Stats 210A, Fall 2023
Homework 10

Due date: Wednesday, Nov. 8

1. Multidimensional testing
Suppose $X \sim N_d(\mu, I_d)$ for unknown $\mu \in \mathbb{R}^d$. Consider testing $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$. You may take as given the fact that if $d = 1$ the UMP test for the Gaussian location family is unique: i.e., it is the only UMPU test for that model up to almost sure equality.

(a) Show that for any $d > 1$ and $\alpha \in (0, 1)$, there exists no UMP or UMPU level-$\alpha$ test.

**Hint:** what would we do if we knew $\mu = (\theta, 0, 0, \ldots, 0)$ for an unknown $\theta \in \mathbb{R}$?

(b) Suppose we have a prior $\Lambda_1$ for the value that $\mu$ takes under the alternative; that is, $\mu \sim \Lambda_1$ if $H_1$ is true and $\mu = 0$ if $H_0$ is true. Define the average power as

$$\int_{\mathbb{R}^d} \mathbb{E}_\mu[\phi(X)] \, d\Lambda_1(\mu).$$

If $\Lambda_1 = N(\nu, \Sigma)$, with positive definite covariance matrix $\Sigma$, find the level-$\alpha$ test that maximizes the average power. Show that the acceptance region is an ellipse centered at 0 if $\nu = 0$.

**Hint:** You can use the result from homework 8.

(c) Optional: Show that if $\Lambda_1$ is rotationally invariant, the $\chi^2$ test that rejects for large $\|X\|_2$ maximizes the average power.

**Moral:** Choosing a test in higher dimensions requires us to think harder about how to compromise across different alternative directions, and Bayesian thinking can give us some guidance.

2. James-Stein estimator with regression-based shrinkage
Consider estimating $\theta \in \mathbb{R}^n$ in the model $Y \sim N_n(\theta, I_n)$. In the standard James-Stein estimator, we shrink all the estimates toward zero, but it might make more sense to shrink them towards the average value $\bar{Y}$, or towards some other value based on observed side information.

(a) Consider the estimator

$$\delta^{(1)}_i(Y) = \bar{Y} + \left(1 - \frac{n-3}{\|Y - \bar{Y}\|_2^2}\right) (Y_i - \bar{Y})$$

Show that $\delta^{(1)}(Y)$ strictly dominates the estimator $\delta^{(0)}(Y) = Y$, for $n \geq 4$.

$$\text{MSE}(\theta; \delta^{(1)}) < \text{MSE}(\theta; \delta^{(0)}), \quad \text{for all } \theta \in \mathbb{R}^n.$$ 

Calculate the MSE of $\delta^{(1)}$ if $\theta_1 = \theta_2 = \cdots = \theta_n$.

**Hint:** Change the basis and think about how the estimator operates on different subspaces.

(b) Now suppose instead that we have side information about each $\theta_i$, represented by covariate vectors $x_1, \ldots, x_n \in \mathbb{R}^d$. Assume the design matrix $X \in \mathbb{R}^{n \times d}$ whose $i$th row is $x_i$ has full column rank. Suppose that we expect $\theta \approx X\beta$ for some $\beta \in \mathbb{R}^d$, but unlike the usual linear regression setup, we will not assume $\theta = X\beta$ with perfect equality.
Find an estimator \( \hat{\theta}^{(2)} \), analogous to the one in part (a), that dominates \( \hat{\theta}^{(0)} \) whenever \( n - d \geq 3 
\)
\[
\text{MSE}(\theta; \hat{\theta}^{(2)}) < \text{MSE}(\theta; \hat{\theta}^{(0)}), \quad \text{for all } \theta \in \mathbb{R}^n,
\]
and for which \( \text{MSE}(X\beta; \hat{\beta}^{(2)}) = d + 2 \), for any \( \beta \in \mathbb{R}^d \).

**Hint:** Think of this setting as a generalization of part (a), which can be considered a special case with \( d = 1 \) and all \( x_i = 1 \).

3. **Confidence regions for regression**

Assume we observe \( x_1, \ldots, x_n \in \mathbb{R} \), which are not all identical (for some \( i \) and \( j, x_i \neq x_j \)). We also observe
\[
Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \text{for } \epsilon_i \sim \text{i.i.d. } N(0, \sigma^2).
\]
\( \beta_0, \beta_1 \in \mathbb{R} \) and \( \sigma^2 > 0 \) are unknown. Let \( \bar{x} \) represent the mean value \( \frac{1}{n} \sum_i x_i \).

(a) Give an explicit expression for the \( t \)-based confidence interval for \( \beta_1 \), in terms of a quantile of a Student’s \( t \) distribution with an appropriate number of degrees of freedom (feel free to break up the expression, for example by first giving an expression for \( \hat{\beta}_1 \) and then using \( \hat{\beta}_1 \) in your final expression). You do not need to show the interval is UMAU.

**Hint:** It may be helpful to consider a translation of the model similar to what we did in Problem 3 of Homework 8.

(b) Invert an \( F \)-test to give a confidence ellipse for \( (\beta_0, \beta_1) \). It may be convenient to represent the set as an affine transformation of the unit ball in \( \mathbb{R}^2 \):
\[
b + A B_1(0) = \{ b + A z : z \in \mathbb{R}^2, ||z|| \leq 1 \}, \quad \text{for } b \in \mathbb{R}^2, A \in \mathbb{R}^{2 \times 2}.
\]

Give explicit expressions for \( b \) and \( A \) in terms of a quantile of an appropriate \( F \) distribution.

**Hint:** Consider the joint distribution of \( (\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1) \).

**Hint:** Use the fact that \( \left( \begin{array}{c} \hat{\beta}_0 \\ \hat{\beta}_1 \end{array} \right) \sim N_2 \left( \left( \begin{array}{c} \beta_0 \\ \beta_1 \end{array} \right), \sigma^2 (X^T X)^{-1} \right) \). You do not need to show that the confidence ellipse you come up with has any optimality properties.

4. **Confidence bands for regression**

The setup for this problem is the same as for Problem 4 only now we are interested in giving confidence bands for the regression line \( f(x) = \beta_0 + \beta_1 x \). In this problem you do not need to give explicit expressions for everything, but you should be explicit enough that someone could calculate the bands based on your description.

(a) For a fixed value \( x_0 \in \mathbb{R} \) (not necessarily one of the observed \( x_i \) values) give a \( 1 - \alpha \) \( t \)-based confidence interval for \( f(x_0) = \beta_0 + \beta_1 x_0 \). That is, we want to find \( C_1^P(x_0), C_2^P(x_0) \) such that
\[
\mathbb{P} \left( C_1^P(x_0) \leq f(x_0) \leq C_2^P(x_0) \right) = 1 - \alpha.
\]

The functions \( C_1^P(x), C_2^P(x) \) that we get from performing this operation on all \( x \) values give a pointwise confidence band for the function \( f(x) \).

(b) Now give a simultaneous confidence band around \( f(x) = \beta_0 + \beta_1 x \). That is, give \( C_1^S(x), C_2^S(x) \) with
\[
\mathbb{P} \left( C_1^S(x) \leq f(x) \leq C_2^S(x) \right), \quad \text{for all } x \in \mathbb{R} = 1 - \alpha,
\]
and show that your confidence band has this property.

**Hint:** If all we know is that \( (\hat{\beta}_0, \hat{\beta}_1) \) is in the confidence ellipse from Problem 4, what can we deduce about \( f(x) \)?
(c) Download the data set in `hw10-4.csv` from the course web site and make a scatter plot of the data. Plot the OLS regression line as well as the two confidence bands. Describe what you see. What do the bands do as $x$ goes away from the data set, and why does this make sense?