Please provide complete and well-written solutions to the following exercises.

Due April 12, at the beginning of class.

Homework 6

Exercise 1. For any random variables X, Y define

$$\mathrm{Var}(X|Y) := \mathbf{E}[(X - \mathbf{E}(X|Y))^2 | Y].$$

In this exercise you can freely use the identity

$$\mathrm{Var}(X) = \mathbf{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}[\mathbf{E}(X|Y)].$$

Using this identity, give a different proof of the Rao-Blackwell Theorem, when the loss function is mean squared error. (In some sense, this new proof is better, since you can explicitly quantify the improvement in the variance that results from conditioning; on the other hand, this proof only seems to work for the quadratic loss function.)

(Hint: starting from the mean squared error $\mathbf{E}(X - g(\theta))^2$, add and subtract the mean of X inside the parentheses.)

Exercise 2. Let $X_1, \ldots, X_n$ be a random sample of size n, so that X_1 is a sample from the uniform distribution on the interval $[\theta - 1/2, \theta + 1/2]$, where $\theta \in \mathbf{R}$ is unknown. From a previous homework, we tried to find a low variance estimator for θ, but the UMVU seemed to not exist. In this exercise, you are asked to show that a UMVU does not exist, using the following outline, in the case $n = 1$. Moreover, if $g(\theta)$ is a nonconstant differentiable function of $\theta \in \mathbf{R}$, show that no UMVU of $g(\theta)$ exists when $n = 1$:

- Let $U = u(X_1)$ be an unbiased estimator of 0, where $u\colon \mathbf{R} \to \mathbf{R}$. By differentiating the definition of unbiasedness with respect to θ, conclude that

  $$u(x + 1) = u(x), \qquad \text{for a.e. } x \in \mathbf{R}.$$

  Give an example of an unbiased estimator U of 0 such that $u(x) \neq 0$ for all $x \in \mathbf{R}$.
- Argue by contradiction. Assume that W is UMVU for $g(\theta)$. Using the characterization from class, conclude that $\mathbf{E}_\theta WU = 0$, so that if $W = w(X_1)$ with $w\colon \mathbf{R} \to \mathbf{R}$, then

  $$w(x + 1)u(x + 1) = w(x)u(x), \qquad \text{for a.e. } x \in \mathbf{R}.$$

  Then conclude that

  $$w(x + 1) = w(x), \qquad \text{for a.e. } x \in \mathbf{R}.$$
- To complete the exercise, what can you say about the condition that W is unbiased for $g(\theta)$?

(Optional) Can you make the same conclusion for a sample of size 2? Hint: Fourier series.

Exercise 3. Let $X_1, \ldots, X_n$ be a random sample of size n, so that X_1 is a sample from the uniform distribution on the interval $[\theta - 1/2, \theta + 1/2]$, where $\theta \in \mathbf{R}$ is unknown. Although the UMVU for θ does not exist and unbiased estimators do exist, if we instead restrict to *location equivariant* estimators, then there *is* a minimum variance estimator of θ among this class. We say that an $W := t(X_1, \ldots, X_n)$ with $t\colon \mathbf{R}^n \to \mathbf{R}$ is location equivariant if

$$t(x_1, \ldots, x_n) + a = t(x_1 + a, \ldots, x_n + a), \qquad \forall\, (x_1, \ldots, x_n) \in \mathbf{R}^n, \quad \forall\, a \in \mathbf{R}.$$

- Using location equivariance for the density $f := 1_{[-1/2,1/2]}$, and letting $f_\theta(x) := f(x - \theta)$, show that

$$\mathbf{E}_\theta (W - \theta)^2 = \int_{\mathbf{R}^n} [t(x)]^2 \prod_{i=1}^{n} f(x_i) dx_1 \cdots dx_n, \qquad \forall\, \theta \in \Theta.$$

  (Note that the expression on the right does not depend on θ.)
- Let $H := \{x \in \mathbf{R}^n \colon \langle x, (1, \ldots, 1)\rangle = 0\}$, where as usual $\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = \sum_{i=1}^n x_i y_i$. Using location equivariance again, show that

$$\mathbf{E}_\theta (W - \theta)^2 = \int_H \Big(\int_{\mathbf{R}} |t(x) - a|^2 \prod_{i=1}^{n} f(x_i - a) da \Big) dH(x).$$

  (Here $dH(x)$ denotes integration on the hypersurface H, i.e. $dH(x)$ is not the same as $dx_1 \cdots dx_n$)
- So, to minimize $\mathbf{E}_\theta (W - \theta)^2$, it suffices to minimize $\int_{\mathbf{R}} [a - t(x)]^2 \prod_{i=1}^n f(x_i - a) da$, for any fixed $x \in H$. What choice of $t(x)$ minimizes $\int_{\mathbf{R}} [a - t(x)]^2 \prod_{i=1}^n f(x_i - a) da$, when $x \in H$ is fixed?
- Conclude that the W minimizing $\mathbf{E}_\theta (W - \theta)^2$ for all $\theta \in \mathbf{R}$, over all location equivariant estimators satisfies

$$W = \frac{\int_{\mathbf{R}} a \prod_{i=1}^n f(X_i - a) da}{\int_{\mathbf{R}} \prod_{i=1}^n f(X_i - a) da}.$$

- So, in our original example when $f = 1_{[-1/2,1/2]}$, show that $W = \frac{X_{(1)} + X_{(n)}}{2}$ achieves the minimum variance among location equivariant estimators, despite the UMVU not existing. This estimator is also unbiased, but this was not guaranteed to occur in our construction.
- (Optional) Perform the above analysis for $f_\theta(x) := \theta^{-1} f(x/\theta)$, $\theta > 0$ to find the variance minimizer among *scale-equivariant* estimators

$$t(ax_1, \ldots, ax_n) = at(x), \qquad \forall\, x = (x_1, \ldots, x_n) \in \mathbf{R}^n, \quad \forall\, a > 0.$$

  You should find the optimal estimator to be

$$t(x) := \frac{\int_{\mathbf{R}} a^n \prod_{i=1}^n f(ax_i) da}{\int_{\mathbf{R}} a^{n+1} \prod_{i=1}^n f(ax_i) da}$$

Exercise 4. Let $f\colon \mathbf{R}^n \to \mathbf{R}$ be a convex function. Let $x \in \mathbf{R}^n$ be a local minimum of f. Show that x is in fact a global minimum of f.

Show also that if f is strictly convex, then there is at most one global minimum of f.

Now suppose additionally that f is a C^1 function (all derivatives of f exist and are continuous), and $x \in \mathbf{R}^n$ satisfies $\nabla f(x) = 0$. Show that x is a global minimum of f.

Exercise 5. Let A be a real $m \times n$ matrix. Let $x \in \mathbf{R}^n$ and let $b \in \mathbf{R}^m$. Show that the function $f \colon \mathbf{R}^n \to \mathbf{R}$ defined by $f(x) = \frac{1}{2} \|Ax - b\|^2$ is convex. Moreover, show that

$$\nabla f(x) = A^T(Ax - b), \qquad D^2 f(x) = A^T A.$$

(Here $D^2 f$ denotes the matrix of second derivatives of f.)

So, if $\nabla f(x) = 0$, i.e. if $A^T Ax = A^T b$, then x is the global minimum of f. And if A has full rank, then $A^T A$ is invertible, so that $x = (A^T A)^{-1} A^T b$ is the global minimum of f.

Exercise 6 (Least Squares/ Ridge Regression)**.** Let $Z_1, \ldots, Z_n$ be independent identically distributed Gaussian random variables with zero mean and known variance $\sigma^2 > 0$. Suppose $w \in \mathbf{R}^k$ is an unknown vector, and for all $1 \leq i \leq n$, there are known vectors $x^{(1)}, \ldots, x^{(n)} \in \mathbf{R}^k$. Our observed data are

$$X_i := \langle x^{(i)}, w \rangle + Z_i, \qquad \forall\, 1 \leq i \leq n.$$

Here $Z_1, \ldots, Z_n$ represent experimental noise. The goal is to determine w.

So, our data are $X = (X_1, \ldots, X_n)^T$. In this exercise we restrict attention to *linear estimators*, i.e. we only consider statistics of the form

$$Y := BX,$$

where B is a $k \times n$ real matrix. (The vectors $x^{(1)}, \ldots, x^{(n)}$ can be thought of as measurement vectors, which are part of the data of the problem, but our observed data from the "experiment" is X.)

- Let A be the $n \times k$ matrix so that the i^{th} row of A is the row vector $x^{(i)}$. Assume that $k \leq n$ and the matrix A has full rank. Find the unbiased estimator of w with minimal variance, among all linear functions. That is, minimize

$$\mathbf{E}\, \|Y - w\|^2 = \mathbf{E} \sum_{j=1}^{k} (Y_j - w_j)^2$$

  over all choices of B such that $\mathbf{E}Y = w$. (The restriction $\mathbf{E}Y = w$ says that Y is unbiased for w.) (Hint: What condition on B guarantees that $\mathbf{E}Y = w$? Compute $\mathbf{E}(Y - w)(Y - w)^T$, where Y and w are column vectors, so that this is the expected value of a matrix. Note that $\mathbf{E}\, \|Y - w\|^2$ is the trace of $\mathbf{E}(Y - w)(Y - w)^T$. Also try doing the second part of the problem first. Whenever possible, write expressions in terms of matrices.)
- Compare your estimator to the value of $w \in \mathbf{R}^k$ that minimizes the quantity

$$\sum_{i=1}^{n} (X_i - \langle x^{(i)}, w \rangle)^2$$

  (considering w as a variable, with *all* other quantities fixed, i.e. for this part of the problem, consider the X_i and $x^{(i)}$ as variables that have no a priori relation to each other.)

Part of the purpose of this problem is to explore that there is more than one way to think about least squares minimization.

Exercise 7. Let $X_1, \ldots, X_n$ be a random sample of size n, so that X_1 has the Poisson distribution with parameter θ, i.e.

$$\mathbf{P}_\theta(X_1 = x) = \theta^x e^{-\theta}/x!, \qquad \forall \text{ nonnegative integers } x.$$

Suppose we want to estimate $\mathbf{P}_\theta(X_1 = 0) = e^{-\theta}$.

- One way we can try to estimate $e^{-\theta}$ is to count the fraction of zeros in the sample of size n. Define
$$Y_n := \frac{1}{n} \left|\{1 \leq i \leq n \colon X_i = 0\}\right|.$$
Find the limiting distribution of Y_n as $n \to \infty$. (That is, find $a_n, b_n \in \mathbf{R}$ such that $a_n(Y_n - b_n)$ converges in distribution to something nonconstant as $n \to \infty$.)
- Give an explicit formula for the MLE Z_n of $e^{-\theta}$. Find the limiting distribution of Z_n as $n \to \infty$.
- Compute the relative efficiency of these two estimators as $n \to \infty$.

Exercise 8. Let $X_1, \ldots, X_n$ be a random sample of size n, so that X_1 has the Laplace density $\frac{1}{2}e^{-|x-\theta|}$ for all $x \in \mathbf{R}$, where $\theta \in \mathbf{R}$ is unknown. Find the MLE of θ.