HW 3.

Problem 1. $f_\theta(x) = \dfrac{\theta^{-x} e^{-\theta^{-1}}}{x!}$, $E(X) = \theta^{-1}$. $Var(X) = \theta^{-1}$.

$\ell(\theta) = \sum_{i=1}^{n} \{-x_i \log\theta - \theta^{-1} - \log(x_i!)\}$, $\ell'(\theta) = \sum_{i=1}^{n} \left(-\frac{x_i}{\theta} + \frac{1}{\theta^2}\right)$

Let $\ell'(\hat\theta) = 0 \Rightarrow \hat\theta = (n^{-1}\sum_i X_i)^{-1}$ is the MLE.

By CLT,

$$\sqrt{n}\,(n^{-1}\sum_i X_i - \theta^{-1}) \xrightarrow{d} N(0, \theta^{-1})$$

Let $g(\theta) = \theta^{-1}$, then $g'(\theta) = -\theta^{-2}$. $\theta^{-1}\cdot\{g'(\theta^{-1})\}^2 = \theta^3$.

By Delta's method,

$$\sqrt{n}\,(\hat\theta - \theta) \xrightarrow{d} N(0, \theta^3).$$

Problem 2. $f_\theta(x) = (2\pi\theta)^{-\frac{1}{2}} \exp\{-\frac{1}{2\theta}(x-\theta)^2\}$. $E(X) = \theta$, $E(X^2) = \theta^2 + \theta$, $Var(X^2) = 4\theta^3 + 2\theta^2$

$\ell(\theta) = \sum_{i=1}^{n} \{-\frac{1}{2}\log\theta - \frac{1}{2}\log(2\pi) - \frac{1}{2\theta}(x_i-\theta)^2\}$, $\ell'(\theta) = \sum_{i=1}^{n}\left(-\frac{1}{2\theta} + \frac{x_i^2}{2\theta^2} - \frac{1}{2}\right)$

Let $\ell'(\hat\theta) = 0 \Rightarrow \hat\theta^2 + \hat\theta - n^{-1}\sum_i X_i^2 = 0 \Rightarrow \hat\theta = \frac{1}{2}\left(\sqrt{1 + 4n^{-1}\sum X_i^2} - 1\right)$ is the MLE.

By CLT,

$$\sqrt{n}\,\{n^{-1}\sum_i X_i^2 - (\theta^2+\theta)\} \xrightarrow{d} N(0, 4\theta^3 + 2\theta^2).$$

Let $g(\theta) = \frac{1}{2}\left(\sqrt{1+4\theta} - 1\right)$, then $g'(\theta) = (1+4\theta)^{-\frac{1}{2}}$. $(4\theta^3 + 2\theta^2)\cdot\{g'(\theta^2+\theta)\}^2 = \dfrac{4\theta^3+2\theta^2}{(2\theta+1)^2} = \dfrac{2\theta^2}{2\theta+1}$

By Delta's method,

$$\sqrt{n}\,(\hat\theta - \theta) \xrightarrow{d} N\left(0, \frac{2\theta^2}{2\theta+1}\right).$$

Problem 3. $f_\theta(x) = \theta^{-1} \cdot \mathbb{1}\{0 \le x \le \theta\}$. $f_\theta(x_1, \cdots x_n) = \theta^{-n} \cdot \mathbb{1}\{\underbrace{x_{(1)}}_{\min(x_1,\cdots x_n)} \ge 0, \underbrace{x_{(n)}}_{\max(x_1,\cdots x_n)} \le \theta\} = \begin{cases} \theta^{-n} & \text{if } \theta \ge x_{(n)} \\ 0 & \text{o/w} \end{cases}$.

$\hat\theta = x_{(n)}$ is the MLE.

$\forall t > 0$,

$$\mathbb{P}(\hat\theta - \theta \le -t) = \mathbb{P}\left(\bigcap_{i=1}^{n} [X_i \le \theta - t]\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \le \theta - t) = \left(1 - \frac{t}{\theta}\right)^n.$$

Hence,

$$\mathbb{P}(|\hat\theta - \theta| < t) = 1 - \mathbb{P}(\hat\theta - \theta \le -t) - \mathbb{P}(\hat\theta - \theta \ge t) = 1 - \left(1 - \frac{t}{\theta}\right)^n \to 1 \text{ as } n \to \infty,$$

since $\mathbb{P}(\hat\theta - \theta \ge t) = 0$.

$\Rightarrow \hat\theta \xrightarrow{P} \theta$

Moreover, $\forall t > 0$,

$$\mathbb{P}(n(\hat\theta - \theta) \le -t) = \left(1 - \frac{t}{n\theta}\right)^n \to e^{-\frac{t}{\theta}} \text{ as } n \to \infty.$$

That is, $-n(\hat\theta - \theta) \xrightarrow{d} \text{Exp}(\theta^{-1})$. Not asymptotically normal.

Problem 4. Step 1: Show $\hat{\theta}_n \xrightarrow{P} \theta_0$ using Lemma 5.10. (Read Example 5.11 & 5.2X)

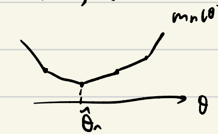
Assume $Y = \theta_0 X + \varepsilon$ with $E(\varepsilon)=0$, $\varepsilon \perp\!\!\!\perp X$, $\varepsilon \sim P_\varepsilon$ has a pdf f_ε satisfying $f_\varepsilon(0)>0$, $E(X^2)<\infty$.

Let $m_n(\theta) := n^{-1}\sum_{i=1}^{n} |Y_i - \theta X_i| = n^{-1}\sum_{i\in I} |Y_i - \theta X_i| + n^{-1}\sum_{i\in I^c} |Y_i|$, where $I := \{i : X_i \neq 0\}$.

Let $\psi_n(\theta) := -n^{-1}\sum_{i=1}^{n} \operatorname{sgn}(Y_i - \theta X_i) X_i = -n^{-1}\sum_{i\in I} \operatorname{sgn}(Y_i-\theta X_i) X_i$, which is nondecreasing w.r.t. θ.

$m_n'(\theta) = \psi_n(\theta)$ for every θ s.t. $Y_i - \theta X_i \neq 0\ \forall i \le n$. $m_n(\theta)$ is continuous.

$m_n(\theta)$ is a piecewise linear function with knots $\left(\frac{Y_i}{X_i}\right)_{i\in I}$. See figure



Showing $\psi_n(\hat{\theta}_n) = o_p(1)$

The increment of slope at any knot is bounded by $2n^{-1}\sup_i |X_i|$.

Hence, $\forall\, \hat{\theta}_n \in \arg\min_\theta m_n(\theta)$,

$$|\psi_n(\hat{\theta}_n)| \le 2n^{-1}\sup_i |X_i|. \quad (*)$$

Note that, $\forall t>0$.

$$\mathbb{P}(n^{-1}\sup_i |X_i| > t) = \mathbb{P}\left(\bigcup_{i=1}^{n} [|X_i| > nt]\right) \le n\,\mathbb{P}(|X_i| > nt) = n\,\mathbb{P}(X_i^2 > n^2t^2)$$

$$\le n \cdot \frac{E(X^2)}{n^2t^2} \quad (\text{Markov's inequality}) = \frac{E(X^2)}{nt} \to 0.$$

That is, $n^{-1}\sup_i |X_i| = o_p(1)$ and together with $(*)$, $\psi_n(\hat{\theta}_n) = o_p(1)$.

By WLLN, we also have $\psi_n(\theta) \to \psi(\theta) := E\operatorname{sgn}(Y-\theta X)X$, since $\operatorname{Var}\{\operatorname{sgn}(Y-\theta X)X\} \le E\{\operatorname{sgn}(Y-\theta X)X\}^2 \le E(X^2) < \infty$.

Additionally,

$$\psi(\theta) = -E[\operatorname{sgn}\{\varepsilon - (\theta-\theta_0)X\}\cdot X] = -E(E[\operatorname{sgn}\{\varepsilon-(\theta-\theta_0)X\} \mid X]\cdot X)$$

$$= -E[X\{\mathbb{P}(\varepsilon > (\theta-\theta_0)X \mid X) - \mathbb{P}(\varepsilon < (\theta-\theta_0)X \mid X)\}] = -E[X\{1 - 2F_\varepsilon((\theta-\theta_0)X)\}] \quad \text{and}$$

$$\dot{\psi}(\theta) = -\frac{\partial}{\partial\theta} E[X\{1-2F_\varepsilon((\theta-\theta_0)X)\}] = 2E\{X^2 f_\varepsilon((\theta-\theta_0)X)\} \ge 0, \text{ with } \dot{\psi}(\theta_0) = 2f_\varepsilon(0)\,E(X^2) > 0.$$

Hence, $\psi(\theta_0 - \varepsilon) < 0 < \psi(\theta_0+\varepsilon)$, $\forall \varepsilon > 0$. By Lemma 5.10, $\hat{\theta} \xrightarrow{P} \theta_0$.

Step 2: Show asymptotic normality using Theorem 5.23

Let $m_\theta(x,y) = -|y-\theta x|$. Then, $\frac{\partial}{\partial\theta} m_\theta(x,y) = -\psi_\theta(x,y) := \mathrm{sgn}(y-\theta x)\cdot x$, $\forall\,\theta$ s.t. $y-\theta x \neq 0$.

$P(Y-\theta_0 X = 0) = P(\varepsilon = (\theta_0-\theta)X) = \int_{\mathbb{R}} P(\varepsilon = (\theta_0-\theta)X \mid X=t)\, dP_X(t) = \int_{\mathbb{R}} P(\varepsilon = (\theta_0-\theta)t)\, dP_X(t)$ (since $\varepsilon \perp\!\!\!\perp X$)

(tower rule)

$= \int_{\mathbb{R}} 0 \, dP_X(t)$ (since ε is a continuous random variable) $= 0$

Hence, $\frac{\partial}{\partial\theta} m_\theta(x,y)$ exists at θ_0 w.p. 1.

$\forall\, \theta_1, \theta_2 \in \mathbb{R}$, by triangular inequality,

$|m_{\theta_1}(x,y) - m_{\theta_2}(x,y)| = \big||y-\theta_2 x| - |y-\theta_1 x|\big| \le |x|\cdot|\theta_1-\theta_2| = \dot m(x)\cdot|\theta_1-\theta_2|$, where $\dot m(x) := |x|$ and $P\dot m^2 = E(X^2) < \infty$.

Moreover, note that

$Pm_\theta = -E|Y-\theta X| = -E|\varepsilon-(\theta-\theta_0)X| = -E[\{\varepsilon-(\theta-\theta_0)X\}\cdot 1\{\varepsilon > (\theta-\theta_0)X\}] + E[\{\varepsilon-(\theta-\theta_0)X\}\cdot 1\{\varepsilon \le (\theta-\theta_0)X\}]$

Here, $E[\{\varepsilon-(\theta-\theta_0)X\}\cdot 1\{\varepsilon>(\theta-\theta_0)X\} \mid X=x] = \int_{(\theta-\theta_0)x}^{\infty} \{t-(\theta-\theta_0)x\} f_\varepsilon(t)\,dt = \int_{(\theta-\theta_0)x}^{\infty} t f_\varepsilon(t)\,dt - (\theta-\theta_0)x\, P(\varepsilon > (\theta-\theta_0)x)$

By Leibniz integral rule, for any $b>0$

$\frac{\partial}{\partial\theta}\int_{(\theta-\theta_0)x}^{\infty} t f_\varepsilon(t)\,dt = \frac{\partial}{\partial\theta}\int_{(\theta-\theta_0)x}^{b} t f_\varepsilon(t)\,dt = -(\theta-\theta_0)x^2 f_\varepsilon((\theta-\theta_0)x)$.

Additionally, $\frac{\partial}{\partial\theta}(\theta-\theta_0)x\, P(\varepsilon>(\theta-\theta_0)x) = xP(\varepsilon>(\theta-\theta_0)x) - (\theta-\theta_0)x^2 f_\varepsilon((\theta-\theta_0)x)$.

Hence, $\frac{\partial}{\partial\theta} E[\{\varepsilon-(\theta-\theta_0)X\}\cdot 1\{\varepsilon>(\theta-\theta_0)X\}\mid X=x] = -(\theta-\theta_0)x^2 f_\varepsilon((\theta-\theta_0)x) - \{xP(\varepsilon>(\theta-\theta_0)x) - (\theta-\theta_0)x^2 f_\varepsilon((\theta-\theta_0)x)\} = -xP(\varepsilon>(\theta-\theta_0)x)$

Similarly, $\frac{\partial}{\partial\theta} E[\{\varepsilon-(\theta-\theta_0)X\}\cdot 1\{\varepsilon\le(\theta-\theta_0)X\}\mid X=x] = xP(\varepsilon<(\theta-\theta_0)x) = x\{1-P(\varepsilon>(\theta-\theta_0)x)\}$

Therefore, $\frac{\partial}{\partial\theta}Pm_\theta = \frac{\partial}{\partial\theta}E\{E[m_\theta(x)\mid X]\} = E\frac{\partial}{\partial\theta}\{E[m_\theta(x)\mid X]\} = E[X\{1-2F_\varepsilon((\theta-\theta_0)X)\}] = -\psi(\theta)$.

and hence $\frac{\partial^2}{\partial\theta^2}Pm_\theta = -\dot\psi(\theta) = -2E\{X^2 f_\varepsilon((\theta-\theta_0)X)\}$ with $V_{\theta_0} = \frac{\partial^2}{\partial\theta^2}Pm|_{\theta=\theta_0} = -2E\{X^2 f_\varepsilon(0)\} = -2E(X^2)f_\varepsilon(0)$

Since $P_n m_{\hat\theta_n} = \sup_\theta P_n m_\theta$ and, as shown in Step 1, $\hat\theta_n \xrightarrow{P} \theta_0$, by Theorem 5.23,

$\sqrt{n}(\hat\theta_n-\theta_0) \xrightarrow{d} N(0, \{4f_\varepsilon^2(0)E(X^2)\}^{-1})$ by the fact that $\{\dot\psi(\theta_0)\}^{-2}\cdot P\dot m_{\theta_0}^2 = \{2f_\varepsilon(0)E(X^2)\}^{-2}\cdot E(X^2) = \{4f_\varepsilon^2(0)E(X^2)\}^{-1}$.

Problem 5. From Problem 3 of HW 1,

$$\hat{\lambda} = \{n^{-1}\sum_i X_i - X_{(1)}\}^{-1} \xrightarrow{P} \lambda \quad \text{and} \quad \hat{a} = X_{(1)} \xrightarrow{P} a.$$

For $\hat{a}$, recall that $\forall t > 0$, $\mathbb{P}(\hat{a} \le a + t) = 1 - e^{-n\lambda t}$.

Hence, $\mathbb{P}(n(\hat{a} - a) \le t) = 1 - e^{-\lambda t}$.

That is, $n(\hat{a} - a) \sim \mathrm{Exp}(\lambda)$.

We also have $\sqrt{n}(\hat{a} - a) = o_p(1)$ since $\mathbb{P}(\sqrt{n}(\hat{a} - a) > t) = e^{-\sqrt{n}\lambda t} \to 0 \quad \forall t > 0$.

By CLT, $\sqrt{n}\{n^{-1}\sum_i (X_i - a) - \lambda^{-1}\} \xrightarrow{d} N(0, \lambda^{-2})$.

Note that $\sqrt{n}(\hat{\lambda}^{-1} - \lambda^{-1}) = \sqrt{n}\{n^{-1}\sum_i (X_i - a) - \lambda^{-1}\} - \sqrt{n}(\hat{a} - a)$.

By Slutsky's theorem, $\sqrt{n}(\hat{\lambda}^{-1} - \lambda^{-1}) \xrightarrow{d} N(0, \lambda^{-2})$.

Let $g(\lambda) = \lambda^{-1}$. Then $g'(\lambda) = -\lambda^{-2}$, $\lambda^{-2} \cdot \{g'(\lambda^{-1})\}^2 = \lambda^2$.

By Delta's method,

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, \lambda^2).$$

Problem 6. $f_{\theta_0}(x) = \pi^{-1}\{1+(x-\theta_0)^2\}^{-1}$.

CF: $\mathbb{E}(e^{itX}) = e^{\theta_0 it - |t|}$, $\forall t \in \mathbb{R}$.

The CF of $\bar{X}$:

$$\mathbb{E}(e^{it\bar{X}}) = \mathbb{E}\prod_{i=1}^{n} e^{itX_i/n} = (\mathbb{E}\, e^{itX/n})^n$$

$$= \{e^{(\theta_0 it - |t|)/n}\}^n = e^{\theta_0 it - |t|} = \mathbb{E}(e^{itX})$$

That is, $\bar{X} \sim \text{Cauchy}(\theta_0)$.

Now, consider the MLE $\hat{\theta} \in \arg\max_\theta f_{\theta_0}(x_1, \dots, x_n)$.

$$\psi_\theta(x) = \frac{2(x-\theta)}{1+(x-\theta)^2}, \quad \dot{\psi}_\theta(x) = \frac{2(x-\theta)^2 - 2}{\{1+(x-\theta)^2\}^2}, \quad \frac{\partial^2}{\partial\theta^2}\psi_\theta(x) = \frac{4(x-\theta)\{(x-\theta)^2-3\}}{\{1+(x-\theta)^2\}^3}.$$

By Example 5.16 (also in the lecture notes), $\hat{\theta} \xrightarrow{P} \theta_0$.

Since $f_{\theta_0}(X)$ is symmetric w.r.t. θ_0, $\mathbb{P}\psi_{\theta_0} = 0$.

$$\mathbb{P}\psi_{\theta_0}^2 = \int_{\mathbb{R}} \frac{4(x-\theta_0)^2}{\{1+(x-\theta_0)^2\}^3}\cdot\pi^{-1}\,dx = \frac{4}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\tan^2 u}{(1+\tan^2 u)^2}\,du \quad (x-\theta_0 = \tan u).$$

$$= \frac{4}{\pi}\int_{-\pi/2}^{\pi/2} \sin^2 u \cos^2 u\, du \quad \left(\text{since } \frac{1}{1+\tan^2 u} = \cos^2 u\right)$$

$$= \pi^{-1}\int_{-\pi/2}^{\pi/2} \sin^2(2u)\,du = (2\pi)^{-1}\int_{-\pi/2}^{\pi/2}\{1-\cos(4u)\}\,du = (2\pi)^{-1}\cdot\pi = \frac{1}{2} < \infty.$$

$$\mathbb{P}\dot{\psi}_{\theta_0} = \frac{2}{\pi}\int_{\mathbb{R}} \frac{(x-\theta_0)^2-1}{\{1+(x-\theta_0)^2\}^3}\,dx = \frac{1}{2}\cdot\frac{1}{2} - \frac{2}{\pi}\int_{\mathbb{R}}(1+x^2)^{-3}\,dx = \frac{1}{4} - \frac{2}{\pi}\left\{\frac{3}{8}\arctan x + \frac{3x^3+5x}{8(x^2+1)^2}\right\}\Big|_{-\infty}^{\infty} = -\frac{1}{2} \neq 0.$$

$$\left|\frac{\partial^2}{\partial\theta^2}\psi_\theta(x)\right| \le \left(4|x-\theta|^{-3} + 3|x-\theta|^{-6}\right) \wedge \left(4|x-\theta|^3 + 3\right)$$

$$\le \{4(|x-\theta_0|-1)^{-3} + 3(|x-\theta_0|-1)^{-6}\} \wedge \{4(|x-\theta_0|+1)^3+3\} := \ddot{\psi}(x)$$ and one can check that $\mathbb{P}\ddot{\psi} < \infty$.

a function whose tail as $x \to \pm\infty$ is of the order $|x|^{-3}$, and bounded on $|x| < M$ for any $M > 0$. Serves as an upper bound for $\frac{\partial^2}{\partial\theta^2}\psi(x)$.

By Theorem 5.41, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, 2)$,

since $(\mathbb{P}\dot{\psi}_{\theta_0})^{-2}\cdot\mathbb{P}\psi_{\theta_0}^2 = 2$.

Lastly, consider the one-step estimator.

Let $\tilde{\theta} = \text{median}(X_1, \cdots, X_n)$ be the initial estimator.

Then, by Example 5.24, or, by Problem 4 with $\mathbb{P}(X=1)=1$,

we have $\tilde{\theta} - \theta_0 = O_p(n^{-\frac{1}{2}})$.

Let $\hat{\theta}^{(1)} = \tilde{\theta} - \{\dot{\psi}_n(\tilde{\theta})\}^{-1} \cdot \psi_n(\tilde{\theta})$, where for any $\theta \in \mathbb{R}$,

$\psi_n(\theta) = n^{-1}\sum_i \psi_\theta(X_i)$, $\dot{\psi}_n(\theta) = n^{-1}\sum_i \dot{\psi}_\theta(X_i)$, $\ddot{\psi}_n = n^{-1}\sum_i \ddot{\psi}(X_i)$ and $\psi_n^{(2)}(\theta) = n^{-1}\sum_i \frac{\partial^2}{\partial\theta^2}\psi_\theta(X_i)$.

the upper bound of $\frac{\partial^2}{\partial\theta^2}\psi_\theta$

By Taylor expansion, for any $\theta \in B(\theta_0, \frac{M}{\sqrt{n}})$, with some θ' between θ and θ_0,

$\left|\sqrt{n}\{\psi_n(\theta) - \psi_n(\theta_0)\} - \sqrt{n}\,\dot{\psi}_n(\theta)(\theta-\theta_0)\right| = \left|\sqrt{n} \cdot \psi_n^{(2)}(\theta')(\theta-\theta_0)^2\right| \le \sqrt{n}\,(\theta-\theta_0)^2 \, |\ddot{\psi}_n|$ since $\left|\frac{\partial^2}{\partial\theta^2}\psi_\theta(x)\right| \le \ddot{\psi}(x)$ when $n \ge M^2$.

Since $\mathbb{P}|\ddot{\psi}| < \infty$, by WLLN, $|\ddot{\psi}_n| = \mathbb{P}|\ddot{\psi}| + o_p(1) = O_p(1)$.

Besides, $\left|\sqrt{n}\{\dot{\psi}_n(\theta) - \mathbb{P}\dot{\psi}_{\theta_0}\}(\theta-\theta_0)\right| \le \sqrt{n}\,|\theta-\theta_0| \cdot |\dot{\psi}_n(\theta) - \dot{\psi}_n(\theta_0)| + \sqrt{n}\,|\theta-\theta_0| \cdot |\dot{\psi}_n(\theta_0) - \mathbb{P}\dot{\psi}_{\theta_0}|$

$\le \sqrt{n}\,(\theta-\theta_0)^2 \cdot \underbrace{|\ddot{\psi}_n|}_{\text{Taylor}} + \sqrt{n}\,|\theta-\theta_0| \cdot \underbrace{o_p(1)}_{\text{WLLN}}$

Check condition (5.44)

Hence, $\sup_{\theta \in B(\theta_0, \frac{M}{\sqrt{n}})} \left|\sqrt{n}\{\psi_n(\theta) - \psi_n(\theta_0)\} - \sqrt{n}\,\underline{\mathbb{P}\dot{\psi}_{\theta_0}} \cdot (\theta-\theta_0)\right| \le 2\sqrt{n}\,(\theta-\theta_0)^2 |\ddot{\psi}_n| + \sqrt{n}\,|\theta-\theta_0| \cdot o_p(1) = o_p(1)$.

corresponds to $\dot{\psi}_0$ in (5.44). should be a non-random matrix / scaler.

Furthermore,

$|\dot{\psi}_n(\tilde{\theta}) - \mathbb{P}\dot{\psi}_{\theta_0}| \le |\dot{\psi}_n(\tilde{\theta}) - \dot{\psi}_n(\theta_0)| + |\dot{\psi}_n(\theta_0) - \mathbb{P}\dot{\psi}_{\theta_0}| \le |\tilde{\theta} - \theta_0| \cdot \underbrace{|\ddot{\psi}_n|}_{\text{Taylor}} + \underbrace{o_p(1)}_{\text{WLLN}} = o_p(1)$.

By Theorem 5.45,

$\sqrt{n}(\hat{\theta}^{(1)} - \theta_0) \to N(0, 2)$,

since $(\mathbb{P}\dot{\psi}_{\theta_0})^{-2}\,\mathbb{P}\psi_{\theta_0}^2 = 2$.

You can also use Thm. 5.46 (Addendum) here. To do so, you need to show the Lipschitz property of $\psi_\theta(x)$.