HW2: Asymptotics 2

Gabriel ROMON

Problem 1

Let $a_1, \ldots, a_q$ and $b_1, \ldots, b_q$ be vectors in $\mathbb{R}^d$ $(d, q \geq 1)$. Suppose that $\sum_{i=1}^q b_i b_i^T$ is invertible. We want to prove that

$$\sum_{i=1}^q a_i b_i^T \left(\sum_{i=1}^q b_i b_i^T \right)^{-1} \sum_{i=1}^q b_i a_i^T \preceq \sum_{i=1}^q a_i a_i^T$$

1. Show that $q \geq d$.
2. Let $C \in \mathbb{R}^{q \times q}$ defined by $C_{ij} = b_i^T \left(\sum_{i=1}^q b_i b_i^T \right)^{-1} b_j$. Show that C is a projection matrix.
3. Let $x \in \mathbb{R}^d$. Show that $x^T \left[\sum_{i=1}^q a_i b_i^T \left(\sum_{i=1}^q b_i b_i^T \right)^{-1} \sum_{i=1}^q b_i a_i^T \right] x$ rewrites as $y^T C y$ for some $y \in \mathbb{R}^q$.
4. Conclude.

1. Since $\sum_{i=1}^q b_i b_i^T \in \mathbb{R}^{d \times d}$ is invertible, $d = \operatorname{rk}\left(\sum_{i=1}^q b_i b_i^T \right) \leq \sum_{i=1}^q \operatorname{rk}(b_i b_i^T) \leq \sum_{i=1}^q 1 = q$.

2. Let us show that $C^2 = C$. Let $B = \sum_{i=1}^q b_i b_i^T$.

$$\begin{aligned}
(C^2)_{ij} &= \sum_{k=1}^q C_{ik} C_{kj} = \sum_{k=1}^q b_i^T B^{-1} b_k b_k^T B^{-1} b_j \\
&= b_i^T B^{-1} \left(\sum_{k=1}^q b_k b_k^T \right) B^{-1} b_j \\
&= b_i^T B^{-1} B B^{-1} b_j \\
&= C_{ij}
\end{aligned}$$

   Hence C is a projection matrix. Besides, since C is symmetric, the projection is orthogonal.

3. Note that

$$\begin{aligned}
x^T \left[\sum_{i=1}^q a_i b_i^T B^{-1} \sum_{k=1}^q b_k a_k^T \right] x &= \sum_{i=1}^q \sum_{k=1}^q x^T a_i C_{ik} a_k^T x \\
&= \sum_{i=1}^q \sum_{k=1}^q y_i C_{ik} y_k \\
&= y^T C y \quad \text{where } y = \begin{pmatrix} a_1^T x \\ \vdots \\ a_q^T x \end{pmatrix}
\end{aligned}$$

4. Note that $x^T \sum_{i=1}^q a_i a_i^T x = y^T I_d y$, hence

$$x^T \sum_{i=1}^q a_i a_i^T x - x^T \left[\sum_{i=1}^q a_i b_i^T B^{-1} \sum_{k=1}^q b_k a_k^T \right] x = y^T (I_d - C) y$$

   Since C satisfies $C^2 = C$ its eigenvalues are either 0 or 1. The eigenvalues of the symmetric matrix $I_d - C$ are thus also 0 or 1, hence $I_d - C \succeq 0$, so $y^T (I_d - C) y \geq 0$. This holds for all x, hence $\sum_{i=1}^q a_i a_i^T \succeq \sum_{i=1}^q a_i b_i^T \left(\sum_{i=1}^q b_i b_i^T \right)^{-1} \sum_{i=1}^q b_i a_i^T$.

Problem 2

Let a, b be random vectors in $\mathbb{R}^d$ such that $E(\|a\|^2 + \|b\|^2) < \infty$.

1. Show that $E(aa^T)$, $E(bb^T)$ and $E(ab^T)$ are well-defined. Show that the transpose of $E(ab^T)$ is $E(ba^T)$.
   In the rest of the problem we assume that $E(bb^T)$ is invertible and we want to show that
   $$E(ab^T)E(bb^T)^{-1}E(ba^T) \preceq E(aa^T)$$
2. Show that the inequality of Problem 1 is a special case of this inequality.
3. Let $M \in \mathbb{R}^{p\times p}$ a block matrix defined by $M = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ where $A \in \mathbb{R}^{k\times k}$, $B \in \mathbb{R}^{k\times l}$, $C \in \mathbb{R}^{l\times l}$ with $k + l = p$ and A and C symmetric. We assume that C is invertible. The Schur complement of C in M is defined as $A - BC^{-1}B^T$. Show that $M \succeq 0$ if and only if C and its Schur complement are $\succeq 0$.
4. Let $M \in \mathbb{R}^{2d\times 2d}$ the block matrix defined by $M = \begin{pmatrix} E(aa^T) & E(ab^T) \\ E(ba^T) & E(bb^T) \end{pmatrix}$
   (a) Show that $M \succeq 0$.
   (b) Conclude.

1. Since $E(a_i^2) \leq E(\sum_{i=1}^d a_i^2) = E(\|a\|_2^2) < \infty$, the coordinates of a are in $L^2(\mathbb{R})$. Thus for each $1 \leq i, j \leq d$, $E(|a_i a_j|) < \infty$. Therefore
   $$E(\|aa^T\|_1) \leq E\left(\sum_{i=1}^d \sum_{j=1}^d |a_i a_j|\right) = \sum_{i=1}^d \sum_{j=1}^d E(|a_i a_j|) < \infty$$
   Similarly one proves that $E(\|bb^T\|_1)$, $E(\|ab^T\|_1)$ and $E(\|ba^T\|_1)$ are finite.
   Note that $(E(ab^T))_{ji} = E(a_j b_i) = E(b_i a_j) = (E(ba^T))_{ij}$, hence $E(ab^T)$ is the transpose of $E(ba^T)$.
2. To avoid confusions, let $c_1, \ldots, c_q$ and $d_1, \ldots, d_q$ denote the vectors in the inequality of Problem 1. Let $\phi : \mathbb{R}^d \to \mathbb{R}^d, c_i \mapsto d_i$. Let a be a random vector with distribution $\frac{1}{q}\sum_{i=1}^q \delta_{c_i}$ and $b = \phi(a)$. a and b are bounded, and $E(ab^T) = E\left(\sum_{i=1}^q c_i d_i^T 1_{a=c_i}\right) = \frac{1}{q}\sum_{i=1}^q c_i d_i^T$. Hence $E(ab^T)E(bb^T)^{-1}E(ba^T) \preceq E(aa^T)$ rewrites as
   $$\sum_{i=1}^q c_i d_i^T \left(\sum_{i=1}^q d_i d_i^T\right)^{-1} \sum_{i=1}^q d_i c_i^T \preceq \sum_{i=1}^q c_i c_i^T$$
3. Let $f : \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}, (u, v) \mapsto \begin{pmatrix} u^T & v^T \end{pmatrix} \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = v^T C v + 2v^T B^T u + u^T A u$
   If $M \succeq 0$, then for all $v \in \mathbb{R}^l$, $f(0, v) \geq 0$ i.e. $v^T C v \geq 0$, thus $C \succeq 0$. Hence $v \mapsto v^T C v$ is convex, thus f is convex as a function of v.
   Let u be fixed. Since $\nabla_v f(u, v) = 2Cv + 2B^T u$, $v = -C^{-1}B^T u$ is a critical point, hence $f(u, \cdot)$ is minimized at $-C^{-1}B^T u$ with $0 \leq f(u, -C^{-1}B^T u) = u^T(A - BC^{-1}B^T)u$. This holds for all $u \in \mathbb{R}^k$, so the Schur complement of C is $\succeq 0$.

   Suppose that $C \succeq 0$ and $A - BC^{-1}B^T \succeq 0$. For any $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^l$, the previous reasoning shows that $f(u, v) \geq f(u, -C^{-1}B^T u) = u^T(A - BC^{-1}B^T)u \geq 0$. Hence $M \succeq 0$.

4. (a) We have

$$\begin{aligned}
\begin{pmatrix} u^T & v^T \end{pmatrix} M \begin{pmatrix} u \\ v \end{pmatrix} &= u^T E(aa^T)u + 2u^T E(ab^T)v + v^T E(bb^T)v \\
&= E\left((u^T a)^2 + 2(u^T a)(b^T v) + (b^T v)^2\right) \\
&= E\left((u^T a + b^T v)^2\right) \\
&\geq 0
\end{aligned}$$

(b) Remember that $E(bb^T)$ is invertible by assumption. By Question 3, the Schur complement of $E(bb^T)$ in M is $\succeq 0$, that is

$$E(aa^T) - E(ab^T)E(bb^T)^{-1}E(ba^T) \succeq 0$$

Problem 3

Let $(X_n)_{n\geq 1}$ a sequence of i.i.d. random variables having distribution $\mathcal{N}(\mu, 1)$ where $\mu \in \mathbb{R}$. Let $\alpha \in (0,1)$. For $n \geq 1$, let $\hat{\mu}_n = \begin{cases} \bar{X}_n & \text{if} |\bar{X}_n| > n^{-1/4} \\ \alpha \bar{X}_n & \text{otherwise} \end{cases}$

We want to show that $\hat{\mu}_n$ is asymptotically normal around μ, with asymptotic variance strictly smaller than to the reciprocal of Fisher's information, for some values of μ.

1. Compute Fisher's information $I(\mu)$.
2. Suppose that $\mu = 0$.
   (a) Show that $P_0(|\bar{X}_n| > n^{-1/4}) \xrightarrow[n\to\infty]{} 0$
   (b) Deduce that $\forall t \in \mathbb{R}$, $P_0(\sqrt{n}\hat{\mu}_n \leq t) - P_0(\sqrt{n}\bar{X}_n \leq \frac{t}{\alpha}) \xrightarrow[n\to\infty]{} 0$.
   (c) Conclude about $\hat{\mu}_n$ and compute its asymptotic variance.
3. Suppose that $\mu > 0$ (the case $\mu < 0$ is similar).
   (a) Show that for n sufficiently large, $P_\mu(|\bar{X}_n| \leq n^{-1/4}) \leq \frac{1}{n(\mu - n^{-1/4})^2}$.
   (b) Deduce that $\forall t \in \mathbb{R}$, $P_\mu(\sqrt{n}(\hat{\mu}_n - \mu) \leq t) - P_\mu(\sqrt{n}(\bar{X}_n - \mu) \leq \frac{t}{\alpha}) \xrightarrow[n\to\infty]{} 0$.
   (c) Conclude about $\hat{\mu}_n$ and compute its asymptotic variance.
4. Conclude about Fisher's programme validity.

1. The likelihood for a single sample is $L(\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu)^2\right)$, thus $(\log L)'' = -1$ and $I(\mu) = 1$.

2. (a) The CLT yields $n^{1/4} n^{1/4} \bar{X}_n \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)$, hence $n^{1/4}\bar{X}_n \xrightarrow[n\to\infty]{P} 0$, hence

$$P_0(n^{1/4}|\bar{X}_n| > 1) \xrightarrow[n\to\infty]{} 0$$

(b) Let $t \in \mathbb{R}$. Note that

$$\begin{aligned}
P_0(\sqrt{n}\hat{\mu}_n \leq t) &= P_0(\sqrt{n}\bar{X}_n \leq t \cap |\bar{X}_n| > n^{-1/4}) + P_0(\sqrt{n}\alpha\bar{X}_n \leq t \cap |\bar{X}_n| \leq n^{-1/4}) \\
&= P_0(\sqrt{n}\bar{X}_n \leq t \cap |\bar{X}_n| > n^{-1/4}) + P_0(\sqrt{n}\alpha\bar{X}_n \leq t) - P(\sqrt{n}\alpha\bar{X}_n \leq t \cap |\bar{X}_n| > n^{-1/4})
\end{aligned}$$

Hence $|P_0(\sqrt{n}\hat{\mu}_n \leq t) - P_0(\sqrt{n}\alpha\bar{X}_n \leq t)| \leq 2P_0(|\bar{X}_n| > n^{-1/4}) \xrightarrow[n\to\infty]{} 0$

(c) The CLT combined with the continuous mapping theorem shows that $\sqrt{n}\alpha\bar{X}_n$ converges in distribution to $\mathcal{N}(0,\alpha^2)$, hence so does $\sqrt{n}\hat{\mu}_n$.
$\hat{\mu}_n$ is thus asymptotically normal with asymptotic variance α^2.

3. (a) Let $n \geq \frac{1}{\mu^4}$. Note that

$$\begin{aligned}
P_\mu(|\bar{X}_n| \leq n^{-1/4}) &\leq P_\mu(\mu - \bar{X}_n \geq \mu - n^{-1/4}) \\
&= P_\mu(\bar{X}_n - \mu \geq \mu - n^{-1/4}) \qquad \bar{X}_n - \mu \text{ is symmetric} \\
&\leq \frac{V_\mu(\hat{X}_n)}{\left(\mu - n^{-1/4}\right)^2} \\
&= \frac{1}{n}\frac{1}{\left(\mu - n^{-1/4}\right)^2}
\end{aligned}$$

(b) Let $t \in \mathbb{R}$. Note that

$$\begin{aligned}
P_\mu(\sqrt{n}(\hat{\mu}_n - \mu) \leq t) &= P_\mu(\sqrt{n}(\bar{X}_n - \mu) \leq t \cap |\bar{X}_n| > n^{-1/4}) + P_\mu(\sqrt{n}(\alpha\bar{X}_n - \mu) \leq t \cap |\bar{X}_n| \leq n^{-1/4}) \\
&= P_\mu(\sqrt{n}(\bar{X}_n - \mu) \leq t) - P_\mu(\sqrt{n}(\bar{X}_n - \mu) \leq t \cap |\bar{X}_n| \leq n^{-1/4}) + o(1) \\
&= P_\mu(\sqrt{n}(\bar{X}_n - \mu) \leq t) + o(1)
\end{aligned}$$

(c) By the CLT, $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution to $\mathcal{N}(0,1)$, hence so does $\sqrt{n}(\hat{\mu}_n - \mu)$. Therefore $\hat{\mu}_n$ is asymptotically normal with asymptotic variance 1.

4. Fisher's programme is not valid. There exists superefficient estimators: for some value of the parameter, the asymptotic variance is stricly smaller than the reciprocal of the information.

Problem 4

Let $(X_n)_{n\geq 1}$ a sequence of i.i.d. random variables and let F denote their cdf. Suppose that the median is unique and note it as m.
Suppose that $F(x) - \frac{1}{2} \sim L_2(x-m)^\alpha$ as $x \underset{>}{\longrightarrow} m$ and $\frac{1}{2} - F(x) \sim L_1(m-x)^\alpha$ as $x \underset{<}{\longrightarrow} m$ where $\alpha \in (0,1]$ and $L_1, L_2 > 0$. Let $\hat{m}_n = X_{(\lceil \frac{n}{2} \rceil)}$.
Show that $n^{\frac{1}{2\alpha}}(\hat{m}_n - m)$ converges in distribution. Interpret this result in terms of the rate of convergence of the empirical median to the median.

Let $t \in \mathbb{R}$ and let us find the limit of $P(n^{\frac{1}{2\alpha}}(\hat{m}_n - m) \leq t)$ as $n \to \infty$.
For $1 \leq i \leq n$, let $Y_i = 1_{X_i \leq m + \frac{t}{n^{\frac{1}{2\alpha}}}}$. Note that $\sum_{i=1}^n Y_i$ follows $\mathcal{B}(n, F(m + \frac{t}{n^{\frac{1}{2\alpha}}}))$ and

$$P(n^{\frac{1}{2\alpha}}(\hat{m}_n - m) \leq t) = P(\hat{m}_n \leq m + \frac{t}{n^{\frac{1}{2\alpha}}}) = P\left(\sum_{i=1}^n Y_i \geq \lceil \frac{n}{2} \rceil\right) = P\left(\sum_{i=1}^n Y_i \geq \frac{n}{2}\right)$$

Let $N = \sum_{i=1}^n Y_i$. Computations analoguous to those carried out in Question 3.c) of Problem 3 in HW1 show that $\frac{1}{\sqrt{n}}\left(N - nF\left(m + \frac{t}{n^{\frac{1}{2\alpha}}}\right)\right)$ converges in distribution to $\mathcal{N}(0, \frac{1}{4})$, regardless of L_1 and L_2.
Let $Z_n = \frac{1}{\sqrt{n}}\left(N - nF\left(m + \frac{t}{n^{\frac{1}{2\alpha}}}\right)\right)$ and note that

$$P\left(\sum_{i=1}^n Y_i \geq \frac{n}{2}\right) = P\left[-Z_n \leq \sqrt{n}\left(F\left(m + \frac{t}{n^{\frac{1}{2\alpha}}}\right) - \frac{1}{2}\right)\right]$$

When $t \geq 0$, the deterministic sequence $\sqrt{n}\left(F\left(m+\frac{t}{n^{\frac{1}{2\alpha}}}\right)-\frac{1}{2}\right)$ converges to $L_2 t^\alpha$ and

$$P\left[-Z_n \leq \sqrt{n}\left(F\left(m+\frac{t}{n^{\frac{1}{2\alpha}}}\right)-\frac{1}{2}\right)\right] = P\Bigg[\underbrace{-2Z_n\frac{L_2 t^\alpha}{\sqrt{n}\left(F\left(m+\frac{t}{n^{\frac{1}{2\alpha}}}\right)-\frac{1}{2}\right)}}_{\xrightarrow[n\to\infty]{(d)}\mathcal{N}(0,1) \text{ by Slutsky}} \leq 2L_2 t^\alpha\Bigg]$$

hence $P(n^{\frac{1}{2\alpha}}(\hat{m}_n - m) \leq t) \xrightarrow[n\to\infty]{} \Phi(2L_2 t^\alpha)$ where Φ is the cdf of $\mathcal{N}(0,1)$.

When $t \leq 0$ the deterministic sequence converges to $-L_1(-t)^\alpha$ and

$$P\left[-Z_n \leq \sqrt{n}\left(F\left(m+\frac{t}{n^{\frac{1}{2\alpha}}}\right)-\frac{1}{2}\right)\right] = P\Bigg[\underbrace{-2Z_n\frac{L_1(-t)^\alpha}{\sqrt{n}\left(\frac{1}{2}-F\left(m+\frac{t}{n^{\frac{1}{2\alpha}}}\right)\right)}}_{\xrightarrow[n\to\infty]{(d)}\mathcal{N}(0,1) \text{ by Slutsky}} \leq -L_1(-t)^\alpha\Bigg]$$

hence $P(n^{\frac{1}{2\alpha}}(\hat{m}_n - m) \leq t) \xrightarrow[n\to\infty]{} \Phi(-2L_1(-t)^\alpha)$

The piecewise function $\varphi : t \mapsto \begin{cases} \Phi(2L_2 t^\alpha) & \text{if } t \geq 0 \\ \Phi(-2L_1(-t)^\alpha) & \text{if } t \leq 0 \end{cases}$ is continuous, increasing, with $\lim_\infty \varphi(x) = 1$ and $\lim_{-\infty} \varphi(x) = 0$. φ is thus the cdf of some r.v. Y, and the cdf of $n^{\frac{1}{2\alpha}}(\hat{m}_n - m)$ converges pointwise to φ. Since φ is continuous, we may conclude that

$$n^{\frac{1}{2\alpha}}(\hat{m}_n - m) \xrightarrow[n\to\infty]{(d)} Y$$

where $F_Y : t \mapsto \begin{cases} \Phi(2L_2 t^\alpha) & \text{if } t \geq 0 \\ \Phi(-2L_1(-t)^\alpha) & \text{if } t \leq 0 \end{cases}$

Problem 5

Let $(X_n)_{n\geq 1}$ be a sequence i.i.d. r.v's with values in some interval $I \subset \mathbb{R}$ having a density f. Suppose that f is positive over I and continuous. Let $\phi : (x,t) \mapsto |x-t|-|x|$, $\Phi : t \mapsto E(\phi(X_1, t))$ and $\Phi_n : t \mapsto \frac{1}{n}\sum_{i=1}^n \phi(X_i, t)$.

1. Show that X_1 has a unique median, say m, and that $m \in \operatorname{int} I$.
2. Show that Φ is well-defined.
3. Show that Φ is twice-differentiable at m with $\Phi''(m) > 0$.
4. Show that Φ has a unique minimum attained at m.
5. Deduce that an empirical median $\hat{m}_n$ computed from $X_1, \ldots, X_n$ is asymptotically normal and compute its asymptotic variance.

1. Since X_1 has a density, its cdf F is continuous. Note that F (as a function defined on $\overline{\mathbb{R}}$) is 0 over $[-\infty, \inf I]$ and 1 over $[\sup I, \infty]$. Since $f > 0$ over I, F is strictly increasing over $\operatorname{int} I$. By the intermediate value theorem, there exists $m \in \operatorname{int} I$ such that $F(m) = \frac{1}{2}$ (which is necessary and sufficient for m to be a median because X_1 is atomless). Since F is strictly increasing over $\operatorname{int} I$, there is no other median.
2. By the reversed triangle inequality, $||X-t|-|X|| \leq |t|$, thus $|X-t|-|X|$ is bounded and its expectation is well-defined.

3. Note that $E(|t-X|-|X|) = E(|-t-(-X)|-|-X|)$, so by replacing t with $-t$ and X with $-X$ we may suppose WLOG that $t \geq 0$ in the expressions to come.

$$\begin{aligned}
\Phi(t) &= E((t-X-X)1_{t-X>0}1_{X>0} + (t-X+X)1_{t-X>0}1_{X<0} + (X-t-X)1_{t-X<0}) \\
&= E(t1_{t>X>0} - 2X1_{t>X>0} + t1_{X<0} - t1_{t<X}) \qquad \text{since } t \geq 0 \\
&= E(t1_{t>X} - 2X1_{t>X>0} - t + t1_{t>X}) \\
&= 2F(t) - t - 2\int_0^t xf(x)dx
\end{aligned}$$

Note how all the inequalities are strict because X is atomless. Since f is continuous, F is differentiable with $F' = f$, thus Φ is differentiable and computations yield $\Phi' = 2F - 1$. Hence Φ is also twice differentiable with $\Phi'' = 2f$. Since f is positive over I and $m \in \operatorname{int} I$, we have $\Phi''(m) > 0$.

4. We have $\Phi'(t) > 0 \iff F(t) > \frac{1}{2} \iff t > m$, hence Φ has a unique minimum and it is attained at m.