1. Suppose that $Z \sim \text{N}(0, 1)$ and, for $\mu \in \mathbb{R}$ and $\sigma > 0$, that $X = \mu + \sigma Z \sim P_{\mu, \sigma} = \text{N}(\mu, \sigma^2)$.
   (a) Compute the likelihood ratio
   $\frac{dP_{\mu, \sigma}}{dP_{0, \sigma}}(x) = \sigma^{-1} \phi((x - \mu)/\sigma) \quad \text{and} \quad Y \equiv \log \frac{dP_{\mu, \sigma}}{dP_{0, \sigma}}(X)$. 
   What is the distribution of $Y$ under $P_{0, \sigma}$ and under $P_{\mu, \sigma}$?
   (b) Plot the function 
   $l(\mu, \sigma; X) \equiv \log \frac{dP_{\mu, \sigma}}{dP_{0, \sigma}}(X)$
   as a function of $\mu$.
   (c) Find the maximum value of the function $l(\mu; X)$ in $B$ (as a function of $\mu$) and the value of $\mu \equiv \hat{\mu}$ which achieves the maximum.
   (d) What is the distribution of $\hat{\mu}$ under $P_{0, \sigma}$ and under $P_{\mu, \sigma}$? What is the distribution of $l(\hat{\mu}; X)$ under $P_{0, \sigma}$ and under $P_{\mu, \sigma}$?

2. Suppose that $\{P_{\theta} : \theta \in \Theta \subset \mathbb{R}^d\}$ is a regular parametric model in the sense of satisfying Cramér’s conditions A0 - A4 of the 581 Chapter 4 notes (page 5) at $\theta_0 \in \Theta$. Show that the LAN condition holds: with $\theta_n = \theta_0 + tn^{-1/2}$ for $t \in \mathbb{R}^d$,
   $\log \frac{\prod_{i=1}^n p_{\theta_n}(X_i)}{\prod_{i=1}^n p_{\theta_0}(X_i)} = l_n(\theta_n) - l_n(\theta_0) \rightarrow_d t^T Z - (1/2)t^T I(\theta_0)t \sim \text{N}_1(-(1/2)\sigma_0^2, \sigma_0^2)$
   where $Z \sim \text{N}_d(0, I(\theta_0))$ and $\sigma_0^2 = t^T I(\theta_0)t$.

3. Suppose that we want to model the survival of twins with a common genetic defect, but with one of the two twins receiving some treatment. Let $X$ represent the survival time of the untreated twin and let $Y$ represent the survival time of the treated twin. One (overly simple) preliminary model might be to assume that $X$
and \( Y \) are independent with Exponential(\( \eta \)) and Exponential(\( \theta \eta \)) distributions, respectively:

\[
f_{\theta,\eta}(x, y) = \eta e^{-\eta x} \eta \theta e^{-\eta \theta y} 1_{(0,\infty)}(x) 1_{(0,\infty)}(y)
\]

Suppose that we observe i.i.d. pairs \((X_i, Y_i)\) with density given by \( f_{\theta_0,\eta_0} \).

(a) One crude approach to estimation in this problem is to reduce the data to \( W = X/Y \), the maximal invariant for the group of scale changes \( g(x, y) = (cx, cy) \) with \( c > 0 \). Find the distribution of \( W \), and compute the Cramér-Rao lower bound for unbiased estimates of \( \theta \) based on \( W \).

(b) Find the information bound for estimation of \( \theta \) based on observation of \((X, Y)\) pairs when \( \eta \) is known and unknown.

(c) Compare the bounds you computed in (a) and (b) and discuss the pros and cons of reducing to estimation based on the \( W \).

4. This is a continuation of the preceding problem. A more realistic model involves assuming that the common parameter \( \eta \) for the two twins varies across sets of twins. There are several different ways of modeling this: one approach involves supposing that each pair of twins observed \((X_i, Y_i)\) has its own fixed parameter \( \eta_i \), \( i = 1, \ldots, n \). In this model we observe \((X_i, Y_i)\) with density \( f_{\nu,\eta_i} \) for \( i = 1, \ldots, n \); i.e.

\[
f_{\nu,\eta_i}(x_i, y_i) = \eta_i e^{-\eta_i x_i} \eta_i \nu e^{-\eta_i \nu y_i} 1_{(0,\infty)}(x_i) 1_{(0,\infty)}(y_i).
\]

This is sometimes called a “functional model” (or model with incidental nuisance parameters).

Another approach is to assume that \( \eta \equiv Z \) has a distribution, and that our observations are from the mixture distribution. Assuming (for simplicity) that \( Z = \eta \sim \text{Gamma}(a, b) \) with density \( g_{a,b}(\eta) \), it follows that the (marginal) distribution of \((X, Y)\) is

\[
p_{\nu,a,b}(x, y) = \int_0^\infty f_{\nu,z}(x,y) g_{a,b}(z) dz
\]

\[
= \frac{\nu}{b^2} \left( \frac{b}{b + x + \nu y} \right)^{a+2} \frac{\Gamma(a+2)}{\Gamma(a)}.
\]

This is sometimes called a “structural model” (or mixture model).

(a) Find the information for \( \nu \) in the functional model.

(b) Find the information for \( \nu \) in the structural model.

(c) Compare the information bounds you computed in (a) and (b). When is the information for \( \nu \) in the functional model larger than the information for \( \nu \) in the structural model?

(d) Find the MLEs of \( \nu \) in the functional model (call it \( \hat{\nu}_f^\nu \)) and in the structural
model (call it $\hat{\nu}_n^*$). Are they both consistent estimators of $\nu$?

**Hint:** this problem is related to the famous examples of Neyman and Scott concerning MLE’s in the presence of nuisance parameters; see e.g. Chapter 4, example 3.7, page 21.

5. (Bonus problem 1:) Suppose that $X_1, \ldots, X_n$ are i.i.d. with the Weibull distribution $F_{\theta}$ given by

\[ 1 - F_{\theta}(x) = \exp(-(x/\alpha)^\beta), \quad x \geq 0 \]

where $\theta = (\alpha, \beta) \in (0, \infty) \times (0, \infty)$.

(a) Find the inverse (or quantile function) $F_{\theta}^{-1}(u)$ corresponding to $F_{\theta}$ in terms of $\alpha$, $\beta$, and $u \in (0, 1)$, and show that

\[ \log F_{\theta}^{-1}(u) = \log \alpha + \frac{1}{\beta} \log \log \left( \frac{1}{1-u} \right). \]

(b) Fix $t \in (0, 1/2)$. Use the $t$–th and $(1-t)$–th quantiles of the $X_i$'s, namely $F_{n}^{-1}(t)$ and $F_{n}^{-1}(1-t)$, to obtain simple consistent estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ of $\alpha$ and $\beta$. Prove that your estimators are consistent.

(c) Prove that your estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ satisfy

\[ \sqrt{n} \left( \begin{array}{c} \hat{\alpha}_n - \alpha \\ \hat{\beta}_n - \beta \end{array} \right) \to_d N_2(0, \Sigma) \]

and identify $\Sigma$ as a function of $\alpha$, $\beta$, and $t$.

(d) How would you choose $t$ to minimize the asymptotic variance of $\hat{\beta}_n$?

6. **Optional bonus problem 2:** Suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors with values in $R^k$ with $E(X_1) = \mu$ and $E(X_1^T X_1) < \infty$ so that $\Sigma = E(X_1 - \mu)(X_1 - \mu)^T$ is well-defined. Thus

\[ Z_n \equiv \sqrt{n}(\overline{X}_n - \mu) \to_d Z \sim N_k(0, \Sigma). \]

Suppose that $g : R^k \to R$ is a function, and suppose that $\nabla g = \dot{g}$ exists at $\mu$. Then the delta-method (or $g'$ theorem) tells us that

\[ \sqrt{n}(g(\overline{X}_n) - g(\mu)) \to_d \nabla g(\mu)^T Z \sim N(0, \nabla g(\mu)^T \Sigma \nabla g(\mu)). \]  \(\text{(3)}\)

(a) Show that we can strengthen (??) as follows: Suppose that $\nabla g = \dot{g}$ is continuous at $\mu$. Then $\sqrt{n}(g(\overline{X}_n) - g(\mu))$ is asymptotically linear at $\mu$:

\[ \sqrt{n}(g(\overline{X}_n) - g(\mu)) = \nabla g(\mu)^T \sqrt{n}(\overline{X}_n - \mu) + o_p(1) \]

\[ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i) + o_p(1) \]
where

$$\psi(x) = \nabla g(\mu)^T(x - \mu)$$  \hspace{1cm} (4)

which is called the influence function of $g(X_n)$ as an estimator of $g(\mu)$, has mean $E\psi(X_i) = 0$ and $Var(\psi(X_i)) = \nabla g(\mu)^T \Sigma \nabla g(\mu)$.

(b) Can the result in (a) be used to establish asymptotic linearity of the empirical or sample quantile $F_n^{-1}(t)$ for $t \in (0, 1)$ if $Q \equiv F^{-1}$ is differentiable at $t$?  (c) Find some other example for which the result of (a) yields asymptotic linearity of some natural (nonlinear) estimator.