ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Information Theory and Signal Processing — Assignment date: January 26th, 2018, 12:15
Fall 2017 — Due date: January 26th, 2018, 15:15

Final Exam

There are six problems. We do not presume that you will finish all of them. Choose the ones you find easiest and collect as many points as possible. Good luck!

Name: ______________________________

Problem 1	/ 20
Problem 2	/ 20
Problem 3	/ 10
Problem 4	/ 10
Problem 5	/ 10
Problem 6	/ 20
Total	/90

Problem 1. *(Hilbert Space Bases)*

The basic function spaces in the development of wavelets are spanned by a basis of the form

$$\{\varphi(t-k)\}_{k\in\mathbb{Z}}\,. \tag{1}$$

1. *8 Points* In our wavelet discussion, we assumed that this is an orthonormal basis. We would like to project a signal $x(t)$ onto the Hilbert space spanned by our orthogonal basis $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$. The claim is that the projection is given by :

$$P_V x(t) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} \langle x(t), \varphi(t-k)\rangle \varphi(t-k), \tag{2}$$

   where, assuming real-valued functions, $\langle x(t), \varphi(t-k)\rangle = \int_{-\infty}^{\infty} x(t)\varphi(t-k)dt$. Show that the orthogonality principle is satisfied, i.e., show that

$$\langle x(t) - P_V x(t), \varphi(t-n)\rangle = 0. \tag{3}$$

   It is sufficient to prove this statement for $n = 0$.

   Remark: This is a special case of Part 2. A correct answer to Part 2 will automatically also get full credit on Part 1.

2. *12 Points* Now, suppose that $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$ is *not* an orthonormal basis. We would like to project a signal $x(t)$ onto the Hilbert space spanned by our (non-orthogonal) basis $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$.

   The claim is that this projection is given by

$$P_V x(t) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} \langle x(t), \tilde{\varphi}(t-k)\rangle \varphi(t-k), \tag{4}$$

   where the function $\tilde{\varphi}(t)$ has Fourier transform

$$\tilde{\Phi}(\omega) = \frac{\Phi(\omega)}{\sum_{k=-\infty}^{\infty} |\Phi(\omega + 2\pi k)|^2}. \tag{5}$$

   Show that the orthogonality principle is satisfied, i.e., show that

$$\langle x(t) - P_V x(t), \varphi(t-n)\rangle = 0. \tag{6}$$

   It is sufficient to prove this statement for $n = 0$.

   Hint: Prove first that the discrete-time Fourier transform of the signal $s(k)$, defined as

$$s(k) = \int_{t=-\infty}^{\infty} f(t)g(t-k)dt, \tag{7}$$

   is given by

$$S(e^{j\omega}) = \sum_{m=-\infty}^{\infty} F(\omega - 2\pi m)G^*(\omega - 2\pi m). \tag{8}$$

Problem 2. *(Moments and Rényi)*

Suppose G is an integer valued random variable taking values in the set $\{1, \dots, K\}$. Let $p_i = \Pr(G = i)$. We will derive bounds on the moments of G, the ρ-th moment of G being $E[G^\rho]$.

1. Show that for any distribution q on $\{1, \dots, K\}$, and any ρ

$$E[G^\rho] = \sum_i q_i \exp\Bigl[\log \frac{p_i i^\rho}{q_i}\Bigr].$$

   (Here and below exp and log are taken to same base.)

2. Show that

$$E[G^\rho] \geq \exp\Bigl[-D(q\|p) + \rho \sum_i q_i \log i\Bigr].$$

   [*Hint:* use Jensen's inequality on Part 1.]

3. Show that

$$\sum_i q_i \log i = H(q) - \sum_i q_i \log \frac{1}{iq_i} \geq H(q) - \log \sum_{i=1}^{K} 1/i.$$

   [*Hint:* use Jensen's inequality.]

4. Using Part 2, Part 3, and the fact that $\sum_{i=1}^K 1/i \leq 1 + \ln K$, show that, for $\rho \geq 0$,

$$E[G^\rho] \geq (1 + \ln K)^{-\rho} \exp[\rho H(q) - D(q\|p)]$$

5. Suppose that $U_1, \dots, U_n$ are i.i.d., each with distribution p. Suppose we try to determine the value of $X = (U_1, \dots, U_n)$ by asking a sequence of questions, each of the type 'Is $X = x$?' until we are answered 'yes'. Let G_n be the number of questions we ask.

   Show that, for $\rho \geq 0$,

$$\liminf_n \frac{1}{n\rho} \log E[G_n^\rho] \geq H_{1/(1+\rho)}(p)$$

   where $H_s(p) = \frac{1}{1-s} \log \sum_u p(u)^s$ is the Rényi entropy of the distribution p.

   [*Hint:* recall from Homework 2 Problem 6 that $\rho H_{1/(1+\rho)}(p) = \max_q \rho H(q) - D(q\|p)$, and that the Rényi entropy of a collection of independent random variables is the sum of their Rényi entropies.]

Problem 3. *(Exponential Families)* What is the maximum entropy distribution, call it $p(x, i)$, on $[0, \infty] \times \mathbb{N}$, both of whose marginals have mean $\mu > 0$. (I.e., in one axis the distribution is over the positive reals, whereas in the other one it is over the natural numbers.)

Problem 4. *(Nonsingular and Uniquely Decodable Codes)*

Recall that for a code $\mathcal{C} : \mathcal{U} \to \{0,1\}^*$ we define $\mathcal{C}^n : \mathcal{U}^n \to \{0,1\}^*$ as $\mathcal{C}^n(u_1, \dots, u_n) = \mathcal{C}(u_1) \dots \mathcal{C}(u_n)$.

If a code $\mathcal{C}$ is uniquely decodable, it is clear that for each n, $\mathcal{C}^n$ is non-singular (indeed $\mathcal{C}^n$ is uniquely decodable).

1. Suppose $\mathcal{C}$ is *not* uniquely decodable. Show that there is an $n \geq 1$ such that $\mathcal{C}^n$ is singular.

2. Suppose $\mathcal{K} : \{0,1,2,\dots\} \to \{0,1\}^*$ is a *prefix-free* code for non-negative integers. Show that for any *non-singular* code $\mathcal{C}$ for any alphabet $\mathcal{U}$, the code $\mathcal{C}' : \mathcal{U} \to \{0,1\}^*$ with

$$\mathcal{C}'(u) = \mathcal{K}(\text{length}(\mathcal{C}(u)))\mathcal{C}(u)$$

   is prefix free.

Recall from Homework 4, Problem 1 that there is a prefix-free $\mathcal{C}_1 : \{1,2,\dots\} \to \{0,1\}^*$ for positive-integers for which $\text{length}(\mathcal{C}_1(n)) \leq 2 + 2\log(1+\log n) + \log n$. Let $\mathcal{K} : \{0,1,\dots\} \to \{0,1\}^*$ be defined as $\mathcal{K}(n) = \mathcal{C}_1(n+1)$.

3. Show that for any non-singular code $\mathcal{C}$ for $\mathcal{U}$ with $E[\text{length}(\mathcal{C}(U))] = L$, there is a prefix-free code $\mathcal{C}'$ for $\mathcal{U}$ with

$$E[\text{length}(\mathcal{C}'(U))] \leq L + 2 + 2\log(1+\log(1+L)) + \log(1+L).$$

Problem 5. *(Missing Data)*

We are given real-valued data with a single missing sample :

$$X_1, X_2, X_3, X_4, X_5, X_6, ?, X_8, X_9, \ldots \tag{9}$$

where we assume that the data is wide-sense stationary with autocorrelation function $R_X[k] = \alpha^{|k|}$, where $0 < \alpha < 1$. We would like to find a meaningful estimate for the missing sample X_7.

1. As a starting point, let us consider the estimate $\hat{X}_7 = wX_6$, where w is a real number. Find the value of w so as to minimize the mean-squared error $\mathbb{E}[(X_7 - \hat{X}_7)^2]$, and determine the incurred mean-squared error.

2. Now, consider the estimate $\hat{X}_7 = w_1 X_6 + w_2 X_8$. Again, find the values of w_1 and w_2 so as to minimize the mean-squared error $\mathbb{E}[(X_7 - \hat{X}_7)^2]$, and determine the incurred mean-squared error.

Problem 6. *(Uniformity Testing)*

Let us reconsider the problem of testing against uniformity. In the lecture we saw a particular *test statistics* that required only $O(\sqrt{k}/\epsilon^2)$ samples where ϵ was the ℓ_1 distance.

Let us now derive a test from scratch. To make things simple let us consider the ℓ_2^2 distance. Recall that the alphabet is $\mathcal{X} = \{1, \cdots, k\}$, where k is known. Let U be the uniform distribution on $\mathcal{X}$, i.e., $u_i = 1/k$. Let P be a given distribution with components p_i. Let X^n be a set of n iid samples. A pair of samples (X_i, X_j), $i \neq j$, is said to *collide* if $X_i = X_j$, if they take on the same value.

1. Show that the expected number of collisions is equal to $\binom{n}{2}\|p\|_2^2$.

2. Show that the uniform distribution minimizes this quantity and compute this minimum.

3. Show that $\|p - u\|_2^2 = \|p\|_2^2 - \frac{1}{k}$.

   NOTE: In words, if we want to distinguish between the uniform distribution and distributions P that have an ℓ_2^2 distance from U of at least ϵ, then this implies that for those distributions $\|p\|_2^2 \geq 1/k + \epsilon$. Together with the first point this suggests the following test: compute the number of collisions in a sample and compare it to $\binom{n}{2}(1/k + \epsilon/2)$. If it is below this threshold decide on the uniform one. What remains is to compute the variance of the collision number as a function of the sample size. This will tell us how many samples we need in order for the test to be reliable.

4. Let $a = \sum_i p_i^2$ and $b = \sum_i p_i^3$. Show that the variance of the collision number is equal to

$$\binom{n}{2}a + \binom{n}{2}\left[\binom{n}{2} - \left(1 + \binom{n-2}{2}\right)\right]b + \binom{n}{2}\binom{n-2}{2}a^2 - \binom{n}{2}^2 a^2$$
$$= \binom{n}{2}\left[b2(n-2) + a(1 + a(3 - 2n))\right]$$

   by giving an interpretation of each of the terms in the above sum.

   NOTE: If you don't have sufficient time, skip this step and go to the last point.

   For the uniform distribution this is equal to

$$\binom{n}{2}\frac{(k-1)(2n-3)}{k^2} \leq \frac{n^2}{2k}.$$

   NOTE: You don't have to derive this from the previous result. Just assume it.

5. Recall that we are considering the ℓ_2^2 distance which becomes generically small when k is large. Therefore, the proper scale to consider is $\epsilon = \kappa/k$. Use the Chebyshev inequality and conclude that if we have $\Theta(\sqrt{k}/\kappa^2)$ samples then with high probability the empirical number of collisions will be less than $\binom{n}{2}(1/k + \kappa/(2k))$ assuming that we get samples from a uniform distribution.

NOTE: The second part, namely verifying that the number of collisions is with high probability no smaller than $\binom{n}{2}(1/k + \kappa/(2k))$ when we get $\Theta(\sqrt{k}/\kappa^2)$ samples from a distribution with ℓ_2^2 distance at least κ/k away from a uniform distribution follows in a similar way.

HINT: Note that if p represents a vector with components p_i then $\|p\|_1 = \sum_i |p_i|$ and $\|p\|_2^2 = \sum_i p_i^2$.