ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 21 Information Theory and Coding
Graded homework, due Monday December 4 Nov. 21, 2017

You are allowed (even encouraged) to discuss the problems on the homework with your colleagues. However, your solutions should be in your own words. If you collaborated on your solution, write down the name of your collaborators and your sources; no points will be deducted. But similarities in solutions beyond the listed collaborations will be considered as cheating.

PROBLEM 1. Let X and Y be random variables with joint distribution p_{XY}. Let p_X and p_Y denote the marginals, and let $p_{Y|X}$ and $p_{X|Y}$ denote the conditional distributions associated with p_{XY}.

(a) Let $q_{X|Y}$ be a conditional distribution. Show that

$$\sum_{x,y} p_{XY}(x,y) \log \frac{q_{X|Y}(x|y)}{p_X(x)} \le I(X;Y)$$

with equality if and only if $q_{X|Y}(x|y)p_Y(y) = p_{X|Y}(x|y)p_Y(y)$ for every (x,y).

(b) Let q_Y be a distribution on $\mathcal{Y}$. Show that

$$I(X;Y) = \sum_{x,y} p_{XY}(x,y) \log \frac{p_{Y|X}(y|x)}{q_Y(y)} - D(p_Y \| q_Y).$$

Suppose we are given a channel $W(y|x)$ and we fix $p_{Y|X} = W$. Let p_X^* be a distribution that maximizes $I(X;Y)$ with value $C(W)$, let $p_{XY}^*(x,y) = p_X^*(x)W(y|x)$, and let p_Y^* denote the Y-marginal of p_{XY}^*.

(c) Show that

$$C(W) = \max_{p_X, q_{X|Y}} \sum_{x,y} p_X(x) W(y|x) \log \frac{q_{X|Y}(x|y)}{p_X(x)}.$$

(d) Show that for any p_X

$$\sum_{x,y} p_{XY}(x,y) \log \frac{W(y|x)}{p_Y^*(y)} \le C(W) \le \sum_{x,y} p_{XY}^*(x,y) \log \frac{W(y|x)}{p_Y(y)}.$$

[Hint: for the left inequality use the KKT conditions we derived in class, for the right inequality use (b).]

Suppose now p_X also maximizes $I(X;Y)$. (It is not necessary that $p_X = p_X^*$.)

(e) Show that $D(p_Y \| p_Y^*) = 0$ (i.e, that $p_Y = p_Y^*$). [Hint: use (b) to write $D(p_Y \| p_Y^*)$ in terms of $C(W)$ and the quantity that appears on the left most term of (d).]

Problem 2. Consider the binary erasure channel W with erasure probability p, that is,

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

As in class, we will be interested in a block code for this channel, but unlike in class we will consider a decoder that declares lists of messages. To be specific, given the encoder $\mathrm{Enc} : \{1, \dots, M\} \to \{0, 1\}^n$, the decoder, upon observing y^n will produce

$$\mathrm{Dec}(y^n) = \begin{cases} \{\} & \text{if } y^n \text{ contains } k \text{ or more erasures,} \\ \{m : W(y^n | \mathrm{Enc}(m)) > 0\} & \text{else,} \end{cases}$$

where k is a parameter that we may adjust. That is, the decoder lists all possible messages that are consistent with the received sequence if there are fewer than k erasures, otherwise its produces the empty set. Let ℓ denote the expected number of incorrect messages on decoders list, averaged over the sent messages

$$\ell = \frac{1}{M} \sum_{m=1}^{M} \sum_{m' \neq m} \Pr\bigl(m' \in \mathrm{Dec}(Y^n) \bigm| m \text{ is the true message}\bigr)$$

(a) In terms of n, p and k, what is the probability that the produced list does not include the transmitted message? Let p_0 denote this probability.

Suppose we construct the encoder randomly, by setting $\mathrm{Enc}(m) = X^n(m)$ where the collection $\{X(m)_i : m = 1, \dots, M,\ i = 1, \dots, n\}$ are i.i.d. and equally likely to be 0 or 1.

(b) Given that Y^n contains j erasures (with $j < k$), what is the expected number (averaged over the ensemble of encoders) of incorrect messages on the decoder's list, expressed in terms of j, M, n, p?

(c) Show that for any $R < 1 - p$ and any $\epsilon > 0$, there is an encoder with rate at least R, $p_0 < \epsilon$ and $\ell < \epsilon$. [Hint: choose $k = nq$ with $p < q < 1 - R$, and then choose n large. Use (a) and (b).]

Suppose now that we want a decoder that declares m only if m is the only message that could have produced the received y^n, otherwise it declares '?'.

(d) Use (c) to show that for any $R < 1 - p$ for any $\epsilon > 0$ there exists an encoder of rate at least R for which the decoder of the kind above declares '?' with probability at most ϵ.

PROBLEM 3. Suppose X and Y are random variables taking values in finite sets $\mathcal{X}$ and $\mathcal{Y}$, with joint distribution p_{XY}. Define the function

$$\lambda(x, y) = \log \frac{p_{XY}(x, y)}{p_X(x) p_Y(y)}.$$

Given $t \in \mathbb{R}$, let $\delta = \Pr\bigl(\lambda(X, Y) < t\bigr)$. For each $x \in \mathcal{X}$ let $T(x) = \bigl\{y : \lambda(x, y) \geq t\bigr\}$ be the set of symbols y such that $\lambda(x, y)$ exceeds the threshold t. Observe that "$\lambda(X, Y) \geq t$" is the same statement as "$Y \in T(X)$." Fix $\epsilon \in [\delta, 1]$.

(a) Show that

$$\sum_x p_X(x) \Pr\bigl(Y \in T(x) \bigm| X = x\bigr) = 1 - \delta,$$

and that there is an x such that $\Pr\bigl(Y \in T(x) \bigm| X = x\bigr) \geq 1 - \epsilon$.

(b) Show that for every x, $\Pr\bigl(Y \in T(x)\bigr) \leq 2^{-t} \Pr\bigl(Y \in T(x) \bigm| X = x\bigr)$.

Consider the following algorithm to construct a collection $x(1), \ldots, x(M)$ in $\mathcal{X}$ and disjoint subsets $D_1, \ldots, D_M$ of $\mathcal{Y}$ such that $\Pr\bigl(Y \in D_i \bigm| X = X(i)\bigr) \geq 1 - \epsilon$ for every $1 \leq i \leq M$.

```
let M = 0, let S = Y.
while (there is an x such that Pr(Y ∈ T(x) ∩ S | X = x) ≥ 1 − ε)
{
      pick an x satisfying the condition above,
      increment M,
      let x(M) = x, let D_M = T(x) ∩ S,
      remove D_M from S
}
print x_1, ..., x_M, D_1, ..., D_M.
```

Clearly the sets $D_1, \ldots, D_M$ are disjoint. When the algorithm terminates the set $\mathcal{S}$ is the complement of $\cup_{i=1}^M D_i$ in $\mathcal{Y}$ and

$$\Pr\bigl(Y \in \mathcal{S} \cap T(x) \bigm| X = x\bigr) < 1 - \epsilon \quad \text{for every } x \in \mathcal{X}. \tag{$*$}$$

(c) Show that $\Pr(\{\lambda(X, Y) \geq t\} \cap \{Y \in \mathcal{S}\}) < 1 - \epsilon$. [Hint: use $(*)$.]

(d) Show that $\Pr\bigl(Y \in \cup_{i=1}^M D_i\bigr) \leq 2^{-t} M$. [Hint: use (b).]

(e) Combine (c) and (d) to show that $1 - \delta < (1 - \epsilon) + 2^{-t} M$, (that is, $M > (\epsilon - \delta) 2^t$).

Given a channel $W(y|x)$, pick $p_X(x)$ and let $\{(X_i, Y_i) : i = 1, \ldots, n\}$ be i.i.d., with distribution $p(x, y) = p_X(x) W(y|x)$. Fix $R < I(X; Y)$ and $\epsilon > 0$. Set $t_n = nR + \log(2/\epsilon)$.

(f) Show that $\delta_n = \Pr(\lambda(X^n, T^n) < t_n)$ approaches zero as n gets large. [Hint: use the law of large numbers on $\frac{1}{n} \lambda(X^n, Y^n)$.]

(g) Assume n is large enough to ensure $\delta_n < \epsilon/2$. Let $x^n(1), \ldots, x^n(M)$ and $D_1, \ldots, D_M$ be the output of the algorithm for the distribution p_{X^n, Y^n} and threshold t_n. Use these to construct an encoder and decoder for the channel W with rate at least R and error probability at most ϵ.