MATH 450: Mathematical statistics

Vu Dinh

Departments of Mathematical Sciences
University of Delaware

August 28th, 2018

General information

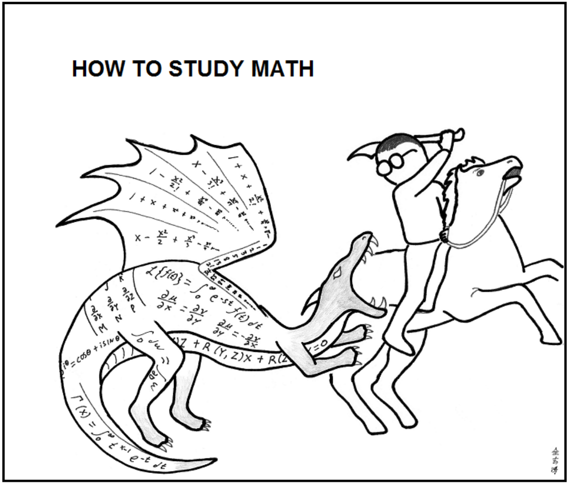
- Classes:

  Tuesday & Thursday 9:30-10:45 am, Gore Hall 115
- Office hours: Tuesday–Wednesday 1-2:30 pm, Ewing Hall 312
- Website: `http://vucdinh.github.io/m450f18`
- Textbook: *Modern mathematical statistics with applications* (Second Edition). Devore and Berk. Springer, 2012.

Evaluation

- Overall scores will be computed as follows:
  25% homework, 10% quizzes, 25% midterm, 40% final
- No letter grades will be given for homework, midterm, or final. Your letter grade for the course will be based on your overall score.
- The lowest homework scores and the lowest quiz score will be dropped.
- Here are the letter grades you can achieve according to your overall score.
  - $\geq 90\%$: At least A
  - $\geq 75\%$: At least B
  - $\geq 60\%$: At least C
  - $\geq 50\%$: At least D

Homework



Don't just read it; fight it!

--- Paul R. Halmos

Homework

- Assignments will be posted on the website every other Tuesday (starting from the second week) and will be due on Thursday of *the following week*, *at the beginning of* lecture.
- No late homework will be accepted.
- The lowest homework scores will be dropped in the calculation of your overall homework grade.

Quizzes and exams

- At the end of some chapter, there will be a short quiz during class.
- The quiz dates will be announced at least one class in advance.
- The lowest quiz score will be dropped.

There will be a midterm on 10/25 and a final exam during exams week.

Data analysis

Open source statistical system R

http://cran.r-project.org/

Tentative schedule

Week	Chapter	Note
1	1	
2	6.1 – 6.2	
3	6.2 – 6.3	HW1 (due 09/13)
4	7.1	
5	7.2	HW2 (due 09/27)
6	7.3 – 7.4	
7	8.1 – 8.2	HW3 (due 10/11)
8	8.3 – 8.5	
9	Review + Exam	Midterm exam (10/25)
10	9.1 – 9.2	HW4 (due 11/01)
11	9.2 – 9.3	
12	10.1	HW5 (due 11/15)
13		Thanksgiving week (no class)
14	11 – 12	HW6 (due 11/29)
15	12 + Review	
16		Final week

Topics

Textbook: *Modern mathematical statistics with applications* (Second Edition).
Devore and Berk. Springer, 2012.

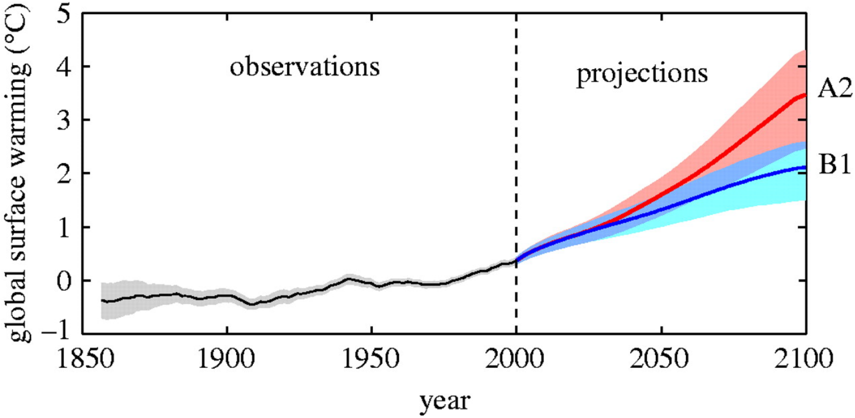
Week 1	*Chapter 6: Statistics and Sampling Distributions*
Week 4	Chapter 7: Point Estimation
Week 7	*Chapter 8: Confidence Intervals*
Week 10	*Chapter 9: Test of Hypothesis*
	Chapter 10: Two-sample inference
Week 14	Regression

Mathematical statistics

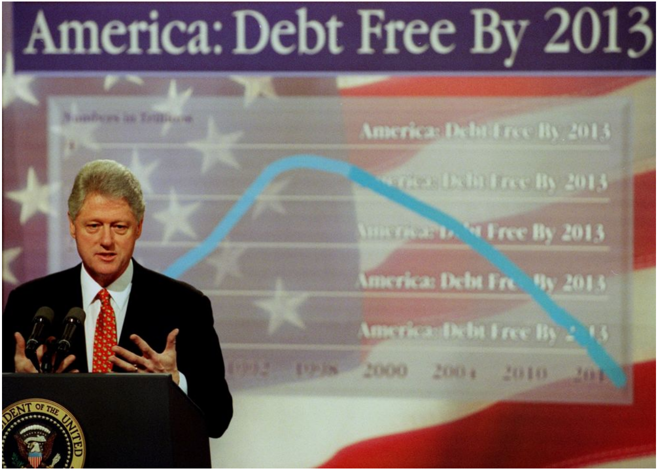
Mathematical statistics

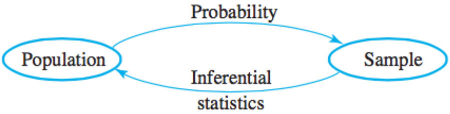
- Statistics ~~is a branch of mathematics that~~ deals with the collection, organization, analysis, interpretation and presentation of data
- "...analysis, interpretation and presentation of data" $\rightarrow$ mathematical statistics
    - descriptive statistics: the part of statistics that describes data
    - inferential statistics: the part of statistics that draws conclusions from data

— Modern statistics is about making prediction in the presence of uncertainties



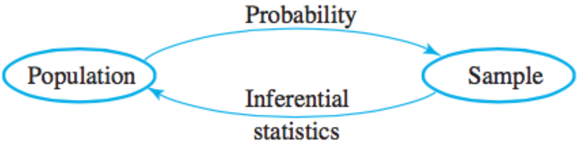
— It is difficult to make predictions, especially about the future.





- *population*: a well-defined collection of objects of interest
- when desired information is available for all objects in the population, we have what is called a *census*
  $\rightarrow$ very expensive
- a *sample*, a subset of the population, is selected

Random sample



Definition

The random variables $X_1, X_2, ..., X_n$ are said to form a (simple) random sample of size n if

1. the X_i's are independent random variables
2. every X_i has the same probability distribution

Review: Probability

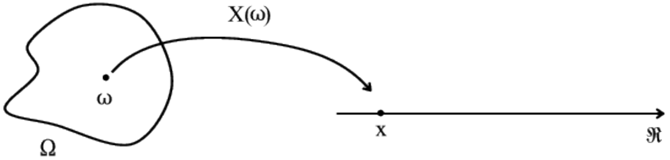
Overview

- Axioms of probability
- Conditional probability and independence
- *Random variables*
- *Special distributions*
- Bivariate and multivariate distributions

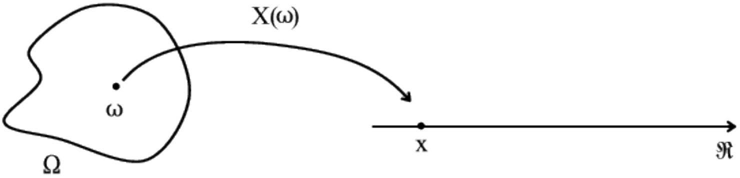
Most important parts

- Expectation and variance of random variables (discrete and continuous)
- Computations with normal distributions
- Bivariate and multivariate distributions

Random variables



- random variables are used to model uncertainties
- Notations:
  - random variables are denoted by uppercase letters (e.g., X);
  - the calculated/observed values of the random variables are denoted by lowercase letters (e.g., x)



Definition

Let S be the sample space of an experiment. A real-valued function $X : S \to \mathbb{R}$ is called a random variable of the experiment.

Discrete random variable

Definition

A random variables X is discrete if the set of all possible values of X

- is finite
- is countably infinite

Note: A set A is countably infinite if its elements can be put in one-to-one correspondence with the set of natural numbers, i.e, we can index the element of A as a sequence

$$A = \{x_1, x_2, \dots, x_n, \dots\}$$

A random variable X is described by its *probability mass function*

Definition *The* ***probability mass function*** *p of a random variable X whose set of possible values is $\{x_1, x_2, x_3, \ldots\}$ is a function from $\mathbf{R}$ to $\mathbf{R}$ that satisfies the following properties.*

(a) $p(x) = 0$ *if* $x \notin \{x_1, x_2, x_3, \ldots\}$.

(b) $p(x_i) = P(X = x_i)$ *and hence* $p(x_i) \geq 0$ $(i = 1, 2, 3, \ldots)$.

(c) $\sum_{i=1}^{\infty} p(x_i) = 1$.

Represent the probability mass function

- As a table

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

- As a function:

$$p(x) = \begin{cases} \frac{1}{2}\left(\frac{2}{3}\right)^x & \text{if } x = 1, 2, 3, \ldots, \\ 0 & \text{elsewhere} \end{cases}$$

Expectation

Definition *The* ***expected value*** *of a discrete random variable X with the set of possible values A and probability mass function $p(x)$ is defined by*

$$E(X) = \sum_{x \in A} x p(x).$$

We say that $E(X)$ exists if this sum converges absolutely.

The expected value of a random variable X is also called the **mean**, or the **mathematical expectation**, or simply the **expectation** of X. It is also occasionally denoted by $E[X]$, $E(X)$, EX, μ_X, or μ.

Exercise

Problem

A random variable X has the following pmf table

X	0	1	2
probability	0.25	0.5	0.25

What is the expected value of X?

Law of the unconscious statistician (LOTUS)

Theorem 4.2 *Let X be a discrete random variable with set of possible values A and probability mass function $p(x)$, and let g be a real-valued function. Then $g(X)$ is a random variable with*

$$E[g(X)] = \sum_{x \in A} g(x)p(x).$$

Exercise

Problem

A random variable X has the following pmf table

X	0	1	2
probability	0.25	0.5	0.25

- *What is $E[X^2 - X]$?*
- *Compute $Var[X]$*

$$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Continuous random variable

Definition

Let X be a random variable. Suppose that there exists a nonnegative real-valued function $f : \mathbb{R} \to [0, \infty)$ such that for any subset of real numbers A, we have

$$P(X \in A) = \int_A f(x)dx$$

Then X is called **absolutely continuous** or, for simplicity, **continuous**. The function f is called the **probability density function**, or simply the **density function** of X.

Whenever we say that X is continuous, we mean that it is absolutely continuous and hence satisfies the equation above.

Properties

Let X be a continuous r.v. with density function f, then

- $\int_{-\infty}^{\infty} f(x)\ dx = 1$
- For any fixed constant a, b,

$$P(a \leq X \leq b) = \int_a^b f(x)\ dx$$

$$P(X = a) = 0$$

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$$

Probability density function

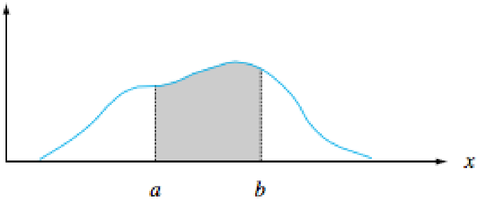


Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Expectation

Definition *If X is a continuous random variable with probability density function f, the* ***expected value*** *of X is defined by*

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx.$$

The expected value of X is also called the **mean**, or **mathematical expectation**, or simply the **expectation** of X, and as in the discrete case, sometimes it is denoted by EX, $E[X]$, μ, or μ_X.

Lotus

Theorem 6.3 *Let X be a continuous random variable with probability density function $f(x)$; then for any function $h : \mathbf{R} \to \mathbf{R}$,*

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x)\, dx.$$

Example

Problem

Let X be a continuous r.v. with density function

$$f(x) = \begin{cases} ce^{-2x} & \textit{if } x \geq 0 \\ 0 & \textit{otherwise} \end{cases}$$

where c is some unknown constant.

- *Compute c*
- *Compite $P(X \in [1, 2])$*
- *Compute $E[X]$ and $Var(X)$.*