Homework 2

Benjamin Noland

1. First, let $\theta_1, \theta_2 \in \Theta$ be such that θ_1 lies on the line segment connecting θ_0 and θ_2; that is, $\theta_1 \in L$, where $L = \{t\theta_0 + (1-t)\theta_2 : t \in [0,1]\}$. I claim that $K(\theta_1) \leq K(\theta_2)$. To see this, note that since K is convex,

$$K(t\theta_0 + (1-t)\theta_2) \leq tK(\theta_0) + (1-t)K(\theta_2) \quad \text{for every } t \in [0,1].$$

In particular, since $\theta_1 \in L$, there exists $t' \in [0,1]$ such that $\theta_1 = t'\theta_0 + (1-t')\theta_2$. Hence,

$$\begin{aligned} K(\theta_1) = K(t'\theta_0 + (1-t')\theta_2) &\leq t'K(\theta_0) + (1-t')K(\theta_2) \\ &\leq t'K(\theta_2) + (1-t')K(\theta_2) = K(\theta_2), \end{aligned}$$

since θ_0 is a minimizer of K. This verifies the claim.

Now for the main result. Let $\epsilon > 0$. Write $B_\epsilon^c = \Theta \setminus B_\epsilon$. Since B_ϵ is a bounded set, and its boundary ∂B_ϵ is closed, it follows that ∂B_ϵ is bounded as well, so that ∂B_ϵ is compact. It is a well-known fact that a convex function on an open convex set is continuous; in particular, K is continuous on Θ. Thus the map $f : \Theta \to \mathbb{R}$ defined by $f(\theta) = K(\theta) - K(\theta_0)$ is continuous as well. Therefore, since ∂B_ϵ is compact, f attains a minimum value $\delta = \min_{\theta \in \partial B_\epsilon} f(\theta)$ on ∂B_ϵ. Moreover, since θ_0 is the unique minimizer of K, and $\theta_0 \notin \partial B_\epsilon$, we see that $\delta > 0$. Next, let $\theta \in B_\epsilon^c$. Let $L = \{t\theta_0 + (1-t)\theta : t \in [0,1]\}$ denote the line segment connecting θ_0 and θ. Then we can choose a point $\theta' \in L$ such that $\theta' \in \partial B_\epsilon$, and by the above claim, $K(\theta') \leq K(\theta)$, so that $K(\theta') - K(\theta) \geq 0$. Therefore,

$$\begin{aligned} K(\theta) - K(\theta_0) &= [K(\theta) - K(\theta')] + [K(\theta') - K(\theta_0)] \\ &= [K(\theta) - K(\theta')] + f(\theta') \\ &\geq [K(\theta) - K(\theta')] + \delta \geq \delta > 0. \end{aligned}$$

Hence,

$$\inf_{\theta \in B_\epsilon^c} [K(\theta) - K(\theta_0)] \geq \delta > 0,$$

so that

$$\inf_{\theta \in B_\epsilon^c} K(\theta) > K(\theta_0),$$

as desired.

2. (i) Let $\epsilon > 0$ and $\delta > 0$, and let $\epsilon_0 = \min\{c\delta^2, \epsilon\}$. Then by hypothesis, there exists $N \geq 1$ such that

$$\mathrm{P}\{|X_n| > \sqrt{\epsilon_0/c}\} \leq \epsilon_0 \quad \text{for every } n \geq N,$$

where $c > 0$ is some constant. Hence for every $n \geq N$, since $\epsilon_0 \leq c\delta^2$ (so that $\delta \geq \sqrt{\epsilon_0/c}$) and $\epsilon_0 \leq \epsilon$, we have

$$\mathrm{P}\{|X_n| > \delta\} \leq \mathrm{P}\{|X_n| > \sqrt{\epsilon_0/c}\} \leq \epsilon_0 \leq \epsilon,$$

and therefore $X_n \xrightarrow{p} 0$, so that $X_n = o_p(1)$.

(ii) Suppose that for every $\epsilon > 0$ there exists $N \geq 1$ such that

$$\mathrm{P}\{|X_n| > M(\epsilon)\} \leq \epsilon \quad \text{for every } n \geq N,$$

where $M : \mathbb{R} \to \mathbb{R}^{>0}$ (here $\mathbb{R}^{>0}$ denotes the set of positive real numbers) is a strictly-increasing function. I claim that $X_n = o_p(1)$.

Let $\epsilon > 0$ and $\delta > 0$. Note that since M is strictly-increasing, it has an inverse $M^{-1} : \mathbb{R}^{>0} \to \mathbb{R}$. Let $\epsilon_0 = \min\{M^{-1}(\delta), \epsilon\}$. Then by hypothesis there exists $N \geq 1$ such that

$$\mathrm{P}\{|X_n| > M(\epsilon_0)\} \leq \epsilon_0 \quad \text{for every } n \geq N.$$

Since $\epsilon_0 \leq M^{-1}(\delta)$, the fact that M is strictly increasing implies that $M(\epsilon_0) \leq \delta$. Moreover, $\epsilon_0 \leq \epsilon$, and therefore, for every $n \geq N$,

$$\mathrm{P}\{|X_n| > \delta\} \leq \mathrm{P}\{|X_n| > M(\epsilon_0)\} \leq \epsilon_0 \leq \epsilon.$$

Thus $X_n \xrightarrow{p} 0$, so that $X_n = o_p(1)$.

3. (i) Let $\epsilon > 0$. Then Markov's Inequality implies that

$$\mathrm{P}\{|Z_n^2| \geq \epsilon\} = \mathrm{P}\{Z_n^2 \geq \epsilon\} \leq \frac{\mathrm{E}(Z_n^2)}{\epsilon}.$$

Thus, since $\mathrm{E}(Z_n^2) \to 0$ as $n \to \infty$ by hypothesis, we see that $\mathrm{P}\{|Z_n^2| \geq \epsilon\} \to 0$ as $n \to \infty$, as well. Thus $Z_n^2 \xrightarrow{p} 0$. Now, define a map $f : \mathbb{R}^{\geq 0} \to \mathbb{R}$ (here $\mathbb{R}^{\geq 0}$ denotes the set of non-negative real numbers) by $f(x) = \sqrt{x}$. Then we can write $Z_n = f(Z_n^2)$. Since f is everywhere continuous, the Continuous Mapping Theorem implies that $f(Z_n^2) \xrightarrow{p} f(0)$, so that $Z_n \xrightarrow{p} 0$.

(ii) Suppose $f : \mathbb{R} \to \mathbb{R}^{\geq 0}$ is a function with a continuous inverse $f^{-1} : \mathbb{R}^{\geq 0} \to \mathbb{R}$ such that $f^{-1}(0) = 0$, and that f satisfies $\mathrm{E}(f(Z_n)) \to 0$ as $n \to \infty$. I claim that $Z_n \xrightarrow{p} 0$.

Let $\epsilon > 0$. Then Markov's Inequality implies that

$$\mathrm{P}\{|f(Z_n)| \geq \epsilon\} = \mathrm{P}\{f(Z_n) \geq \epsilon\} \leq \frac{\mathrm{E}(f(Z_n))}{\epsilon}.$$

Thus since $\mathrm{E}(f(Z_n)) \to 0$ as $n \to \infty$ by hypothesis, we see that $\mathrm{P}\{|f(Z_n)| \geq \epsilon\} \to 0$ as $n \to \infty$, as well. Thus $f(Z_n) \xrightarrow{p} 0$. Moreover, since f has a continuous inverse f^{-1}, the Continuous Mapping Theorem implies that $f^{-1}(f(Z_n)) \xrightarrow{p} f^{-1}(0)$. Thus, since $f^{-1}(0) = 0$ by hypothesis, we see that $Z_n \xrightarrow{p} 0$.

4. Since K is twice-differentiable, it is possible to expand it into a second-order Taylor series about θ_0. Specifically, we can write

$$\begin{aligned} K(\theta_0 + \alpha/\sqrt{n}) = K(\theta_0) &+ (\theta_0 + \alpha/\sqrt{n} - \theta_0)^T \left(\frac{\partial}{\partial\theta} K(\theta_0) \right) \\ &+ \frac{1}{2}(\theta_0 + \alpha/\sqrt{n} - \theta_0)^T \left(\frac{\partial^2}{\partial\theta\partial\theta^T} K(\tilde{\theta}_n) \right) (\theta_0 + \alpha/\sqrt{n} - \theta_0), \end{aligned}$$

where $\tilde{\theta}_n$ lies on the line segment between $(\theta_0 + \alpha/\sqrt{n})$ and θ_0. Since θ_0 is a minimizer of K, $\frac{\partial}{\partial\theta}K(\theta_0) = 0$. Thus we can simplify the above expression to yield

$$\begin{aligned} K(\theta_0 + \alpha/\sqrt{n}) &= K(\theta_0) + \frac{1}{2}(\alpha/\sqrt{n})^T \left(\frac{\partial^2}{\partial\theta\partial\theta^T} K(\tilde{\theta}_n) \right) (\alpha/\sqrt{n}) \\ &= K(\theta_0) + \frac{1}{2n}\alpha^T \left(\frac{\partial^2}{\partial\theta\partial\theta^T} K(\tilde{\theta}_n) \right) \alpha. \end{aligned}$$

Therefore,

$$n[K(\theta_0 + \alpha/\sqrt{n}) - K(\theta_0)] = \frac{1}{2}\alpha^T \left(\frac{\partial^2}{\partial\theta\partial\theta^T} K(\tilde{\theta}_n) \right) \alpha.$$

Now, since $(\theta_0 + \alpha/\sqrt{n}) \to \theta_0$ as $n \to \infty$, and $\tilde{\theta}_n$ lies on the line segment between $(\theta_0 + \alpha/\sqrt{n})$ and θ_0, it follows that $\tilde{\theta}_n \to \theta_0$ as $n \to \infty$, as well. It follows that, as $n \to \infty$,

$$\begin{aligned} n[K(\theta_0 + \alpha/\sqrt{n}) - K(\theta_0)] &= \frac{1}{2}\alpha^T \left(\frac{\partial^2}{\partial\theta\partial\theta^T} K(\tilde{\theta}_n) \right) \alpha \\ &\to \frac{1}{2}\alpha^T \left(\frac{\partial^2}{\partial\theta\partial\theta^T} K(\theta_0) \right) \alpha = \frac{1}{2}\alpha^T H \alpha. \end{aligned}$$

5. (i) Define

$$\begin{aligned} f(x, y; \beta) = \mathrm{P}\{Y = y|x\} &= \mathrm{P}\{Y = 1|x\}^y \mathrm{P}\{Y = 0|x\}^{1-y} \\ &= \left(\frac{e^{x^T\beta}}{1 + e^{x^T\beta}} \right)^y \left(\frac{1}{1 + e^{x^T\beta}} \right)^{1-y}, \end{aligned}$$

where $y \in \{0,1\}$. The log-likelihood is therefore given by

$$\begin{aligned}
\log L(\beta; x, y) &= \sum_{i=1}^{n} \log f(x, y; \beta) \\
&= \sum_{i=1}^{n} \left[y_i \left(\frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}} \right) + (1 - y_i) \left(\frac{1}{1 + e^{x_i^T \beta}} \right) \right] \\
&= \sum_{i=1}^{n} \left\{ y_i \left[x_i^T \beta - \log(1 + e^{x_i^T \beta}) \right] - (1 - y_i) \log(1 + e^{x_i^T \beta}) \right\} \\
&= \sum_{i=1}^{n} \left[y_i x_i^T \beta - \log(1 + e^{x_i^T \beta}) \right] = \sum_{i=1}^{n} [-\rho(x_i, y_i; \beta)],
\end{aligned}$$

where

$$\rho(x, y; \beta) = -y x^T \beta + \log(1 + e^{x^T \beta}).$$

Take the parameter space Θ to be a convex set. Note that $\rho(x, y; \beta)$ is twice-differentiable in β. Define

$$h(x, y; \beta) = \frac{\partial}{\partial \beta} \rho(x, y; \beta) = -yx + \frac{e^{x^T \beta}}{1 + e^{x^T \beta}} x.$$

Thus $h(x, y; \beta)$ is a subgradient of $\rho(x, y; \beta)$. Also define

$$H(\beta; x, y) = \frac{\partial^2}{\partial \beta \partial \beta^T} \rho(x, y; \beta) = \frac{e^{x^T \beta}}{(1 + e^{x^T \beta})^2} x x^T.$$

Note that for any $\beta \in \Theta$, since the matrix xx^T is positive semidefinite, it follows that the Hessian $H(\beta; x, y)$ is positive semidefinite, and thus $\rho(x, y; \beta)$ is convex in β. (In particular, this implies that the log-likelihood $\log L(\beta; x, y)$ is concave in β, and hence has a unique maximizer). Define

$$K(\beta) = \mathrm{E}[\rho(X, Y; \beta)] = \mathrm{E}[-YX + \log(1 + e^{X^T \beta})].$$

Assume that $K(\beta) < \infty$ for every $\beta \in \Theta$, and that $K(\beta)$ has a unique minimizer β_0. Assume further that $\mathrm{E}[|h(X, Y; \beta)|^2] < \infty$ for every β in a neighborhood of β_0. Assuming it is admissible to swap expectation with differentiation in this context, $K(\beta)$ is twice-differentiable, and so we have

$$\begin{aligned}
H = \frac{\partial^2}{\partial \beta \partial \beta^T} K(\beta_0) &= \frac{\partial^2}{\partial \beta \partial \beta^T} \mathrm{E}[\rho(X, Y; \beta_0)] = \mathrm{E}\left[\frac{\partial^2}{\partial \beta \partial \beta^T} \rho(X, Y; \beta_0) \right] \\
&= \mathrm{E}\left[\frac{e^{X^T \beta_0}}{(1 + e^{X^T \beta_0})^2} X X^T \right] = \mathrm{E}[H(X, Y; \beta_0)].
\end{aligned}$$

As established, the Hessian $H(x, y; \beta)$ of $\rho(x, y; \beta)$ is positive semidefinite since the matrix xx^T is positive semidefinite. Moreover, if xx^T is non-singular, then it must be positive definite. In particular, H is positive definite if $\mathrm{E}(XX^T)$ is non-singular.

Thus, under the conditions outlined above, $\hat{\beta}_{\mathrm{ML}}$ is consistent and asymptotically normal. To summarize (the definitions are as in the discussion above), the conditions are:

(1) The parameter space Θ is convex.

(2) $K(\beta) < \infty$ for every $\beta \in \Theta$.

(3) $K(\beta)$ has a unique minimizer β_0.

(4) $\mathrm{E}[|h(X, Y; \beta)|^2] < \infty$ for every β in a neighborhood of β_0.

(5) $\mathrm{E}(XX^T)$ is non-singular.

Therefore, under these conditions, $\hat{\beta}_{\mathrm{ML}} \xrightarrow{p} \beta_0$, and

$$\sqrt{n}(\hat{\beta}_{\mathrm{ML}} - \beta_0) = -H^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(X_i, Y_i; \beta_0) + o_p(1),$$

so that

$$\sqrt{n}(\hat{\beta}_{\mathrm{ML}} - \beta_0) \xrightarrow{\mathcal{D}} N(0, H^{-1}JH^{-1}),$$

where J is given by

$$\begin{aligned} J &= \mathrm{E}[h(X, Y; \beta_0)h(X, Y; \beta_0)^T] \\ &= \mathrm{E}\left[\left(-YX + \frac{e^{X^T\beta_0}}{1 + e^{X^T\beta_0}}X\right)\left(-YX + \frac{e^{X^T\beta_0}}{1 + e^{X^T\beta_0}}X\right)^T\right]. \end{aligned}$$

(ii) Define

$$\rho(x, y; \beta) = ye^{-\beta^T x} + (1 - y)\beta^T x.$$

Take the parameter space Θ to be a convex set. Note that $\rho(x, y; \beta)$ is twice-differentiable in β. Define

$$h(x, y; \beta) = \frac{\partial}{\partial\beta}\rho(x, y; \beta) = -ye^{-\beta^T x}x + (1 - y)x.$$

Thus $h(x, y; \beta)$ is a subgradient of $\rho(x, y; \beta)$. Also define

$$H(\beta; x, y) = \frac{\partial^2}{\partial\beta\partial\beta^T}\rho(x, y; \beta) = ye^{-\beta^t x}xx^T.$$

Note that for any $\beta \in \Theta$, since the matrix xx^T is positive semidefinite, it follows that the Hessian $H(\beta; x, y)$ is positive semidefinite, and thus $\rho(x, y; \beta)$ is convex in β. Define

$$K(\beta) = \mathrm{E}[\rho(X, Y; \beta)] = \mathrm{E}(Ye^{-\beta^T X} + (1 - Y)\beta^T X).$$

Assume that $K(\beta) < \infty$ for every $\beta \in \Theta$, and that $K(\beta)$ has a unique minimizer β_0. Assume further that $\mathrm{E}[|h(X, Y; \beta)|^2] < \infty$ for every β in a neighborhood of β_0. Assuming it is admissible to swap expectation with differentiation in this context, $K(\beta)$ is twice-differentiable, and so we have

$$\begin{aligned} H = \frac{\partial^2}{\partial\beta\partial\beta^T} K(\beta_0) = \frac{\partial^2}{\partial\beta\partial\beta^T} \mathrm{E}[\rho(X, Y; \beta_0)] &= \mathrm{E}\left[\frac{\partial^2}{\partial\beta\partial\beta^T} \rho(X, Y; \beta_0)\right] \\ &= \mathrm{E}(Y e^{-\beta_0^t X} X X^T) = \mathrm{E}[H(X, Y; \beta_0)]. \end{aligned}$$

As established, the Hessian $H(x, y; \beta)$ of $\rho(x, y; \beta)$ is positive semidefinite since the matrix xx^T is positive semidefinite. Moreover, if xx^T is non-singular, then it must be positive definite. In particular, H is positive definite if $\mathrm{E}(XX^T)$ is non-singular.

Thus, under the conditions outlined above, $\hat{\beta}_{\mathrm{CAL}}$ is consistent and asymptotically normal. To summarize (the definitions are as in the discussion above), the conditions are:

(1) The parameter space Θ is convex.

(2) $K(\beta) < \infty$ for every $\beta \in \Theta$.

(3) $K(\beta)$ has a unique minimizer β_0.

(4) $\mathrm{E}[|h(X, Y; \beta)|^2] < \infty$ for every β in a neighborhood of β_0.

(5) $\mathrm{E}(XX^T)$ is non-singular.

Therefore, under these conditions, $\hat{\beta}_{\mathrm{CAL}} \xrightarrow{p} \beta_0$, and

$$\sqrt{n}(\hat{\beta}_{\mathrm{CAL}} - \beta_0) = -H^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i, Y_i; \beta_0) + o_p(1),$$

so that

$$\sqrt{n}(\hat{\beta}_{\mathrm{CAL}} - \beta_0) \xrightarrow{\mathcal{D}} N(0, H^{-1} J H^{-1}),$$

where J is given by

$$\begin{aligned} J &= \mathrm{E}[h(X, Y; \beta_0) h(X, Y; \beta_0)^T] \\ &= \mathrm{E}[(-Y e^{-\beta^T X} X + (1 - Y)X)(-Y e^{-\beta^T X} X + (1 - Y)X)^T] \\ &= \mathrm{E}[(-YX(1 - e^{\beta^T X - 1}) + X)(-YX(1 - e^{\beta^T X - 1}) + X)^T]. \end{aligned}$$

(iii) The two estimators yield different expressions for their respective asymptotic variances. A precise comparison seems computationally arduous, however, and I will not undertake it due to time constraints.