Course 048703:
Noise Removal - An Information Theoretic View
Final Assignment

Spring 2008

This exercise sheet contains 16 questions, most but not all of which were posed during the lectures. You are required to complete all of them.[1] Submission deadline is Monday, August 18th (can submit either to my mailbox on 8th floor or by email if you have it all electronically).

Lecture 2

1. Suggest a Hebrew word or phrase for "denoiser", "denoising", etc.

2. Recall our definition of the Bayes envelope U and show that:

   (a) U is concave

   (b) U satisfies a "data processing" inequality:

   $$E[U(P_{X|Z})] \leq E[U(P_{X|Y})]$$

   if $Y = f(Z)$, where f is a deterministic function.

   (c) Generalize the previous part to the case where $X - Z - Y$ (i.e., X, Z, Y form a Markov triple).

3. For jointly stationary processes $(\mathbf{X}, \mathbf{Z})$, we defined the "denoisability of $\mathbf{X}$ based on $\mathbf{Z}$" as

   $$\mathbb{D}(\mathbf{X}, \mathbf{Z}) = \lim_{n \to \infty} \min_{\hat{X}^n \in \mathcal{D}_n} EL_{\hat{X}^n}(X^n, Z^n), \tag{1}$$

   where $\mathcal{D}_n$ denotes the set of all n-block denoisers and (X^n, Z^n) on the right side of (1) are the first n symbols of the pair $(\mathbf{X}, \mathbf{Z})$. Show that the limit in (1) exists.

[1] Note that only a subset of the questions posed in the lectures are included here.

4. The goal of this exercise is to show that for jointly stationary processes $(\mathbf{X}, \mathbf{Z})$, the denoisability defined in (1) satisfies

$$\mathbb{D}(\mathbf{X}, \mathbf{Z}) = E[U(P_{X_0|\mathbf{Z}})]. \tag{2}$$

You can do it in the following steps:

(a) Show that for all $m \geq 1$

$$\mathbb{D}(\mathbf{X}, \mathbf{Z}) \leq E[U(P_{X_0|Z_{-m}^m})] \tag{3}$$

and therefore

$$\mathbb{D}(\mathbf{X}, \mathbf{Z}) \leq \lim_{m\to\infty} E[U(P_{X_0|Z_{-m}^m})]. \tag{4}$$

Why does the limit on the right side of (4) exist ?

(b) Conclude[2] that

$$\mathbb{D}(\mathbf{X}, \mathbf{Z}) \leq E[U(P_{X_0|\mathbf{Z}})] \tag{5}$$

by showing that the limit and expectation in the right side of (4) can be switched, i.e.,

$$\lim_{m\to\infty} E[U(P_{X_0|Z_{-m}^m})] = E[U(P_{X_0|\mathbf{Z}})] \tag{6}$$

(c) On the other hand, show that

$$\mathbb{D}(\mathbf{X}, \mathbf{Z}) \geq E[U(P_{X_0|\mathbf{Z}})]. \tag{7}$$

5. Let $\mathbf{X}$ be a binary symmetric first-order Markov process with transition probability (from 0 to 1 and from 1 to 0) equal to $0 < \alpha < 1$. Let $\mathbf{Z}$ be the output of a memoryless erasure channel with erasure probability ε. Let the loss function be Hamming. Express $\mathbb{D}(\mathbf{X}, \mathbf{Z})$ explicitly[3] as a function of the pair (α, ε). Fix a numerical value of $0 < \alpha < 1$ and plot your expression, as a function of ε. Comment on the form of the curve obtained.

Lecture 3

1. Let $\mathbf{X}$ be a binary symmetric first-order Markov process as in the previous question, with transition probability $0 < \alpha < 1/2$. Let $\mathbf{Z}$ be the output of a memoryless binary symmetric channel (BSC) with channel crossover probability $0 \leq \delta \leq 1/2$.

(a) Develop explicitly the forward-backward recursions for obtaining the conditional distributions $P_{X_i|Z^n}$, $1 \leq i \leq n$.

[2]This part can be done by those who have taken measure theoretic probability. If you have not, you may assume (6) without proof.

[3]An infinite sum is considered explicit.

(b) Show that, regardless of what the loss function may be, for every $m \geq 1$,

$$\mathbb{D}(\mathbf{X}, \mathbf{Z}) \geq E[U(P_{X_0|Z_{-m}^m, X_{-m-1}, X_{m+1}})] \tag{8}$$

(c) Assuming Hamming loss, compute the lower bound in (8) and the upper bound in (3) for $m = 1, 2$. Fix a numerical value of $0 < \alpha < 1/2$ and plot the four bounds obtained, as a function of $0 \leq \delta \leq 1/2$. Comment on the form of these bounds and their tightness.

Lecture 4

1. Recall the mapping used by the DUDE

$$\phi(\Lambda, \Pi, v, z) = \arg\min_{\hat{x}} v^T \Pi^T \left(\Pi\Pi^T\right)^{-1} (\boldsymbol{\lambda}_{\hat{x}} \odot \boldsymbol{\pi}_z),$$

where Λ is the $\mathcal{X} \times \hat{\mathcal{X}}$ loss matrix, Π is the $\mathcal{X} \times \mathcal{Z}$ channel matrix, v is a $|\mathcal{Z}|$-dimensional column vector, $z \in \mathcal{Z}$, $\boldsymbol{\lambda}_{\hat{x}}$ is the column of the loss matrix associated with the symbol $\hat{x}$, and $\boldsymbol{\pi}_z$ is the column of the channel matrix associated with symbol z. Assuming Hamming loss, find $\phi(\Lambda, \Pi, v, z)$ explicitly for

(a) The BSC(δ) (assuming $0 \leq \delta \leq 1/2$).

(b) The Erasure channel with erasure probability ε.

(c) The "Z-channel" with parameter p (probability of output 1 when the input is 0).

Give intuitive explanations for the expressions obtained.

2. Consider the setting of the fifth question in the section of Lecture 2. Fix a numerical value for $0 < \alpha < 1$ and for $0 < \varepsilon < 1$. Simulate the processes $\mathbf{X}$ and $\mathbf{Z}$ on a computer, and implement and employ the DUDE for denoising X^n on the basis of Z^n under Hamming loss criterion. Do this for $n = 10^2, 10^3, 10^4, 10^5, 10^6$. For every such value of n, compare between:

(a) the Bit Error Rate (BER) of the DUDE with $k = \lfloor \frac{1}{2} \log_3 n \rfloor$

(b) the BER of the DUDE with the best choice of k (as can be selected by a genie with access to the noisefree as well as the noisy data)

(c) the BER of the DUDE which chooses the k that results in the most compressible reconstruction sequence, using the lossless compression standard of your choice (can take an off-the-shelf scheme)

Also, compare the resulting BERs to the value of $\mathbb{D}(\mathbf{X}, \mathbf{Z})$ for the chosen numerical values of α and ε (as obtained in the fifth question of Lecture 2).

Lecture 7

In the questions of this section, assume the semi-stochastic analog denoising setup, with its associated regularity assumptions and notation, as described in Lecture 7.

1. Recall that we defined, for every individual sequence x^n,

$$\mathbb{D}_0(x^n) = \min_{\phi} E\left[\frac{1}{n}\sum_{i=1}^{n}\Lambda(x_i, \phi(Z_i))\right], \tag{9}$$

   where the minimum is over all measurable functions $\phi : \mathbb{R} \to \mathbb{R}$. Show that the minimum in (9) is achieved by the function

$$\phi(z) = \hat{X}_{Bayes}\left([P^{emp}(x^n) \odot \Pi]_{X|Z=z}\right). \tag{10}$$

2. Consider the kernel density estimate

$$\hat{f}_Z(z) = \hat{f}_Z[Z^n](z) \triangleq \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{z - Z_i}{h}\right), \tag{11}$$

   where K is a kernel (as defined in Lecture 6) and $h = h_n$ is a sequence of positive reals satisfying $h_n \to 0$ but sufficiently slowly that $nh_n \to \infty$. Define the estimation error by

$$J_n = \int \left|\hat{f}_Z - \Pi \circ P^{emp}(x^n)\right|, \tag{12}$$

   where $\Pi \circ P_X$ denotes the distribution of a channel output symbol when the input symbol is distributed according to P_X. Show that, regardless of what the underlying individual sequence $\mathbf{x}$ may be,

   (a) $J_n \to 0$ in probability

   (b) $J_n \to 0$ a.s.

   (c) $\forall \varepsilon > 0$ there exists $r > 0$ such that $P(J_n > \varepsilon) \leq e^{-rn}$ for all n.

   Hint: understand first how the classical analogue of this result (of density estimation for i.i.d. random variables) is proven. Also note that (c)⇒(b)⇒(a) (why ?), so can focus on (c).

Lecture 9

1. Assume in what follows that the process components take values in a finite alphabet.

   (a) Show that if X_i are i.i.d. then $\mathbf{X}$ is totally ergodic.

   (b) Give an example of a stationary process which is not ergodic.

(c) Give an example of a stationary ergodic process which is not totally ergodic.

2. Let $d : \{0, 1, \ldots, M-1\} \to [0, \infty)$ satisfy $d(v) = 0$ if and only if $v = 0$. Define the maximum entropy function ϕ_d by

$$\phi_d(D) = \max H(V), \tag{13}$$

where the maximization is over all random variables V that take values in $\{0, 1, \ldots, M-1\}$ and satisfy $Ed(V) \leq D$. Establish the following properties of $\phi_d(D)$, for $D > 0$:

- $\phi_d(0) = 0$
- $\phi_d(D) = \log M$ for $D \geq \frac{1}{M}\sum_{i=1}^{M-1} d(i)$
- ϕ_d is concave
- $\phi_d(D)$ is strictly increasing in the range $[0, \frac{1}{M}\sum_{i=1}^{M-1} d(i)]$
- $\phi_d(D)$ is attained *uniquely* in distribution by V_D whose PMF is of the form

$$P(V_D = v) = \frac{e^{-\beta d(v)}}{\sum_{v'=0}^{M-1} e^{-\beta d(v')}}, \tag{14}$$

  for some $\beta \in [0, \infty]$.

Lecture 11

1. Let N be a random variable taking values in $\{0, 1, \ldots, M-1\}$ and define the loss function d by $d(a) = -\log P(N = a)$. Use your finding from the previous exercise to show that the maximum in the definition of $\phi_d(H(N))$ is uniquely achieved by N.

2. Recall that we defined

$$\Psi(P) = \max_{X \sim P, \hat{X} \sim P} E\Lambda(X, \hat{X}) \tag{15}$$

   Show that when Λ is Hamming and the alphabet is binary we have

$$\Psi(P) = 2U(P), \tag{16}$$

   where $U(P)$ is the Bayes envelope.

3. Establish constructively the existence of a sequence of rate-distortion block codes that achieves the point $(R, D(R))$ for every stationary and ergodic source. You may assume finite alphabets.

Lecture 12

1. Do the second question in

   `http://www.stanford.edu/class/ee477/midterm.pdf`