Stats 110 Notes

Daniel Flannery

July 28, 2015

Contents

1 Lecture 1: Probability and Counting

1.1 Notes

- A **Sample Space** is the set of all possible outcomes of an experiment
- An **event** is a subset of the sample space.
- **Naive Definition of Probability**:

$$P(A) = \frac{\#\ of\ favorable\ outcomes}{\#\ possible\ outcomes}$$

- **The Multiplication Rule** - if you have experiment with n possible outcomes, and for each outcome of the first experiment there are n outcomes for the second experiments, ..., for each r^{th} experiment there are n_r outcomes, then there are $n_1 \cdot n_2 \cdots n_r$ overall possible outcomes. Assumes all outcomes are equally likely within a finite sample space

- **Binomial Coefficient**:
$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$
0 if $k > n$

  How do you get there? Number of subsets of size k of group n. Choosing k from n in order: $n \cdot (n-1)(n-2) \cdot ... \cdot (n-k+1)$ Now if order does not matter: $\frac{n \cdot (n-1)(n-2) \cdot ... \cdot (n-k+1)}{k!}$ Now through canceling we reach the formula for the binomial coefficient.

1.2 Example: Full House, 5 Card Hand

There are 13 suits. We must choose one suit then 3 of the 4 possible cards. Then we must choose another suit of the remaining 12, then select 2 of the 4 possible cards. Then divide by the sample space which is all possible 5 card hands from a deck of 52 cards.

$$\frac{13\binom{4}{3} \cdot 12\binom{4}{2}}{\binom{52}{5}}$$

1.3 Sampling Table

Sampling Table		
	Order Matters	**Order Doesn't Matter**
Replace	n^k	$\binom{n+k-1}{k}$
Don't Replace	$n(n-1)(n-2)\cdots(n-k+1)$	$\binom{n}{k}$

2 Lecture 2: Story Proofs, Axioms of Probability

2.1 Notes

- Classic and Fundamental Relationship:

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

- Exploring picking k times from set of n objects with replacement where order doesn't matter. Let's try out a few cases:

  Exploration

  - Case 1: $k = 0 \rightarrow \binom{n-1}{0} = 1$
  - Case 2: $k = 1 \rightarrow \binom{n}{1} = n$
  - Case 3 (Simplest non-trivial example): $n = 2 \rightarrow \binom{k+1}{k} = \binom{k+1}{1} = k + 1$
  - This is equivalent to the number of ways to put k indistinguishable particles into n distinguishable boxes, which can be visualized as such:

    $$\cdots || \cdot\cdot | \cdot$$

    The $|$ represent partitions for the 'boxes' and the $\cdot$ is a particle. There are then k $\cdot$'s and $n-1$ $|$'s. This can be simplified again to be thought of just arranging symbols, and you are choosing to place the symbols. There are $k+n-1 = n+k-1$ symbols, so choose where to place the separators or the particles. Choose where the $\cdot$'s are, and the $|$'s are then positioned, and vice versa.

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

- **Story Proof**: proof by interpretation.

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

    Pick k people out of n with designated as the president, so you can first pick the president out of n, then select who can be in the club. Or you can pick the club k out of n then choose the president from the k people.

- **Vandermonde Identity**:

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}$$

You have two groups with labeled people containing m and n people. You need to select k people total from both groups. The summation refers to the combinations of selecting j from m and $k-j$ from n such that you have $j+k-j = k$ total people from the $m+n$ total people in two groups. Each instance is multiplied together due to the multiplication rule.

- **The Non-Naive Definition of Probability**
  - A probability sample consists of **S** and **P** where **S** is a sample space and **P**, a function which takes an event $A \subseteq S$ as input and returns $P(A) \in [0,1]$ as output such that:
    - (1) $P(\emptyset) = 0$, $P(S) = 1$
    - (2)

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$$

if $A_1, A_2, ..., A_n$ are disjoint.

3 Lecture 3: Birthday Problem, Properties Probability

3.1 Notes

- **Birthday Problem**: k people, find the probability that 2 have the same birthday. Exclude February 29, assume other 365 days are equally likely. If $k > 365$, the probability is 1 by the pigeonhole principle. If $k \leq 365$, the probability of no match:

$$P(M^c) = \frac{365 \cdot 364 \cdot 363 \cdots (365-k+1)}{365^k}$$

The probability of a match is: $1 - P(M^c)$

$$P(M) = 1 - P(M^c) = 1 - \frac{365 \cdot 364 \cdot 363 \cdots (365-k+1)}{365^k}$$

Imagine pairing up the people. There are $\binom{k}{2}$ pairs.

- **Axioms**:

1. $P(\emptyset) = 0$, $P(S) = 1$

2. $P(\cup_{n=1}^{\infty} = \sum_{n=1}{}^{\infty} P(A_n)$ if A_i are disjoint.

- **Properties**:
  1. $P(A^c) = 1 - P(A)$ Proof: $1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$ because $A \cap A^c = \emptyset$
  2. If $A \subseteq B$, then $P(A) \leq P(B)$. Proof: $B = A \cup (B \cap A^c)$, disjoint, then $P(B) = P(A) + P(B \cap A^c) \geq P(A)$
  3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Proof: $P(A \cup B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c)$

  $$? = P(A) + P(B) - P(A \cap B), P(A \cap B) + P(B \cap A^c) = P(B)$$

  $A \cap B$ and $A^c \cap B$ are disjoint and the union is B

- **Inclusion-Exclusion**: Expanding upon most recent item. Example with 3 sets: A, B, C.

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

  Generalized:

  $$P(A_1 \cup A_2 \cup \cdots \cup A_n)$$

  $$= \sum_{j=1}^{n} P(A_j) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n)$$

- **de Montmort's Matching Problem** (link). What is the probability that the j^{th} card of n shuffled cards has the label j?

  $$= 1/e$$

4 Lecture 4: Conditional Probability

4.1 Notes

- Independence
- $P(A \cap B) = P(A) \cdot P(B), P(B) > 0$
- Events A, B, C are independent if they are pairwise independent and completely independent:
  - $P(A \cap B) = P(A) \cdot P(B)$ and $P(A \cap C) = P(A) \cdot P(C)$ and $P(B \cap C) = P(B) \cdot P(C)$ and $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$
- Binomial Probability: Newton-Pepys (1693)
  - What is more likely? At least one six when rolling 6 dice, at least 2 sixes when rolling 12 dice, or at least 3 sixes when rolling 18 dice?
  - For at least 1 in 6, the probability is 1 - the probability of no six.

$$P(A) = 1 - \left(\frac{5}{6}\right)^6$$

  - For at least 2 in 12, the probability is 1 - the probability of no six - the probability of one six

$$P(B) = 1 - \left(\frac{5}{6}\right)^6 - \binom{12}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{11}$$

  - For at least 3 in 18, the probability is 1 - the probability of no six - the probability of one six - the probability of two sixes

$$P(C) = 1 - \left(\frac{5}{6}\right)^6 - \binom{12}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{15} - \binom{12}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{16}$$
$$= 1 - \sum_{k=0}^{2}\binom{18}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{18-k}$$

- Conditional Probability
  -

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

  - Pebble World: Remove B^c and renormalize to 1
  - Frequentist: Circle reps where B occurred, find the fraction of those where A occurred.

– Theorems:

1. $$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

2. $$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

3. Bayes Rule
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

5 Lecture 5: Conditioning Continued, Law of Total Probability

5.1 Notes

- How to Solve a Problem:
  1. Try simple and extreme cases
  2. Law of Total Probability: Divide and conquer - break up problem into simpler pieces. (Partition - disjoint and union is all of sample space S)

$$P(B) = P(B \cap A_1) + \cdots + P(B \cap A_n)$$

$$= P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)$$

- Conditional probability can be used to break a problem up into simpler pieces
- Example: random 2 card hand from standard deck
  - Find $P(\text{both aces}|\text{have ace})$. You have at least one ace.

$$\frac{P(\text{both aces})}{P(\text{have ace}} = \frac{\binom{4}{2}/\binom{52}{2}}{1 - \binom{48}{2}/\binom{52}{2}}$$

  - Find $P(\text{both aces}|\text{have Ace of Spades})$. You have a specific ace.

$$= \frac{3}{51}$$

- How does more information change the probability?
- Disease testing - good test on what specifying what your goal is.
  - Disease afflicts 1% of population, suppose patient tests positive, means the test is asserting that the patient has the disease. Suppose the test is advertised as 95% accurate.
  - $P(D) = .01$
  - **D: Patient has disease**: $P(T|D) = 0.95 = P(T^c|D^c)$
  - **T: Patient tests positive**: $P(D|T) = \frac{P(T|D)P(D)}{P(T)}$
  - Expand the denominator using the law of total probability because the denominator is usually the tricky part:

$$\frac{P(T|D) \cdot P(D)}{P(T|D) \cdot P(D) + P(T|D^c) \cdot P(D^c)} = .16$$

- Probability of having the disease is rare already.

- Common Mistakes with Conditional Probability
  1. Prosecutor's Fallacy - confusing $P(A|B)$ and $P(B|A)$ - focusing on probability of innocence given evidence, focus on evidence given innocence - sad case of Sally Clark and SIDS
  2. Confusing $P(A)$ prior with $P(A|B)$ posterior
  3. Confusing independence with conditional independence
  4. Conditional Independence: Events A, B are conditionally independent given C if:
  $$P(A \cap B \cap C) = P(A|C) \cdot P(B|C)$$
  5. Does conditional independence given C imply independence? No. Example: chess opponent of unknown strength - earlier games given understanding of how strong opponent is, may be that game outcomes are conditionally independent given strength of opponent but not independent unconditionally. Earlier trials give data that can be used - conditioning.
  6. Does (unconditional) independence imply conditional independence given C? No. Common for an event with multiple causes. Counterexample: Fire alarm goes off, with two causes: (1.) fire, (2.) student making popcorn. Suppose then (1.) (2.) are independent. But $P(F|A, C^c) = 1$ not conditionally independent given alarm.