Exercises January 19 **STA 4508H (Spring, 2022)**

1. The Kullback-Leibler divergence between F and G is given by

$$KL(F : G) = \int \log \frac{f(y)}{g(y)} f(y) dy, \tag{1}$$

where f and g are and density functions with respect to Lebesgue measure. Note that the divergence is not symmetric in its arguments. This is called the directed information distance from F to G in Barndorff-Nielsen and Cox (1994) where the more general definition $KL(F : G) = \int \log(dF/dG)dF$ is used, assuming F and G are mutually absolutely continuous.[1]

   (a) In the canonical exponential family model with density $f(s; \varphi) = \exp\{\varphi^T s - k(\varphi)\}h(s)$, $s \in \mathbb{R}^p$, find an expression for the KL divergence from the model with parameter φ_1 to that with parameter φ_2.

   (b) Show that for a sample of observations from a model with density $f(y; \theta)$ the maximum likelihood estimator minimizes the KL divergence from $G_n(\cdot)$ to $F(\cdot; \theta)$, where $G_n(\cdot)$ is the empirical distribution function putting mass $1/n$ at each observation y_i.

2. Read "A Conversation with Sir David Cox" *Statistical Science*, 1994. It is posted on the course web page.

[1] Terminology varies from source to source. Notionally we think of F as the true distribution and G an approximation. Wikipedia calls KL(F:G) "the divergence of F from G" and "the divergence from G to F"; Mackay [ref needed] refers to the relative entropy of F with respect to G. Davison's *Statistical Models* seems to write it backwards, i.e. he defines $D(f, g) = \int \log(g/f) g dy$.