Note: The problems are taken from the Exercises in Casella and Berger (2002) unless otherwise noted. For each problem, please explain your reasoning clearly. It is not acceptable to only provide your final result.

**Homework 1** (Due Wed, Sept 18):
6.9, 6.10, 6.11, 6.12, 6.15, 6.20, 6.23, 6.30, 6.31(b,c)

The question 6.31(b)(ii) should be corrected as follows. Suppose $X_1, \ldots, X_N$ are $N$ Monte Carlo samples, where each $X^j$ consists of an iid sample of size $n$ from $N(\mu, \sigma^2)$. Let $M^j$ be the sample median and $\bar{X}^j$ the sample mean, both computed from $X^j$. Let $\bar{M}$ be the average of $(M^1, \ldots, M^N)$, and $\bar{\bar{X}}$ be the average of $(\bar{X}^1, \ldots, \bar{X}^N)$. A naive estimator of the variance of the sample median is then

$$v_1 = \frac{1}{N-1} \sum_{j=1}^N (M^j - \bar{M})^2.$$

The swindle estimator of the variance of the sample median is

$$v_2 = \frac{\sigma^2}{n} + \frac{1}{N-1} \sum_{j=1}^N \{M^j - \bar{X}^j - (\bar{M} - \bar{\bar{X}})\}^2.$$

Show that the variance of $v_1$ is approximately $2[\text{var}(M)]^2/(N-1)$, and the variance of $v_2$ is approximately $2[\text{var}(M - \bar{X})]^2/(N-1)$.

Additional exercise I (optional): Read the paper, “Models as Approximations, Part I: A Conspiracy of Random Regressors and Model Deviations Against Classical Inference in Regression” by Buja et al., currently available at https://www.imstat.org/journals-and-publications/statistical-science/statistical-science-future-papers/. Think about three questions you may have about the paper.

Additional exercise II (optional): The following steps can be used to show that if $T$ is a complete, sufficient statistic, then $T$ is minimal sufficient. For simplicity, assume that the parameter $\theta$ is a scalar. Let $U$ be another sufficient statistic. Define

$$h(u) = E_{\theta}(T|U = u),$$
$$g(t) = E_{\theta}(h(U)|T = t).$$

1) Show that $T = g(T)$ almost surely. [Hint: Use iterated expectations.]
2) Show that $T = h(U)$ almost surely. [Hint: Use decomposition of variance in terms of conditional expectation and variance.]
Additional exercise III (optional):
1) Show that if a statistician follows the (weak) likelihood principle, then s/he also follows the sufficiency principle.
2) Show that if a statistician follows the sufficiency principle, then s/he also follows the (weak) likelihood principle provided that there exists a minimal sufficient statistic $T(\tilde{x})$ such that for every two samples $\tilde{x}$ and $\tilde{y}$, the ratio $p(\tilde{x}; \theta)/p(\tilde{y}; \theta)$ does not depend on $\theta$ if and only if $T(\tilde{x}) = T(\tilde{y})$.

Homework 2 (Due Wed, Oct 9):
6.37, 7.2, 7.6, 7.12, 7.13, 7.18, 7.23, 7.24,
7.37, 7.38, 7.42, 7.44, 7.45, 7.46, 7.49, 7.52

Additional problem I:
Let $(X_1, \ldots, X_n)$ be an iid sample from $N(\mu, \sigma^2)$ with unknown $(\mu, \sigma^2)$.
1) Find the canonical parameters for $N(\mu, \sigma^2)$ as an exponential family.
2) Derive the maximum likelihood estimators of $(\mu, \sigma^2)$ by solving the score equation in terms of the canonical parameters.

Additional problem II:
Let $X$ be an observation from $N(\mu,1)$ with unknown $\mu$. Suppose that the prior on $\mu$ is $N(0, \tau^2)$ with a fixed $\tau > 0$.
1) Show that the maximum likelihood estimator of $\mu$ is $\hat{\mu} = X$.
2) Find the posterior mean of $\mu$, denoted by $\tilde{\mu}$, as a point estimator.
3) Find the bias, variance, and mean squared error of $\tilde{\mu}$.
4) Determine the condition, depending on $\mu$ and $\tau$, such that the mean square error of $\tilde{\mu}$ is no greater than that of $\hat{\mu}$.

Additional problem III:
Let $(X_1, \ldots, X_n)$ be an iid sample from Uniform $(0, \theta)$ with unknown $\theta > 0$. Consider the estimator $\hat{\theta} = \frac{n+1}{n}X_{(n)}$.
1) Show that $\hat{\theta}$ is unbiased for $\theta$.
2) Show that
$$\text{var}_\theta(\hat{\theta}) = \frac{\theta^2}{n(n+2)}.$$
Let \((X_1, \ldots, X_n)\) be an iid sample from Bernoulli \((\theta)\) with \(0 < \theta < 1\) unknown. For estimation of \(\theta\), suppose that the loss is squared error, and hence the risk is mean squared error. Consider the estimator \(\bar{X}\) (sample proportion) for \(\theta\).

1) Is \(\bar{X}\) admissible? Justify your answer.
2) Is \(\bar{X}\) minimax? Justify your answer.

Additional problem II:
1) For a finite random variable \(Y\), i.e., \(P(|Y| < \infty) = 1\), show that for any \(\epsilon > 0\), there exists \(M > 0\) such that \(P(|Y| \leq M) > 1 - \epsilon\).
2) Let \(X\) and \((X_n : n \geq 1)\) be (finite) random variables. Show that if \(X_n \rightarrow_{D} X\), then \(X_n = O_p(1)\).

Additional problem III:
1) Show that \(o_p(1) + O_p(1) = O_p(1)\).
2) For any constant \(c > 0\), show that \((c + o_p(1))^{-1} = c^{-1} + o_p(1)\).

Additional problem IV:
Let \((x_{11}, \ldots, x_{1n_1})\) be an iid sample of size \(n_1\) from a distribution with probability density function \(q_1(x)/Z_1\), and independently, \((x_{21}, \ldots, x_{2n_2})\) be an iid sample of size \(n_2\) from a distribution with probability density function \(q_2(x)/Z_2\), where \(Z_1 = \int q_1(x)dx\) and \(Z_2 = \int q_2(x)dx\). Denote \(r^* = Z_2/Z_1\). For asymptotic evaluation, assume that \(n = n_1 + n_2 \rightarrow \infty\) and \(n_1/n \rightarrow \rho\), a constant in \((0, 1)\).

1) For any function \(\alpha(x)\) such that \(0 < \int \alpha(x)q_1(x)q_2(x)dx < \infty\), consider
\[
\hat{r}_\alpha = \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\alpha(x_{1i})q_2(x_{1i})}{\alpha(x_{2i})q_1(x_{2i})}.
\]
Show that \(\hat{r}_\alpha\) is a consistent estimator of \(r^*\).
2) Find \(V_\alpha\) such that \(\sqrt{n}(\hat{r}_\alpha - r^*) \rightarrow_{D} N(0, V_\alpha)\).
3) Find the optimal choice of \(\alpha(\cdot)\) which minimizes \(V_\alpha\). Denote the optimal choice as \(\alpha^+(:, r^*)\), depending on \(r^*\). Find the minimum variance \(V_{\alpha^+}\).
4) Consider the estimator \(\tilde{r}\), defined as a solution to
\[
\tilde{r} = \frac{1}{n_1} \sum_{i=1}^{n_1} \alpha^+(x_{1i}; r)q_2(x_{1i})
\frac{1}{n_2} \sum_{i=1}^{n_2} \alpha^+(x_{2i}; r)q_1(x_{2i})
\]
Express \(\tilde{r}\) as an \(Z\)-estimator.
5) Can \(\tilde{r}\) be expressed as an \(M\)-estimator? If so, find the objective function.
6) Find \( U \) such that \( \sqrt{n}(\tilde{r} - r^*) \rightarrow_{D} N(0, U) \).

7) Compare \( U \) and \( V_{\alpha^1} \), which is the minimum possible variance found in 3).

Note: The problem is motivated from a method in Monte Carlo integration. Suppose that 
\( q_1(x) = \exp(-x) \) and \( q_2(x) = x^{\alpha-1}\exp(-x) \) for \( x > 0 \), with \( \alpha > 0 \) fixed. Hence \( Z_1 = 1 \) and \( Z_2 = \Gamma(\alpha) \). But pretend we have no access to statistical software which can compute \( \Gamma(\alpha) \) as an integral \( \int_{0}^{\infty} q_2(x)dx \). Instead, we only have access to a Monte Carlo simulator which generates the two samples \( (x_{11}, \ldots, x_{1n_1}) \) and \( (x_{21}, \ldots, x_{2n_2}) \). The objective is then to use these samples to “estimate” \( r^* = Z_2/Z_1 = \Gamma(\alpha) \). Here is a paper which you may consult: Meng, X.-L. and Wong, W.H. (1996), Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Explanation, Statistica Sinica, 6, 831–860.

Additional problem V:
Let \( (X_1, \ldots, X_n) \) be an iid sample from \( \Gamma(\nu^*/2, 2) \). Let \( \hat{\nu}_0 \) be the MLE. In addition, consider the two estimators \( \hat{\nu}_1 = \bar{X} \), which is a solution to 
\[
\tau_{1n}(\nu) = \sum_{i=1}^{n} (X_i - \nu) = 0,
\]
and \( \hat{\nu}_2 \) defined as a solution
\[
\tau_{2n}(\nu) = \sum_{i=1}^{m} (X_i^2 - \nu^2 - 2\nu) = 0.
\]

1) Find \( V_0 \) such that \( \sqrt{n}(\hat{\nu}_0 - \nu^*) \rightarrow_{D} N(0, V_0) \).
2) Show that \( \hat{\nu}_1 \) and \( \hat{\nu}_2 \) are consistent.
3) Find \( V_1 \) and \( V_2 \) such that \( \sqrt{n}(\hat{\nu}_1 - \nu^*) \rightarrow_{D} N(0, V_1) \) and \( \sqrt{n}(\hat{\nu}_2 - \nu^*) \rightarrow_{D} N(0, V_2) \).
4) For any constant \( b \), consider the estimator \( \hat{\nu}^b \) defined as a solution to
\[
(1 - b)\tau_{1n}(\nu) + b\tau_{2n}(\nu) = 0.
\]

From the lectures, we found the optimal choice of \( b \), which minimizes \( V^b \) such that \( \sqrt{n}(\hat{\nu}^b - \nu^*) \rightarrow_{D} N(0, V_b) \). Denote the optimal choice as \( b^*(\nu^*) \), depending on \( \nu^* \). Find the minimum variance \( V_{b^*} \) and compare it with \( V_0, V_1, \) and \( V_2 \).
5) Consider the two-step estimator \( \hat{\nu}_3 \) defined as a solution to
\[
(1 - b(\hat{\nu}_0))\tau_{1n}(\nu) + b(\hat{\nu}_0)\tau_{2n}(\nu) = 0,
\]
and the simultaneous estimator \( \hat{\nu}_4 \) defined as a solution to
\[
(1 - b(\nu))\tau_{1n}(\nu) + b(\nu)\tau_{2n}(\nu) = 0.
\]

What are \( V_3 \) and \( V_4 \) such that \( \sqrt{n}(\hat{\nu}_3 - \nu^*) \rightarrow_{D} N(0, V_3) \) and \( \sqrt{n}(\hat{\nu}_4 - \nu^*) \rightarrow_{D} N(0, V_4) \) ?
6) Conduct a simulation study which evaluates the biases, variances, and mean squared errors.
of $\hat{\nu}_1, \hat{\nu}_2, \hat{\nu}_3$, and $\hat{\nu}_4$ for sample sizes $n = 10, 20, 50, 100$, with $\nu^* = 1, 5, 10, 20$. Compare the (finite-sample) simulation variances with asymptotic variances $V_1, \ldots, V_4$. Use your preferred programming language such as R.

Additional problem VI:
Let $Y$ be a $q \times 1$ random vector with nonsingular variance matrix $V = \text{var}(Y)$. For any $q \times d$ matrices $U$ and $B$ with $B^tV B$ nonsingular, show that

$$U^tV^{-1}U - U^tB(B^tV B)^{-1}B^tU = \text{var}(U^tV^{-1}Y - \Gamma B^tY),$$

where $\Gamma = \text{cov}(U^tV^{-1}Y, B^tY)\text{var}^{-1}(B^tY)$.


Homework 4 (Due Mon, Nov 25):
7.62, 7.65, 8.5, 8.6, 8.7(a), 8.11, 8.12, 8.15, 8.22, 8.23, 8.26 (MLR means nondecreasing MLR), 8.28, 8.29, 8.31, 8.40, 8.41, 8.47, 8.52, 8.53

Additional problem I:
Let $(X_i, Y_i), 1 \leq i \leq n$, be an iid sample. Consider a logistic regression model $P(Y_i = 1 | X_i) = \mu(X_i; \beta)$ with

$$\mu(X_i; \beta) = \frac{e^{X_i^t\beta}}{1 + e^{X_i^t\beta}},$$

where $\beta$ is a vector of unknown coefficients.

1) The MLE of $\beta$ is defined as a maximizer to the log likelihood, conditionally on $(X_1, \ldots, X_n)$,

$$\sum_{i=1}^{n} \{Y_i \log \mu(X_i; \beta) + (1-Y_i) \log(1 - \mu(X_i; \beta))\}.$$

Derive the score equation which satisfied by the MLE.

2) Consider an estimator of $\beta$, $\hat{\beta}_h$, defined as a solution to

$$0 = \sum_{i=1}^{n} \{Y_i - \mu(X_i; \beta)\}h(X_i),$$

where $h(x)$ is a vector of functions of $x$, of the same dimension as $\theta$. Derive the optimal choice of $h(x)$ in minimizing the asymptotic variance of $\hat{\beta}_h$. Compare the resulting estimator with the MLE.
Additional problem II:
Let \((X_1, \ldots, X_n)\) be an iid sample from \(N(\mu_x, \sigma_x^2)\) and, independently, \((Y_1, \ldots, Y_m)\) be an iid sample from \(N(\mu_y, \sigma_y^2)\), where \(\sigma_x^2\) and \(\sigma_y^2\) are known. Consider testing

\[
H_0 : \mu_x = \mu_y = 0, \\
H_1 : \mu_x \neq 0 \text{ or } \mu_y \neq 0.
\]

1) Find a likelihood ratio test (LRT) rejection region and a union-intersection test (UIT) rejection region to achieve size \(\alpha\).
For the remaining questions, set \(\alpha = 10\%\).
2) Compute and plot the power function for LRT and UIT rejection regions along the direction \(\mu_x = 0\) and \(\mu_y \in \mathbb{R}\). You may use R or other language.
3) Compute and plot the power function for LRT and UIT rejection regions along the direction \(\mu_x = \mu_y \in \mathbb{R}\).
4) Compute and plot the power function for LRT and UIT rejection regions for \((\mu_x, \mu_y) \in \mathbb{R}^2\).
Choose a plotting method that you think most appropriate.

Additional problem III:
Let \(X \sim \text{Binomial}(2, \theta)\). Consider \(H_0 : \theta = 1/2\) vs \(H_1 : \theta = 3/4\).
1) Describe the randomized test of size 30\%, derived in the class.
2) Use Neyman–Pearson lemma to show that this test is UMP level 30\%, i.e., UMP among all tests of level 30\% (which depend only on \(X\)).
Note: To use Neyman–Pearson lemma, we need to expand the observation \(x\) to \((x,u)\), where \(u \sim \text{Uniform}(0,1)\).