STOR 654 Homework 13

1. Consider the Gaussian sequence model with observations $Y \sim \mathcal{N}_n(\theta, I)$.

   a. Show that for linear estimators of the form $\hat{\theta}_C(y) = Cy$ where $C \in \mathbb{R}^{n \times n}$ the bias-variance decomposition yields

   $$r(\hat{\theta}_C, \theta) = \sigma^2 \text{tr}(C^t C) + ||(I - C)\theta||^2$$

   b. Argue that SURE for $\hat{\theta}_C$ is given by $U_C(y) = -n + 2\,\text{tr}(C) + ||(I - C)y||^2$.

   c. Show that $\mathbb{E}_\theta\, U_C(Y)$ is equal to the expression for $r(\hat{\theta}_C, \theta)$ above.

2. Let $\mathcal{X}$ be a finite set and let p and q be pmfs on $\mathcal{X}$.

   a. Show that $\text{KL}(p : q)$ is infinite if and only if there is some $x \in \mathcal{X}$ with $q(x) = 0$ and $p(x) > 0$. (This simple relation does *not* hold when $\mathcal{X}$ is infinite.)

Let $f : \mathcal{X} \to \mathbb{R}$ be any function on $\mathcal{X}$. Define a new pmf on $\mathcal{X}$ by "exponentially tilting" q according to f as follows

$$q_f(x) \;=\; \frac{e^{f(x)}\, q(x)}{Z_f}$$

where $Z_f = \sum_{x \in \mathcal{X}} e^{f(x)}\, q(x) > 0$ is the normalizing constant needed to make $q_f(x)$ sum to one.

   b. Show that if $\text{KL}(p : q)$ is finite then we have the elementary identity ($\mathbb{E}_p$ denotes expectation under p)

   $$\text{KL}(p : q) - \mathbb{E}_p(f) \;=\; \text{KL}(p : q_f) - \log(Z_f)$$

   c. Use the previous identity to show that for all p, q we have the following variational expression for the KL divergence:

   $$\text{KL}(p : q) \;=\; \sup_{f : \mathcal{X} \to \mathbb{R}} \left[\, \mathbb{E}_p(f) - \log(Z_f) \,\right]$$

   Hint: consider separately the case where $\text{KL}(p : q) < \infty$ and $\text{KL}(p : q) = \infty$. In the former case, first establish an inequality, and then find a function f achieving equality.

   d. Use the variational expression above to show that the KL divergence is convex, namely, for all pmfs p, q and all $\alpha \in [0, 1]$ we have

   $$\text{KL}(\alpha\, p_1 + (1 - \alpha)\, p_2 : \alpha\, q_1 + (1 - \alpha)\, q_2) \;\le\; \alpha\, \text{KL}(p_1 : q_1) + (1 - \alpha)\, \text{KL}(p_2 : q_2)$$

   Hint: Extensive calculations are not necessary.