EE376B/Stat 376B
Information Theory
Prof. T. Cover

Handout #6
Tuesday, April 14, 2009

Solutions to Homework Set #1

1. **Differential entropy.**
Evaluate the differential entropy $h(X) = -\int f \ln f$ for the following:

   (a) The Laplace density, $f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$. Relate this to the entropy of the exponential density $\lambda e^{-\lambda x},\ x \geq 0$.

   (b) The sum of X_1 and X_2, where X_1 and X_2 are independent normal random variables with means μ_i and variances $\sigma_i^2, \quad i = 1, 2$.

Solution: Differential entropy.

(a) Laplace density.

Note that the Laplace density is a two sided exponential density, so each side has a differential entropy of the exponential and one bit is needed to specify which side. So for $f_0(x) = \lambda e^{-\lambda x}$, $x \geq 0$ we have,

$$h(f) = \frac{1}{2}h(f_0(x)) + \frac{1}{2}h(f_0(-x)) + H\left(\frac{1}{2}\right) \tag{1}$$

$$= \log\frac{e}{\lambda} + \log 2 \text{ bits.} \tag{2}$$

$$= \log\frac{2e}{\lambda} \text{ bits.} \tag{3}$$

(b) Sum of two independent normal distributions.

The sum of two independent normal random variables is also normal, so applying the result derived the class for the normal distribution, since $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

$$h(f) = \frac{1}{2}\log 2\pi e(\sigma_1^2 + \sigma_2^2) \text{ bits.} \tag{4}$$

2. **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

$$H(X, Y) = H(X) + H(Y|X) \leq H(X) + H(Y), \tag{5}$$

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 from Theorem 11.1.1 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.

3. **Maximum entropy of atmosphere.**
Maximize $h(Z, V_x, V_y, V_z)$, $\quad Z \geq 0$, $\quad (V_x, V_y, V_z) \in R^3$, subject to the energy constraint $E(\frac{1}{2}m \parallel V \parallel^2 + mgZ) = E_0$. Show that the resulting distribution yields

$$E\frac{1}{2}m \parallel V \parallel^2 = \frac{3}{5}E_0$$

$$EmgZ = \frac{2}{5}E_0.$$

Thus $\frac{2}{5}$ of the energy is stored in the potential field, regardless of its strength g.

Solution: Maximum entropy of atmosphere.

As derived in class, the maximum entropy distribution subject to the constraint

$$E\left(\frac{1}{2}m\|v\|^2 + mgZ\right) = E_0 \tag{6}$$

is of the form $f(z, v_x, v_y, v_z) = Ce^{-\lambda\left(\frac{1}{2}m\|v\|^2+mgZ\right)} = Ce^{-\frac{\lambda m}{2}v_x^2}e^{-\frac{\lambda m}{2}v_y^2}e^{-\frac{\lambda m}{2}v_z^2}e^{-\lambda mgZ}$. We recognize this as a product distribution of four independent random variables with $V_x, V_y, V_z \sim \mathcal{N}(0, \frac{1}{\lambda m})$ and $Z \sim \exp(\lambda mg)$. Therefore,

$$\begin{aligned} E(mgZ) &= mg\left(\frac{1}{\lambda mg}\right) \\ &= \frac{1}{\lambda} \qquad (7) \\ E(\frac{1}{2}mv_x^2) &= \frac{1}{2}m\left(\frac{1}{\lambda m}\right) \\ &= \frac{1}{2\lambda} \qquad (8) \end{aligned}$$

The constraint on energy yields $\frac{1}{\lambda} = \frac{2}{5}E_0$. This immediately gives $E(mgZ) = \frac{2}{5}E_0$ and $\mathbf{E}\left(\frac{1}{2}m\|v\|^2\right) = \frac{3}{5}E_0$. The split of energy between kinetic energy and potential energy is $\frac{2}{3}$ regardless of the strength of gravitational field g.

4. **Gaussian mutual information.**
Suppose that (X, Y, Z) are jointly Gaussian and that $X \to Y \to Z$ forms a Markov chain. Let X and Y have correlation coefficient ρ_1 and let Y and Z have correlation coefficient ρ_2. Find $I(X; Z)$.

Solution: Gaussian mutual information.

First note that we may without any loss of generality assume that the means of X, Y and Z are zero. If in fact the means are not zero one can subtract the vector of means without affecting the mutual information or the conditional independence of X, Z given Y. Similary we can also assume the variances of X, Y, and Z to be 1. (The scaling may change the differential entropy, but not the mutual information.)

Let

$$\Sigma = \begin{pmatrix} 1 & \rho_{xz} \\ \rho_{xz} & 1 \end{pmatrix},$$

be the covariance matrix of X and Z. From Eqs. (9.93) and (9.94)

$$\begin{aligned} I(X;Z) &= h(X)+h(Z)-h(X,Z) \\ &= \frac{1}{2}\log(2\pi e)+\frac{1}{2}\log(2\pi e)-\frac{1}{2}\log(2\pi e\det(\Sigma)) \\ &= -\frac{1}{2}\log(1-\rho_{xz}^2) \end{aligned}$$

Now from the conditional independence of X and Z given Y, we have

$$\begin{aligned} \rho_{xz} &= \mathbf{E}[XZ] \\ &= \mathbf{E}\left[\mathbf{E}[XZ|Y]\right] \\ &= \mathbf{E}\left[\mathbf{E}[X|Y]\cdot\mathbf{E}[Z|Y]\right] \\ &= \mathbf{E}[\rho_1 Y\cdot\rho_2 Y] \\ &= \rho_1\rho_2. \end{aligned}$$

We can thus conclude that

$$I(X;Z) = -\frac{1}{2}\log(1-\rho_1^2\rho_2^2)$$

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

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

$$\int xf(x)\,dx=\alpha_1 \tag{9}$$

and

$$\int(\ln x)f(x)\,dx=\alpha_2 \tag{10}$$

is of the form

$$f(x)=e^{\lambda_0+\lambda_1 x+\lambda_2\ln x}=cx^{\lambda_2}e^{\lambda_1 x}, \tag{11}$$

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

6. **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

$$P^*(x) = Q(x) e^{\sum_{i=1}^{\infty} \lambda_i g_i(x) + \lambda_0} \tag{12}$$

   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

$$J(P) = \int P(x) \ln \frac{P(x)}{Q(x)} + \sum_i \lambda_i \int P(x) h_i(x) + \lambda_0 \int P(x). \tag{13}$$

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

$$\frac{\partial J}{\partial P} = \ln \frac{P(x)}{Q(x)} + 1 + \sum_i \lambda_i h_i(x) + \lambda_0 = 0, \tag{14}$$

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

$$P^*(x) = Q(x) e^{\lambda_0 + \sum_i \lambda_i h_i(x)}. \tag{15}$$

(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

$$\begin{aligned}
D(P||Q) \;&-\; D(P^*||Q) && (16)\\
&= \int P(x) \ln \frac{P(x)}{Q(x)} - \int P^*(x) \ln \frac{P^*(x)}{Q(x)} && (17)\\
&= \int P(x) \ln \frac{P(x)}{Q(x)} - \int P^*(x) [\lambda_0 + \sum_i \lambda_i h_i(x)] && (18)\\
&= \int P(x) \ln \frac{P(x)}{Q(x)} - \int P(x) [\lambda_0 + \sum_i \lambda_i h_i(x)] && \text{(since both } P \text{ and } P^* \text{ satisfy the constraints)}\\
&= \int P(x) \ln \frac{P(x)}{Q(x)} - \int P(x) \ln \frac{P^*(x)}{Q(x)} && (19)
\end{aligned}$$

$$= \int P(x) \ln \frac{P(x)}{P^*(x)} \tag{20}$$
$$= D(P||P^*) \tag{21}$$
$$\geq 0, \tag{22}$$

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.

7. **Every density is a maximum entropy density for some constraint.**
We wish to show that any density f_0 can be considered to be a maximum entropy density. Let $f_0(x)$ be a density and consider the problem of maximizing $h(f)$ subject to the constraint

$$\int f(x) r(x)\, dx = \alpha$$

where $r(x) = \ln f_0(x)$. Show that there is a choice of α, $\alpha = \alpha_0$, such that the maximizing distribution is $f^*(x) = f_0(x)$. Thus $f_0(x)$ is indeed a maximum entropy density under the constraint $\int f \ln f_0 = \alpha_0$.

Solution: Every density is a maximum entropy density for some constraint.
Given the constraints that

$$\int r(x) f(x) = \alpha \tag{23}$$

the maximum entropy density is

$$f^*(x) = e^{\lambda_0 + \lambda_1 r(x)} \tag{24}$$

With $r(x) = \log f_0(x)$, we have

$$f^*(x) = \frac{f_0^{\lambda_1}(x)}{\int f_0^{\lambda_1}(x)\, dx} \tag{25}$$

where λ_1 is chosen to satisfy the constraint. We can choose the value of the constraint to correspond to the value $\lambda_1 = 1$, i.e., $\alpha_0 = \int f \ln f_0$, in which case $f^* = f_0$. So f_0 is a maximum entropy density under appropriate constraints.