

520.447: Information Theory and Coding

Final Examination

9:00 AM — 12:00 PM, Dec 18, 2002.

Read these instructions before starting the examination.

- (i) You are advised to carefully read all the questions before you start answering any of them, prioritize the order in which you will attempt to answer them, and then answer as many as time permits. Use the check-list on the next page to keep track of your progress. Skipping over questions with which you encounter unexpected difficulties and returning to answer them later is usually a good exam strategy.
- (ii) This is an open-book examination. Use of any *one* textbook and a copy of Chapter 3 from Lin and Costello is permitted. Additional books, handwritten notes, solutions to problems obtained via the web, *etc.*, are **not** permitted.
- (iii) Use of electronic calculators is permitted for numeric calculations only.
- (iv) Show all your work clearly and concisely. Points may be deducted for illegible or unclear answers.
- (v) Write your answers in the space provided. Additional pages have been provided in this booklet for rough-work.
- (vi) There are 12 questions for a total of 120 points. You are expected to answer *any subset of questions worth 100 points*. If you answer additional questions, you can earn up to 20 *bonus* points! If you earn more than 100 points on this exam, excess points will be credited to your cumulative score in the two mid-term exams.
- (vii) If you want your exam score and final grade sent to you by electronic mail, please enter your e-mail address in the space provided below. If you do not wish your grade to be sent via electronic mail, please write “do not post grade” in this space.

Best of luck!

Name: _____

E-mail: _____

Check-list of Questions

No 1: Fano's Inequality for Continuous Random Variables	/10 Points
No 2: Data Processing for a Gaussian Markov Chain	/10 Points
No 3: Uniquely Decodable Codes	/5 Points
No 4: Encoding Equiprobable Sources	/15 Points
No 5: Doubling Rates of Portfolios with Volatile Stocks	/10 Points
No 6: Binary Symmetric Erasure Channel	/10 Points
No 7: Linear Block Codes	/10 Points
No 8: Capacity and Fading Channels	/10 Points
No 9: Parallel Gaussian Channels	/10 Points
No 10: Rate-Distortion for a Binary Source with Erasure Distortion	/10 Points
No 11: Shannon Bounds on the Rate-Distortion Function	/10 Points
No 12: Type Counting and Large Deviations	/10 Points
<hr/>	
TOTAL	/100 Points
<hr/>	

Question No 1: Fano's inequality for the probability of error P_e in predicting a discrete random variable $X \in \mathcal{X}$ from side information Y states that

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|}.$$

This problem concerns an analogue of Fano's inequality for a \mathbb{R} -valued random variable.

Let X be a \mathbb{R} -valued random variable with differential entropy $h(X)$ and let \hat{x} be an estimate of X . Let $E[(X - \hat{x})^2]$ be the expected prediction error.

1(a) Show that $E[(X - \hat{x})^2] \geq \frac{1}{2\pi e} e^{2h(X)}$. (5 points)

1(b) Given side information Y and an estimator $\hat{X}(Y)$, show that (5 points)

$$E \left[(X - \hat{X}(Y))^2 \right] \geq \frac{1}{2\pi e} e^{2h(X|Y)}.$$

Hint: Since e^z is a convex function, by Jensen's inequality, $E[e^Z] \geq e^{E[Z]}$.

Question No 2: Suppose that (X, Y, Z) are jointly Gaussian and that $X \rightarrow Y \rightarrow Z$ forms a Markov chain. Assume that they each have zero mean. Let X and Y have correlation coefficient ρ_1 and Y and Z have correlation coefficient ρ_2 .

2(a) Show that X and Z have correlation coefficient $\rho_1\rho_2$ (3 points)

Hint: Recall that $E[AB] = E[E[AB|C]]$ and that for jointly Gaussian zero-mean r.v.'s X, Y , the conditional mean of X given Y is $\hat{X}(Y) = \frac{\sigma_{xy}}{\sigma_y^2}Y$.

2(b) Show that $I(X; Z) = -\frac{1}{2} \log(1 - \rho_1^2\rho_2^2)$. (4 points)

2(c) Show that the data processing inequality holds by comparing $I(X; Z)$ with $I(X; Y)$ and $I(Y; Z)$. (3 points)

Question No 3: Find a uniquely decodable code which satisfies neither the prefix condition nor the suffix condition. (5 points)

Hint: A binary code for $\mathcal{X} = \{a, b\}$ may suffice.

Extra Work Space 1

Question No 4: *Encoding Equiprobable Sources.*

4(a) Show that for any optimal binary prefix code for a source $\mathcal{X} = \{1, 2, \dots, n\}$ with n equally likely symbols, the codeword lengths differ by at most 1-bit. (7 points)

Hint: If a codeword is 2 or more bits shorter than some other codeword(s), can a shorter code be created? Draw the code-tree and ponder.

4(b) Find the value of n which maximizes the redundancy

(5 points)

$$E[\ell^*(X)] - H(X) = E[\ell^*(X)] - \log_2 n$$

of the optimal binary code for a source with n symbols.

Hint: Let $2^m \leq n < 2^{m+1}$ and $k = n - 2^m$. Is it true that the optimal code has $2k$ symbols with codeword length $m + 1$ and $n - 2k$ symbols with codeword length m ? Find the maximum redundancy by varying k for fixed m , and then vary m .

- 4(c) What is the limiting value of this worst-case redundancy as $n \rightarrow \infty$? (3 points)
- Hint: $\log_2 \left(\frac{2}{e} \log_2 e \right) = 0.0861$.

Question No 5: *Doubling Rate of a Portfolio.* Let the random vector

$$X = \begin{cases} \begin{bmatrix} 1 \\ a \end{bmatrix} & \text{with probability } \frac{1}{2} \text{ and} \\ \begin{bmatrix} 1 \\ \frac{1}{a} \end{bmatrix} & \text{with probability } \frac{1}{2}, \end{cases} \quad \text{for some } a > 1,$$

designate a stock market vector of cash v/s a “hot” stock. In other words, the value of the cash asset remains unchanged from day to day, while the value of the stock goes up or down by a factor a with equal probability. Note that over a long run, the “hot” stock by itself goes nowhere: by the law of large numbers, it goes up by a roughly half the days and down by $\frac{1}{a}$ the remaining days, resulting in no net gain or loss from start to finish. Yet, an investor who allocates his assets wisely between cash and the stock stands to gain in the long run.

Let $\mathbf{b} = \begin{bmatrix} 1-b \\ b \end{bmatrix}$ be the (constant) fractional allocation of a portfolio, $0 \leq b \leq 1$, and define the expected doubling rate as

$$W(\mathbf{b}) = E[\log \mathbf{b}^T X].$$

5(a) Find and describe in words the log-optimal portfolio $\mathbf{b}^* = \arg \max_{\mathbf{b}} W(\mathbf{b})$. (4 points)

5(b) Find the optimal doubling rate $W^* = \max_{\mathbf{b}} W(\mathbf{b})$. Explain why the investment grows in an i.i.d. stock market X_1, X_2, \dots , when neither asset can grow by itself. (4 points)

5(c) Find the asymptotic behavior of $S_n = \prod_{t=1}^n \mathbf{b}^T X_t$.

Hint: Look at $\lim_{n \rightarrow \infty} \frac{1}{n} \log S_n$.

(2 points)

Extra Work Space 2

Question No 6: *Binary Symmetric Erasure Channel.*

6(a) Find the capacity of the binary symmetric erasure channel

$$\mathcal{X} = \{0, 1\}, \quad W = \begin{bmatrix} 1 - \epsilon - \rho & \rho \\ \epsilon & 1 - \epsilon - \rho \end{bmatrix}, \quad \mathcal{Y} = \{0, e, 1\},$$

with cross-over probability ϵ , and erasure probability ρ .

(5 points)

6(b) Interpret the capacity as the cascade of two channels: a BSC with cross-over probability $\delta = \frac{\epsilon}{1-\rho}$ followed by a binary erasure channel with erasure probability ρ . (5 points)

Question No 7: Consider the (5, 3) binary linear block code with generator matrix

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

for mapping 3-bit messages into 5-bit code vectors.

7(a) Write down all the codewords of this code. (3 points)

Message	Codeword	Message	Codeword

7(b) Write down the 2×5 parity check matrix H for this code. (3 points)

$$H = \begin{bmatrix} & & & & \\ & & & & \end{bmatrix}$$

7(c) There are 4 possible syndromes (including the all-zero syndrome) for a 5-bit received vector. Write down the 5-bit error pattern(s) with the smallest Hamming weight that can cause each syndrome. (4 points)

Syndrome	"Lightest" Error Pattern(s)

Hint: Is the smallest-weight error pattern resulting in any particular syndrome unique?

Extra Work Space 3

Question No 8: Consider the additive noise fading channel

$$Y = VX + Z,$$

where the channel input X is attenuated by a random factor $V \geq 0$ and is then corrupted by an additive Gaussian noise Z . Assume that V and Z are mutually independent and also jointly independent of X .

- 8(a) Argue, by showing that $I(X;Y|V) \geq I(X;Y)$, that knowledge of the fading factor V increases capacity. In practice, V can be estimated by measuring the received signal to noise ratio. (6 points)

8(b) Show that, in general, conditioning does not increase mutual information. Construct binary random variables U, V, W such that $I(U; W|V) < I(U; W)$. (4 points)

Question No 9: *Two Parallel Gaussian Channels.* Consider a pair of parallel Gaussian channels. *i.e.*

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right),$$

and there is a power constraint $E[X_1^2 + X_2^2] \leq 2P$. Assume that $\sigma_1^2 < \sigma_2^2$.

9(a) Write down a succinct formula for the channel capacity. (2 points)

9(b) Discuss the allocation of the total power $2P$ between X_1 and X_2 in the two limiting cases of very small and very large signal-to-noise ratios, *i.e.*

$$2P \ll \sigma_1^2 \quad \text{and} \quad 2P \gg \sigma_2^2,$$

by drawing the appropriate “water-filling” diagrams. (4 points)

9(c) At what signal power does the channel stop behaving like a single channel with noise power σ_1^2 and begin behaving like a pair of channels? (4 points)

Question No 10: Consider $X \sim \text{Bernoulli}(\frac{1}{2})$, and let the distortion measure be given by

$$\mathcal{X} = \{0, 1\}, \quad d(x, \hat{x}) = \begin{bmatrix} 0 & 1 & \infty \\ \infty & 1 & 0 \end{bmatrix}, \quad \hat{\mathcal{X}} = \{0, e, 1\},$$

where the reproduction allows for erasure of some of the source symbols.

10(a) Specify the “interesting” range $[D_{\min}, D_{\max}]$ of distortions D where the rate-distortion function $R(D)$ is nontrivial. (2 points)

10(b) Calculate $R(D)$ for this source-distortion pair (4 points)

10(c) Suggest a simple block code $g \circ f : \mathcal{X}^n \rightarrow \hat{\mathcal{X}}^n$ with distortion close to D , whose rate can be made close to $R(D)$ for sufficiently large block-lengths n . (4 points)

Question No 11: For a continuous random variable with zero mean and variance σ^2 , and squared error distortion,

11(a) show that the rate distortion function is bounded below as (6 points)

$$R(D) \geq h(X) - \frac{1}{2} \log(2\pi e)D.$$

11(b) Interpret whether Gaussian random variables with variance σ^2 are easier or harder to quantize at a given D than other random variables with the same variance. (4 points)

Question No 12: Let $\mathcal{X} = \{1, 2, \dots, m\}$. Show that for large n , the number of sequences $x^n \in \mathcal{X}^n$ satisfying $\frac{1}{n} \sum_{t=1}^n g(x_t) \geq \alpha$ is approximately 2^{nH^*} , where (10 points)

$$H^* = \max_{P: \sum_{x=1}^m P(x)g(x) \geq \alpha} H(P).$$

Hint: Use Sanov's theorem for large deviations of the empirical measure, and the fact that under the uniform distribution, there is a simple relationship between the probability of a set and the number of points in the set.

Extra Work Space 4