Ronald A. Fisher.

Long run frequency ...

The statistics student tossed a coin throughout the Boston Marathon to study the behavior of a chance process in the long run.

Interpretation of Probability

Bayesian interpretation of Probability

Probability of an event is a degree of belief in whether the event will happen

- This is a subjective interpretation: Two people could have different probabilities for the same event
- Contributed to a 18th century Presbyterian Minister and mathematician Thomas Bayes.
- Avoids the philosophical difficulties of the frequency interpretation (e.g. repeatability).
- Bayesian statistics has been gaining popularity over the last few decades, largely due to advances in computing.

Remember: Regardless of interpretation, the mathematics of probability are the same - no controversy there.

Experiments / observational studies

Experiments (from book)

An experiment is any process were the possible outcomes can be identified ahead of time

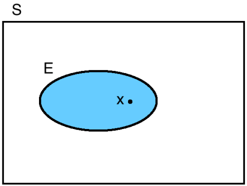
- Sometimes we distinguish between *experiments* and *observational studies*:
  - In experiments we assign treatments to objects
  - In observational studies we collect data but do not interfere

Examples of outcomes of experiments / observational studies:

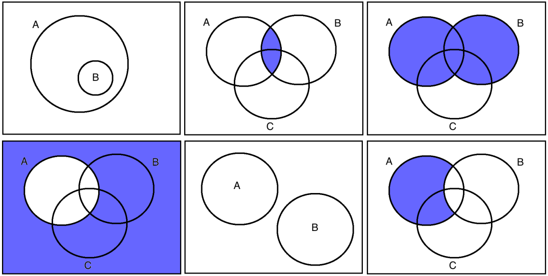
- Measurements from a laboratory experiment
- Results from a clinical trial
- Survival time of a cancer patient
- Collection of measurements from a sampled population

Set Theory – Notation

- Sample Space S: the collection (set) of possible outcomes of an experiment
  - Example: Rolling a die. $S = \{1, 2, 3, 4, 5, 6\}$
- Event E: a subset of S (a collection of outcomes)
  - Example: We roll an odd number on the die. $E = \{1, 3, 5\}$
- Outcome x: If the outcome x satisfies the conditions that define the event E we write $x \in E$
- Empty set: $\emptyset$



Interpretation of set operations



1. Subset $B \subset A$: Occurrence of B implies occurrence of A
2. Intersection $A \cap B$: Both events A and B occur
3. Union $A \cup B$: At least one of A or B occur
4. Complement A^c (or $\overline{A}$): the event A does not occur
5. A and B are disjoint if $A \cap B = \emptyset$. The two events cannot occur at the same time
6. A but not B

Some properties of set operations

- If $A \subset B$ and $B \subset A$ then $A = B$
- If $A \subset B$ and $B \subset C$ then $A \subset C$
- $(A^c)^c = A$ and $\Omega^c = \emptyset$
- Commutativity:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

- Associativity

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

Some properties of set operations

- Distributive Law:

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

- De Morgan's laws

$$(A \cup B)^c = A^c \cap B^c$$ prove this one...
$$(A \cap B)^c = A^c \cup B^c$$

- Union / intersection of many sets:

$$A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i$$
$$A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i$$

Example (1.10 from the book)

Three six-sided dice are rolled. We define the following events:

- $A =$ The first die shows an even number
- $B =$ The second die shows an even number
- $C =$ The third die shows an even number
- A_i, B_i and C_i: The first/second/third die shows the number i

Express each of the following events in terms of the named events above:

1. All three dice show even numbers
2. No die shows an even number
3. At least one die shows an odd number
4. At most two dise show odd numbers
5. The sum of the three dice is no greater than 5

Example: Typical weekend

In typical year, out of 52 weekends

- You have money exactly 39 weekends
- You have the weekend off 17 weekends
- You have somewhere to go 37 weekends
- You have money and the weekend off 7 weekends
- You have the weekend off and somewhere to go 7 weekends
- You have money but have to work and besides, you have nowhere to go 3 weekends
- You have the weekend off, you have money but nowhere to go 5 weekends

How many weekends a year do you actually end up going somewhere?

Typical Situation