COMP2610/COMP6261 - Information Theory

Lecture 22: Review and Exam Preparation

Robert C. Williamson

Research School of Computer Science

Australian National University

22 October, 2018

Outline

Exam: Basic Stats

3 hours writing time + 15 minutes reading time

20 page writing book provided

You can bring in:

- **One A4 page** with notes on both sides

Worth **60%** of final grade and is a hurdle assessment.
Questions will be designed with this in mind.

Exam: Basic Comments

Four sections, each worth 25 marks, on the four main parts of the course

- Basic probability
- Entropy, mutual information and inequalities
- Compression (source coding)
- Error correction (channel coding)

COMP2610 and COMP6261 will sit the exact same exam

- no “bonus” or separate questions

Deliberately variable hardness of questions reflecting its hurdle status.

Exam: Style Comments

Focus is on understanding key concepts

- being apply to apply key theorems (but **not** to prove them in full from scratch!)
- But in order to apply them, you need to understand them!

2013 – 2015 exams are reflective of what to expect

- please attempt all of them...
- ...but **don't overfit** to them!
  - just because e.g. there were questions on arithmetic coding previously, doesn't mean there has to be one this year (although there might be!)

Style of questions should be similar to previous years, though substance may be different

Exam: Some Tips

Make sure you understand all tutorial and assignment questions. Also, go through the notes!

Expect some "conceptual" questions

- Not just a bunch of calculations
- e.g. "Is it possible to have a code with expected length less than entropy?"

There will be an ordering of the questions, but difficulty is a personal thing — what is hard for one person is easy for the other, and vice versa.

- Don't get stuck on a question for too long
- "Easy" questions can take time under exam conditions

Leave numbers in terms of fractions/logs if easier

- Need to provide a correct, unambiguous answer

What follows is minimal knowledge you will need.

Outline

Probability Theory: Overview

Expect to be able to compute probabilities given either explicit distributions, or English descriptions

Be very comfortable with Bayes' rule!

- Always write down formally the random variables you are using
- Translating an English description into conditionals is tricky

Be familiar with maximum likelihood and Bayesian estimation of probabilities

- What are they?
- Why do we need them?
- How do we use them?

Understand Bernoulli and Binomial random variables

Probability Theory - I

- Different types of probability distribution
  - Joint distribution: $p(X, Y)$
  - Conditional distribution: $p(X|Y) = \frac{p(X, Y)}{p(Y)}$
  - Marginal distribution: $p(Y) = \sum_x p(X = x, Y)$
- Basic rules of probability
  - Sum rule (marginalisation)
  - Product rule
- Statistical (marginal and conditional) independence:

$$X \perp\!\!\!\perp Y \leftrightarrow p(X, Y) = p(X)p(Y)$$
$$X \perp\!\!\!\perp Y|Z \leftrightarrow p(X, Y|Z) = p(X|Z)p(Y|Z)$$

Probability Theory - II

- Bayes' theorem: posterior $\propto$ prior $\times$ likelihood:

$$
\begin{aligned}
p(X|Y) &= \frac{p(X)p(Y|X)}{p(Y)} \\
&= \frac{p(X)p(Y|X)}{\sum_x p(X=x, Y)}
\end{aligned}
$$

- Application of Bayes' theorem
  - Translating description to conditional probability
  - Natural frequency interpretation

Probability Theory - III

- Basic probability distributions
  - Bernoulli: for a single flip of a coin, will it land heads?
  - Binomial: out of N flips of a coin, how many land heads?
  - see lectures for other distributions

- Estimating probabilities from observations
  - The maximum likelihood principle

Outline

1. Exam
2. Probability Theory
3. Entropy, Mutual Information and Inequalities
4. Coding
    - For Compression
    - For Communication

Entropy, Mutual Information and Inequalities: Overview

Be familiar with the various forms of entropy (marginal, conditional, ...) and mutual information

- Be prepared to explain their intuitive meaning
- Be comfortable with computing such quantities given either explicit distributions, or English descriptions

Understand definition and meaning of typical sequences, AEP

Apply Markov and Chebyshev to specific problems

Entropy

- Entropy, joint entropy and conditional entropy:

$$H(X) = \sum_x p(x) \underbrace{\log \frac{1}{p(x)}}_{h(x)}$$

$$H(X, Y) = \sum_{x,y} p(x, y) \log \frac{1}{p(x, y)}$$

$$H(Y|X) = \sum_x p(x) \underbrace{\sum_y p(y|x) \log \frac{1}{p(y|x)}}_{H(Y|X=x)}$$

- Intuitive meaning
  - Inherent / conditional uncertainty in a random variable
  - Relation to average-length code
  - Relation to minimum number of binary questions

KL and Mutual information

- Definition and properties of KL divergence

$$D_{\mathrm{KL}}(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)}.$$

- Definition and understanding of mutual information:

$$\begin{aligned} I(X;Y) &= D_{\mathrm{KL}}\big(p(X,Y)\|p(X)p(Y)\big) \\ &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \end{aligned}$$

  - Relation to statistical independence:

$$X \perp\!\!\!\perp Y \leftrightarrow I(X;Y) = 0$$

  - Relation to entropy:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Inequalities

- Statement and applications of Jensen's inequality: for convex f, and any rv X,
$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$
- Statement and applications of Markov's inequality: for nonnegative rv X,
$$p(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$
- Statement and applications of Chebyshev's inequality: for any rv X with finite expectation,
$$p(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}[X]}{a^2}.$$

Typicality and AEP

- Definition and computation of typical set:

$$T_{N\beta} = \left\{ x : \left| -\frac{1}{N} \log P(x) - H(X) \right| < \beta \right\}$$

- Statement of AEP: if $x_1, \ldots, x_N$ are iid with distribution P,

$$-\frac{1}{N} \log P(x_1, \ldots, x_N) \to H(X)$$

Outline

Source Coding: Overview

Review basic notions of (extended) ensembles, δ-sufficient subsets

Be prepared to apply SCT for block codes

Be prepared to compute Shannon-Fano-Elias/Huffman/Arithmetic

- Arithmetic only for simpler cases

Be very comfortable with basic properties of prefix codes (definition, expected length, Kraft's inequality, ...)

Source Coding

Uniform Lossy Codes

- Extended Ensembles X^N
- Smallest δ-Sufficient Sets S_δ
- Source Coding Theorem I

$$\frac{1}{N} H_\delta\left(X^N\right) \to H(X)$$

Stream Codes

- Interval Coding (Shannon-Fano-Elias)
- Arithmetic Coding

Variable-Length Lossless Codes

- Prefix & Uniquely Decodable
- Expected Code Length $L(C, X)$
- Kraft's Inequality

$$\sum_i 2^{-\ell_i} \le 1 \iff \text{prefix/U.D.}$$

- Shannon Codes $\ell_i = \left\lceil \log \frac{1}{p_i} \right\rceil$
- Source Coding Theorem II

$$H(X) \le L(C, X) \le H(X) + 1$$

- Huffman Codes

Noisy-Channel Coding

Noisy Channels $X \stackrel{Q}{\rightarrow} Y$

- Transition Matrix Q
  $Q_{i,j} = P(Y = y_i | X = x_j)$
- Binary Symmetric Channel

$$\begin{bmatrix} 1-f & f \\ f & 1-f \end{bmatrix}$$

- Z Channel

$$\begin{bmatrix} 1 & f \\ 0 & 1-f \end{bmatrix}$$

- Noisy Typewriter Channel
- Channel Capacity

$$C = \max_{\mathbf{p}_X} I(X; Y)$$

Noisy Channel Codes

- (N, K) Block Codes
- Repetition Code
- Block Code Rate $R = \frac{K}{N}$
- Probability of Block Error

$$p_B = P(\mathbf{s}_{in} \neq \mathbf{s}_{out})$$

- Noisy-Channel Coding Theorem
  Arbitrarily good codes with rate R if and only if $R \leq C$

Channel Coding: Overview

Understand notion of rate for a block code

Recall steps to computing channel capacity (easier to do with $H(Y)$ and $H(Y|X)$ rather than $H(X)$ and $H(X|Y)$)

Be prepared to apply NCCT

Final Thoughts

- Explain what you are trying to do
- I will be looking for understanding the material
- Your job: make my job easier! Where explanations are required of you, the *quality* of your explanation matters!

Thanks & Good Luck!