ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Information Theory and Signal Processing
Fall 2017

Assignment date: January 26th, 2018, 12:15
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 :

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

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

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

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

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

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

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

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

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

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

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

is given by

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

Solution:

1. A direct solution can be given by noting that, by the linearity of the inner product,

$$\begin{aligned}\langle x(t) - \sum_{k=-\infty}^{\infty} \langle x(t), \varphi(t-k)\rangle \varphi(t-k), \varphi(t-n)\rangle \\ = \quad \langle x(t), \varphi(t-n)\rangle - \sum_{k=-\infty}^{\infty} \langle x(t), \varphi(t-k)\rangle \langle \varphi(t-k), \varphi(t-n)\rangle. \end{aligned} \tag{9}$$

However, by the orthonormality of the basis, we know that $\langle \varphi(t-k), \varphi(t-n)\rangle = \delta(n-k)$, and hence, in the sum, the only term surviving is when $k = n$. That is, we find that the above is equal to

$$\begin{aligned}\langle x(t), \varphi(t-n)\rangle - \sum_{k=-\infty}^{\infty} \langle x(t), \varphi(t-k)\rangle \langle \varphi(t-k), \varphi(t-n)\rangle \\ = \quad \langle x(t), \varphi(t-n)\rangle - \langle x(t), \varphi(t-n)\rangle, \end{aligned} \tag{10}$$

which proves the claim. Note that for this part of the proof, we have *only* used the orthonormality of the basis; no other properties were needed.

Finally, to see that Part 1 is indeed a special case of Part 2, recall from Homework 1, Problem 3, that when $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis, we have that

$$\sum_{k=-\infty}^{\infty} |\Phi(\omega + 2\pi k)|^2 = 1, \tag{11}$$

and thus,

$$\tilde{\Phi}(\omega) \quad = \quad \frac{\Phi(\omega)}{\sum_{k=-\infty}^{\infty} |\Phi(\omega + 2\pi k)|^2} = \Phi(\omega). \tag{12}$$

Thus, Part 1 is a special case of Part 2.

2. As suggested in the hint, let us first prove the Fourier pair

$$s(k) \quad = \quad \int_{t=-\infty}^{\infty} f(t) g(t-k) dt \tag{13}$$

with discrete Fourier transform

$$S(e^{j\omega}) \quad = \quad \sum_{m=-\infty}^{\infty} F(\omega - 2\pi m) G^*(\omega - 2\pi m). \tag{14}$$

To see this, simply define the signal

$$s(t) \quad = \quad \int_{t=-\infty}^{\infty} f(\tau) g(\tau - t) d\tau, \tag{15}$$

that is, $s(t)$ is the convolution of $f(t)$ with $g(-t)$. Therefore,

$$S(\omega) \;=\; F(\omega)G^*(\omega). \tag{16}$$

Now, $s(k)$ is simply the sampled version of $s(t)$, and therefore,

$$S(e^{j\omega}) \;=\; \sum_{m=-\infty}^{\infty} S(\omega - 2\pi m) \tag{17}$$

$$\;=\; \sum_{m=-\infty}^{\infty} F(\omega - 2\pi m)G^*(\omega - 2\pi m). \tag{18}$$

Now for the tough part of the proof...[1] we have to establish that with the suggested choice of $\tilde{\varphi}(t)$, we satisfy the orthogonality principle:

$$\langle x(t) - P_V x(t), \varphi(t-n)\rangle \;=\; 0. \tag{19}$$

Including the remark, it suffices to prove

$$\langle x(t) - P_V x(t), \varphi(t)\rangle \;=\; 0. \tag{20}$$

where now

$$P_V x(t) \;\stackrel{\text{def}}{=}\; \sum_{k=-\infty}^{\infty} \langle x(t), \tilde{\varphi}(t-k)\rangle \varphi(t-k). \tag{21}$$

Substituting, we can write the claim as

$$\langle x(t) - \sum_{k=-\infty}^{\infty} \langle x(t), \tilde{\varphi}(t-k)\rangle \varphi(t-k), \varphi(t)\rangle \;=\; 0. \tag{22}$$

Rewriting the sum over k, we obtain

$$\langle x(t), \varphi(t)\rangle - \sum_{k=-\infty}^{\infty} \langle \langle x(t), \tilde{\varphi}(t-k)\rangle \varphi(t-k), \varphi(t)\rangle$$

$$\;=\; \langle x(t), \varphi(t)\rangle - \sum_{k=-\infty}^{\infty} \langle x(t), \tilde{\varphi}(t-k)\rangle \langle \varphi(t-k), \varphi(t)\rangle \tag{23}$$

Now, defining

$$y(k) \;=\; \langle x(t), \tilde{\varphi}(t-k)\rangle \tag{24}$$

$$z(k) \;=\; \langle \varphi(t-k), \varphi(t)\rangle, \tag{25}$$

[1] Or can someone come up with a simple proof?

we can express

$$\sum_{k=-\infty}^{\infty} \langle x(t), \tilde{\varphi}(t-k)\rangle\langle \varphi(t-k), \varphi(t)\rangle \;=\; (y*z)(n=0) \tag{26}$$

In the Fourier domain, we can write this as

$$(y*z)(n=0) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} Y(e^{j\omega})Z(e^{j\omega})d\omega, \tag{27}$$

or, using the argument given in the hint, we can express this as

$$(y*z)(n=0) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{m=-\infty}^{\infty} X(\omega-2\pi m)\tilde{\Phi}^*(\omega-2\pi m)\right)\left(\sum_{k=-\infty}^{\infty}|\Phi(\omega-2\pi k)|^2\right)d\omega \tag{28}$$

Moreover, we also rewrite the other inner product:

$$\begin{aligned}\langle x(t), \varphi(t)\rangle &= \int_{-\infty}^{\infty} x(t)\varphi(t)dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\Phi^*(\omega)d\omega\end{aligned} \tag{29}$$

by the same exact trick: interpret the first line as the convolution of $x(t)$ with $\varphi(-t)$, evaluated at $t=0$. We will rewrite this integral in a slightly strange fashion as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\Phi^*(\omega)d\omega \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m=-\infty}^{\infty} X(\omega-2\pi m)\Phi^*(\omega-2\pi m)d\omega. \tag{30}$$

With this, we can rewrite

$$\begin{aligned}&\langle x(t), \varphi(t)\rangle - \sum_{k=-\infty}^{\infty}\langle\langle x(t), \tilde{\varphi}(t-k)\rangle\varphi(t-k), \varphi(t)\rangle \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m=-\infty}^{\infty} X(\omega-2\pi m)\Phi^*(\omega-2\pi m)d\omega. \\ &\quad -\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{m=-\infty}^{\infty} X(\omega-2\pi m)\tilde{\Phi}^*(\omega-2\pi m)\right)\left(\sum_{k=-\infty}^{\infty}|\Phi(\omega-2\pi k)|^2\right)d\omega.\end{aligned} \tag{31}$$

Merging terms, we find

$$\begin{aligned}&\langle x(t), \varphi(t)\rangle - \sum_{k=-\infty}^{\infty}\langle\langle x(t), \tilde{\varphi}(t-k)\rangle\varphi(t-k), \varphi(t)\rangle \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m=-\infty}^{\infty} X(\omega-2\pi m)\left(\Phi^*(\omega-2\pi m) - \tilde{\Phi}^*(\omega-2\pi m)\left(\sum_{k=-\infty}^{\infty}|\Phi(\omega-2\pi k)|^2\right)\right)d\omega.\end{aligned}$$

It is now obvious that if we select

$$\tilde{\Phi}(\omega) \;=\; \frac{\Phi(\omega)}{\sum_{k=-\infty}^{\infty}|\Phi(\omega+2\pi k)|^2}, \tag{32}$$

the integrand becomes zero, thus proving the claim.

Problem 2. *(Moments and Rényi)*

Suppose G is an integer valued random variable taking values in the set $\{1, \ldots, K\}$. 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 $\{1, \ldots, K\}$, and any ρ

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

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

2. Show that

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

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

3. Show that

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

   [*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 $U_1, \ldots, U_n$ are i.i.d., each with distribution p. Suppose we try to determine the value of $X = (U_1, \ldots, U_n)$ 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.]

Solution:

1. Simplifying the right hand side of the equation, we can get

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

2. By Jensen's inequality and $\exp(x)$ is a convex function

$$\begin{aligned} E[G^\rho] = \sum_i q_i \exp\Big[\log \frac{p_i i^\rho}{q_i}\Big] &\geq \exp\Big[\sum_i q_i \log \frac{p_i i^\rho}{q_i}\Big] \\ &= \exp\Big[\sum_i q_i \log \frac{p_i}{q_i} + \rho \sum_i q_i \log i\Big] \\ &= \exp\Big[-D(q\|p) + \rho \sum_i q_i \log i\Big] \end{aligned}$$

3.

$$\begin{aligned} \sum_i q_i \log i &= \sum_i q_i (\log \frac{1}{q_i} - \log \frac{1}{iq_i}) \\ &= H(q) - \sum_i q_i \log \frac{1}{iq_i} \\ &\geq H(q) - \log \sum_i \frac{1}{i} \end{aligned}$$

where the last inequality is obtained by apply Jensen's inequality on concave function $\log(x)$.

4. Using previous results, we have

$$\begin{aligned} E[G^\rho] &\geq \exp\Big[-D(q\|p) + \rho \sum_i q_i \log i\Big] \\ &\geq \exp\Big[-D(q\|p) + \rho(H(q) - \log \sum_i \frac{1}{i})\Big] \\ &= \exp\Big[\rho H(q) - D(q\|p) - \rho \log \sum_i \frac{1}{i}\Big] \\ &= \exp\Big[\rho H(q) - D(q\|p) - \rho \log(1 + \ln K)\Big] \\ &= (1 + \ln K)^{-\rho} \exp\Big[\rho H(q) - D(q\|p)\Big] \end{aligned}$$

5. Since U_i's are i.i.d with distribution p, $X = (U_1, \ldots, U_n)$ is the joint distribution p^n. If each U_i has K distinct values, then X has K^n values. Thus, $G_n \in \{1, \ldots, K^n\}$

   Recall from Homework 2 Problem 6 that

$$\max_q \rho H(q) - D(q\|p_X) = \rho H_{1/(1+\rho)}(p_X) = \rho \sum_{i=1}^n H_{1/(1+\rho)}(p_{U_i}) = n\rho H_{1/(1+\rho)}(p)$$

   Since the result of Part 4 holds for any q, it also holds for the q which maximizes

$\rho H(q) - D(q\|p_X)$. Hence, we have

$$\begin{aligned}
\liminf_n \frac{1}{n\rho} \log E[G_n^\rho] &\geq \liminf_n \max_q \frac{1}{n\rho} \log \left((1 + \ln K^n)^{-\rho} \exp \left[\rho H(q) - D(q\|p_X) \right] \right) \\
&= \liminf_n \frac{1}{n\rho} [-\rho \log(1 + \ln K^n) + \max_q \rho H(q) - D(q\|p_X)] \\
&= \liminf_n -\frac{1}{n} \log(1 + n \ln K) + H_{1/(1+\rho)}(p) \\
&= H_{1/(1+\rho)}(p)
\end{aligned}$$

In the last step, $\liminf_n -\frac{1}{n} \log(1 + n \ln K) = 0$.

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.)

Solution:

We know that entropy of a joint distribution is at most as large as the sum of the entropies of the marginals and we have equality if the two random variables are independent. Hence the solution will be the product distribution of two independent random variables. It hence suffices to find these two marginal distributions.

1. Maximum entropy distribution on $[0, \infty)$ with mean μ: The answer is the *exponential* distribution with density $p(x) = \frac{1}{\mu} e^{-x/\mu}$. To see this note that the general form is an exponential distribution (i.e., a member of the exponential family) with the form $p(x) = e^{\theta x - A(\theta)}$ since the condition is that $\mathbb{E}[X] = \mu$, i.e., $\phi(x) = x$. The normalization constraint $\int_0^\infty p(x)dx = 1$ requires $e^{-A(\theta)} = -\theta$ and $\theta < 0$ so that $p(x) = -\theta e^{\theta x}$ for $\theta < 0$. And the mean constraint gives us that $\theta = -\frac{1}{\mu}$.
2. In a similar manner the maximum entropy distribution on $\mathbb{N}$ with mean μ is equal to $p(i) = (1 - \frac{1}{\mu})^{i-1} \frac{1}{\mu} = (\mu - 1)^{i-1}/\mu^i$. Note that the maximum entropy of $p(i)$ is Geometric distribution, not Poisson distribution, since the support set is natural numbers $\{1, 2, 3, \dots\}$.

Therefore, the answer is $p(x, i) = e^{-x/\mu} \frac{(\mu-1)^{i-1}}{\mu^{i+1}}$.

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 $\mathcal{C}^n(u_1, \ldots, u_n) = \mathcal{C}(u_1) \ldots \mathcal{C}(u_n)$.

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 $\mathcal{K} : \{0, 1, 2, \ldots\} \to \{0,1\}^*$ 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 $\mathcal{C}_1 : \{1, 2, \ldots\} \to \{0,1\}^*$ for positive-integers for which $\text{length}(\mathcal{C}_1(n)) \leq 2 + 2\log(1 + \log n) + \log n$. Let $\mathcal{K} : \{0, 1, \ldots\} \to \{0,1\}^*$ 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

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

Solution:

1. Suppose $\mathcal{C}$ is not uniquely decodable. This means that there is u^n and v^m such that $u^n \neq v^m$ but $\mathcal{C}^*(u^n) = \mathcal{C}^*(v^m)$. Now consider $a = u^n v^m$ and $b = v^m u^n$. Both are in U^{n+m}, $a \neq b$, and $\mathcal{C}^{n+m}(a) = \mathcal{C}^{n+m}(b)$; consequently $\mathcal{C}^{n+m}$ is singular (i.e. not injective).

2. For any two letters u_i and u_j, we have $\mathcal{C}'(u_i) = \mathcal{K}(\text{length}(\mathcal{C}(u_i)))\mathcal{C}(u_i)$ and $\mathcal{C}'(u_j) = \mathcal{K}(\text{length}(\mathcal{C}(u_j)))\mathcal{C}(u_j)$. If $\text{length}(\mathcal{C}(u_i)) \neq \text{length}(\mathcal{C}(u_j))$, since $\mathcal{K}$ is a prefix-free code, $\mathcal{K}(\text{length}(\mathcal{C}(u_i)))$ and $\mathcal{K}(\text{length}(\mathcal{C}(u_j)))$ are not prefix of each other. Hence $\mathcal{C}'(u_i)$ and $\mathcal{C}'(u_j)$ are not prefix of each other. If $\text{length}(\mathcal{C}(u_i)) = \text{length}(\mathcal{C}(u_i))$, we have $\mathcal{K}(\text{length}(\mathcal{C}(u_i))) = \mathcal{K}(\text{length}(\mathcal{C}(u_j)))$. $\mathcal{C}'(u_i)$ and $\mathcal{C}'(u_j)$ have the same first part. However, since $\mathcal{C}$ is non-singular, $\mathcal{C}(u_i)$ and $\mathcal{C}(u_j)$ are different. And since they have the same length, they are not prefix of each other. Therefore, we can conclude that $\mathcal{C}'$ is prefix-free.

3. According to the definition of $\mathcal{K}(n) = \mathcal{C}_1(n+1)$ and $\text{length}(\mathcal{C}_1(n)) \leq 2 + 2\log(1 + \log n) + \log n$, we have

$$\text{length}(\mathcal{K}(n)) = \text{length}(\mathcal{C}_1(n+1)) \leq 2 + 2\log(1 + \log(n+1)) + \log(n+1)$$

For any non-singular code $\mathcal{C}$ with $E[\text{length}(\mathcal{C}(U))] = L$, we have

$$\begin{aligned} E[\text{length}(\mathcal{C}'(U))] &= E[\text{length}(\mathcal{K}(\text{length}(\mathcal{C}(U)))) + \text{length}(\mathcal{C}(U))] \\ &= E[\text{length}(\mathcal{K}(\text{length}(\mathcal{C}(U))))] + E[\text{length}(\mathcal{C}(U))] \\ &\leq E[2 + 2\log(1 + \log(\text{length}(\mathcal{C}(U)) + 1)) + \log(\text{length}(\mathcal{C}(U)) + 1)] + L \\ &\leq L + 2 + 2\log(1 + \log(1 + L)) + \log(1 + L). \end{aligned}$$

where the last step is obtained by using Jensen's inequality for concave and increasing function $\log(x)$.

Problem 5. *(Missing Data)*

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

$$X_1, X_2, X_3, X_4, X_5, X_6, ?, X_8, X_9, \ldots \tag{33}$$

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 $\hat{X}_7 = w_1X_6 + w_2X_8$. 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.

Solution:

1. $\hat{X}_7 = \alpha X_6$ and $\mathcal{E} = 1 - \alpha^2$. The corresponding Wiener filter has

$$R_x = \alpha^0 = 1 \qquad\qquad r_{dx} = \mathbb{E}[X_6X_7] = \alpha$$

2. $\hat{X}_7 = \frac{\alpha}{1+\alpha^2}(X_6 + X_8)$ and $\mathcal{E} = \frac{1-\alpha^2}{1+\alpha^2}$ (better than Part 1). The corresponding Wiener filter has

$$R_x = \begin{bmatrix} 1 & \alpha^2 \\ \alpha^2 & 1 \end{bmatrix} \qquad\qquad r_{dx} = \begin{bmatrix} \alpha \\ \alpha \end{bmatrix}$$

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

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

   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 $\|p\|_2^2 = \sum_i p_i^2$.

Solution:

1. There are $\binom{n}{2}$ pairs. For each pair the chance that both values agree is equal to $\sum_i p_i^2 = \|p\|_2^2$.

2. Let u be the vector of length k with all-one entries. Then, by using the Cauchy-Schwartz inequality, $\|p\|_2^2 = \langle p, p\rangle \geq \langle p, u\rangle^2/\langle u, u\rangle = 1/k$.

3. Expanding the expression, we get

$$\|p - u\|_2^2 = \|p\|_2^2 - 2\langle p, u\rangle + \|u\|_2^2 = \|p\|_2^2 - 2/k + 1/k = \|p\|_2^2 - 1/k.$$

4. Recall that in order to count collisions we look at pairs of indices in our samples. Let (i, j), $1 \leq i < j \leq n$, be one such pair. When computing the variance we are looking at *pairs of pairs*. E.g., (i, j) and (u, v). There are four parts in the expression for the variance. These have the following interpretation. The first part comes from all pairs with *total* overlap, i.e., $(i, j) = (u, v)$. There are $\binom{n}{2}$ such cases. The second part comes from pairs where exactly one index is repeated. The third term comes from pairs with no overlap. And the fourth term is the mean squared so that we convert from the second moment to the variance.

5. By the Chebyshev's inequality, if $C(X^n)$ counts the number of collisions in our sample then, assuming that the sample comes from the uniform distribution,

$$\Pr\{C(X^n) - \binom{n}{2}\frac{1}{k} \geq \binom{n}{2}\frac{\kappa}{2k}\} \leq \frac{n^2/(2k)}{\binom{n}{2}^2\frac{\kappa^2}{4k^2}} \leq \frac{k}{n^2\kappa^2}.$$

   Therefore, as long as n is large compared to $\sqrt{k/\kappa^2}$ the right-hand side goes to zero. In other words, we need $\Theta(\sqrt{k}/\kappa^2)$ samples.