ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 24 — Information Theory and Coding
Homework 10 — Nov. 24, 2015

PROBLEM 1. Suppose we have a source that produces an independent and identically distributed sequence $U_1U_2\ldots$ according to p_U. We design a source coder in the following fashion:

- generate $M = 2^{nR}$ sequences
  $U(1) = U(1)_1 \ \ldots \ U(1)_n$
  $\vdots$
  $U(M) = U(M)_1 \ \ldots \ U(M)_n$
  by drawing $\{U(m)_i : 1 \le i \le n, 1 \le m \le M\}$ independently according to p_U.
- encode $U_1 \ldots U_n$ as follows:
  if there exists m such that $U_1 \ldots U_n = U(m)$ send the $\log_2 M = nR$ bit representation of m else declare encoding failure.

(a) Conditioned on $U^n = u^n$, what is the probability that $U(1) \neq U^n$?

(b) Conditioned on $U^n = u^n$, what is the probability of encoding failure?

(c) Show that $\Pr(\text{“failure”} | U^n \in \mathcal{T}_\epsilon^n(p_U)) \le \exp\left(-2^{nR-nH(U)(1+\epsilon)}\right)$.
Hint: $(1-x)^M \le \exp(-Mx)$

(d) Show that if $R > H(U)$ then $\Pr(\text{error}) \to 0$ as n gets large.

PROBLEM 2. A discrete memoryless channel has three input symbols: $\{-1, 0, 1\}$, and two output symbols: $\{1, -1\}$. The transition probabilities are

$$p(-1|-1) = p(1|1) = 1, \quad p(1|0) = p(-1|0) = 0.5.$$

Find the capacity of this channel with cost constraint β, if the cost function is $b(x) = x^2$.

PROBLEM 3. Random variables X and Y are correlated Gaussian variables:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; K = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}\right).$$

Find $I(X;Y)$.

PROBLEM 4. Consider a vector Gaussian channel described as follows:

$$Y_1 = x + Z_1$$
$$Y_2 = Z_2$$

where x is the input to the channel constrained in power to P; Z_1 and Z_2 are jointly Gaussian random variables with $E[Z_1] = E[Z_2] = 0$, $E[Z_1^2] = E[Z_2^2] = \sigma^2$ and $E[Z_1Z_2] = \rho\sigma^2$, with $\rho \in [-1, 1]$, and independent of the channel input.

(a) Consider a receiver that discards Y_2 and decodes the message based only on Y_1. What rates are achievable with such a receiver?

(b) Consider a receiver that forms $Y = Y_1 - \rho Y_2$, and decodes the message based only on Y. What rates are achievable with such a receiver?

(c) Find the capacity of the channel and compare it to the part (b).

PROBLEM 5. Consider an additive noise channel $Y = X + Z$ where X is the input of the channel, Y is the output of the channel and Z is the noise. The set of inputs to the channel are *non-negative* real numbers. Furthermore, the channel input is constrained in its average value: a codeword $\mathbf{x} = (x_1, \ldots, x_n)$ has to satisfy

$$\frac{1}{n}\sum_{i=1}^{n} x_i \leq P.$$

The noise Z is independent of the input X, and has the exponential distribution with $E[Z] = 1$, i.e.,

$$f_Z(z) = \begin{cases} \exp(-z) & z \geq 0 \\ 0 & \text{else.} \end{cases}$$

(a) The capacity of this channel is given by

$$C = \max_{\substack{X: E[X] \leq P \\ X \text{ is non-negative}}} I(X;Y).$$

Express the mutual information in terms of the differential entropy of Y and the differential entropy of Z.

(b) What is the differential entropy of Z?

(c) For a random variable X that satisfies the input constraints, what are the constraints on the range and the expectation of Y? Find the maximum possible differential entropy of Y subject to these constraints. Hence show that the capacity is upper bounded by

$$C \leq \log(1 + P).$$

(d) Find the distribution on X that gives an exponential distribution for $Y = X + Z$

$$f_Y(y) = \mu e^{-\mu y} \quad \text{for } y \geq 0$$

[Use Laplace transforms to compute this distribution.]

(e) Conclude that the upper bound of part (c) is actually an equality, i.e.,

$$C = \log(1 + P).$$

PROBLEM 6. Let $P(y|x)$ be a channel of input alphabet $\mathcal{X}$ and of output alphabet $\mathcal{Y}$, and let $p(x)$ be a distribution on $\mathcal{X}$. Let $r(x|y)$ be a conditional distribution on $\mathcal{X}$ given $\mathcal{Y}$, i.e., for each $x \in \mathcal{X}$ and each $y \in \mathcal{Y}$, $r(x|y) \geq 0$ and $\sum_{x' \in \mathcal{X}} r(x'|y) = 1$. Define the functional $F(p, r)$ as follows:

$$F(p, r) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x) P(y|x) \log_2 \frac{r(x|y)}{p(x)}.$$

Now for each input distribution p on $\mathcal{X}$, define the conditional distribution r_p as

$$r_p(x|y) = \frac{p(x)P(y|x)}{\sum_{x' \in \mathcal{X}} p(x')P(y|x')}.$$

I.e., r_p is the "true" conditional distribution of $\mathcal{X}$ given $\mathcal{Y}$ when p is the input distribution.

(a) Use the positivity of divergence to show that for all conditional distributions r we have $F(p, r) \leq F(p, r_p) = I(X; Y)$, and deduce that $I(X; Y) = \max_r F(p, r)$.

(b) Show that $F(p, r)$ is concave in both p and r.

The fact that the capacity C is equal to $\max_p \max_r F(p, r)$ suggests the following algorithm to compute the capacity of the channel P:

1. Set p_0 to be uniform in $\mathcal{X}$, and set $k = 0$.
2. Set $r_k = \underset{r}{\operatorname{argmax}}\ F(p_k, r) = r_{p_k}$.
3. Set $p_{k+1} = \underset{p}{\operatorname{argmax}}\ F(p, r_k)$.
4. Set $k = k + 1$.
5. Go to step 2.

(c) Use the Kuhn-Tucker conditions to show that $p_{k+1}(x) = \dfrac{\alpha_k(x)}{\sum_{x' \in \mathcal{X}} \alpha_k(x')}$, where

$$\log_2 \alpha_k(x) = \sum_{y \in \mathcal{Y}} P(y|x) \log_2 r_k(x|y).$$

This shows how to do step 3 of the algorithm.

(d) Show that $C \geq F(p_{k+1}, r_k) = \log_2 \sum_{x \in \mathcal{X}} \alpha_k(x)$.

(e) Show that $\log_2 \dfrac{\alpha_k(x)}{p_k(x)} = \sum_{y \in \mathcal{Y}} P(y|x) \log_2 \dfrac{P(y|x)}{\sum_{x' \in \mathcal{X}} P(y|x') p_k(x')}$.

(f) Let p^* be the input distribution that achieves the capacity C of the channel P. Use the result of Homework 8 Problem 5 to show that

$$C \leq \sum_x p^*(x) \log_2 \frac{\alpha_k(x)}{p_k(x)}.$$

(g) Show that

$$C - F(p_{k+1}, r_k) \leq \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{p_{k+1}(x)}{p_k(x)} \leq \max_{x \in \mathcal{X}} \log_2 \frac{p_{k+1}(x)}{p_k(x)}.$$

This upper bound provides us with a stopping condition for the algorithm. I.e., we can run the algorithm until $\max_{x \in \mathcal{X}} \log_2 \dfrac{p_{k+1}(x)}{p_k(x)} \leq \epsilon$, where ϵ is some desired accuracy.

(h) Show that

$$\sum_{k=0}^{n} (C - F(p_{k+1}, r_k)) \leq \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{p_{n+1}(x)}{p_0(x)} \leq \log |\mathcal{X}|.$$

Hint: p_0 was chosen to be uniform.

(i) Deduce that the sequence $F(p_{k+1}, r_k)$ converges to C and that the stopping condition $\max_{x \in \mathcal{X}} \log_2 \dfrac{p_{k+1}(x)}{p_k(x)} \leq \epsilon$ is guaranteed to be met eventually.