00137960: Statistical Thinking
Homework 3

1. If X and Y are independent random variables, prove the convolution inequality for Fisher information:
$$\frac{1}{I(X+Y)} \geq \frac{1}{I(X)} + \frac{1}{I(Y)},$$
with equality only when X and Y are normally distributed.

2. Consider the linear model
$$Y_i = \beta x_i + u_i, \qquad i = 1, \dots, n,$$
where $x_i > 0$, $u_i = \rho u_{i-1} + \varepsilon_i$, and ε_i are independent $N(0, \sigma^2)$. If $\hat{\beta}$ is the least squares estimate of β, show that $\mathrm{Var}(\hat{\beta})$ is inflated when $\rho > 0$.

3. For a sample from the Laplace distribution with density
$$f(x) = \frac{1}{2} e^{-|x-\mu|},$$
show that the MLE of μ is the sample median. Derive its asymptotic distribution.

4. For a one-parameter exponential family $f_\mu(y) = e^{\eta y - \psi(\eta)} f_0(y)$, show that the MLE of $\mu = E_\eta y$ is y, and Hoeffding's formula holds:
$$f_\mu(y) = f_y(y) e^{-D(y,\mu)/2}.$$

5. (Programming) Reproduce the Poisson regression analysis for the galaxy data from Table 8.5 of Efron and Hastie (available at https://hastie.su.domains/CASI_files/DATA/galaxy.txt). Calculate the Poisson deviance residuals and the goodness-of-fit test statistic.

6. Derive the distributions of the following statistics for combining p-values under $H_0 : p_i \sim U(0, 1)$, $i = 1, \dots, n$:
   (a) $S_{\mathrm{F}} = \sum_{i=1}^n \log p_i$;
   (b) $S_{\mathrm{P}} = -\sum_{i=1}^n \log(1 - p_i)$;
   (c) $S_{\mathrm{G}} = \sum_{i=1}^n \log\{p_i/(1 - p_i)\}$;
   (d) $S_{\mathrm{E}} = \sum_{i=1}^n p_i$;
   (e) $S_{\mathrm{S}} = \sum_{i=1}^n \Phi^{-1}(p_i)$, where Φ is the standard normal cumulative distribution function;
   (f) $S_{\mathrm{T}} = \min(p_1, \dots, p_n)$.
   State for each case whether the distribution is exact or asymptotic.

7. Complete the proof of Benjamini and Hochberg's FDR control theorem by showing that $A(t) = a(t)/t$ is a martingale as t decreases from 1 to 0, where $a(t)$ is the number of null cases with $p_i \leq t$.