CSE 312

Course Overview

Counting

- Permutations, Combination. Taking order into consideration or not?
- Inclusion-Exclusion Principle. Computing the size of the union of complicated sets.
- Pigeonhole Principle

Basic Probability

- Conditional Probability
- Random Variables
- Bayes' Rule: $$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}$$

Linearity of Expectation

- E[X+Y] = E[X] + E[Y], even when X,Y are dependent!
- Many applications and uses in problem solving. Often indicator random variables are useful here.

Variance

- Intuition: measures typical deviation from the mean.
- Several formulas for computing it: $E[(X - E[X])^2]$, $E[X^2] - E[X]^2$
- If X,Y independent, then Var(X+Y) = Var(X) + Var(Y)

Concentration Inequalities

- Markov's Inequality. Works for non-negative random variables and only needs a bound on expectation.
- Chebyshev's Inequality. Works for all random variables and needs a bound on variance.
- Chernoff Bounds. Exponentially small bounds for tails of sums of independent random variables. In class we only discussed the case of sums of Bernoulli random variables. Applications like polling and load balancing.

Several Distributions

- Bernoulli: coin flip
- Binomial: # of heads in n coin flips
- Geometric: # flips to see first heads
- Poisson process: Good for modeling arrivals. Limit of Binomial for large n, while keeping expected number of heads the same.
- Exponential: Distribution of first arrival in Poisson.
- Normal distribution: very important because of CLT.

Poisson

- Good for modeling arrivals at traffic lights etc
- Limit of binomial
- Good for approximating binomial when n is large and expected number of heads is small.

Continuous Random Variables

- PDFs, CDFs, going back and forth between them using integrals and derivatives
- Calculating the probabilities of events
- Calculating a new pdf/cdf from an old pdf/cdf.

Normal Distribution

- Pdf has formula, no formula for CDF, but you will get a table if you need it.

Central Limit Theorem

- The sum independent random variables converges to a normal (after appropriate normalization of parameters)
- This is the reason why Normal distribution is so important
- CLT approximation: If we have the sum of many independent samples and we know the mean variance, you can use normal to approximate the sum.

Maximum Likelihood Estimators

- A method to use samples from a distribution to estimate the parameters of the distribution: pick the params that maximize the “likelihood” of seeing the given samples.

- Unbiased estimators: An estimator whose expectation is the parameter we seek.

Randomized Algorithms

- Min-cut
- Max-cut
- Primality testing
- NO RANDOMIZED ALGORITHMS ON THE FINAL

Where will you use this stuff?

- Machine learning / AI
- Distributed computing
- Algorithm design
- Complexity theory
- Basic math that is useful in many, many other places: e.g. stock market, polling, survey design etc etc.

The Final

- Monday, 2:30 - 4:20 in THO 101
- One cheat sheet, calculator allowed. No other materials allowed.
- Standard normal table will be included.
- Good luck!

Review Session with Joshua

- In EEB 045, from 2:30-4:20 after this class.