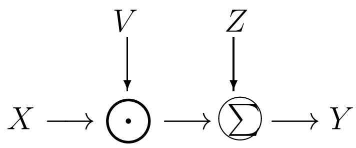
2nd Semester 2010

Solutions to Homework Set #6
Source channel separation, Max entropy principle, and channel coding with side information

1. **Fading channel.**
Consider an additive noise fading channel



$$Y = XV + Z,$$

where Z is additive noise, V is a random variable representing fading, and Z and V are independent of each other and of X.

(a) Argue that knowledge of the fading factor V improves capacity by showing
$$I(X;Y|V) \geq I(X;Y).$$

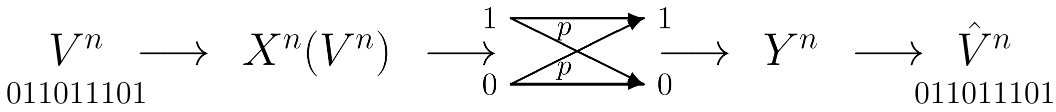
(b) Incidentally, conditioning does not always increase mutual information. Give an example of $p(u,r,s)$ such that $I(U;R|S) < I(U;R)$.

Solution: Fading channel
We may show the inequality as follows:

$$\begin{aligned} I(X;Y|V) &= h(X|V) - h(X|Y,V) \\ &= h(X) - h(X|Y,V) && (1) \\ &\geq h(X) - h(X|Y) && (2) \\ &= I(X;Y) \end{aligned}$$

where (1) follows from the independence of X and V, and (2) is true because conditioning reduces entropy.

2. **Source and channel.**
We wish to encode a Bernoulli(α) process $V_1, V_2, \ldots$ for transmission over a binary symmetric channel with error probability p.



Find conditions on α and p so that the probability of error $P(\hat{V}^n \neq V^n)$ can be made to go to zero as $n \longrightarrow \infty$.

Solution: Source and channel.
Suppose we want to send a binary i.i.d. Bernoulli(α) source over a binary symmetric channel with error probability p.

By the source-channel separation theorem, in order to achieve the probability of error that vanishes asymptotically, i.e. $P(\hat{V}^n \neq V^n) \to 0$, we need the entropy of the source to be less than the capacity of the channel. Hence,

$$H(\alpha) + H(p) < 1,$$

or, equivalently,

$$\alpha^\alpha (1-\alpha)^{1-\alpha} p^p (1-p)^{1-p} < \frac{1}{2}.$$

3. **Maximum entropy.**
Find the maximum entropy density f satisfying $EX = \alpha_1, E \ln X = \alpha_2$. That is,

$$\text{maximize} \quad h(f)$$

subject to $\int x f(x)\,dx = \alpha_1, \int (\ln x) f(x)\,dx = \alpha_2$. What family of densities is this?

Solution: Maximum entropy.
As derived in class, the maximum entropy distribution subject to constraints

$$\int x f(x)\,dx = \alpha_1 \tag{3}$$

and

$$\int (\ln x) f(x)\, dx = \alpha_2 \tag{4}$$

is of the form

$$f(x) = e^{\lambda_0 + \lambda_1 x + \lambda_2 \ln x} = c x^{\lambda_2} e^{\lambda_1 x}, \tag{5}$$

which is a Gamma distribution. The constants should be chosen to satisfy the constraints.

4. **Minimum relative entropy $D(P \parallel Q)$ under constraints on P.** We wish to find the (parametric form) of the probability mass function $P(x), x \in \{1, 2, \ldots\}$ that minimizes the relative entropy $D(P \parallel Q)$ over all P such that $\sum P(x) g_i(x) = \alpha_i, i = 1, 2, \ldots$.

   (a) Use Lagrange multipliers to guess that

   $$P^*(x) = Q(x) e^{\sum_{i=1}^{\infty} \lambda_i g_i(x) + \lambda_0}$$

   achieves this minimum if there exist λ_i's satisfying the α_i constraints. This generalizes the theorem on maximum entropy distributions subject to constraints.

   (b) Verify that P^* minimizes $D(P \parallel Q)$.

   Solution: Minimize relative entropy $D(P||Q)$ under constraints on P.

   (a) We construct the functional using Lagrange multipliers

   $$J(P) = \int P(x) \ln \frac{P(x)}{Q(x)} + \sum_i \lambda_i \int P(x) h_i(x) + \lambda_0 \int P(x). \tag{6}$$

   'Differentiating' with respect to $P(x)$, we get

   $$\frac{\partial J}{\partial P} = \ln \frac{P(x)}{Q(x)} + 1 + \sum_i \lambda_i h_i(x) + \lambda_0 = 0, \tag{7}$$

   which indicates that the form of $P(x)$ that minimizes the Kullback Leibler distance is

   $$P^*(x) = Q(x) e^{\lambda_0 + \sum_i \lambda_i h_i(x)}. \tag{8}$$

(b) Though the Lagrange multiplier method correctly indicates the form of the solution, it is difficult to prove that it is a minimum using calculus. Instead we use the properties of $D(P||Q)$. Let P be any other distribution satisfying the constraints. Then

$$\begin{aligned}
D(P||Q) &- D(P^*||Q) && (9)\\
&= \int P(x)\ln\frac{P(x)}{Q(x)} - \int P^*(x)\ln\frac{P^*(x)}{Q(x)} && (10)\\
&= \int P(x)\ln\frac{P(x)}{Q(x)} - \int P^*(x)[\lambda_0 + \sum_i \lambda_i h_i(x)] && (11)\\
&= \int P(x)\ln\frac{P(x)}{Q(x)} - \int P(x)[\lambda_0 + \sum_i \lambda_i h_i(x)] && \text{(since both } P \text{ and } P^* \text{ satisfy the constraints)}\\
&= \int P(x)\ln\frac{P(x)}{Q(x)} - \int P(x)\ln\frac{P^*(x)}{Q(x)} && (12)\\
&= \int P(x)\ln\frac{P(x)}{P^*(x)} && (13)\\
&= D(P||P^*) && (14)\\
&\geq 0, && (15)
\end{aligned}$$

and hence P^* uniquely minimizes $D(P||Q)$.

In the special case when Q is a uniform distribution over a finite set, minimizing $D(P||Q)$ corresponds to maximizing the entropy of P.

5. **Maximum entropy with marginals.**
What is the maximum entropy probability mass function $p(x, y)$ with the following marginals? You may wish to guess and verify a more general result.

	y_1	y_2	y_3	
x_1	p_{11}	p_{12}	p_{13}	1/2
x_2	p_{21}	p_{22}	p_{23}	1/4
x_3	p_{31}	p_{32}	p_{33}	1/4
	2/3	1/6	1/6	

Solution: Maximum entropy with marginals.
Given the marginal distributions of X and Y, $H(X)$ and $H(Y)$ are fixed. We may write

$$H(X,Y) = H(X) + H(Y|X) \leq H(X) + H(Y), \tag{16}$$

with equality if and only if X and Y are independent. Hence the maximum value of $H(X,Y)$ is $H(X)+H(Y)$, and is attained by choosing the joint distribution to be the product distribution, i.e.,

	y_1	y_2	y_3	
x_1	1/3	1/12	1/12	1/2
x_2	1/6	1/24	1/24	1/4
x_3	1/6	1/24	1/24	1/4
	2/3	1/6	1/6	

This problem can also be solved by using the maximum entropy distribution with the $r_i(x,y)$ as indicator functions on x and y for each of the six constraints, and recognizing that the solution is the product distribution.