STAT 538 Homework 5
Out February 16, 2012
Due Thursday, February 23, 2012

mmp@stat.washington.edu

For the problems in this homework, which all deal with convexity, try to find the most elegant solution. Elegant can mean that you give a short proof, based on an example in the textbook, or a property proven in the text, instead of a long proof starting from the definitions.

Problem 1 – Some functions on the probability simplex. B&V problem 3.24

(Transferred from previous homework) Only **a, b, e, f**.

Problem 2 – Log-concavity (Transferred from previous homework) Do one of **BV 3.52, 3.53**

Solve either 3A or 3B Bring all expression to a simple, intelligible form for full credit.

Problem 3A – The conjugate of the entropy

In this problem you will compute the conjugate of the entropy of a discrete distribution. Reading B&V Chapter 3 and Lecture 7, Section 1. Bring all your results to a simple interpretable form.

1. Binary domain Assume that the domain of X is $\mathcal{X} = \{0, 1\}$, so that any distribution on $\mathcal{X}$ is parametrized by $p = Pr[X = 1]$. The entropy is a concave function, so we will compute the conjugate of the negative entropy $f(p) = -H(p)$ with $H(p) = -p \ln p - (1-p)\ln(1-p)$.

1.1 Compute the gradient of f

1.2 Solve the equation $f'(p) = q$. What is the domain of q, for which a solution exists?

1.3 Now compute the expression of the conjugate $f^*(q)$ as a function of q and verify that it is convex.

2. Arbitrary finite domain Now let $\mathcal{X} = \{1, \ldots, m\}$ and let $p = (p_1, \ldots p_m)$ parametrize a distribution on it.

2.1 Note that one of the variables p_i is redundant being completely determined by the other $m-1$. Therefore you need to express one of the variables as a function of the others, e.g $p_m = 1 - \sum_{j=1:m-1} p_j$. Write the expression of the entropy $H(p)$ as a function of the $m-1$ free variables $p_1, \dots p_{m-1}$.

2.2 Evaluate the conjugate of $f(p) = -H(p)$ in this case. To get full credit you need to arrive at a simple expression, that is symmetric in all the variables.

2.3 Optional-for extra credit The expression of the conjugate function often depends on the domain of f. Re-evaluate the conjugate of $f(p)$ as a function of m free variables and compare with the result in **2.2**.

Problem 3B – Entropies, KL divergences, mutual informations

a. Calculate the entropy of a univariate normal distribution with mean μ and variance σ^2. Explain why the expression you obtain does not depend on μ.

[**b. - Optional, extra credit**] Calculate the entropy of a multivariate normal distribution with mean 0 and covariance matrix Σ.

c. Calculate the KL divergence between two univariate normal distributions $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$.

d. Let $X, Y \in \mathbb{R}$ be jointly normal with means 0. Find the expression of the mutual information $I(X, Y)$.

e. $X, Y \in \{0, 1\}$ have joint distribution given by $P[X = 1, Y = 1] = P[X = 0, Y = 0] = p/2$, $P[X = 1, Y = 0] = P[X = 0, Y = 1] = q/2$, $p + q = 1$. Find the expression of the mutual information $I(X, Y)$ as a function of p, q. For what values of p, q is $I(X, Y)$ maximized and minimized? What other information theoretic expression is represented by $I(X, Y)$?

f. Let $X \in \{0, 1\}$, $Y \in \mathbb{R}$ be two random variables, with $P_X(1) = p$, $P_{Y|X} = N(X, \sigma^2)$. In other words, P_Y is a mixture of two normal distributions.

For $p = 0.25$ and $p = 0.5$ calculate (numerically) the entropy $H(Y; \sigma)$ as a function of σ. Plot the values obtained for $\sigma \in [0.05, 1]$.

g. Explain how you can obtain the mutual information $I(X, Y)$ using the numerical values computed above. Plot $I(X, Y)$ as a function of σ for $p = 0.25$ and $p = 0.5$, $\sigma \in [0.05, 1]$.

Explain in a sentence or two the behavior observed towards small σ values. What is the value towards which $I(X, Y)$ converges when $\sigma \to 0$? (No rigurous proof required for this question.)