HW1: Asymptotics

Gabriel ROMON

Problem 1

Let $(T_n)_{n\geq 1}$ be a sequence of random vectors of $\mathbb{R}^d$ ($d \geq 1$). T_n is said to be *bounded in probability* or *tight* (denoted $T_n = O_{\mathbb{P}}(1)$) if for any $\varepsilon > 0$, there is some $A > 0$ and $n_0 \geq 1$ such that $n \geq n_0 \implies P(\|T_n\|_2 \geq A) \leq \varepsilon$. More generally, if $(s_n)_{n\geq 1}$ is a sequence of real random variables we write $T_n = O_{\mathbb{P}}(s_n)$ if for any $\varepsilon > 0$, there is some $A > 0$ and $n_0 \geq 1$ such that $n \geq n_0 \implies P(\|T_n\|_2 \geq A s_n) \leq \varepsilon$.

1. Show that if $T_n = o_{\mathbb{P}}(1)$ then $T_n = O_{\mathbb{P}}(1)$.
2. Show that if T_n converges in probability, T_n is tight.
3. Show that if T_n converges in distribution, T_n is tight.
4. Show that if $(\rho_n)_{n\geq 1}$ is a sequence that goes to ∞ and $\rho_n T_n$ converges in distribution, then $T_n = o_{\mathbb{P}}(1)$.
5. Suppose that T_n goes to 0 in probability. Let $g : \mathbb{R}^d \to \mathbb{R}$ be such that $g(x) = o(\|x\|_2^p)$ as $x \to 0$. Show that $g(T_n) = o_{\mathbb{P}}(\|T_n\|_2^p)$.
6. Suppose that T_n goes to 0 in probability. Let $g : \mathbb{R}^d \to \mathbb{R}$ be such that $g(x) = O(\|x\|_2^p)$ as $x \to 0$. Show that $g(T_n) = O_{\mathbb{P}}(\|T_n\|_2^p)$
7. Let $(X_n)_{n\geq 1}$ be a sequence of r.v.'s such that $X_n \sim \mathcal{P}(\frac{1}{n})$.
   (a) Show that $X_n = o_{\mathbb{P}}(1)$.
   (b) Show that for any sequence $(u_n)_{n\geq 1}$ of positive reals, $X_n = o_{\mathbb{P}}(u_n)$.
   (c) Does X_n converge almost surely ?

1. Let $\varepsilon > 0$. Since $T_n = o_{\mathbb{P}}(1)$, the sequence $P(\|T_n\|_2 \geq \varepsilon)$ goes to 0. There exists n_0 such that $n \geq n_0 \implies P(\|T_n\|_2 \geq \varepsilon) \leq \varepsilon$ and we're done.

2. Lemma 1: If $X_n = O_{\mathbb{P}}(1)$ and $Y_n = O_{\mathbb{P}}(1)$ then $X + Y = O_{\mathbb{P}}(1)$.
   Proof: Let $\varepsilon > 0$. There are some $A_1, A_2 > 0$ and $n_1, n_2 \geq 1$ such that $n \geq n_1 \implies P(\|X_n\|_2 \geq A_1) \leq \frac{\varepsilon}{2}$ and $n \geq n_2 \implies P(\|Y_n\|_2 \geq A_2) \leq \frac{\varepsilon}{2}$. Let $A = \max(A_1, A_2)$, $n_0 = \max(n_1, n_2)$ and note that for $n \geq n_0$:

$$
\begin{aligned}
P(\|X_n + Y_n\|_2 \geq 2A) &\leq P(\|X_n\|_2 + \|Y_n\|_2 \geq 2A) \\
&\leq P(\|X_n\|_2 \geq A) + P(\|Y_n\|_2 \geq A) \\
&\leq P(\|X_n\|_2 \geq A_1) + P(\|Y_n\|_2 \geq A_2) \\
&\leq \varepsilon
\end{aligned}
$$

$\square$

   Let T denote the limit of T_n in probability. We have $T_n - T = o_{\mathbb{P}}(1)$, hence 1. yields $T_n - T = O_{\mathbb{P}}(1)$. Since $T_n = T_n - T + T$ and $T = O_{\mathbb{P}}(1)$, the lemma yields $T_n = O_{\mathbb{P}}(1)$.

3. Lemma 2: Let $(X_n^1, \ldots, X_n^d)$ be a random vector. $(X_n^1, \ldots, X_n^d)$ is tight if and only if each of the X_n^i is tight.
   Proof: $\Rightarrow$ Let $\varepsilon > 0$. There are some $A > 0$ and $n_0 \geq 1$ such that $n \geq n_0 \implies P(\|X_n\|_2 \geq A) \leq \varepsilon$. For $n \geq n_0$, $P(|X_n^i| \geq A) \leq P(\|X_n\|_\infty \geq A) \leq P(\|X_n\|_2 \geq A) \leq \varepsilon$ and we're

done.
$\Leftarrow$ Let $\varepsilon > 0$. There are some $A_1, \ldots, A_d$ and $n_1, \ldots, n_d$ linked to the $\dfrac{\varepsilon}{d}$ tightness of each X^i. Let $A = \sqrt{d}\max_{1\leq i\leq d} A_i$ and $n_0 = \max_{1\leq i\leq d} n_i$. For $n \geq n_0$ we have

$$\begin{aligned}
P(\|X_n\|_2 \geq A) &\leq P(\sqrt{d}\|X_n\|_\infty \geq A) \\
&= P(\bigcup_{i=1}^{d} |X_n^i| \geq \frac{A}{\sqrt{d}}) \\
&\leq \sum_{i=1}^{d} P(|X_n^i| \geq \frac{A}{\sqrt{d}}) \\
&\leq \varepsilon
\end{aligned}$$

□

Let us prove the result for $d = 1$. By Lemma 1 it suffices to prove that if $T_n \in \mathbb{R}$ converges to 0 in distribution, then T_n is tight. The cdf of 0 is continuous everywhere except at 0, so $P(T_n \leq 1)$ goes to 1 as $n \to \infty$. Consequently $P(T_n \geq 2) \to 0$, hence T_n is tight.

Let $d \geq 2$ and T_n be a sequence that converges in distribution. By the continous mapping theorem, each T_n^i converges in distribution, hence each T_n^i is tight. By Lemma 2, T_n is tight.

4. Let $\varepsilon > 0$ and X denote a random variable having the distribution of the limit of $\rho_n T_n$. By the continuous mapping theorem $\|\rho_n T_n\|_2$ converges in distribution to $\|X\|_2$. Since $\rho_n \to \infty$, we may assume WLOG that $\rho_n \geq 0$. Let $A > 0$ be fixed. There exists some n_0 such that $n \geq n_0 \implies \rho_n \geq A$. For $n \geq n_0$,

$$\begin{aligned}
P(\|T_n\|_2 \geq \varepsilon) &= P(\rho_n\|T_n\|_2 \geq \varepsilon\rho_n) \\
&\leq P(\rho_n\|T_n\|_2 \geq \varepsilon A)
\end{aligned}$$

Taking the lim sup on both side yields $\limsup_n P(\|T_n\|_2 \geq \varepsilon) \leq \limsup_n P(\rho_n\|T_n\|_2 \geq \varepsilon A)$. The portmanteau theorem applied to $\rho_n\|T_n\|_2$ and the closed set $[\varepsilon A, \infty)$ gives

$$\limsup_n P(\rho_n\|T_n\|_2 \geq \varepsilon A) \leq P(\|X\|_2 \geq \varepsilon A)$$

hence $\limsup_n P(\|T_n\|_2 \geq \varepsilon) \leq P(\|X\|_2 \geq \varepsilon A)$.
Letting $A \to \infty$ yields $\limsup_n P(\|T_n\|_2 \geq \varepsilon) = 0$, hence $P(\|T_n\|_2 \geq \varepsilon) \to 0$ and $T_n = o_{\mathbb{P}}(1)$.

5. Let $\varepsilon > 0$. Since $g(x) = o(\|x\|_2^p)$ as $x \to 0$, there exists some R such that $x \in B_2(0, R) \setminus \{0\} \implies \frac{|g(x)|}{\|x\|_2^p} \leq \frac{\varepsilon}{2}$. This implies

$$P(g(T_n) \geq \varepsilon\|T_n\|_2^p) \leq P(\|T_n\|_2 > R) \xrightarrow[n\to\infty]{} 0$$

and we're done.

6. Let $\varepsilon > 0$. Since $g(x) = O(\|x\|_2^p)$ as $x \to 0$, there exists some $A > 0$ and $R > 0$ such that $x \in B_2(0, R) \setminus \{0\} \implies \frac{|g(x)|}{\|x\|_2^p} \leq A$. This implies

$$P(g(T_n) \geq 2A\|T_n\|_2^p) \leq P(\|T_n\|_2 > R) \xrightarrow[n\to\infty]{} 0$$

Consequently there is some n_0 such that $n \geq n_0 \implies P(\|T_n\|_2 > R) \leq \varepsilon$. Hence $n \geq n_0 \implies P(g(T_n) \geq 2A\|T_n\|_2^p) \leq \varepsilon$ and we're done.

7. (a) Let $\varepsilon \in (0, 1)$. Since X_n is integer-valued, we have

$$P(X_n \geq \varepsilon) \leq P(X_n \geq 1) = 1 - P(X_n = 0) = 1 - \exp(-\frac{1}{n}) \xrightarrow[n\to\infty]{} 0$$

For $\varepsilon \geq 1$, $P(X_n \geq \varepsilon) \leq P(X_n \geq \frac{1}{2}) \xrightarrow[n\to\infty]{} 0$, hence X_n converges in probability to 0.

(b) Let $(u_n)_{n\geq 1}$ be a sequence of positive reals and $\varepsilon > 0$. Note that

$$\begin{aligned}
P(X_n \geq \varepsilon u_n) &= P(X_n \geq \varepsilon u_n) 1_{u_n > \frac{1}{\varepsilon}} + \sum_{k=0}^{\infty} P(X_n \geq \varepsilon u_n) 1_{\frac{1}{2^k \varepsilon} \geq u_n > \frac{1}{2^{k+1}\varepsilon}} \\
&\leq P(X_n \geq 1) + \sum_{k=0}^{\infty} P\left(X_n \geq \frac{1}{2^{k+1}}\right) 1_{\frac{1}{2^k \varepsilon} \geq u_n > \frac{1}{2^{k+1}\varepsilon}} \\
&\leq P(X_n \geq 1) + \sum_{k=0}^{\infty} \left(1 - \exp(-\frac{1}{n})\right) 1_{\frac{1}{2^k \varepsilon} \geq u_n > \frac{1}{2^{k+1}\varepsilon}} \\
&\leq P(X_n \geq 1) + \sum_{k=0}^{\infty} \frac{1}{n} 1_{\frac{1}{2^k \varepsilon} \geq u_n > \frac{1}{2^{k+1}\varepsilon}} \\
&\leq P(X_n \geq 1) + \frac{1}{n} \\
&\xrightarrow[n\to\infty]{} 0
\end{aligned}$$

Hence $X_n = o_{\mathbb{P}}(u_n)$.

(c) Note that $P(X_n > \frac{1}{2}) = 1 - \exp(-\frac{1}{n}) \sim \frac{1}{n}$, hence $\sum_n P(X_n > \frac{1}{2}) = \infty$.
If the X_i are (at least pairwise) **independent**, Borel-Cantelli lemma yields $P\left(\limsup_n \left(X_n > \frac{1}{2}\right)\right) = 1$. For almost all w, we have $X_n(w) > \frac{1}{2}$ infinitely often, hence X_n cannot converge almost surely to 0. Thus X_n does not converge almost surely.

Problem 2

Let $(T_n)_{n\geq 1}$ be a sequence of random vectors of $\mathbb{R}^d$ ($d \geq 1$) and T a random vector.

1. Show that if T_n converges almost surely to T, then T_n converges in probability to T.
2. Show that if T_n converges in probability to T, then T_n converges in distribution to T.
3. Show that if T is constant almost surely, convergence in distribution implies convergence in probability.

1. Let $A = \{w \in \Omega,\ T_n(w) \xrightarrow[n\to\infty]{} T(w)\}$. Note that $A = \bigcap_{m\geq 1} \bigcup_{n\geq 1} \bigcap_{k\geq n} \|T_k - T\| < \frac{1}{m}$.
By assumption, $P(A^c) = 0$, hence $P(\bigcup_{m\geq 1} \bigcap_{n\geq 1} \bigcup_{k\geq n} \|T_k - T\| \geq \frac{1}{m}) = 0$, which implies $\forall m \geq 1, P(\bigcap_{n\geq 1} \bigcup_{k\geq n} \|T_k - T\| \geq \frac{1}{m}) = 0$.
Let $\varepsilon > 0$. There is some $m \geq 1$ such that $\varepsilon \geq \frac{1}{m}$, hence

$$P(\|T_n - T\| \geq \varepsilon) \leq P(\|T_n - T\| \geq \frac{1}{m}) \leq P(\bigcup_{k\geq n} \|T_k - T\| \geq \frac{1}{m})$$

$\bigcup_{k\geq n} \|T_k - T\| \geq \frac{1}{m}$ is a decreasing sequence of events, hence

$$P(\bigcup_{k\geq n} \|T_k - T\| \geq \frac{1}{m}) \xrightarrow[n\to\infty]{} P(\bigcap_{n\geq 1} \bigcup_{k\geq n} \|T_k - T\| \geq \frac{1}{m}) = 0$$

Squeezing thus yields $P(\|T_n - T\| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$, hence T_n converges to T in probability.

2. Let f be a K-Lipschitz function bounded by some $A \geq 0$. Note that for any $\varepsilon > 0$,

$$|f(T_n) - f(T)| \leq K\varepsilon 1_{\|T_n - T\|_2 \leq \varepsilon} + 2A\varepsilon 1_{\|T_n - T\|_2 > \varepsilon}$$

Thus $|E(f(T_n)) - E(f(T))| \leq E(|f(T_n) - f(T)|) \leq K\varepsilon + 2A\varepsilon P(\|T_n - T\|_2 > \varepsilon)$.
Taking $\limsup$ on both side yields $\limsup_n |E(f(T_n)) - E(f(T))| \leq K\varepsilon$. Letting $\varepsilon \to 0$ proves that $\lim_n E(f(T_n)) = E(f(T))$. By the portmanteau theorem, T_n converges to T in distribution.

3. If the random vector $(T_n^1, \dots, T_n^d)$ converges in distribution to some T with $T = (t_1, \dots, t_d)$ a.s., then by the continuous mapping theorem each T_n^i converges in distribution to δ_{t_i}.

   We recall a useful lemma about convergence:
   Lemma 3: $(X_n^1, \dots, X_n^d)$ converges in probability to $(X^1, \dots, X^d)$ if and only if each real r.v. X_n^i converges in probability to X^i.
   By Lemma 3 it suffices to prove the claim in the case $d = 1$.
   By the continuous mapping theorem applied with $x \mapsto |x - t|$, $|T_n - t|$ converges in distribution to 0. Let $\varepsilon > 0$ and note that

$$P(|T_n - t| \geq \varepsilon) \leq P(|T_n - t| > \frac{\varepsilon}{2}) = 1 - P(|T_n - t| \leq \frac{\varepsilon}{2})$$

   Since the cdf of 0 is continuous at $\frac{\varepsilon}{2}$, the convergence of $|T_n - t|$ implies $P(|T_n - t| \leq \frac{\varepsilon}{2}) \xrightarrow[n\to\infty]{} 1$ and we're done.

Problem 3

Let $\alpha \in (0,1)$, $\ell_\alpha : t \mapsto (1-\alpha)t^+ + \alpha t^-$ and $\phi : (x,t) \mapsto \ell_\alpha(x - t)$. Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. r.v.'s with positive density. For $n \geq 1$, let $\hat{q}_n$ an M-estimator associated to ϕ.

1. Show that $\hat{q}_n$ is an α-quantile of the sample $X_1, \dots, X_n$. To simplify matters, $\hat{q}_n$ will be chosen to be maximal.
2. Find k such that $\hat{q}_n = X_{(k)}$ where $X_{(1)} \leq \dots \leq X_{(n)}$ are the order statistics. Show that the inequalities are strict almost surely.
3. We want to prove that $\hat{q}_n$ is asymptotically normal.
   (a) Show that X_1 has a unique α-quantile, say q.
   (b) For $t \in \mathbb{R}$, show that $P(\sqrt{n}\,(\hat{q}_n - q) \leq t) = P(N \geq n\alpha)$ where $N \sim \mathcal{B}(n, F(q + t/\sqrt{n}))$ where F is the cdf of X_1.
   (c) What is the limiting distribution of $\frac{1}{\sqrt{n}}(N - nF(q + t/\sqrt{n}))$ as $n \to \infty$?
   (d) Use Slutsky's theorem to conclude.

1. Let $x_1, \dots, x_n$ be fixed real numbers and $g : t \mapsto \frac{1}{n}\sum_{i=1}^n \ell_\alpha(x_i - t)$.
   By definition, $\hat{q}_n \in \arg\min_t g(t)$. Each $t \mapsto \ell_\alpha(x_i - t)$ is a convex function, so g is convex and t is minimal if and only if $0 \in \partial g(t)$.
   Let $p(t) = |\{i \in [\![1,n]\!],\ x_i < t\}|$ and $q(t) = |\{i \in [\![1,n]\!],\ x_i > t\}|$. Subgradient calculus yields

$$\begin{aligned} \partial g(t) &= -\frac{1}{n}\sum_{i=1}^n \begin{cases} \{\alpha - 1\} & \text{if } x_i < t \\ [\alpha - 1, \alpha] & \text{if } x_i = t \\ \{\alpha\} & \text{if } x_i > t \end{cases} \quad \text{where the summation is over sets} \\ &= \{p(t)(\alpha - 1) + q(t)\alpha\} + [(n - p(t) - q(t))(\alpha - 1), (n - p(t) - q(t))\alpha] \end{aligned}$$

   Thus
$$\begin{aligned} 0 \in \partial g(t) &\iff -p(t)(\alpha - 1) - q(t)\alpha \in [(n - p(t) - q(t))(\alpha - 1), (n - p(t) - q(t))\alpha] \\ &\iff 0 \leq n\alpha - p(t) \leq n - (p(t) + q(t)) \\ &\iff \frac{n - p(t)}{n} \geq 1 - \alpha \quad \text{and} \quad \frac{n - q(t)}{n} \geq \alpha \end{aligned}$$
   Given the definition of $p(t)$ and $q(t)$, this can be rephrased as: t is minimal if and only if it is an α-quantile of $x_1, \dots, x_n$.

2. Let $x_1, \dots, x_n$ be fixed real numbers and $p(t), q(t)$ be defined as above. We want to find the greatest t such that $p(t) \leq n\alpha$ and $q(t) \leq n(1-\alpha)$ both hold. Let us show that $t^* = x_{\lfloor n\alpha \rfloor + 1}$ fits the bill. By definition, $p(t^*) \leq \lfloor n\alpha \rfloor \leq n\alpha$ and

$$q(t^*) \leq n - (\lfloor n\alpha \rfloor + 2) + 1 = n(1-\alpha) + \{n\alpha\} - 1 \leq n(1-\alpha)$$

If $t > x_{\lfloor n\alpha \rfloor + 1}$, then $p(t) \geq \lfloor n\alpha \rfloor + 1 > n\alpha$, hence t^* is the maximal t such that $p(t) \leq n\alpha$ and $q(t) \leq n(1-\alpha)$. Thus $\hat{q}_n = x_{\lfloor n\alpha \rfloor + 1}$.

Remark: $x_{\lceil n\alpha \rceil}$ is another α-quantile, but it is not maximal (consider $n = 6$ and $\alpha = \frac{1}{2}$). To check that it is a quantile note that $p(x_{\lceil n\alpha \rceil}) \leq \lceil n\alpha \rceil - 1 < n\alpha$ and

$$q(x_{\lceil n\alpha \rceil}) \leq n - (\lceil n\alpha \rceil + 1) + 1 = n - \lceil n\alpha \rceil \leq n(1-\alpha)$$

Let $1 \leq i \neq j \leq n$ and f denote the density of X_i.
Note that
$$\begin{aligned} P(X_i = X_j) = E(1_{X_i = X_j}) &= \int 1_{x=y} dP_{(X_i, X_j)}(x,y) \\ &= \int 1_{x=y} dP_{X_i} \otimes dP_{X_j}(x,y) \quad \text{by independence} \\ &= \int \int 1_{x=y} f(x) f(y) dx dy \\ &= \int \left(\int 1_{x=y} f(x)^2 dx \right) dy \quad \text{by Fubini} \end{aligned}$$
For fixed y, the function $x \mapsto 1_{x=y} f(x)^2$ is 0 almost everywhere, thus $\int 1_{x=y} f(x)^2 dx = 0$, hence $P(X_i = X_j) = \int 0 dy = 0$.

3. (a) q is an α-quantile of X_1 if and only if $P(X_1 \leq q) \geq \alpha$ and $P(X_1 \geq q) \geq 1 - \alpha$. Since X_1 has a density, its cdf F is continuous. Since the density is > 0 everywhere, F is also strictly increasing, so F is a continuous increasing bijection from $\mathbb{R}$ to $(0, 1)$.

   Consequently there exists $q \in \mathbb{R}$ such that $F(q) = \alpha$, hence $P(X_1 \leq q) = \alpha$ and since X_1 is atomless, $P(X_1 \geq q) = P(X_1 > q) = 1 - \alpha$. Hence q is an α-quantile of X_1.

   If q is an α-quantile of X_1, we have both $P(X_1 \leq q) \geq \alpha$ and $P(X_1 < q) \leq \alpha$. Since X_1 is atomless $P(X_1 \leq q) = P(X_1 < q) \leq \alpha$, hence $P(X_1 \leq q) = \alpha$ and q is unique by the injectivity of F.

   (b) **In this question it is essential that** $\hat{q}_n = X_{\lceil n\alpha \rceil}$, contrary to what's stated in Question 1.
   For $i \in [\![1, n]\!]$, let $Y_i = 1_{X_i \leq \frac{t}{\sqrt{n}} + q}$ and note that

$$\begin{aligned} P(\sqrt{n}\,(\hat{q}_n - q) \leq t) &= P(X_{\lceil n\alpha \rceil} \leq \frac{t}{\sqrt{n}} + q) \\ &= P(\sum_{i=1}^{n} Y_i \geq \lceil n\alpha \rceil) \\ &= P(\sum_{i=1}^{n} Y_i \geq n\alpha) \end{aligned}$$

   $\sum_{i=1}^{n} Y_i$ has distribution $\mathcal{B}(n, F(t/\sqrt{n} + q))$ as a sum of n i.i.d. Bernoulli r.v.'s.

   If $\hat{q}_n = X_{\lfloor n\alpha \rfloor + 1}$, one gets $P(\sqrt{n}\,(\hat{q}_n - q) \leq t) = P(\sum_{i=1}^{n} Y_i \geq \lfloor n\alpha \rfloor + 1)$ but the last term isn't necessarily equal to $P(\sum_{i=1}^{n} Y_i \geq n\alpha)$ (if $n\alpha \in \mathbb{N}$ and $m \in \mathbb{N}$, $m \geq n\alpha$ does not imply $m \geq \lfloor n\alpha \rfloor + 1$)

(c) Note that

$$\begin{aligned}
&E\left[\exp\left(it\frac{1}{\sqrt{n}}\left(N-nF\left(q+\frac{t}{\sqrt{n}}\right)\right)\right)\right]\\
&=E\left[\exp\left(\frac{itN}{\sqrt{n}}\right)\right]\exp\left(-it\sqrt{n}F\left(q+\frac{t}{\sqrt{n}}\right)\right)\\
&=\left[1+F\left(q+\frac{t}{\sqrt{n}}\right)\left(\exp\left(\frac{it}{\sqrt{n}}-1\right)\right)\right]^n\exp\left(-it\sqrt{n}F\left(q+\frac{t}{\sqrt{n}}\right)\right)
\end{aligned}$$

Since the density of X_1 is continuous, F is differentiable everywhere with $F'=f$. This provides the following asymptotic expansion for F:

$$\begin{aligned}
F\left(q+\frac{t}{\sqrt{n}}\right)&=F(q)+\frac{t}{\sqrt{n}}f(q)+o\left(\frac{1}{\sqrt{n}}\right)\\
&=\alpha+\frac{t}{\sqrt{n}}f(q)+o\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

Let Log denote the principal branch of the logarithm. For $|z|<1$,

$$\mathrm{Log}(1+z)=\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}z^n$$

thus $\dfrac{|\mathrm{Log}(1+z)-z+\frac{z^2}{2}|}{|z^2|}=|z|\left|\sum_{n=3}^{\infty}\frac{(-1)^{n+1}}{n}z^{n-3}\right|$. $z\mapsto\sum_{n=3}^{\infty}\frac{(-1)^{n+1}}{n}z^{n-3}$ is a power series with radius ≥ 1, it is therefore bounded over $\overline{B}(0,\frac{1}{2})$. As a result

$$\lim_{z\to 0}\frac{\mathrm{Log}(1+z)-z+\frac{z^2}{2}}{z^2}=0$$

and $\mathrm{Log}(1+z)=z-\frac{z^2}{2}+o(z^2)$. A bit of algebra shows that

$$\begin{aligned}
\mathrm{Log}\left[1+F\left(q+\frac{t}{\sqrt{n}}\right)\left(\exp\left(\frac{it}{\sqrt{n}}-1\right)\right)\right]&=\mathrm{Log}[1+\frac{i\alpha t}{\sqrt{n}}+\frac{t^2}{n}\left(if(q)-\frac{\alpha}{2}\right)+o\left(\frac{1}{n}\right)]\\
&=\frac{i\alpha t}{\sqrt{n}}+\frac{t^2}{n}\left(if(q)-\frac{\alpha}{2}\right)+\frac{\alpha^2t^2}{2n}+o\left(\frac{1}{n}\right)\\
&=\frac{i\alpha t}{\sqrt{n}}+\frac{t^2}{n}\left(if(q)-\frac{\alpha}{2}+\frac{\alpha^2}{2}\right)+o\left(\frac{1}{n}\right)
\end{aligned}$$

The original expectation turns into

$$\begin{aligned}
E\left[\exp\left(it\frac{1}{\sqrt{n}}\left(N-nF\left(q+\frac{t}{\sqrt{n}}\right)\right)\right)\right]&=\exp\left[n\left(\frac{i\alpha t}{\sqrt{n}}+\frac{t^2}{n}\left(if(q)-\frac{\alpha}{2}+\frac{\alpha^2}{2}\right)+o\left(\frac{1}{n}\right)\right)\right]\\
&\quad\cdot\exp\left(-it\alpha\sqrt{n}-it^2f(q)+o(1)\right)\\
&=\exp\left[-\frac{\alpha(1-\alpha)}{2}t^2+o(1)\right]\\
&\xrightarrow[n\to\infty]{}\exp\left[-\frac{\alpha(1-\alpha)}{2}t^2\right]
\end{aligned}$$

The characteristic function of $\frac{1}{\sqrt{n}}\left(N-nF\left(q+\frac{t}{\sqrt{n}}\right)\right)$ converges pointwise to that of a $\mathcal{N}(0,\alpha(1-\alpha))$, hence $\frac{1}{\sqrt{n}}\left(N-nF\left(q+\frac{t}{\sqrt{n}}\right)\right)$ converges in distribution to $\mathcal{N}(0,\alpha(1-\alpha))$.

(d) Let $Z_n = \frac{1}{\sqrt{n}}\left(N - nF\left(q + \frac{t}{\sqrt{n}}\right)\right)$. Note that

$$P(N \geq n\alpha) = P\left(Z_n \geq \sqrt{n}\left(\alpha - F\left(q + \frac{t}{\sqrt{n}}\right)\right)\right) = P\left(-Z_n \leq \sqrt{n}\left(F\left(q + \frac{t}{\sqrt{n}}\right) - F(q)\right)\right)$$

$\sqrt{n}\left(F\left(q + \frac{t}{\sqrt{n}}\right) - F(q)\right)$ is a deterministic sequence that converges (everywhere, hence almost surely, thus in probability) to $tf(q)$. We have

$$P(N \geq n\alpha) = P\Bigg(-Z_n \underbrace{\frac{tf(q)}{\sqrt{n}\left(F\left(q + \frac{t}{\sqrt{n}}\right) - F(q)\right)} \frac{1}{f(q)}}_{\text{converges in probability to } \frac{1}{f(q)}} \leq t\Bigg)$$

By Slutsky's theorem, the random variable on the left of the $\leq$ sign converges in distribution to $-\frac{1}{f(q)}\mathcal{N}(0, \alpha(1-\alpha)) = \mathcal{N}(0, \frac{\alpha(1-\alpha)}{f(q)^2})$.
Hence $P(\sqrt{n}\,(\hat{q}_n - q) \leq t) = P(N \geq n\alpha)$ converges to the cdf of a $\mathcal{N}(0, \frac{\alpha(1-\alpha)}{f(q)^2})$ evaluated at t (and this cdf is continuous).
This proves that $\sqrt{n}\,(\hat{q}_n - q)$ converges in distribution to a $\mathcal{N}(0, \frac{\alpha(1-\alpha)}{f(q)^2})$.

Problem 4

Let $\theta > 0$ and $(X_n)_{n\geq 1}$ be i.i.d. random variables following $\mathcal{U}([0,\theta])$. Show that the MLE $\hat{\theta}_n$ of θ is asymptotically exponential with convergence rate $\frac{1}{n}$.

Let $x_1, \dots, x_n$ be an n-sample. The likelihood of the model writes as

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} 1_{[0,\theta]}(x_i) = \frac{1}{\theta^n} 1_{\min x_i \geq 0} 1_{\max x_i \leq \theta}$$

If $\min x_i < 0$, $L = 0$ and the MLE is not defined, so we may assume WLOG that $\min x_i \geq 0$.
L is 0 when $\theta < \max x_i$ and positive decreasing for $\theta \geq \max x_i$. Thus L has a unique maximum at $\theta = \max x_i$, hence $\hat{\theta}_n = \max x_i$.

Let us compute the cdf of $\hat{\theta}_n$. Let F denote the cdf of X_1.

$$\begin{aligned} P(\max X_i \leq t) &= P(\bigcap_{i=1}^{n} X_i \leq t) = F(t)^n \\ &= \begin{cases} 0 & \text{if } t < 0 \\ \frac{t^n}{\theta^n} & \text{if } t \in [0,\theta) \\ 1 & \text{if } t \geq \theta \end{cases} \end{aligned}$$

The cdf is continuous so the distribution of $\max X_i$ is atomless.

Let $t \geq 0$. Since $\hat{\theta}_n$ is atomless,

$$\begin{aligned} P(n(\theta - \hat{\theta}_n) \leq t) &= P(\hat{\theta}_n \geq \theta - \frac{t}{n}) = 1 - P(\hat{\theta}_n \leq \theta - \frac{t}{n}) \\ &= 1 - \left(\theta - \frac{t}{n}\right)^n \frac{1}{\theta^n} = 1 - \left(1 - \frac{t}{\theta n}\right)^n \\ &\xrightarrow[n\to\infty]{} 1 - \exp(-\frac{t}{\theta}) \end{aligned}$$

If $t < 0$ similar computations show that $P(n(\theta - \hat{\theta}_n) \leq t) \xrightarrow[n\to\infty]{} 0$
The limiting cdf is that of a $\mathcal{E}(\frac{1}{\theta})$ (and it is continuous), so $n(\theta - \hat{\theta}_n)$ converges in distribution to $\mathcal{E}(\frac{1}{\theta})$.

Problem 5

Let $a \in \mathbb{R}$, $\lambda > 0$ and $f : x \mapsto \lambda e^{-\lambda(x-a)} 1_{x \geq a}$. Let $(X_n)_{n\geq 1}$ be an i.i.d sequence of r.v.'s with density f. For $n \geq 1$, let $(\hat{a}_n, \hat{\lambda}_n)$ the MLE of (a, λ). Show that $\hat{a}_n$ is asymptotically exponential with convergence rate $\frac{1}{n}$ and $\hat{\lambda}_n$ is asymptotically normal.

Let $x_1, \ldots, x_n$ be an n-sample. The likelihood of the model writes as

$$L(a, \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda(x-a)} 1_{x \geq a} = \lambda^n 1_{\min x_i \geq a} e^{-\lambda \sum_{i=1}^{n}(x_i - a)}$$

When $a > \min x_i$, $L(a, \lambda) = 0$ and the likelihood is minimized. We may therefore assume that $a \leq \min x_i$. If $a = \frac{1}{n}\sum_{i=1}^{n} x_i$, then $a = x_1 = \ldots = x_n$ and $L(a, \lambda) = \lambda^n \xrightarrow[\lambda \to \infty]{} \infty$, so the MLE does not exist. We may thus assume additionally that the equalities $x_1 = \ldots = x_n$ do not hold, so $a < \frac{1}{n}\sum_{i=1}^{n} x_i$.

We have $\log L(a, \lambda) = n \log \lambda - \lambda \sum_{i=1}^{n}(x_i - a)$. Studying the derivative w.r.t λ shows that $\lambda \mapsto \log L(a, \lambda)$ reaches a unique maximum at $\lambda^*(a) = \dfrac{n}{\sum_{i=1}^{n}(x_i - a)}$ (which is well-defined given the previous assumption). Since log is stricty monotonic, $\lambda \mapsto L(a, \lambda)$ also has its unique maximum at $\lambda^*(a)$.
Consider $(-\infty, \min x_i] \to \mathbb{R}, a \mapsto L(a, \lambda(a^*)) = \frac{n^n}{\left[\sum_{i=1}^{n}(x_i - a)\right]^n}$. This function is increasing in a, so it reaches its maximum at $a = \min x_i$.
Thus $\hat{a}_n = \min x_i$ and $\hat{\lambda}_n = \lambda(\hat{a}_n) = \frac{n}{\sum_{i=1}^{n}(x_i - \min x_i)}$.

The cdf of X_1 is given by $P(X_1 \leq t) = \begin{cases} 0 & \text{if } t < a \\ 1 - e^{-\lambda(t-a)} & \text{if } t \geq a \end{cases}$ and the cdf of $\min X_i$ by $P(\min X_i \leq t) = 1 - (1 - P(X_1 \leq t))^n$.

Let $t \geq 0$. We have $P(n(\min X_i - a) \leq t) = 1 - (1 - (1 - e^{-\lambda \frac{t}{n}}))^n = 1 - e^{-\lambda t}$. For $t < 0$ we get $P(n(\min X_i - a) \leq t) = 0$ in a similar fashion. The cdf of $n(\min X_i - a)$ is that of a $\mathcal{E}(\lambda)$, hence $n(\min X_i - a) \sim \mathcal{E}(\lambda)$ (and remarkably this holds for finite n).

X_1 is square-integrable with $E(X_1) = \frac{1}{\lambda} + a$ and $V(X_1) = \frac{1}{\lambda^2}$. Note that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \min X_i - \frac{1}{\lambda}\right) = \sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i)) + \sqrt{n}\,(a - \min X_i)$$

Since $\sqrt{n} \cdot \sqrt{n}(\min X_i - a)$ converges in distribution, Question 4 from Problem 1 implies that $\sqrt{n}(\min X_i - a) = o_{\mathbb{P}}(1)$. The CLT yields the convergence in distribution of $\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))$ to $\mathcal{N}(0, \frac{1}{\lambda^2})$. By Slutsky's theorem $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \min X_i - \frac{1}{\lambda}\right)$ converges in distribution to $\mathcal{N}(0, \frac{1}{\lambda^2})$.
The Delta Method applied with $x \mapsto \frac{1}{x}$ yields the convergence in distribution of

$$\sqrt{n}\left(\frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i - \min X_i} - \lambda\right)$$

to $\mathcal{N}(0, \frac{1}{\lambda^2} \cdot \lambda^4) = \mathcal{N}(0, \lambda^2)$.

Problem 6

Let $\theta \in \mathbb{R}$ and $(X_n)_{n\geq 1}$ a sequence of i.i.d. r.v.'s following $\mathcal{N}(\theta^3, 1)$.

1. For $n \geq 1$ compute $\hat{\theta}_n$ the MLE of θ.
2. Show that $\hat{\theta}_n$ is consistent.
3. For what values of θ is $\hat{\theta}_n$ asymptotically normal ?
4. Depending on θ find $\alpha > 0$ such that $|\hat{\theta}_n - \theta| = O_{\mathbb{P}}\left(\frac{1}{n^\alpha}\right)$

1. Let $x_1, \dots, x_n$ be an n-sample. The likelihood of the model writes as

$$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp(-\frac{(x_i - \theta^3)^2}{2})$$
$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n (x_i - \theta^3)^2\right)$$

Thus $\log L(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n (\theta^3 - x_i)^2$ which is a degree 6 polynomial in θ with leading coefficient $-\frac{n}{2}$. It is therefore coercive and reaches a global maximum at a critical point. We have

$$(\log L)'(\theta) = 0 \iff 6\theta^2 \sum_{i=1}^n (\theta^3 - x_i) = 0 \iff \theta = 0 \text{ or } \theta = \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^{1/3} = \overline{x}^{1/3}$$

Up to a constant we have $(\log L)(0) = -\frac{1}{2}\sum_{i=1}^n x_i^2$ and

$$(\log L)\left(\overline{x}^{1/3}\right) = -\frac{1}{2}\sum_{i=1}^n (x_i - \overline{x})^2 = -\frac{1}{2}\left[\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right] \geq (\log L)(0)$$

The MLE is thus $\hat{\theta}_n = \overline{x}^{1/3}$.

2. By the weak Law of Large Numbers $\overline{X}$ converges in probability to θ^3. The continuous mapping theorem applied with $x \mapsto x^{1/3}$ yields convergence in probability of $\overline{X}^{1/3}$ to θ, thus $\hat{\theta}_n$ is consistent.

3. By the CLT $\sqrt{n}(\overline{X} - \theta^3)$ converges in distribution to $\mathcal{N}(0,1)$. If $\theta \neq 0$ the function $x \mapsto x^{1/3}$ is differentiable at θ and the Delta Method yields convergence in distribution of $\sqrt{n}(\overline{X}^{1/3} - \theta)$ to $\mathcal{N}(0, \frac{1}{9\theta^4})$.
Let Y be a r.v. with distribution $\mathcal{N}(0,1)$. When $\theta = 0$, combining the CLT with the continuous mapping theorem gives convergence in distribution of $n^{1/6}\overline{X}^{1/3}$ to $Y^{1/3}$, which rewrites as $\left[n^{1/2}\overline{X}^{1/3}\right]\frac{1}{n^{1/3}} \to Y^{1/3}$. If $n^{1/2}\overline{X}^{1/3}$ converged in distribution, Slutksy's theorem would imply that $\left[n^{1/2}\overline{X}^{1/3}\right]\frac{1}{n^{1/3}} \to 0$ in distribution, a contradiction. Consequently, when $\theta = 0$, $\hat{\theta}_n$ is not asymptotically normal.

4. For $\theta \neq 0$ we proved that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \frac{1}{9\theta^4})$. Question 3 of Problem 1 implies that $\sqrt{n}(\hat{\theta}_n - \theta)$ is tight, hence $\hat{\theta}_n - \theta = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$.

For $\theta = 0$, $n^{1/6}\hat{\theta}_n \xrightarrow[n\to\infty]{\mathcal{L}} Y^{1/3}$, and by the same argument $\hat{\theta}_n - \theta = O_{\mathbb{P}}\left(\frac{1}{n^{1/6}}\right)$.