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 $\eta$, $I_X (\eta)$.
   (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 $\theta$, if and only if $T(X)$ is sufficient. You may use the fact that for $W$ with finite expected square, $EW^2 = (EW)^2$ 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 $X \sim \text{Poisson}(\lambda)$? For this "prior," what is the posterior distribution of $\lambda|X$?

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

   (a) Consider the two-dimensional parameter, $\theta = (\mu, \sigma)$. Find the Fisher information matrix $I_X (\theta_0)$.
   (b) Then consider the reparameterization in exponential family form with

   \[
   \eta = \left( \begin{array}{c} \eta_1 \\ \eta_2 \end{array} \right) = \left( \begin{array}{c} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{array} \right) = \left( \begin{array}{c} \frac{\theta_1}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{array} \right)
   \]

   What is $I_X (\eta_0)$?
   (c) If

   \[
   g (\theta) = \left( \begin{array}{c} \frac{\theta_1}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{array} \right)
   \]

   and $\theta_0 = g^{-1} (\eta_0)$, can one simply plug $\theta_0$ 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.

<table>
<thead>
<tr>
<th></th>
<th>$x=1$</th>
<th>$x=2$</th>
<th>$x=3$</th>
<th>$x=4$</th>
<th>$x=5$</th>
<th>$x=6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta=3$</td>
<td>0</td>
<td>.1</td>
<td>.1</td>
<td>.6</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>$\theta=2$</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
</tr>
<tr>
<td>$\theta=1$</td>
<td>0</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
</tr>
</tbody>
</table>

(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 $T(X)$ corresponding to the $f$'s).