Homework 1

Benjamin Noland

1. (i) Let $x \in \mathbb{R}^d$. Then by Cauchy-Schwartz,

$$\|x\|_1 = \sum_{i=1}^{d} |x_i| = \left| \sum_{i=1}^{d} |x_i| \cdot 1 \right| \leq \left(\sum_{i=1}^{d} |x_i| \sum_{i=1}^{d} 1 \right)^{1/2} = \sqrt{d}\|x\|_2.$$

I will take $c_2 = \sqrt{d}$. To see that the bound $\|x\|_1 \leq c_2\|x\|_2$ is as tight as possible, it suffices to show that equality holds for a single $x \in \mathbb{R}^d$ with $x \neq 0$. If we take $x = (1, 1, \ldots 1) \in \mathbb{R}^d$ (i.e., a tuple where each component is a 1), then $\|x\|_1 = d$ and $\|x\|_2 = \sqrt{d}$, so that $\|x\|_1 = c_2\|x\|_2$. This bound is therefore as tight as possible.

Now for the lower bound. Let $x \in \mathbb{R}^d$, and assume that $x \neq 0$. Then

$$\frac{|x_i|}{\|x\|_1} = \frac{|x_i|}{\sum_{j=1}^{d} |x_j|} \leq 1 \quad \text{for any } 1 \leq i \leq d.$$

Hence,

$$\left(\frac{|x_i|}{\|x\|_1} \right)^2 = \frac{x_i^2}{\|x\|_1^2} \leq \frac{|x_i|}{\|x\|_1} \quad \text{for any } 1 \leq i \leq d,$$

so that

$$\frac{\|x\|_2}{\|x\|_1} = \sum_{i=1}^{d} \frac{x_i^2}{\|x\|_1^2} \leq \sum_{i=1}^{d} \frac{|x_i|}{\|x\|_1} = \frac{\|x\|_1}{\|x\|_1} = 1.$$

Thus $\|x\|_2 \leq \|x\|_1$. Moreover, this bound holds trivially when $x = 0$, and therefore holds for any $x \in \mathbb{R}^d$. I will take $c_1 = 1$. Note that if we take $x = (1, 0, \ldots, 0) \in \mathbb{R}^d$ (i.e., a tuple consisting of 1 followed by $(d-1)$ 0's), then $c_1\|x\|_1 = \|x\|_2$. This bound is therefore as tight as possible.

To summarize, we therefore have the inequality

$$c_1\|x\|_2 \leq \|x\|_1 \leq c_2\|x\|_2 \quad \text{for every } x \in \mathbb{R}^d,$$

where the constants c_1 and c_2, chosen above, make the bounds as tight as possible.

(ii) We know from the argument in part (i) of this problem that $\|x\|_2 \leq \|x\|_1$ for any $x \in \mathbb{R}^d$, for any choice of $d \geq 1$. Thus we can take $c_1 = 1$, and this choice is independent of d.

However, for the upper bound, no such constant c_2 exists. To see this, note first that, as established in part (i),

$$\|x\|_1 \leq \sqrt{d}\|x\|_2 \quad \text{for any } x \in \mathbb{R}^d, \tag{1}$$

and this bound is as tight as possible for any fixed value of $d \geq 1$. Now suppose a choice of c_2 exists that is independent of d. Then since the bound (1) is as tight as possible for any fixed $d \geq 1$, we must have

$$\sqrt{d}\|x\|_2 \leq c_2\|x\|_2 \quad \text{for any } x \in \mathbb{R}^d \text{ and any } d \geq 1. \tag{2}$$

In particular, this implies that we must have $\sqrt{d} \leq c_2$ for any $d \geq 1$ (take $x \neq 0$ in (2) and divide through by $\|x\|_2$). However, this does not hold if $d > c_2^2$. Therefore no choice of c_2 exists that is independent of d.

2. First we need to establish that any norm $\|\cdot\| : \mathbb{R}^d \to \mathbb{R}$ is continuous. To see this, let $x \in \mathbb{R}^d$, and let $\{x_n\}$ be a sequence in $\mathbb{R}^d$ such that $x_n \to x$. Let $\epsilon > 0$. Then there exists $N \geq 1$ such that for every $n \geq N$,

$$\|x_n - x\| < \epsilon.$$

Hence for every $n \geq N$,

$$|\|x_n\| - \|x\|| \leq \|x_n - x\| < \epsilon$$

through an application of the triangle inequality. Therefore $\|x_n\| \to \|x\|$, so that the norm $\|\cdot\|$ is continuous.

Now for the main result. Let $\epsilon > 0$. Note that we can choose a constant $M > 0$ satisfying $\mathrm{P}\{\|X\| > M\} < \epsilon/2$ (this follows from basic properties of CDFs). Then since $X_n \xrightarrow{\mathcal{D}} X$, the Continuous Mapping Theorem implies that

$$\mathrm{P}\{\|X_n\| > M\} \to \mathrm{P}\{\|X\| > M\} \quad \text{as } n \to \infty,$$

since norms are everywhere continuous, as established above. Thus there exists $N \geq 1$ such that for every $n \geq N$,

$$\mathrm{P}\{\|X_n\| > M\} - \mathrm{P}\{\|X\| > M\} \leq |\mathrm{P}\{\|X_n\| > M\} - \mathrm{P}\{\|X\| > M\}| < \frac{\epsilon}{2},$$

and so rearrangement yields,

$$\mathrm{P}\{\|X_n\| > M\} \leq \mathrm{P}\{\|X\| > M\} + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

since M was chosen to satisfy $\mathrm{P}\{\|X\| > M\} < \epsilon/2$. Therefore $\mathrm{P}\{\|X_n\| > M\} < \epsilon$ for every $n \geq N$, so that $X_n = O_p(1)$.

3. Let $X_n = o_p(1)$ and $Y_n = O_p(1)$. First, we want to show that $X_n + Y_n = O_p(1)$. Let $\epsilon > 0$. Then since $Y_n = O_p(1)$, there exists $N \geq 1$ and a constant $M > 0$ such that

$$\mathrm{P}\left\{\|Y_n\| > \frac{M}{2}\right\} < \frac{\epsilon}{2} \quad \text{for every } n \geq N.$$

In addition, since $X_n = o_p(1)$, there exists $N' \geq 1$ such that

$$\mathrm{P}\left\{\|X_n\| > \frac{M}{2}\right\} < \frac{\epsilon}{2} \quad \text{for every } n \geq N'.$$

Thus for any $n \geq \max\{N, N'\}$,

$$\begin{aligned}\mathrm{P}\{\|X_n + Y_n\| > M\} &\leq \mathrm{P}\{\|X_n\| + \|Y_n\| > M\} \\ &\leq \mathrm{P}\left\{\|X_n\| > \frac{M}{2}\right\} + \mathrm{P}\left\{\|Y_n\| > \frac{M}{2}\right\} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,\end{aligned}$$

through an application of the triangle inequality. Therefore $\mathrm{P}\{\|X_n + Y_n\| > M\} < \epsilon$ for every $n \geq \max\{N, N'\}$, so that $X_n + Y_n = O_p(1)$.

Next, we want to show that $X_n Y_n = o_p(1)$. I will prove this under the assumption that X_n and Y_n are both scalar random variables (the result does not make sense for random vectors in general). Let $\epsilon > 0$. Then since $Y_n = O_p(1)$ there exists $N \geq 1$ and a constant $M > 0$ such that

$$\mathrm{P}\{|Y_n| > M\} < \frac{\epsilon}{2} \quad \text{for every } n \geq N.$$

Let $\delta > 0$. Then since $X_n = o_p(1)$, there exists $N' \geq 1$ such that

$$\mathrm{P}\left\{|X_n| > \frac{\delta}{M}\right\} < \frac{\epsilon}{2} \quad \text{for every } n \geq N'.$$

Therefore, for every $n \geq \max\{N, N'\}$,

$$\begin{aligned}\mathrm{P}\{|X_n Y_n| > \delta\} &\leq \mathrm{P}\left\{|X_n| > \frac{\delta}{M}\right\} + \mathrm{P}\{|Y_n| > M\} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

Hence $\mathrm{P}\{|X_n Y_n| > \delta\} \to 0$ as $n \to \infty$, so that $X_n Y_n = o_p(1)$.

4. Let $\Theta = [0, 1)$, and define maps $g_n : \Theta \to \mathbb{R}$ (for every $n \geq 1$) and $g : \Theta \to \mathbb{R}$ by

$$g_n(\theta) = \theta^n \quad \text{and} \quad g(\theta) = 0 \quad \text{for every } \theta \in \Theta.$$

Then g and g_n are continuous on Θ for every $n \geq 1$, and the sequence $\{g_n\}$ converges pointwise to g; that is

$$\lim_{n\to\infty} g_n(\theta) = \lim_{n\to\infty} \theta^n = 0 = g(\theta) \quad \text{for every } \theta \in \Theta.$$

However,

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} |g_n(\theta) - g(\theta)| = \lim_{n\to\infty} \sup_{\theta\in\Theta} \theta^n = 1,$$

and therefore the sequence $\{g_n\}$ does not converge uniformly to g.

5. Let $X_1, \ldots, X_n$ be IID random variables, and let $N = \{\theta : \|\theta - \theta_0\| < \delta\}$, where $\delta > 0$, denote a neighborhood of θ_0 where the given bound on the second partial derivatives holds. Let $\theta \in N$. Let $\|\cdot\|_{\max}$ denote the max norm on the space of $d \times d$ matrices. Then norm equivalence implies that there exists a constant $r > 0$ such that

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \psi(X_i;\theta) \right\| &\leq r \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \psi(X_i;\theta) \right\|_{\max} \\
&\leq r \frac{1}{n} \sum_{i=1}^{n} \left\| \frac{\partial^2}{\partial\theta\partial\theta^T} \psi(X_i;\theta) \right\|_{\max} \\
&= r \frac{1}{n} \sum_{i=1}^{n} \max_{j,k} \left\{ \left| \frac{\partial^2}{\partial\theta_j\partial\theta_k} \psi(X_i;\theta) \right| \right\} \\
&\leq r \frac{1}{n} \sum_{i=1}^{n} M(X_i).
\end{aligned} \tag{1}$$

Next we want to show that $\tilde{\theta}_n \xrightarrow{p} \theta_0$. Let $\epsilon > 0$. Recall that by definition, $\tilde{\theta}_n$ lies on the line segment between $\hat{\theta}_n$ and θ_0. Thus if $\|\tilde{\theta}_n - \theta_0\| > \epsilon$, then $\|\hat{\theta}_n - \theta_0\| > \epsilon$. Hence,

$$\mathrm{P}\{\|\tilde{\theta}_n - \theta_0\| > \epsilon\} \leq \mathrm{P}\{\|\hat{\theta}_n - \theta_0\| > \epsilon\} \to 0 \quad \text{as } n \to \infty,$$

and hence $\tilde{\theta}_n \xrightarrow{p} \theta_0$. Therefore,

$$\mathrm{P}\{\tilde{\theta}_n \in N\} = \mathrm{P}\{\|\tilde{\theta}_n - \theta_0\| < \delta\} \to 1 \quad \text{as } n \to \infty.$$

In other words, $\tilde{\theta}_n$ will lie in N with high probability if n is chosen large enough. In particular, this means that the bound on the second partial derivatives holds with high probability if n is chosen large enough. Thus if we let

$$\Psi = \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta\partial\theta^T} \psi(X_i;\tilde{\theta}_n) \right\|,$$

then, using (1),

$$\mathrm{P}\left\{ \|\Psi\| \leq r \frac{1}{n} \sum_{i=1}^{n} M(X_i) \right\} \to 1 \quad \text{as } n \to \infty. \tag{2}$$

Furthermore, note that since $\mu = \mathrm{E}[M(X_1)] < \infty$, the WLLN implies that

$$\frac{1}{n}\sum_{i=1}^{n} M(X_i) \xrightarrow{p} \mu. \tag{3}$$

Let $\epsilon > 0$, and assume without loss of generality that $\epsilon < 1$, so that $\epsilon/2 < 1 - \epsilon/2$. Then by (3) there exists $N \geq 1$ such that for every $n \geq N$,

$$\mathrm{P}\left\{\mu - \frac{1}{n}\sum_{i=1}^{n} M(X_i) > \frac{\mu}{2}\right\} \leq \mathrm{P}\left\{\left|\mu - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| > \frac{\mu}{2}\right\} < 1 - \frac{\epsilon}{2}.$$

In addition, by (2) there exists $N' \geq 1$ such that

$$\mathrm{P}\left\{\|\Psi\| > r\frac{1}{n}\sum_{i=1}^{n} M(X_i)\right\} < \frac{\epsilon}{2} \quad \text{for every } n \geq N'.$$

Thus, for every $n \geq \max\{N, N'\}$,

$$\begin{aligned}
&\mathrm{P}\left\{\|\Psi\| > r\frac{\mu}{2}, \mu - \frac{1}{n}\sum_{i=1}^{n} M(X_i) > \frac{\mu}{2}\right\} \\
&\qquad = \mathrm{P}\left\{\|\Psi\| + r\frac{\mu}{2} > r\mu, \mu - \frac{1}{n}\sum_{i=1}^{n} M(X_i) > \frac{\mu}{2}\right\} \\
&\qquad = \mathrm{P}\left\{\|\Psi\| > r\frac{1}{n}\sum_{i=1}^{n} M(X_i), \mu - \frac{1}{n}\sum_{i=1}^{n} M(X_i) > \frac{\mu}{2}\right\} \\
&\qquad \leq \mathrm{P}\left\{\|\Psi\| > r\frac{1}{n}\sum_{i=1}^{n} M(X_i)\right\} < \frac{\epsilon}{2},
\end{aligned}$$

and so upon taking complements,

$$\mathrm{P}\left\{\|\Psi\| \leq r\frac{\mu}{2}\right\} + \mathrm{P}\left\{\mu - \frac{1}{n}\sum_{i=1}^{n} M(X_i) \leq \frac{\mu}{2}\right\} \geq 1 - \frac{\epsilon}{2}.$$

Finally, rearrangement yields

$$\mathrm{P}\left\{\|\Psi\| \leq r\frac{\mu}{2}\right\} \geq 1 - \frac{\epsilon}{2} - \mathrm{P}\left\{\mu - \frac{1}{n}\sum_{i=1}^{n} M(X_i) \leq \frac{\mu}{2}\right\} \geq 1 - \frac{\epsilon}{2} - \frac{\epsilon}{2} = 1 - \epsilon.$$

Therefore, upon taking complements again,

$$\mathrm{P}\left\{\|\Psi\| > r\frac{\mu}{2}\right\} < \epsilon.$$

Therefore $\Psi = O_p(1)$.

6. Note that since H is constant, $H = O_p(1)$ and $H^{-1} = O_p(1)$. Therefore,

$$\begin{aligned}(H + o_p(1))(H^{-1} + o_p(1)) &= HH^{-1} + Ho_p(1) + o_p(1)H^{-1} + o_p(1)p_p(1) \\ &= I + O_p(1)o_p(1) + o_p(1)O_p(1) + o_p(1) \\ &= I + o_p(1) + o_p(1) + o_p(1) \\ &= I + o_p(1).\end{aligned}$$

Thus $(H + o_p(1))$ is non-singular with inverse $(H^{-1} + o_p(1))$ if and only if $I + o_p(1) = I$ identically. We have

$$\mathrm{P}\{I + o_p(1) = I\} = \mathrm{P}\{o_p(1) = O\} = \mathrm{P}\{\|o_p(1)\| = 0\},$$

where O denotes the square zero matrix of appropriate dimension. But since, for every $\epsilon > 0$,

$$\mathrm{P}\{\|o_p(1)\| > \epsilon\} \to 0 \quad \text{as } n \to \infty,$$

we see that the probability that $(H + o_p(1))$ is non-singular with inverse $(H^{-1} + o_p(1))$ tends to 1 as $n \to \infty$.

7. If we define

$$\begin{aligned}\hat{G}_n &= \frac{1}{n}\sum_{i=1}^{n} \psi(X_i; \hat{\theta}_n)^2 \\ \hat{H}_n &= \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\psi(X_i, \hat{\theta}_n),\end{aligned}$$

then $\hat{V}_n = \hat{H}_n^{-1}\hat{G}_n\hat{H}^{-T}$ seems like a reasonable candidate for a consistent estimator of V. (*I ran out of time before being able to give this problem much attention*).

8. (i) Let $x = (x_1, \ldots, x_n)$ be a realization of the sample $(X_1, \ldots, X_n)$. Define a sequence $\{\theta_k\}$ of estimates by $\theta_k = (1/2, 0, 1, x_1, 1/k)$ for every $k \geq 1$. Then, for every $k \geq 1$,

$$\begin{aligned}f(x_1, \theta_k) &= \frac{1/2}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_1^2\right) + \frac{1/2}{\sqrt{2\pi/k^2}}\exp\left(-\frac{1}{2/k^2}(x_1 - x_1)^2\right) \\ &= \frac{1/2}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_1^2\right) + \frac{1/2}{\sqrt{2\pi}}k \geq \frac{1/2}{\sqrt{2\pi}}k,\end{aligned}$$

and for any $i \geq 2$,

$$\begin{aligned}f(x_i, \theta_k) &= \frac{1/2}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_i^2\right) + \frac{1/2}{\sqrt{2\pi/k^2}}\exp\left(-\frac{1}{2/k^2}(x_i - x_1)^2\right) \\ &\geq \frac{1/2}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_i^2\right).\end{aligned}$$

The log-likelihood therefore satisfies

$$\log L(\theta_k; x) = \sum_{i=1}^{n} \log f(x_i; \theta_k) \geq \log\left(\frac{1/2}{\sqrt{2\pi}} k\right) - \frac{1}{2}\sum_{i=1}^{n} x_i^2 + n \log\left(\frac{1/2}{\sqrt{2\pi}}\right).$$

Hence $\log L(\theta_k; x) \to \infty$ as $k \to \infty$, so that the log-likelihood is unbounded. In particular,

$$\hat{\theta}_n = \lim_{k\to\infty} \theta_k = \left(\frac{1}{2}, 0, 1, x_1, 0\right)$$

defines a "degenerate" MLE, in the sense that $L(\hat{\theta}_n; x) = \infty$. Moreover, $\hat{\theta}_n$ lies on the boundary of the parameter space Θ, but not in Θ itself. The corresponding estimator

$$\hat{\theta}_n = \left(\frac{1}{2}, 0, 1, X_1, 0\right)$$

is inconsistent, since if $\theta_0 = (\alpha, \mu_1, \sigma_1, \mu_2, \sigma_2)$ denotes the true value of the parameter θ, the condition $\hat{\theta}_n \xrightarrow{p} \theta_0$ implies that $0 \xrightarrow{p} \sigma_2$ by the Continuous Mapping Theorem, and hence that $\sigma_2 = 0$, which is impossible.

(ii) Define

$$\Theta_1 = \{(\alpha, \mu_1, \sigma_1, \mu_2, \sigma_2) : 0 \leq \alpha \leq 1, 0 \leq \mu_1, \mu_2 \leq 1, 1 \leq \sigma_1, \sigma_2 \leq 2\}.$$

Then $\Theta_1 \subset \Theta$. Thus if the true value, say θ_0, of the parameter θ lies in Θ_1, the following conditions ensure that the MLE of θ in Θ_1 is consistent (see Theorem 2.5 in Newey and McFadden (1994)):

(1) Θ_1 is compact (i.e., it is closed and bounded).

(2) The pdf $f(x; \theta)$ is identifiable.

(3) $\log f(x; \theta)$ is continuous as a function of θ on Θ_1 for any fixed x.

(4) $\mathrm{E}(\sup_{\theta\in\Theta_1} |\log f(X; \theta)|) < \infty$.

Assume that $\theta_0 \in \Theta_1$. Conditions (1)-(3) are easily seen to be satisfied. To see that condition (4) also holds, note that since Θ_1 is compact, the fact that $\log f(x; \theta)$, and hence $|\log f(x; \theta)|$, is continuous as a function of θ on Θ_1 for any fixed x implies that $|\log f(x; \theta)|$ attains its global maximum at a point $\theta' \in \Theta_1$. Hence,

$$\mathrm{E}\left(\sup_{\theta\in\Theta_1} |\log f(X; \theta)|\right) = \mathrm{E}\left(|\log f(X; \theta')|\right).$$

To show that $\mathrm{E}\left(|\log f(X; \theta')|\right) < \infty$, it suffices to show that $\mathrm{E}\left(\log f(X; \theta')\right) < \infty$. We have the following:

$$\begin{aligned}
\mathrm{E}\left(\log f(X; \theta')\right) &= \int_{-\infty}^{\infty} \log f(x; \theta') f(x; \theta)\, dx \\
&\leq \int_{-\infty}^{\infty} (f(x; \theta') - 1) f(x; \theta)\, dx < \infty,
\end{aligned}$$

using the fact that $\log f(x;\theta') \leq f(x;\theta') - 1$ for any x, and that $(f(x;\theta') - 1)$ is bounded as a function of x. Thus condition (4) is satisfied, so that the MLE of θ in Θ_1 is consistent.