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

Due April 26, at the beginning of class.

Homework 7

Exercise 1. Consistency of a continuous method of moments estimator follows from the following statement, which you are required to prove.

Fix $k \geq 1$. For any $1 \leq j \leq k$, let $M_{j,1}, M_{j,2}, \ldots$ be real-valued random variables that converge in probability to a constant $c_j \in \mathbf{R}$. Let $h \colon \mathbf{R}^k \to \mathbf{R}$ be continous. Then, as $n \to \infty$,

$$h(M_{1,n}, \ldots, M_{j,n})$$

converges in probability to the constant $h(c_1, \ldots, c_j)$.

Exercise 2. This exercise demonstrates that the MLE might not be consistent.

Let Z be a Gaussian random variable with mean $\mu \in \mathbf{R}$ and variance $\sigma^2 > 0$. Then $X := e^Z$ has the lognormal distribution with parameters μ and σ^2. Let $\gamma \in \mathbf{R}$ and define

$$X' := \gamma + e^Z.$$

In this case X' is said to have the three-parameter lognormal distribution with parameters $\gamma, \mu \in \mathbf{R}$, and $\sigma^2 > 0$. Let $X_1, \ldots, X_n$ be i.i.d. from this three-parameter lognormal distribution.

- Find the density of X_1.
- Suppose γ is known. Find the maximum likelihood estimator (M, T) of (μ, σ^2). (Assume $\gamma < X_{(1)}$.)
- Let $\ell(\gamma, \mu, \sigma^2)$ denote the log-likelihood function. The MLE of (γ, μ, σ^2) if it exists, will maximize $\ell(\gamma, M, T)$ over γ. Determine

  $$\lim_{\gamma \uparrow X_{(1)}} \ell(\gamma, M, T).$$

  Hint: Show first that as $\gamma \uparrow X_{(1)}$,

  $$M = M(\gamma) \sim \frac{1}{n} \log(X_{(1)} - \gamma), \qquad \text{and } T = T(\gamma) \sim \frac{n-1}{n^2} \log^2(X_{(1)} - \gamma),$$

  where the notation $f(\gamma) \sim g(\gamma)$ means $f(\gamma)/g(\gamma) \to 1$ as $\gamma \uparrow X_{(1)}$.

Why does the last conclusion violate consistency of the MLE? (Note that the point achieving the maximum of ℓ might not be unique.) What assumption of the MLE Consistency Theorem does not hold in this case?

Exercise 3 (Least Squares/ Ridge Regression, Part 2)**.** 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 $X_1, \ldots, X_n \in \mathbf{R}$. In linear least squares regression, we try to determine the best linear relationship between the vectors $x^{(1)}, \ldots, x^{(n)}$ and the data $X_1, \ldots, X_n$. 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. In a previous homework, we found $w \in \mathbf{R}^k$ that minimizes the quantity

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

We also interpreted the minimal w as an estimator. In some cases, the estimator for w could have large variance, which is undesirable. To deal with this issue, let $c > 0$ and consider the quantity

$$\sum_{i=1}^{n} (X_i - \langle x^{(i)}, w \rangle)^2 + c \left|\left|w\right|\right|^2 . \qquad (*)$$

Find the value of $w \in \mathbf{R}^k$ that minimizes this quantity.

The term $\left|\left|w\right|\right|^2$ penalizes w from having large entries. By Lagrange Multipliers, a critical point w of the constrained minimization problem

$$\text{minimize } \sum_{i=1}^{n} (X_i - \langle x^{(i)}, w \rangle)^2 \quad \text{subject to } \left|\left|w\right|\right|^2 \leq 1$$

is equivalent to the existence of a $c \in \mathbf{R}$ such that w is a critical point of $(*)$.

The L_2 penalization term in $(*)$ sometimes still allows w to have large entries. So, let $c > 0$ and consider the quantity

$$\sum_{i=1}^{n} (X_i - \langle x^{(i)}, w \rangle)^2 + c \sum_{i=1}^{n} |w_i| . \qquad (**)$$

Prove that there exists a $w \in \mathbf{R}^k$ that minimizes this quantity (this w is known as the LASSO, or least absolute shrinkage and selection operator). The L_1 penalization term in $(**)$ is better at penalizing large entries of w (a similar observation applies in the compressed sensing literature). Unfortunately, there is no closed form solution to $(**)$ in general. The constrained minimization problem

$$\text{minimize } \sum_{i=1}^{n} (X_i - \langle x^{(i)}, w \rangle)^2 \quad \text{subject to } \sum_{i=1}^{n} |w_i| \leq 1$$

is morally equivalent to $(**)$, but technically Lagrange Multipliers does not apply since the constraint is not differentiable everywhere.

Exercise 4 (Second Order Jackknife)**.** Let $X_1, X_2, \ldots \colon \Omega \to \mathbf{R}^n$ be i.i.d random variables so that X_1 has distribution $f_\theta \colon \mathbf{R}^n \to [0, \infty)$, $\theta \in \Theta$. Let $Y_1, Y_2, \ldots$ be a sequence of estimators for θ so that for any $n \geq 1$, $Y_n = t_n(X_1, \ldots, X_n)$ for some $t_n \colon \mathbf{R}^{n^2} \to \Theta$. For any $n \geq 1$, define the **second order jackknife estimator** of Y_n to be

$$Z_n := \frac{n^2}{2}Y_n - \frac{(n-1)^2}{n}\sum_{i=1}^{n} t_{n-1}(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$$
$$+ \frac{(n-2)^2}{n(n-1)}\sum_{1\leq i<j\leq n}^{n} t_{n-2}(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_{j-1},X_{j+1},\ldots,X_n).$$

Assume that $Y_1, Y_2, \ldots$ are asymptotically unbiased, so that there exists $a, b, c, d \in \mathbf{R}$ such that

$$\mathbf{E}Y_n = \theta + a/n + b/n^2 + \frac{c}{n^3} + \frac{d}{n^4} + O(1/n^5), \qquad \forall\, n \geq 1. \qquad (*)$$

Show that

$$\mathbf{E}Z_n = \theta + O(1/n^3).$$

And if $c = d = 0$ and the $O(1/n^4)$ term is zero in $(*)$, then Z_n is unbiased.

For more on the jackknife, see here

Exercise 5. Do Question 1 on the Fall 2011 qualifying exam here:

https://dornsife.usc.edu/mgsa/statistics-a/

Exercise 6. Take another qual exam.

[Please submit your solutions together with the homework.]

Remark 1. For a discussion of the estimation of a covariance matrix in the context of rotationally equivariant estimation, see http://www.econ.uzh.ch/static/wp/econwp122.pdf