Solutions to Homework Set #2

1. **Maximum entropy processes.**
   Find the maximum entropy rate stochastic processes $\{X_i\}_{-\infty}^{\infty}$ subject to the constraints:

   (a) $EX_i^2 = 1, \quad i = 1, 2, \ldots,$
   (b) $EX_i^2 = 1,\ EX_iX_{i+1} = \frac{1}{2}, \quad i = 1, 2, \ldots.$

   Find the maximum entropy spectrum for the processes in parts (a) and (b).

Solution: Maximum entropy processes

(a) If the only constraint is $EX_i^2 = 1$, by Burg's theorem, it is clear that the maximum entropy process is a 0-th order Gauss-Markov, i.e., X_i i.i.d. $\sim \mathcal{N}(0,1)$.
From Eq. (11.53) in the textbook, the maximum entropy spectrum is

$$S(\omega) = \frac{1}{|1|^2} = 1.$$

(b) If the constraints are $EX_i^2 = 1, EX_iX_{i+1} = \frac{1}{2}$, the maximum entropy process is a first order Gauss-Markov process of the form

$$X_i = -aX_{i-1} + Z_i, \quad Z_i \sim \mathcal{N}(0, \sigma^2).$$

To determine a and σ^2, we use the Yule-Walker equations

$$\begin{aligned} R_0 &= -aR_1 + \sigma^2 \\ R_1 &= -aR_0 \end{aligned}$$

Substituting $R_0 = 1$ and $R_1 = \frac{1}{2}$, we get $a = -\frac{1}{2}$ and $\sigma^2 = \frac{3}{4}$. Hence the maximum entropy process is

$$X_i = \frac{1}{2}X_{i-1} + Z_i, \quad Z_i \sim \mathcal{N}(0, \frac{3}{4}).$$

The maximum entropy spectrum is

$$S(\omega) = \frac{3}{4} \cdot \frac{1}{|1 + \frac{1}{2}e^{-i\omega}|^2}.$$

2. **Maximum entropy discrete processes.**

   (a) Find the maximum entropy rate binary stochastic process $\{X_i\}_{i=-\infty}^{\infty}$, $X_i \in \{0,1\}$, satisfying $\Pr\{X_i = X_{i+1}\} = \frac{1}{3}$, for all i.

   (b) What is the resulting entropy rate?

Solution: Maximum entropy discrete processes

Our first hope may be an i.i.d. Bern(p) process that is consistent with the constraint. Unfortunately, there is no such process. (Check!) However, we can still construct an independent *non-identically* distributed sequence of Bernoulli r.v.'s, such that the entropy rate exists, and the constraints are met. (For example, $X_i \sim$ Bern(1) for odd i and $X_i \sim$ Bern(1/3) for even i.) This process does not yield the maximum entropy rate.

This problem is, in fact, a discrete (or more precisely, binary) analogue of Burg's maximum entropy theorem and we can obtain the maximum entropy process from a similar argument.

(a) Let X_i be any binary process satisfying the constraint $\Pr\{X_i = X_{i+1}\} = 1/3$. Let Z_i be a first order stationary Markov chain, that stays at 0 with probability 1/3, jumps to 1 with probability 2/3, and vice versa. This process obviously meets the constraint. With a slight abuse of notation, we have

$$\begin{aligned} H(X_i|X_{i-1}) &= \mathbf{E}H(X_i|X_{i-1}=x) \\ &= \mathbf{E}H(\Pr(X_i = x|X_{i-1}=x)) \\ &\le H(\mathbf{E}\Pr(X_i = x|X_{i-1}=x)) \\ &= H(1/3) \\ &= H(Z_i|Z_{i-1}), \end{aligned}$$

where the inequallity follows from the concavity of the binary entropy function. Since $H(X_1) \le 1 = H(Z_1)$,

$$\begin{aligned} H(X_1,\ldots,X_n) &= H(X_1) + \sum_{i=2}^{n} H(X_i|X^{i-1}) \\ &\le H(X_1) + \sum_{i=2}^{n} H(X_i|X_{i-1}) \\ &\le H(Z_1) + \sum_{i=2}^{n} H(Z_i|Z_{i-1}) \\ &= H(Z_1,\ldots,Z_n), \end{aligned}$$

whence $\{Z_i\}$ is the maximum entropy process under the given constraint.

(b) The maximum entropy rate $H(\mathcal{Z}) = H(Z_2|Z_1) = H(1/3) = \log 3 - 2/3$.

3. **Processes With Fixed Marginals**
Consider the set of all densities with fixed pairwise marginals

$$f_{12}(x_1, x_2), f_{23}(x_2, x_3), \ldots, f_{n-1,n}(x_{n-1}, x_n).$$

Show that the maximum entropy process with these marginals is the first-order (possibly time-varying) Markov process with these marginals. Identify the maximizing $f^*(x_1, x_2, \ldots, x_n)$.

Solution: Processes With Fixed Marginals
By the chain rule,

$$h(X_1, X_2, \ldots, X_n) = h(X_1) + \sum_{i=2}^{n} h(X_i|X_{i-1}, \ldots, X_1) \quad (1)$$

$$\leq h(X_1) + \sum_{i=2}^{n} h(X_i|X_{i-1}), \quad (2)$$

since conditioning reduces entropy. The quantities $h(X_1)$ and $h(X_i|X_{i-1})$ depend only on the second order marginals of the process and hence the upper bound is true for all processes satisfying the second order marginal constraints.

Define

$$f^*(x_1, x_2, \ldots, x_n) = f_0(x_1) \prod_{i=2}^{n} \frac{f_0(x_{i-1}, x_i)}{f_0(x_{i-1})}. \quad (3)$$

We will show that f^* maximizes the entropy among all processes with the same second order marginals. To prove this, we just have to show that this process has the same second order marginals and that this process achieves the upper bound (2). The fact that the process satisfies the marginal constraints can be easily proved by induction. Clearly, it is true for $f^*(x_1, x_2)$ and if $f^*(x_{i-1}, x_i) = f_0(x_{i-1}, x_i)$, then $f^*(x_i) = f_0(x_i)$ and by the definition of f^*, it follows that $f^*(x_i, x_{i+1}) = f_0(x_i, x_{i+1})$. Also, since by definition, f^* is first order Markov, $h(X_i|X_{i-1}, \ldots, X_1) = h(X_i|X_{i-1})$ and we have equality in (2). Hence f^* has the maximum entropy of all processes with the same second order marginals.

4. **Maximum entropy of sums.**
Let $Y = X_1 + X_2$. Find the maximum entropy (over all distributions on X_1 and X_2) of Y under the constraint $EX_1^2 = P_1, \quad EX_2^2 = P_2$,

(a) if X_1 and X_2 are independent.

(b) if X_1 and X_2 are allowed to be dependent.

Solution: Maximum entropy of sums
Assume that $Y = X_1 + X_2$ has some distribution with the variance σ^2. Since the Gaussian distribution maximizes the entropy under the variance constraint,

$$h(Y) \le h(Y') = \frac{1}{2}\log(2\pi e\sigma^2),$$

where $Y' = X_1' + X_2'$ is a Gaussian random variable with the same variance. Therefore, to maximize the entropy, we should maximize the variance σ^2 over all jointly Gaussian X_1 and X_2 with $EX_1^2 = P_1, \quad EX_2^2 = P_2$.

(a) Independence implies that

$$\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = P_1 - \mu_1^2 + P_2 - \mu_2^2.$$

To maximize the variance of Y, we take $\mu_1 = \mu_2 = 0$. This yields the variance $P_1 + P_2$ and the entropy $(1/2)\log 2\pi e(P_1 + P_2)$.

(b) In this case,

$$\mathrm{Var}(Y) = P_1 + P_2 + 2\rho\sqrt{P_1 P_2} - \mu_1^2 - \mu_2^2,$$

where $\rho \in [-1, 1]$ is the correlation coefficient between X_1 and X_2. The variance of Y is maximized for $\mu_1 = \mu_2 = 0$, and $\rho = 1$. Note that $\rho = 1$ means that the two random variables are adding coherently. The maximum entropy is then $(1/2)\ln 2\pi e(P_1 + P_2 + 2\sqrt{P_1 P_2})$.

5. **Hadamard.**
Let K be a $2n \times 2n$ nonnegative definite symmetric matrix. Show

$$\det(K) \leq \prod_{i=1}^{n} \det\left(K(2i-1, 2i)\right),$$

where $K(i,j)$ denotes the 2×2 submatrix

$$\begin{pmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{pmatrix}.$$

Solution to: Hadamard.
Since K is a nonnegative definite symmetric matrix, we can think of K as the covariance matrix of a Gaussian vector $(X_1, \ldots, X_{2n})$. Therefore

$$\begin{aligned}
\frac{1}{2}\log\big((2\pi e)^{2n}\det(K)\big) &= h(X_1, \ldots, X_{2n}) \\
&\leq \sum_{i=1}^{n} h(X_{2i-1}, X_{2i}) \\
&= \sum_{i=1}^{n} \frac{1}{2}\log\big((2\pi e)^2 \det(K(2i-1, 2i))\big) \\
&= \frac{1}{2}\log\left(\prod_{i=1}^{n}(2\pi e)^2 \det(K(2i-1, 2i))\right) \\
&= \frac{1}{2}\log\left((2\pi e)^{2n}\prod_{i=1}^{n}\det(K(2i-1, 2i))\right).
\end{aligned}$$

Since log is an increasing function we get

$$\det(K) \leq \prod_{i=1}^{n} \det(K(2i-1, 2i)).$$

6. **Maximum entropy.**

(a) What is the parametric form maximum entropy density $f(x)$ satisfying the two conditions

$$EX^8 = a$$
$$EX^{16} = b.$$

Don't solve for the λ's.

(b) What is the maximum entropy density satisfying the condition

$$E(X^8 + X^{16}) = a + b \quad ?$$

Solution: Maximum entropy.

(a) The maximun entropy distribution is given by

$$f(x) = \frac{e^{\lambda_1 x^8 + \lambda_2 x^{16}}}{\int e^{\lambda_1 x^8 + \lambda_2 x^{16}} dx},$$

where λ_1 and λ_2 are chosen so that $EX^8 = a$ and $EX^{16} = b$.

(b) Now the maximun entropy distribution is given by

$$g(x) = \frac{e^{\lambda(x^8 + x^{16})}}{\int e^{\lambda(x^8 + x^{16})} dx},$$

where λ is chosen so that $E(X^8 + X^{16}) = a + b$. Note that the distribution $f(x)$ obtained in part (a) also satisfies the given condition, but in general, it does not result in the maximum entropy.

7. **Mutual information.**
Recall that

$$I(X;Y) = \sup_{[\ ],[\ ]} I([X];[Y])$$

to calculate $I(X;Y)$, where the distribution of (X,Y) is the $(\lambda, 1-\lambda)$ mixture of the density $f(x,y)$ and the probability mass function

$$(X,Y) = \begin{cases} (1,1), & \frac{1}{2}q \\ (1,2), & \frac{1}{2}p \\ (2,1), & \frac{1}{2}p \\ (2,2), & \frac{1}{2}q \end{cases}.$$

Solution: Mutual information.
Since the distribution of (X,Y) is a mixture and can't be represented with either a mass or a density function, it will be helpful to construct a variable $Z \sim Bern(\lambda)$, and let the conditional distribution of (X,Y) given $Z=0$ be $f(x,y)$, and let the conditional distribution of (X,Y) given $Z=1$ be the probability mass function given above. After marginalizing out Z, we are left with the appropriate joint distribution on (X,Y).

For any triple of random variables we can derive the following equality based on expanding mutual information using the chain rule in two different ways:

$$\begin{aligned} I(X;Y,Z) &= I(X;Y) + I(X;Z|Y) \\ &= I(X;Z) + I(X;Y|Z). \end{aligned}$$

Therefore,

$$I(X;Y) = I(X;Y|Z) + I(X;Z) - I(X;Z|Y).$$

The key here is that $I(X;Y|Z)$ is easy to evaluate. But what about the rest? We can write the equality again in a way that will separate out the terms that need special care:

$$I(X;Y) = I(X;Y|Z) + H(Z) - (H(Z|X) + I(X;Z|Y)).$$

We will show that the quantities in the parenthesis on the right-hand side are equal to zero. This is because Z is completely known by observing X or Y. If X or Y corresponds to one of the mass points than $Z=1$, otherwise $Z=0$. However, since the two conditional supports are not actually disjoint, we will approach this carefully using the partition definition of mutual information. Notice that the upper bound on $I(X;Y)$

follows immediately from the non-negativity of the quantities in the parenthesis. It remains to provide the lower bound.

Let X_Δ and Y_Δ be quantized versions of X and Y under a partitioning that divides the real numbers into intervals of size Δ. Then,

$$\begin{aligned}
\sup_{[\ ],[\ ]} I([X];[Y]) &\geq \lim_{\Delta\to 0} I(X_\Delta;Y_\Delta) \\
&= \lim_{\Delta\to 0} I(X_\Delta;Y_\Delta|Z) + H(Z) - (H(Z|X_\Delta) + H(Z|Y_\Delta) - H(Z|X_\Delta,Y_\Delta)) \\
&= I(X;Y|Z) + H(Z) - \lim_{\Delta\to 0}(H(Z|X_\Delta) + H(Z|Y_\Delta) - H(Z|X_\Delta,Y_\Delta)).
\end{aligned}$$

Consider $H(Z|X_\Delta)$. We can infer the distribution $p(z|x_\Delta)$ using Bayes rule. If x_Δ doesn't contain a mass point, then

$$P(Z=0|X_\Delta = x_\Delta) = 1.$$

If x_Δ does contain mass points of total probability α, then

$$P(Z=0|X_\Delta = x_\Delta) = \frac{\lambda \int_\Delta f(x)dx}{\lambda \int_\Delta f(x)dx + (1-\lambda)\alpha},$$

which converges to zero as Δ converges to zero.

Therefore, $H(Z|X_\Delta) \to 0$, as does $H(Z|Y_\Delta)$ and $H(Z|X_\Delta,Y_\Delta)$.

We conclude that

$$\begin{aligned}
I(X;Y) &= I(X;Y|Z) + H(Z) \\
&= \lambda I(f(x,y)) + (1-\lambda)(1-H(p)) + H(\lambda).
\end{aligned}$$