CSE 312 Final Review: Section AA

CSE 312 TAs

December 8, 2011

General Information

General Information

- Comprehensive Midterm

General Information

- Comprehensive Midterm
- Heavily weighted toward material after the midterm

Pre-Midterm Material

Pre-Midterm Material

- Basic Counting Principles

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle
  - Inclusion Exclusion

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle
  - Inclusion Exclusion
  - Counting the Complement

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle
  - Inclusion Exclusion
  - Counting the Complement
  - Using symmetry

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle
  - Inclusion Exclusion
  - Counting the Complement
  - Using symmetry
- Conditional Probability

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle
  - Inclusion Exclusion
  - Counting the Complement
  - Using symmetry
- Conditional Probability
  - $P(A \mid B) = \frac{P(AB)}{P(B)}$

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle
  - Inclusion Exclusion
  - Counting the Complement
  - Using symmetry
- Conditional Probability
  - $P(A \mid B) = \frac{P(AB)}{P(B)}$
  - Law of Total Probability:

$$P(A) = P(A \mid B) \cdot P(B) + P(A \mid \overline{B}) \cdot P(\overline{B})$$

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle
  - Inclusion Exclusion
  - Counting the Complement
  - Using symmetry
- Conditional Probability
  - $P(A \mid B) = \frac{P(AB)}{P(B)}$
  - Law of Total Probability:

$$P(A) = P(A \mid B) \cdot P(B) + P(A \mid \overline{B}) \cdot P(\overline{B})$$

  - Bayes' Theorem:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Pre-Midterm Material

- Basic Counting Principles
  - Pigeonhole Principle
  - Inclusion Exclusion
  - Counting the Complement
  - Using symmetry
- Conditional Probability
  - $P(A \mid B) = \frac{P(AB)}{P(B)}$
  - Law of Total Probability:

$$P(A) = P(A \mid B) \cdot P(B) + P(A \mid \overline{B}) \cdot P(\overline{B})$$

  - Bayes' Theorem:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

  - Network Failure Questions

Pre-Midterm Material

Pre-Midterm Material

- Independence

Pre-Midterm Material

- Independence
  - E and F are independent if $P(EF) = P(E)P(F)$

Pre-Midterm Material

- Independence
  - E and F are independent if $P(EF) = P(E)P(F)$
  - E and F are independent if $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$

Pre-Midterm Material

- Independence
  - E and F are independent if $P(EF) = P(E)P(F)$
  - E and F are independent if $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$
  - Events $E_1, \ldots E_n$ are independent if for every subset S of events

$$P\left(\bigcap_{i \in S} E_i\right) = \prod_{i \in S} P(E_i)$$

Pre-Midterm Material

- Independence
  - E and F are independent if $P(EF) = P(E)P(F)$
  - E and F are independent if $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$
  - Events $E_1, \ldots E_n$ are independent if for every subset S of events

$$P\left(\bigcap_{i \in S} E_i\right) = \prod_{i \in S} P(E_i)$$

  - Biased coin example from Lecture 5, slide 7

Pre-Midterm Material

- Independence
  - E and F are independent if $P(EF) = P(E)P(F)$
  - E and F are independent if $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$
  - Events $E_1, \ldots E_n$ are independent if for every subset S of events

$$P\left(\bigcap_{i \in S} E_i\right) = \prod_{i \in S} P(E_i)$$

  - Biased coin example from Lecture 5, slide 7
  - If E and F are independent and G is an arbitrary event then in general

$$P(EF \mid G) \neq P(E \mid G) \cdot P(F \mid G)$$

Pre-Midterm Material

- Independence
  - E and F are independent if $P(EF) = P(E)P(F)$
  - E and F are independent if $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$
  - Events $E_1, \ldots E_n$ are independent if for every subset S of events

$$P\left(\bigcap_{i \in S} E_i\right) = \prod_{i \in S} P(E_i)$$

  - Biased coin example from Lecture 5, slide 7
  - If E and F are independent and G is an arbitrary event then in general

$$P(EF \mid G) \neq P(E \mid G) \cdot P(F \mid G)$$

  - For any given G, equality in the above statement means that E and F are **Conditionally Independent given** G

Distributions

Distributions

- Know the mean and variance for:

Distributions

- Know the mean and variance for:
  - Uniform distribution

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution
  - Geometric distribution

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution
  - Geometric distribution
  - Binomial distribution

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution
  - Geometric distribution
  - Binomial distribution
  - Poisson distribution

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution
  - Geometric distribution
  - Binomial distribution
  - Poisson distribution
  - Hypergeometric distribution

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution
  - Geometric distribution
  - Binomial distribution
  - Poisson distribution
  - Hypergeometric distribution
- Remember Linearity of Expectation and other useful facts (e.g. $Var[aX + b] = a^2 Var[X]$; in general $Var[X + Y] \neq Var[X] + Var[Y]$).

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution
  - Geometric distribution
  - Binomial distribution
  - Poisson distribution
  - Hypergeometric distribution
- Remember Linearity of Expectation and other useful facts (e.g. $Var[aX + b] = a^2 Var[X]$; in general $Var[X + Y] \neq Var[X] + Var[Y]$).
- Remember: For any a, $P(X = a) = 0$ (the *probability* that a continuous R.V. falls at a specific point is 0!)

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution
  - Geometric distribution
  - Binomial distribution
  - Poisson distribution
  - Hypergeometric distribution
- Remember Linearity of Expectation and other useful facts (e.g. $Var[aX + b] = a^2 Var[X]$; in general $Var[X + Y] \neq Var[X] + Var[Y]$).
- Remember: For any a, $P(X = a) = 0$ (the *probability* that a continuous R.V. falls at a specific point is 0!)
- Expectation is now an integral: $E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$

Distributions

- Know the mean and variance for:
  - Uniform distribution
  - Normal distribution
  - Geometric distribution
  - Binomial distribution
  - Poisson distribution
  - Hypergeometric distribution
- Remember Linearity of Expectation and other useful facts (e.g. $Var[aX + b] = a^2 Var[X]$; in general $Var[X + Y] \neq Var[X] + Var[Y]$).
- Remember: For any a, $P(X = a) = 0$ (the *probability* that a continuous R.V. falls at a specific point is 0!)
- Expectation is now an integral: $E[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx$
- Use normal approximation when applicable.

Central Limit Theorem

Central Limit Theorem

- **Central Limit Theorem:** Consider i.i.d. (independent, identically distributed) random variables $X_1, X_2, \ldots$. Xi has $\mu = E[X_i]$ and $\sigma^2 = Var[X_i]$. Then, as $n \to \infty$

  $$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to N(0,1)$$

  Alternatively

  $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Tail Bounds

Tail Bounds

- **Markov's Inequality**: If X is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Tail Bounds

- **Markov's Inequality**: If X is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Corollary

$$P(X \geq \alpha E[X]) \leq \frac{1}{\alpha}$$

Tail Bounds

- **Markov's Inequality**: If X is a non-negative random variable, then for every $\alpha > 0$, we have

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

Corollary

$$P(X \geq \alpha E[X]) \leq \frac{1}{\alpha}$$

- **Chebyshev's Inequality**: If Y is an arbitrary random variable with $E[Y] = \mu$, then, for any $\alpha > 0$,

$$P(|Y - \mu| \geq \alpha) \leq \frac{Var[Y]}{\alpha^2}$$

Tail Bounds

Tail Bounds

- **Chernoff Bounds**: Suppose X is drawn from $Bin(n, p)$ and $\mu = E[X] = pn$ Then, for any $0 < \delta < 1$

$$P(X > (1+\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

$$P(X < (1-\delta)\mu) \leq e^{-\frac{\delta^2 \mu}{3}}$$

Law of Large Numbers

- **Weak Law of Large Numbers**: Let $\overline{X}$ be the empirical mean of i.i.d.s $X_1, \ldots, X_n$. For any $\epsilon > 0$, as $n \to \infty$

$$Pr\left(\left|\overline{X} - \mu\right| > \epsilon\right) \to 0$$

Law of Large Numbers

- **Weak Law of Large Numbers**: Let $\overline{X}$ be the empirical mean of i.i.d.s $X_1, \ldots, X_n$. For any $\epsilon > 0$, as $n \to \infty$

$$Pr\left(\left|\overline{X} - \mu\right| > \epsilon\right) \to 0$$

- **Strong Law of Large Numbers**: Same hypotheses

$$Pr\left(\lim_{n\to\infty}\left(\frac{X_1 + \cdots + X_n}{n} = \mu\right)\right) = 1$$

Random Facts

Random Facts

- If X and Y are R.V.s from the same distribution, then $Z = X + Y$ isn't necessarily from the same distribution as X and Y.

Random Facts

- If X and Y are R.V.s from the same distribution, then $Z = X + Y$ isn't necessarily from the same distribution as X and Y.
- If X and Y are both normal, then so is $X + Y$.

MLEs

MLEs

- Write an expression for the likelihood.

MLEs

- Write an expression for the likelihood.
- Convert this into the log likelihood.

MLEs

- Write an expression for the likelihood.
- Convert this into the log likelihood.
- Take derivatives to find the maximum.

MLEs

- Write an expression for the likelihood.
- Convert this into the log likelihood.
- Take derivatives to find the maximum.
- Verify that this is indeed a maximum.

MLEs

- Write an expression for the likelihood.
- Convert this into the log likelihood.
- Take derivatives to find the maximum.
- Verify that this is indeed a maximum.
- See **Lecture 11** for worked examples.

Expectation-Maximization

Expectation-Maximization

- **E-step:** Computes the log-likelihood using current parameters.

Expectation-Maximization

- **E-step:** Computes the log-likelihood using current parameters.
- **M-step:** Maximize the expected log-likelihood, changing the parameters.

Expectation-Maximization

- **E-step:** Computes the log-likelihood using current parameters.
- **M-step:** Maximize the expected log-likelihood, changing the parameters.
- Iterated until convergence is achieved.

Hypothesis Testing

Hypothesis Testing

- H_0 is the **null hypothesis**

Hypothesis Testing

- H_0 is the **null hypothesis**
- H_1 is the **alternative hypothesis**

Hypothesis Testing

- H_0 is the **null hypothesis**
- H_1 is the **alternative hypothesis**
- Likelihood Ratio $= \frac{L_1}{L_0}$ where L_0 is the likelihood of H_0 and L_1 is the likelihood of H_1.

Hypothesis Testing

- H_0 is the **null hypothesis**
- H_1 is the **alternative hypothesis**
- Likelihood Ratio $= \frac{L_1}{L_0}$ where L_0 is the likelihood of H_0 and L_1 is the likelihood of H_1.
- Saying that alternative hypothesis is 5 times more likely than the null hypothesis means that $\frac{L_1}{L_0} \geq 5$.

Hypothesis Testing

- H_0 is the **null hypothesis**
- H_1 is the **alternative hypothesis**
- Likelihood Ratio $= \frac{L_1}{L_0}$ where L_0 is the likelihood of H_0 and L_1 is the likelihood of H_1.
- Saying that alternative hypothesis is 5 times more likely than the null hypothesis means that $\frac{L_1}{L_0} \geq 5$.
- Decision rule: When to reject H_0.

Hypothesis Testing

- H_0 is the **null hypothesis**
- H_1 is the **alternative hypothesis**
- Likelihood Ratio $= \frac{L_1}{L_0}$ where L_0 is the likelihood of H_0 and L_1 is the likelihood of H_1.
- Saying that alternative hypothesis is 5 times more likely than the null hypothesis means that $\frac{L_1}{L_0} \geq 5$.
- Decision rule: When to reject H_0.
- $\alpha = P(\text{rejected } H_0 \text{ but } H_0 \text{ was true})$

Hypothesis Testing

- H_0 is the **null hypothesis**
- H_1 is the **alternative hypothesis**
- Likelihood Ratio $= \frac{L_1}{L_0}$ where L_0 is the likelihood of H_0 and L_1 is the likelihood of H_1.
- Saying that alternative hypothesis is 5 times more likely than the null hypothesis means that $\frac{L_1}{L_0} \geq 5$.
- Decision rule: When to reject H_0.
- $\alpha = P(\text{rejected } H_0 \text{ but } H_0 \text{ was true})$
- $\beta = P(\text{accept } H_0 \text{ but } H_1 \text{ was true})$

Algorithms

Algorithms

- Better algorithms usually trump better hardware, but both are needed for progress

Algorithms

- Better algorithms usually trump better hardware, but both are needed for progress
- Some problems cannot be solved (e.g. the Halting Problem)

Algorithms

- Better algorithms usually trump better hardware, but both are needed for progress
- Some problems cannot be solved (e.g. the Halting Problem)
- Other (**intractable**) problems cannot (yet?) be solved in a reasonable amount of time (e.g. Integer Factorization)

Sequence Alignment

Sequence Alignment

- Brute force solution take at least $\binom{2n}{n}$ computations of the sequence score, which is **exponential time**.

Sequence Alignment

- Brute force solution take at least $\binom{2n}{n}$ computations of the sequence score, which is **exponential time**.
- **Dynamic Programming** decrease computation to $O(n^2)$.

Sequence Alignment

- Brute force solution take at least $\binom{2n}{n}$ computations of the sequence score, which is **exponential time**.
- **Dynamic Programming** decrease computation to $O(n^2)$.
- IDEA:

Sequence Alignment

- Brute force solution take at least $\binom{2n}{n}$ computations of the sequence score, which is **exponential time**.
- **Dynamic Programming** decrease computation to $O(n^2)$.
- IDEA: Store prior computations so that future computations can do table look-ups.

Sequence Alignment

- Brute force solution take at least $\binom{2n}{n}$ computations of the sequence score, which is **exponential time**.
- **Dynamic Programming** decrease computation to $O(n^2)$.
- IDEA: Store prior computations so that future computations can do table look-ups.
- Backtrace Algorithm: Start at the bottom right of the matrix and find which neighboring cells could have transitioned to the current cell under the cost function, σ. Time Bound ($O(n^2)$)

Sequence Alignment

- Brute force solution take at least $\binom{2n}{n}$ computations of the sequence score, which is **exponential time**.
- **Dynamic Programming** decrease computation to $O(n^2)$.
- IDEA: Store prior computations so that future computations can do table look-ups.
- Backtrace Algorithm: Start at the bottom right of the matrix and find which neighboring cells could have transitioned to the current cell under the cost function, σ. Time Bound ($O(n^2)$)
- See **Lecture 15** for a worked example.

P vs. NP

P vs. NP

- Some problems cannot be computed (e.g. The Halting Problem).

P vs. NP

- Some problems cannot be computed (e.g. The Halting Problem).
- Some problems can be computed but take a long time (e.g. SAT, 3-SAT, 3-coloring).

P vs. NP

- Some problems cannot be computed (e.g. The Halting Problem).
- Some problems can be computed but take a long time (e.g. SAT, 3-SAT, 3-coloring).
- **P** - a solution is computable in polynomial time

P vs. NP

- Some problems cannot be computed (e.g. The Halting Problem).
- Some problems can be computed but take a long time (e.g. SAT, 3-SAT, 3-coloring).
- **P** - a solution is computable in polynomial time
- **NP** - a solution can be verified in polynomial time, given a hint that is polynomial in the input length

P vs. NP

- Some problems cannot be computed (e.g. The Halting Problem).
- Some problems can be computed but take a long time (e.g. SAT, 3-SAT, 3-coloring).
- **P** - a solution is computable in polynomial time
- **NP** - a solution can be verified in polynomial time, given a hint that is polynomial in the input length
- $A \leq_p B$ means that if you have a fast algorithm for B, you have a fast algorithm for A.

P vs. NP

- Some problems cannot be computed (e.g. The Halting Problem).
- Some problems can be computed but take a long time (e.g. SAT, 3-SAT, 3-coloring).
- **P** - a solution is computable in polynomial time
- **NP** - a solution can be verified in polynomial time, given a hint that is polynomial in the input length
- $A \leq_p B$ means that if you have a fast algorithm for B, you have a fast algorithm for A.
- NP-complete - In NP and as hard as the hardest problem in NP

P vs. NP

- Some problems cannot be computed (e.g. The Halting Problem).
- Some problems can be computed but take a long time (e.g. SAT, 3-SAT, 3-coloring).
- **P** - a solution is computable in polynomial time
- **NP** - a solution can be verified in polynomial time, given a hint that is polynomial in the input length
- $A \leq_p B$ means that if you have a fast algorithm for B, you have a fast algorithm for A.
- NP-complete - In NP and as hard as the hardest problem in NP
- Any fast solution to an NP-complete problem would yield a fast solution to all problems in NP.