STAT 450 Lecture 1
Course outline

Reading for Today's Lecture: Chapter 1 and Sections 1, 2 and 3 of Chapter 2 of Mood, Graybill and Boes.
Goals of Today's Lecture:

- Give an overview of the course.
- Define
  - Probability Space, sample space.
  - Random variables (in R^p)
  - The distribution of a random variable
  - Cumulative Distribution Function (in R^1)
  - Discrete and Absolutely Continuous Distribution
  - Probability Mass Function
  - Probability Density
- Introduce the problem of distribution theory.

Today's notes
Course outline:

- Distribution Theory. Chapters 1 through 6 of Mood, Graybill and Boes (but Chapter 3 will be regarded as review and not covered in class).
  - Basic concepts of probability.
  - Distributions
  - Expectation and moments
  - Moment generating functions
  - Distribution of transformations
- Point estimation (Chapter 7 in Mood, Graybill and Boes.)

  - Maximum likelihood estimation.
  - Method of moments.
  - Optimality Theory.
  - Bias, mean squared error.
  - Sufficiency.
  - Uniform Minimum Variance Unbiased Estimators.

- Hypothesis Testing (Cf. Chapter 9 in Mood, Graybill and Boes.)
  - Neyman Pearson optimality theory.
  - Most Powerful, Uniformly Most Powerful, Unbiased tests.

- Confidence sets (Cf. Chapter 8 in Mood, Graybill and Boes.)
  - Pivots
  - Associated Hypothesis Tests
  - Inversion of hypothesis tests to get confidence sets.

- Decision Theory. (Cf. Section 7 of Chapter 7 in Mood, Graybill and Boes.)

Statistics versus Probability

Standard view of scientific inference has a set of theories which make predictions about the outcomes of an experiment:

Theory	Prediction
A	1
B	2
C	3

If we conduct the experiment and see outcome 2 we **infer** that Theory B is correct (or at least that A and C are wrong).

Add **Randomness**

Theory	Prediction
A	Usually 1 sometimes 2 never 3
B	Usually 2 sometimes 1 never 3
C	Usually 3 sometimes 1 never 2

Now if we actually see outcome 2 we infer that Theory B is probably correct, that Theory A is probably not correct and that Theory C is wrong.

Probability Theory is concerned with constructing the table just given: computing the likely outcomes of experiments.

Statistics is concerned with the inverse process of using the table to draw inferences from the outcome of the experiment. How should we do it and how wrong are our inferences likely to be?

In this course we will be working on Probability for quite a while before we do Statistics.

Probability Definitions

I am going to begin the course with some very formal mathematical definitions. These definitions need not be memorized and I won't expect you to use them in the homework. I want you to see that the mathematics does not really define the idea of random; instead we just give computational rules which match, we think, our intuitive notion of what probability ought to mean.

Definition: A **Probability Space** is an ordered triple $(\Omega, \mathcal{F}, P)$. The idea is that Ω (called the **Sample Space**) is the set of possible outcomes of a random experiment, $\mathcal{F}$ is the set of those events, or subsets of Ω whose probability is defined and P is the rule for computing probabilities. [The book uses the jargon **Event Space** for $\mathcal{F}$ but this is not standard and I won't be using the term.] Formally:

- Ω is a set.
- $\mathcal{F}$ is a family of subsets of Ω with the property that $\mathcal{F}$ is a σ-field (or Borel field or σ-algebra):
  1. The empty set $\emptyset$ and Ω are members of $\mathcal{F}$.
  2. $\mathcal{F}$ is closed under complementation. That is, if A is in $\mathcal{F}$ (meaning $P(A)$ is defined) then $A^c = \{\omega \in \Omega : \omega \notin A\}$ is in $\mathcal{F}$ (because we want to be able to say $P(A^c) = 1 - P(A)$).
  3. If $A_1, A_2, \cdots$ are all in $\mathcal{F}$ then so is $A = \cup_{i=1}^{\infty} A_i$. ($A$ is the event that at least one of the A_i happens and we want to be sure that if each of the A_i has a probability then so does this event A.)

- P is a function whose domain is $\mathcal{F}$ and whose range is a subset of $[0,1]$ which satisfies the axioms for a probability:

  1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
  2. If $A_1, A_2, \cdots$ are **pairwise disjoint** (or **mutually exclusive**) (meaning for any $j \neq k$ $A_j \cap A_k = \emptyset$) then
  $$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$
  This property is called **countable additivity**.

These axioms guarantee that as we compute probabilities by the usual rules, including approximation of an event by a sequence of others we don't get caught in any logical contradictions. The symbol σ in the definition of a σ-field means we allow countably infinite unions and intersections. The book discusses the situation with only finite unions and intersections; those examples where we have finite additivity but not countable additivity are mathematical pathologies, in my view. I won't be discussing such things.
NOTE: This definition is included principally to make clear that the mathematics of probability theory is not mystical in any way. You may well feel that the definition completely fails to capture the spirit of "randomness". The definitions merely provide the rules for manipulating probabilities; they don't provide any intuition. There will be no problems in this course using the definitions just given.
Definition: A **real random variable** is a function X whose domain is Ω and whose range is in the real line R^1 with the property that

$$P(\{\omega \in \Omega; X(\omega) \le x\})$$

is defined for each $x \in R$ (that is, $\{\omega \in \Omega; X(\omega) \le x\}$ is in $\mathcal{F}$).
Notation: we will write $P(X \le x)$ for $P(\{\omega \in \Omega; X(\omega) \le x\})$

Idea: defined events in terms of numerical quantities determined by the outcome of a random experiment.
Definition: A **R^p- valued random variable** X is just $X = (X_1, \ldots, X_k)$ where each X_i is a real valued random variable.
Definition: The **Cumulative Distribution Function** (or CDF) of a real random variable X is given by

$$F_X(x) = P(X \le x)$$

Definition: The **Cumulative Distribution Function** (or CDF) of an R^p valued random variable X is the function F_X on R^p defined by

$$F_X(x_1, \ldots, x_p) = P(X_1 \le x_1, \ldots, X_p \le x_p)$$

In this course we will usually study F only for $p = 1$.

Properties of F_X (or just F when there's only one CDF under consideration) in the case $p = 1$:

1. $0 \le F(x) \le 1$.
2. $x > y \Rightarrow F(x) \ge F(y)$ (F is monotone non-decreasing).
3. $\lim_{x \to -\infty} F(x) = 0$
4. $\lim_{x \to \infty} F(x) = 1$
5. $\lim_{x \searrow y} F(x) = F(y)$ (F is right continuous).
6. $\lim_{x \nearrow y} F(x) \equiv F(y-)$ exists (the jargon is that F has left limits).
7. $F(x) - F(x-) = P(X = x)$.
8. $F_X(t) = F_Y(t)$ for all t implies that X and Y have the same distribution, that is, $P(X \in A) = P(Y \in A)$ for any (Borel) set A.

As an illustration of the role of the axioms of probability I will prove the second and fourth of these assertions.

Proof of 2: Let $A = \{\omega \in \Omega : X(\omega) \le y\}$ and $B = \{\omega \in \Omega : X(\omega) \le x\}$. Let $C = \{\omega \in \Omega : y < X(\omega) \le x\}$. Then an individual ω in B must be in either A or C and an ω in either A or C will be in B so $B = A \cup C$. If ω belongs to C then it does not belong to A and vice-versa; that is, $A \cap C = \emptyset$. The second axiom of probability tells us that

$$P(B) = P(A) + P(C) .$$

Now you have to recognize from the definition that $P(B) = F(x)$ and $P(A) = F(y)$ so

$$F(x) = F(y) + P(C) .$$

Since $P(C) \ge 0$ we get $F(x) \ge F(y)$ which ends the proof. •

Proof of 4: Here we are assuming that X has values in R; that is we are assuming that $P(-\infty < X < \infty) = 1$. Let $A_n = \{\omega \in \Omega : X(\omega) \leq n\}$ for $n = 0, 1, \ldots$ and define

$$B_n = A_n \setminus A_{n-1} = \{\omega \in \Omega : n - 1 < X(\omega) \leq n\}$$

for $n = 1, 2, \ldots.$ Let $B_0 = A_0$. Then we have

$$\{-\infty < X < \infty\} = \cup_{n=0}^{\infty} B_n\,.$$

Moreover, the B_n are pairwise disjoint so

$$\begin{aligned}
1 &= P(-\infty < X < \infty) \\
&= P(\cup_{n=0}^{\infty} B_n) \\
&= \sum_{n=0}^{\infty} P(B_n) \\
&= \lim_{m\to\infty} \sum_{n=0}^{m} P(B_n) \\
&= \lim_{m\to\infty} P(\cup_{n=0}^{m} B_n) \\
&= \lim_{m\to\infty} P(A_m) \\
&= \lim_{m\to\infty} F(m)
\end{aligned}$$

You should make sure you understand why each line is correct. The first is our assumption that X is real valued. The second rewrites the event $-\infty < X < \infty$ as I did above. The third line uses the second axiom in the definition of probability. The fourth line uses the fact that the definition of an infinite sum is that it is the limit of partial sums. The fifth line uses the second axiom for probabilities again but with a finite union. The sixth line recognizes that if $X(\omega) \leq m$ then either $X(\omega) \leq 0$ or $X(\omega)$ is between some pair of integers $n - 1$ and n with $1 \leq n \leq m$. The last line is just the definition of $F(m)$. •

Definition: The distribution of a random variable X is **discrete** (we also call the random variable discrete) if there is a countable set $x_1, x_2, \cdots$ such that

$$P(X \in \{x_1, x_2 \cdots\}) = 1 = \sum_i P(X = x_i)$$

In this case the **discrete density** or **probability mass function** of X is

$$f_X(x) = P(X = x)$$

The distribution of a random variable X is **absolutely continuous** if there is a function f such that

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

for any set A. This is a p dimensional integral in general. This condition is equivalent (for $p = 1$) to

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

We call f the **density** of X. For most values of x we then have F is differentiable at x and (for $p = 1$)

$$F'(x) = f(x)\,.$$

Notation: Some students will not be used to the notation $\int_A f(x)dx$ for a multiple integral. For instance if $p = 2$ and the set A is the disk of radius 1 centered at the origin then

$$\int_A f(x)dx \equiv \int_{-1}^{1} \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} f(x_1, x_2)\, dx_2\, dx_1$$

Notation: X is exponential.

$$F(x) = \begin{cases} 1 - e^{-x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ \text{undefined} & x = 0 \\ 0 & x < 0 \end{cases}$$

Notation: The function

$$f(u, v) = \begin{cases} u + v & 0 < u, v < 1 \\ 0 & \text{otherwise} \end{cases}$$

is a density. The corresponding cdf is

$$F(x, y) = \begin{cases} 0 & x \le 0 \text{ or } y \le 0 \\ xy(x + y)/2 & 0 < x, y < 1 \\ x(1 + x)/2 & y \ge 1 \text{ and } 0 < x < 1 \\ y(1 + y)/2 & x \ge 1 \text{ and } 0 < y < 1 \\ 1 & x, y \ge 1 \end{cases}$$