Stat 543 Assignment 3 (due Monday February 15, 2016)
Exponential Families, Fisher and Kullback-Leibler Information

1. Problems: 1.6.2, 1.6.5, 1.6.6, 1.6.34 of Bickel and Doksum

2. In a 1-parameter exponential family of distributions

   (a) give a simple representation for the Fisher Information in X about η, $I_X(\eta)$.

   (b) give a simple representation for the K-L Information in X, $I_X(\eta', \eta)$.

3. On the course web page, there is an argument that (for discrete cases) for statistic $T(X)$, $I_X(\theta) \geq I_{T(X)}(\theta)$. Argue carefully that in the context considered on that handout $I_X(\theta) = I_{T(X)}(\theta)$ for all θ, if and only if $T(X)$ is sufficient. You may use the fact that for W with finite expected square, $\mathrm{E}W^2 = (\mathrm{E}W)^2$ if and only if the distribution of W is degenerate. It may also be helpful to realize that for differentiable functions $h(\theta)$ and $k(\theta)$, the logarithms $\ln h(\theta)$ and $\ln k(\theta)$ have the same derivative functions if and only if they differ by at most a constant.

4. Problems 3.3.14, 3.3.15 of Bickel and Doksum. (For 3.3.14 just find the posterior densities up to a constant of proportionality. You don't need to find posterior means if it's not obvious to you what "standard" form a posterior takes. Regarding 3.3.15, see Problem 8.9 of Young and Smith for interest.)

5. What is the "Jefferys prior" for a model with $X \sim$Poisson(λ)? For this "prior," what is the posterior distribution of $\lambda|X$?

6. Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$.

   (a) Consider the two-dimensional parameter, $\boldsymbol{\theta} = (\mu, \sigma)$. Find the Fisher information matrix $I_X(\boldsymbol{\theta}_0)$.

   (b) Then consider the reparameterization in exponential family form with

   $$\boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{\theta_1}{\theta_2^2} \\ -\frac{1}{2\theta^2} \end{pmatrix}$$

   What is $I_X(\boldsymbol{\eta}_0)$?

   (c) If

   $$g(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\theta_1}{\theta_2^2} \\ -\frac{1}{2\theta^2} \end{pmatrix}$$

   and $\boldsymbol{\theta}_0 = g^{-1}(\boldsymbol{\eta}_0)$, can one simply plug $\boldsymbol{\theta}_0$ into matrix from **a)** to get the matrix from **b)**?

   The complete story hinted at here is told in Problem 3.4.3 of B&D.

7. Below are pmfs $f(x|\theta)$ for $\theta = 1, 2, 3$. Use them in what follows.

	$x=1$	$x=2$	$x=3$	$x=4$	$x=5$	$x=6$
$\theta=3$	0	.1	.1	.6	.1	.1
$\theta=2$	.1	.1	.2	.2	.2	.2
$\theta=1$	0	.2	.2	.2	.2	.2

(a) Find the K-L Information values $I_X(f(x|1), f(x|2))$ and $I_X(f(x|2), f(x|1))$. Are these the same? Offer an interpretation regarding what their magnitudes say about one's ability to detect the possibility that the first argument of $I(\cdot,\cdot)$ governs the behavior of X if "the other possibility" is that the second argument governs the behavior of X.

(b) Find $I_X(f(x|1), f(x|3))$ and compare it to $I_X(f(x|1), f(x|2))$. From the point of view of K-L information, is X more informative for discriminating $\theta = 1$ from $\theta = 2$, or for discriminating $\theta = 1$ from $\theta = 3$?

(c) Find a minimal sufficient statistic, $T(X)$, for the 2-class model $\{f(x|1), f(x|3)\}$ and verify that $I_X(f(x|1), f(x|3)) = I_{T(X)}(f^*(t|1), f^*(t|3))$ (where the f^*'s are pmfs for $T(X)$ correpsonding to the f's).