Stat 110

Notes by Edward Chen

October 2022

Preface

These are my notes to Harvard's Stat 110 class, taught by Joseph Blitzstein and Jessica Hwang. The class covers all the basics of probability–counting principles, probabilistic events, random variables, distributions, conditional probability, expectation, and Bayesian inference.

Compared to other notes, this should not be used as a whole subsititue for the class, rather as an indication of the more complex topics covered. Put harsher, I will likely spend less time note taking on things I already know and more on things I'm confused out. For clarity's sake, I am pretty comfortable with probability from my competition math background.

Lecture videos are freely available at

1 Chapter 1

Probability gives us a logical framework to view and anlyze uncertaintiy. It is the foundation and language for statistic ass well as the basis for topics such as Statistics, Physics, Biology, and Computer science.

A **sample space** S is the set of all theoretically possible outcomes.

Let D be the set of all coin flips with at least two consecutive heads. The sameple space, expressed as a set would be:

$$D = \cup_{j=1}^{9}(A_j \cap A_{j+1})$$

Some more set notation:
sameple space: S
s is a possible outcome: $s \in S$
A is an event: $A17S$
A implies B:

A and B are mutually exclusive: $A \cap B = \emptyset$

Example 1.4.10: Birthday Problem
Theorem 1.4.15 (Binomial coefficient formula). For $k \leq n$, we have:

$$\binom{n}{k} = \frac{n(n-1)...(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$$

Some proof methods covered in the chapter include complementary counting, stars and bars, and story proofs(1.5).

There are a lot of formulas with binomial coefficients:

$$\binom{n}{k} = \binom{n-k}{k}$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$n\binom{n-1}{k-1} = k\binom{n}{k}$$

$$\text{Vandermonde's}: \binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}$$

Example 1.5.4(Partnerships). Let's prove

$$\frac{(2n)!}{2^n \cdot n!} = (2n-1)(2n-3)...3 \cdot 1$$

Story proof:: We will show that both sides count the number of ways to break $2n$ people into n partnerships. Take $2n$ people, and give them ID numbers from 1 to $2n$. We can form partnerships by lining up the people in some order and then saying the first two are a pair, the next two are a pair, etc. This overcounts by a factor of $n! \cdot 2^n$ since the order of pairs doesn't matter, nor does the order within each pair. Alternaitvely, count the number of possibilities by noting that there are $2n-1$ choices for the parnter of person 1, then $2n-3$ choices for person 2, and so on.

Definition 1.6.1 (General definition of probability). A *probability space* consists of a sample space S and a *probability function* P and returns a real number between 0 and 1, $P(A)$, where A is the event it takes in.

Unlike the naive definition, here we can have events with different probabilities.

The *frequentist* view of probability is that it represents a long-run frequency over a large number of repetitions of an experiment. The *Bayesian view* is that it represents a degree of belief about the event in question, so we can assign probabilities to hypotheses.

Inclusion-exclusion example: With a triple venn diagram, we can write:

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

Generally, we can write:

Example 1.6.4 (de Montmort's matching problem). Consider a shuffled deck of n cards, from 1 to n. You flip each one over one by one. What is the probability the ith index card you turn over has value i?

With PIE, we get

$$P(\cup_{i=1}^{n} A_i) = \frac{n}{n} - \frac{\binom{n}{2}}{n(n-1)} + \frac{\binom{n}{3}}{n(n-1)(n-2)} - ... + (-1)^{n+1} \cdot \frac{1}{n!}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - ... + (-1)^{n+1} \cdot \frac{1}{n!}$$

With large n, this approaches the Taylor series for $\frac{1}{e}$:

$$e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + ...$$

R: Please read section 1.8 for an introduction to R. R allows us to simulate and deal with large sets of data.

2 Chapter 2: Conditional Probability

Conditional Probability explores how we should update our probability when we receive new information.

Definition 2.2.1. If A and B are events with $P(B) > 0$, then conditional probability can be expressed as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example 2.2.5(Two Children): Martin Gardner posed the following puzzle:
Mr. Jones has two children. The older hild is a girl. What is the probability that both children are girls?
Mr. Smith has two children. At least one of them is aboy. What is the probability that both children are boys?

The intuition is that they should both be $\frac{1}{2}$, but the respesctive probabilities are actually $\frac{1}{2}$ and $\frac{1}{3}$.

Theorem 2.23 (Bayes' rule).

$$P(A|B) = \frac{P(A \cap B)}{B} = \frac{P(B|A)P(A)}{P(B)}$$

Definition 2.3.4 (Odds). The *odds* of an event A are

$$\text{odds(A)} = P(A)/P(A^c)$$

. Ex.: if $P(A) = 2/3$, we say the odds in favor of A are 2 to 1.

Theorem 2.3.5 (Odds form of Bayes' rule)

Example 2.4.5 (Unanimous agreement). The artaicle "Why too much evidence can be a bad thing" by Lisa Zyga says:

Under ancient Jewish law, if a suspect on trial was unanimously found guilty by all judges, then the suspe

This is because a systemic error occurs independet of the conviction, so there may be an unseen bias at play here.

Independence

Events A and B are *independent* if $P(A \cap B) = P(A)P(B)$.

For infinitely many events, we say that they are independent if every finite subset of the events is independent.

Definition 2.5.7 (Conditional independence). Events A and B are said to be *conditionally independent* given E if $P(A \cap B|E) = P(A|E)P(B|E)$.

> It is easy to make terrible blunders stemming from confusing independence and conditional independence. Two events can be conditionally independence given E, but not independent given E^c. Two events can be conditionally independent given E, but not independent. Two events can be independent, but not conditionally independent given E. Great care is needed

todo: 65-67

An important property of Bayes' rule is that is is *coherent*: for the same updates, no matter the update size, the final state should be the same.

todo: 68-75