CS3203 #6

6/14/04
Janak J Parekh

Administrivia

- Exam today
  - Second half
  - Any last-minute questions?

Discrete probability

- Extension of combinatorics we discussed last time; theory developed by Pascal 300 years ago
- Studied in many fields, such as genetics, gambling, etc.
- Probability theory used in CS for average-case complexity, nondeterministic algorithms, etc.

Definition of probability

- Laplace: The *probability* of an event *E*, which is a subset of a finite sample space *S* of equally likely outcomes, is p(E) = |E| / |S|.
  - Use combinatorics!
  - What's p(E) where E is the fact that a password begins with "c"?
    - Capitals/lowercase? Numbers? Symbols?
  - Find p(F) where F is the event that the password contains no vowels.
    - Let's assume alphabetic characters only.

Combining events

- Probability of a *complementary* event is $p(\overline{E}) = 1 - p(E)$
  - A sequence of 10 bits is randomly generated. What is the probability that at least one of these bits is zero?
  - Do the opposite, and subtract.
  - 1023/1024
- Union of two events:
  $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$
  - If they're disjoint events, then you don't need to worry about the third term

Examples

- You pick five numbers, without replacement, from {1, 2, 3, …, 25}.
  - What is the probability that the sum of numbers chosen is odd?
    - If we don't care about ordering of numbers, then $p(E \cup F \cup G)$ where E is 1 odd chosen, F is 3 odd chosen, and G is 5 odd chosen. Since disjoint, add $p(E) + p(F) + p(G) = C(13,1)C(12,4)/C(25,5) + C(13,3)C(12,2)/C(25,5) + C(13,5)C(12,0)/C(25,5) \sim .501$
    - If we care about ordering, then $p(E) + p(F) + p(G) = C(5,1)P(13,1)P(12,4) / P(25,5) + C(5,3)P(13,3)P(12,2) / P(25,5) + C(5,5)P(13,5) / P(25,5) \sim .501$
  - What is the probability that the product of the numbers chosen is odd?
    - Only happens if every number is odd, so $C(13,5)/C(25,5) \sim .024$.
    - If ordered, $P(13,5)/P(25,5) \sim .024$.

Probabilistic reasoning

- Often, you want to determine which of two events is more likely
- Monty Hall Three Door Puzzle
  - Is it better to switch or not?
  - P(not changing the door) = 1/3, and *does not change* if the host opens the other door
  - Therefore, P(incorrect) = 2/3, and *that* doesn't change either. Since the host always opens an incorrect door, if you were incorrect in the first place, you will be guaranteed by switching. Therefore, the probability of winning when you switch is 2/3.
  - Weird, huh?

But…

- What if each outcome *isn't* equally likely?
  - Biased coin
- Need to *assign* probabilities
- If S is the sample space of an experiment with a finite/countable number of outcomes, and p(s) a probability for each outcome *s*, we require that
  - 0 <= p(s) <= 1 for each s in S;
  - Sum of all p(s), s in S = 1.
- Generalization of Laplace's definition
- Function p from set of all events in S is a **probability distribution**

Distributions and examples

- For a fair coin, what should we assign probabilities?
  - *Uniform distribution* assigns probability 1/n to each element of S
  - Probability of an event E is now the sum of the probabilities of the outcomes in E, e.g., p(E) = sum of p(s), s in E.
- If the coin is biased such that it shows tails twice as often as heads?
- For a loaded die such that 3 appears twice as often as each other number, what's the probability that an odd number appears when we roll this die?

Combinations of events

- Definition of complement remains the same
- E and E-bar now partition S such that p(E) + p(E-bar) still = 1
- Also remains same for the probability of combined events
- Theorem 1, page 366, holds for *disjoint* events

Conditional probability

- Let's say that F means that the first flip of three flips comes out tails, and that E is that an odd number of tails over the three flips.
  - We have four possibilities, right? TTT, TTH, THT, THH, and two of them would leave us with a total odd (TT and HH), so probability of F *given* E is 2/4
- Conditional probability $p(E|F) = p(E \cap F) / P(F)$.

Example

- Probability that a family with two children has two boys, given they have at least one?
  - E = “has two boys”, while F = “two children with at least one boy”
  - E = {BB}, F = {BB, BG, GB}, so $E \cap F$ = {BB}; p(F) = ¾ and p(E) = ¼, so 1/4/3/4 = 1/3.
- Can often do by inspection
  - You draw 23 cards, one at a time without replacement, at random from a deck of 52 cards. Find
  - p(second card is a Jack | first card is a Jack).
    - If first is a Jack, then there are 3 Jacks left, so 3/51.
  - p(second card is red | first card is black).
    - Still 26 cards left that are read, so 26/51

Independence

- Does knowing that a first flip come up tails have any implication on the likelihood of tails or heads for future flips? In other words, does $p(E|F) = p(E)$? For coins, yes; $p(E|F) = p(E) = ½$, so these events are **independent**
- Events E and F are independent if and only if $p(E \cap F) = p(E)p(F)$.
- You write a string of letters of length 3 from the usual alphabet, with no repeated letters allowed. Let E_1 be the event that the string begins with a vowel and E_2 be the event that the string ends with a vowel. Determine whether E_1 and E_2 are independent.
  – Sample space size is 26*25*24.
  – $|E_1| = 5*25*24$. $|E_2| = 25*24*5$.
  – $p(E_1) * p(E_2) = 5/26 * 5/26$
  – $E_1 \cap E_2$ = strings of length three where first and last are vowels, which is 5*24*4, so $p(E_1 \cap E_2) = 5*24*4 / 26*25*24$ or 2/65.
  – Therefore, **not** independent.
  – Can often look at the problem abstractly
  – What **would** be independent?

Bernoulli trial/binomial distribution

- For experiments with only two possible outcomes, e.g., coin flips, each performance is called a **Bernoulli trial**
- "Success" vs. "Failure"
- The probability of exactly *k* successes in *n* independent Bernoulli trials, with probability of success *p* and failure $q = 1-p$, is $\mathbf{C(n,k)p^k q^{n-k}}$.
  - "Binomial distribution", or b(k; n, p)
  - A fair coin is flipped five times. Find the probability of obtaining exactly four heads.
  - $C(5,4)(0.5)^4(0.5)^1 = 0.156$
- A die is rolled six times in a row. Find
  - P(exactly four 1s are rolled).
  - P(no 6s are rolled).

Random variables

- If we make a function that maps from the sample space of an experiment to the set of real numbers, we have a **random variable**. In other words, a random variable assigns a real number to each possible outcome.
  - Sadly, random variables are **not** random, **nor** a variable. ☹
  - Example: map three coin flips, such that X(t) is the number of heads
  - Useful way of “counting” for probability’s sake
- The **distribution** of a random variable X on a sample space S is the set of pairs (r, p(X = r)) for all $r \in X(S)$, where p(X = r) is the probability that X takes the value r.
  - Described by specifying each probability for each r.
  - For example, P(X = 3) = 1/8, P(X = 2) = 3/8, P(X = 1) = 3/8, P(X = 0) = 1/8
  - Used in *expected values*, which we may or may not get to.

Birthday Paradox

- How many people are needed in the room such that it's more likely than not (e.g., greater than .5 probability) that two people have the same birthday?
  - We assume that birthdays are independent, equally likely, and 366 birthdays per year.
  - If p_n = probability have all different birthdays, then $1 - p_n$ = probability two peopl ehave the same birthday
  - Compute probability has a different birthday as people "walk in the room".
  - First person $p_1 = 1$, second is 365/366, third is 364, 366, etc.
  - p_n is therefore 1 * 365/366 * 364/366 * 363/366 * … * 367-n/366, and $1 - p_n$ is 1 – same thing.
  - Use formula for $1 - p_n$ until it becomes greater than ½, and we have our value n. $1 - p_n \sim 0.475$ for n = 22, $1 - p_n \sim 0.506$ for n = 23.
- Should we try for months in this room?

Next time

- Finish up probability
- Recurrence relations