Stat 553 Homework Solutions

Zhanxiong Xu

Spring 2017

Homework 1

Problem 1 *Proof.* If $\omega \in \liminf_n A_n$, then there exists $n_0 \geq 1$, such that $\omega \in A_k$ for all $n \geq n_0$, that is, ω belongs to all but finitely many A_n's.

Conversely, assume $\omega \in A_k$ for all $k \geq n_0$, for some $n_0 \geq 1$, then $\omega \in \bigcap_{k=n_0}^{\infty} A_k \subset \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$.

The other statement can be proved similarly. $\qquad\square$

1.17 *Proof.* (a) This statement is false. The counterexample is $a_n = b_n = (-1)^n$, for which the limit of a_n (b_n) does not exist.

(b) The statement is true, as by the algebra of limit operation (Theorem 3.3, [5]), we have:

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n} = 1.$$

(c) The statement is false. We can use the same counterexample in part (a). $\qquad\square$

1.18 *Proof.* (a) Since $\dfrac{a_n c_n}{b_n d_n} = \dfrac{a_n}{b_n} \cdot \dfrac{c_n}{d_n} \to 1 \times 1 = 1$ as $n \to \infty$, $a_n c_n \sim b_n d_n$.

(b) Take $a_n = n + 1, b_n = n + 2, c_n = d_n = -n$, then $\lim_{n\to\infty} \dfrac{a_n + c_n}{b_n + d_n} = 1/2 \neq 1$, hence $a_n + c_n \nsim b_n + d_n$, although $a_n \sim b_n$, $c_n \sim d_n$.

(c) For simplicity, denote $|a_n|, |b_n|, |c_n|, |d_n|$ by A_n, B_n, C_n, D_n respectively. Consider the difference $\left| \dfrac{A_n + C_n}{B_n + D_n} - 1 \right|$:

$$\left| \frac{A_n + C_n}{B_n + D_n} - 1 \right| = \left| \frac{A_n - B_n}{B_n + D_n} - \frac{C_n - D_n}{B_n + D_n} \right| \leq \frac{|A_n - B_n|}{B_n + D_n} + \frac{|C_n - D_n|}{B_n + D_n}$$

$$\leq \frac{|A_n - B_n|}{B_n} + \frac{|C_n - D_n|}{D_n} = \left| \frac{A_n}{B_n} - 1 \right| + \left| \frac{C_n}{D_n} - 1 \right| \to 0$$

as $n \to \infty$. Therefore $A_n + C_n \sim B_n + D_n$.

(d) Take $f(x) = e^x, x \in \mathbb{R}, a_n = n, b_n = n + 1$. Obviously, $a_n \sim b_n$, while $f(b_n)/f(a_n) = e^{n+1}/e^n = e \not\to 1$ as $n \to \infty$. Therefore $f(b_n)$ is not equivalent to $f(a_n)$. $\qquad\square$

1.25 *Proof.* Fix c that is in the domain of f. By condition, for all sufficiently large n, we have $b_n > a_n > c$, as well as $f(a_n) > 0, f(b_n) > 0$.

Write a_n as the convex combination of b_n and c as follows:

$$a_n = \frac{a_n - c}{b_n - c} b_n + \frac{b_n - a_n}{b_n - c} c.$$

By convexity of f and the expression above, it follows that

$$0 < f(a_n) \leq \frac{a_n - c}{b_n - c} f(b_n) + \frac{b_n - a_n}{b_n - c} f(c).$$

Or equivalently,

$$0 < \frac{f(a_n)}{f(b_n)} \leq \frac{a_n - c}{b_n - c} + \frac{b_n - a_n}{b_n - c} \frac{f(c)}{f(b_n)}$$

A passage to limit and squeeze principle then shows that $f(a_n)/f(b_n) \to 0$ as $n \to \infty$, i.e., $f(a_n) = o(f(b_n))$. $\qquad\square$

1.26 *Proof.* (a) $a_n \equiv 0, b_n \equiv 1, f(x) = 1 + x^2, x \in \mathbb{R}$.

(b) $a_n = n, b_n = n^2, f(x) \equiv 1, x \in \mathbb{R}$ (f is a constant function).

2

(c) $a_n = n, b_n = n^2, f(x) = \log x, x \in (0, \infty)$. $\qquad\square$

1.27 *Proof.* (a) The correct order is:

$$(\log n)^{\log \log n} \ll \sqrt{\log(n!)} \ll 2^{\log n} \ll n \ll \sum_{i=1}^{n} i^{1/3}.$$

Here the operator "$\ll$" bears the meaning: $a_n \ll b_n \iff a_n = o(b_n)$. In the remaining of this part we show all these relations from left to right.

Clever application of Theorem 1.31 (which has been proved in Exercise 1.25) could substantially simplify some proofs, as will be shown below.

- Since the function $f(x) = e^x, x \in \mathbb{R}$ is convex, to show the first relation, it suffices to show $(\log \log n)^2 = o(\log \log(n!)/2)$ by Theorem 1.31.

  By the Stirling's formula (Theorem 8.22, [5]), we have

  $$\log(n!) = n \log n - n + \theta_n, \tag{1}$$

  where $\theta_n = O(\log n)$. Therefore $\log \log(n!) = \log(n \log n - n + \theta_n) = \log n + \log(\log n - 1 + \theta_n/n) > \log n$ for all sufficiently large n. [1] Hence

  $$0 < \frac{(\log \log n)^2}{\frac{1}{2} \log \log(n!)} < \frac{2(\log \log n)^2}{\log n} \to 0 \tag{2}$$

  as $n \to \infty$. The last assertion may be shown by letting $x_n = \log \log n$. Clearly $x_n \to \infty$, thereby the inequality $e^x > 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$ implies that

  $$\frac{2(\log \log n)^2}{\log n} = \frac{2x_n^2}{e^{x_n}} < \frac{2x_n^2}{1 + x_n + \frac{1}{2}x_n^2 + \frac{1}{6}x_n^3} = \frac{2}{x_n^{-2} + x_n^{-1} + \frac{1}{2} + \frac{1}{6}x_n} \to 0$$

  as $n \to \infty$.

---

[1] For a shorter proof without bothering Stirling, note that for sufficiently large n,
$$\log(n!) = \sum_{i=1}^{n} \log i > (n-2) \log 3 > n.$$

3

- It would be helpful to recognize that $2^{\log n} = n^{\log 2}$. By Stirling's formula (1), it follows that

$$\frac{\log(n!)}{n^2 \log 2} = \frac{n \log n - n + \theta_n}{n^2 \log 2} = \frac{\log n}{n^2 \log 2 - 1} - \frac{1}{n^2 \log 2 - 1} + \frac{\theta_n}{n^2 \log 2} \to 0$$

as $n \to \infty$, where we used the well-known limit:

$$\lim_{n \to \infty} \frac{\log n}{n^\varepsilon} = 0$$

for any $\varepsilon > 0$ (which can be proved by the similar argument I used to prove (2) or L'Hospital's rule.). Therefore,

$$\frac{\sqrt{\log(n!)}}{n^{\log 2}} = \sqrt{\frac{\log(n!)}{n^2 \log 2}} \to 0$$

as $n \to \infty$, by the continuity of the function $x \mapsto \sqrt{x}$. (Note such argument is essentially different from applying Theorem 1.31.).

- The relation $2^{\log n} = n^{\log 2} \ll n$ is trivial.

- Use the result of Exercise 1.19, we have $\sum_{i=1}^{n} i^{1/3} \sim \frac{3}{4} n^{4/3}$, hence $n \ll \sum_{i=1}^{n} i^{1/3}$.

This completes the proof. $\qquad\square$

(b) The correct order is given as follows:

$$\log \log n \ll \log n \ll n \ll \log(n!) \ll n^2 \ll 2^{3 \log n}$$
$$\ll n^{\log n} \ll 3^n \ll (\log n)^n \ll n^{n/2} \ll n! \ll n^n \ll 2^{2^n}.$$

The verification for every relation seems too tedious, and actually very similar to what we did for part (a). Let me just mention that some tips:

1. Use Stirling's formula (1) to handle expressions containing $n!$.

2. Use Theorem 1.31 or the continuity argument when necessary.

3. $2^{2^n} \neq 4^n$.

1.32 It's straightforward to obtain the following answer by evaluating partial derivatives:

$$U(\mu, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) \\ \frac{1}{2\sigma^4} \left[\sum_{i=1}^{n} (X_i - \mu)^2 - n\sigma^2 \right] \end{bmatrix}.$$

Homework 2

Problem 1 *Proof.* If either $E[X^2] = 0$ or $E[Y^2] = 0$ then $X = 0$ almost surely or $Y = 0$ almost surely, which implies $E[XY] = 0$, hence the inequality holds.

Now assume $E[X^2] \neq 0$ and $E[Y^2] \neq 0$, then for each $t \in \mathbb{R}^1$,

$$0 \leq E[(tX + Y)^2] = t^2 E[X^2] + 2tE[XY] + E[Y^2]$$

implies that $\Delta = \{2E[XY]\}^2 - 4E[X^2]E[Y^2] \leq 0$. Rearrangement gives the desired inequality. $\square$

Problem 2 *Proof.* If $t = 0$, there is nothing to prove. So let's assume $t > 0$, and denote X/t by Y, then $E[Y] = E[X]/t = 0$. It is therefore easy to see that the inequality to be proved is equivalent to

$$P[Y \geq 1] \leq \frac{E[Y^2]}{1 + E[Y^2]}.$$

For simplicity, denote $E[Y^2]$ by μ_2. By Markov's inequality, it follows that

$$P[Y \geq 1] = P[Y + \mu_2 \geq 1 + \mu_2] \leq \frac{E[(Y + \mu_2)^2]}{(1 + \mu_2)^2}$$
$$= \frac{E[Y^2] + 2\mu_2 E[Y] + \mu_2^2}{(1 + \mu_2)^2} = \frac{\mu_2 + \mu_2^2}{(1 + \mu_2)^2} = \frac{\mu_2}{1 + \mu_2}.$$

In the penultimate equality above, we used the condition $E[Y] = 0$. This completes the proof. $\square$

Problem 3 Before starting the proof, keep in mind that the notations $O_P(1)$ and $o_P(1)$ need to be understood as shorthands for sequences of random variables. For example, "$o_P(1)$" stands for a sequence $\{X_n\}$ such that X_n converges to 0 in probability.

- *Proof.* Suppose $X_n = O_P(1)$ and $Y_n = o_P(1)$. Given $\varepsilon > 0$, since $X_n = O_P(1)$, there exists $K > 0$ such that $P[|X_n| \geq K] < \frac{\varepsilon}{2}$ for all $n \in \mathbb{N}$. In addition, $Y_n = o_P(1)$ implies that there exists $N \in \mathbb{N}$ such that $P[|Y_n| \geq \varepsilon/K] < \frac{\varepsilon}{2}$ for all

6

$n > N$. It then follows that for each $n > N$,

$$P[|X_n Y_n| \geq \varepsilon] = P[|X_n Y_n| \geq \varepsilon, |X_n| \geq K] + P[|X_n Y_n| \geq \varepsilon, |X_n| < K]$$
$$\leq P[|X_n| \geq K] + P[|Y_n| \geq \varepsilon/K] < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Therefore $X_n Y_n \to_P 0$, i.e., $X_n Y_n = o_P(1)$. □

- *Proof.* Suppose $X_n = O_P(1)$ and $Y_n = o_P(1)$. It is clear that if $Y_n = o_P(1)$, then $Y_n = O_P(1)$. Therefore, this statement can be treated as a corollary of the next statement so we can take a rest here. □

- *Proof.* Suppose $X_n = O_P(1)$ and $Y_n = O_P(1)$. Given $\varepsilon > 0$, by condition, there exist $K_1 > 0$ and $K_2 > 0$ such that $P[|X_n| \geq K_1] < \frac{\varepsilon}{2}$ and $P[|Y_n| \geq K_2] < \frac{\varepsilon}{2}$ for all $n \in \mathbb{N}$. Noticing that the event $\{\omega : |X_n(\omega)| + |Y_n(\omega)| \geq K_1 + K_2\}$ is a subevent of $\{\omega : |X_n(\omega)| \geq K_1\} \cup \{\omega : |Y_n(\omega)| \geq K_2\}$, we have for each $n \in \mathbb{N}$,

$$P[|X_n + Y_n| \geq K_1 + K_2]$$
$$\leq P[|X_n| + |Y_n| \geq K_1 + K_2]$$
$$\leq P[\{|X_n| \geq K_1\} \cup \{|Y_n| \geq K_2\}]$$
$$\leq P[|X_n| \geq K_1] + P[|Y_n| \geq K_2]$$
$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This shows that $X_n + Y_n = O_P(1)$. □

- *Proof.* Suppose $X_n = o_P(1)$. Given $\varepsilon > 0$, by condition, there exists $N \in \mathbb{N}$ such that $P[|X_n| \geq 1/2] < \varepsilon$ for all $n > N$. It then follows that for each $n > N$,

$$P\left[\left|\frac{1}{1 + X_n}\right| \geq 2\right]$$
$$= P[|X_n + 1| \leq 1/2]$$
$$= P[|X_n + 1| \leq 1/2, |X_n| \geq 1/2] + P[|X_n + 1| \leq 1/2, |X_n| < 1/2]$$
$$\leq P[|X_n| \geq 1/2] + P[\varnothing]$$

7

$$=P[|X_n| \geq 1/2] < \varepsilon.$$

In addition, since $(1+X_1)^{-1}, \ldots, (1+X_N)^{-1}$ are finitely many random variables, there exists $K > 0$ such that $P[|(1+X_n)^{-1}| \geq K] < \varepsilon$ for $n = 1, \ldots, N$. Take $K' = \max\{2, K\}$, it then holds that $P[|(1+X_n)^{-1}| \geq K'] < \varepsilon$ for all $n \in \mathbb{N}$, i.e.,

$$\frac{1}{1+X_n} = O_P(1). \qquad \qquad \square$$

1.33 *Proof.* Using the notations adopted in Chapter 9 of [5], straightforward calculation shows that

$$(D_1 f)(x, y) = \begin{cases} \dfrac{x^4 y + 4x^2 y^3 - y^5}{(x^2 + y^2)^2} & (x, y) \neq (0, 0), \\ 0 & (x, y) = (0, 0). \end{cases}$$

$$(D_2 f)(x, y) = \begin{cases} \dfrac{-xy^4 - 4x^2 y^3 + x^5}{(x^2 + y^2)^2} & (x, y) \neq (0, 0), \\ 0 & (x, y) = (0, 0). \end{cases}$$

Therefore,

$$(D_{21} f)(0, 0) = \lim_{t \to 0} \frac{(D_1 f)(0, t) - (D_1 f)(0, 0)}{t} = -1,$$
$$(D_{12} f)(0, 0) = \lim_{t \to 0} \frac{(D_2 f)(t, 0) - (D_2 f)(0, 0)}{t} = 1.$$

That shows $\nabla^2 f(0, 0)$ is not symmetric, hence f is not twice differentiable. $\qquad \square$

1.39 *Proof.* I shall prove the following more standard, yet trivially equivalent version to (1.39):

$$E[|XY|] \leq \{E[|X|^p]\}^{1/p} \{E[|Y|^q]\}^{1/q}, \text{ if } \frac{1}{p} + \frac{1}{q} = 1. \tag{3}$$

The reason I prefer (3) to (1.39) is that the term $\{E[|X|^p]\}^{1/p}$ can be interpreted as the L^p norm of X provided $X \in L^p(\Omega, \mathscr{F}, P)$, the space of all p-integrable random variables. In this way, the Hölder's inequality can be more succinctly expressed as

8

$\|XY\|_1 \leq \|X\|_p \|Y\|_q$, which may help you memorize it more easily.

To show (3), let's prove the following *Young's inequality* first: For any two positive reals $a > 0$ and $b > 0$,

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q, \text{ if } p > 0, q > 0, \frac{1}{p} + \frac{1}{q} = 1. \tag{4}$$

To prove (4), reparametrize a and b as $a = e^{x/p}$ and $b = e^{y/q}$ for some $x \in \mathbb{R}$ and $y \in \mathbb{R}$. By the convexity of the function $x \mapsto e^x$ and Jensen's inequality, it follows that

$$ab = e^{x/p+y/q} \leq \frac{1}{p}e^x + \frac{1}{q}e^y = \frac{1}{p}a^p + \frac{1}{q}b^q,$$

which shows (4). To show the Hölder's inequality, take $a = |X|/E^{1/p}[|X|^p]$, $b = |Y|/E^{1/q}[|Y|^q]$ in the Young's inequality to obtain (To be complete, again you need to justify why $E[|X|^p]$ and $E[|Y|^q]$ can be placed as denominators.):

$$\frac{|XY|}{E^{1/p}[|X|^p]E^{1/q}[|Y|^q]} \leq \frac{1}{p}\frac{|X|^p}{E[|X|^p]} + \frac{1}{q}\frac{|Y|^q}{E[|Y|^q]}.$$

Then take expectations on both sides in the above inequality, it follows that

$$\frac{E[|XY|]}{E^{1/p}[|X|^p]E^{1/q}[|Y|^q]} \leq \frac{1}{p} + \frac{1}{q} = 1.$$

Hence $E[|XY|] \leq E^{1/p}[|X|^p]E^{1/q}[|Y|^q]$. This completes the proof. $\square$

1.41 *Proof.* First note that without loss of generality, we can assume $E[X_1] = \cdots = E[X_n]$, for which case $S_n = \sum_{i=1}^n X_i, \mathrm{Var}(S_n) = E[S_n^2]$.

Following the hint, let A_k be the set such that $|S_k| \geq a$ and $|S_j| < a$ for $j < k$, $k = 1, \ldots, n$ (define $S_0 = 0$.). Since $A_1, \ldots, A_n$ are disjoint, it follows that

$$E[S_n^2] \geq \sum_{k=1}^n E[S_n^2 I_{A_k}] = \sum_{k=1}^n E\{[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2]I_{A_k}\}$$

9

$$\geq \sum_{k=1}^{n} E\{[S_k^2 + 2S_k(S_n - S_k)]I_{A_k}\}.$$

Note that $S_k I_{A_k} = (X_1 + \cdots + X_k)I_{A_k}$ and $S_n - S_k = X_{k+1} + \cdots + X_n$ are independent, therefore $E[S_k(S_n - S_k)I_{A_k}] = E[S_n - S_k]E[S_k I_{A_k}] = 0$. Hence,

$$\sum_{k=1}^{n} E\{[S_k^2 + 2(S_n - S_k)S_k]I_{A_k}\} = \sum_{k=1}^{n} E[S_k^2 I_{A_k}] \geq a^2 \sum_{k=1}^{n} E[I_{A_k}]$$

$$= a^2 \sum_{k=1}^{n} P(A_k) = a^2 P\left(\bigcup_{k=1}^{n} A_k\right) = a^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq a\right).$$

This completes the proof. $\qquad\square$

1.45 *Proof.* See [1], Theorem 3.2.1 for a more general statement:

$$\sum_{n=1}^{\infty} P[|X| \geq n] \leq E[|X|] \leq 1 + \sum_{n=1}^{\infty} P[|X| \geq n].$$

See also [2], pp.79, 275 for establishing the general equality $E[X] = \int_0^\infty P[X \geq t]\,dt$ when X is nonnegative. $\qquad\square$

Quiz 2

In this problem, x is tacitly assumed to be positive.

- By independence assumption, it follows that

$$P[S_n \geq nx] \leq \min_{t>0} \frac{\prod_{i=1}^{n} E[e^{tX_i}]}{e^{ntx}}.$$

- By condition and the generic inequality derived in the first part, we have

$$P[S_n \geq nx] \leq \min_{t>0} \exp\left(\frac{1}{2}n\sigma^2 t^2 - ntx\right).$$

The expression on the right hand side has a unique minimizer $t_0 = x/\sigma^2$, hence the Chernoff bound is $\boxed{\exp(-nx^2/2\sigma^2)}$.

- By hint and the generic inequality derived in the first part, we have

$$P[S_n \geq nx] \leq \min_{t>0} \exp\left(npe^t - ntx - np\right) := \min_{t>0} g(t).$$

It can be shown that $g(t)$ is a strictly convex function (of t) on $\mathbb{R}$, therefore it has a unique minimizer. If $x > p$, the minimum is attained at $t_0 = \log(x/p) > 0$. If $0 < x \leq p$, then $g(t)$ is nondecreasing on $[0, +\infty)$, hence the minimum is attained at $t_0 = 0$. In summary,

$$P[S_n \geq nx] \leq \begin{cases} \exp[nx - np - nx\log(x/p)] & x > p, \\ 1 & 0 < x \leq p. \end{cases}$$

The bound corresponds to $0 < x \leq p$ is trivial and useless.

Homework 3

Problem 1 *Proof.* Problem 1 can be proved in a similar way to Problem 3, Homework 2, i.e., decomposing the probability of interest into two parts by introducing some event with small probability, then controlling the upper bound.

As an alternative way, let's try to apply the following handy theorem to prove the two statements in this problem.

Theorem ([3], Section 2, Theorem 2 (d)). *A necessary and sufficient condition for $X_n \to_P X$ is that for any subsequence $\{X_{n_k}\}$ of $\{X_n\}$, there exists a further subsequence $\{X_{n_{k_i}}\}$ of $\{X_{n_k}\}$ such that $\{X_{n_{k_i}}\}$ converge to X almost surely as $i \to \infty$.*

Why this theorem can make things easier? Because in general convergence almost surely is easier to handle than convergence in probability (measure), as the former one is essentially a pointwise convergence mode. With the help of this theorem, we can give a neat proof for the first statement as follows:

Given any subsequence $\{X_{n_k} + Y_{n_k}\}$ of $\{X_n + Y_n\}$. Since $X_n \to_P X$, by the necessity part of the above theorem, there exists a set $\mathscr{K}_1$ of natural numbers, which is a subset of $\{n_k\}$ (this is a succinct way to express a further subsequence of some subsequence), such that $X_\ell \to X$ almost surely as $\ell \in \mathscr{K}_1$ and $\ell \to \infty$. Now consider the Y-subsequence $\{Y_\ell\}_{\ell \in \mathscr{K}_1}$, again by necessity, there exists $\mathscr{K}_2 \subset \mathscr{K}_1$ such that $Y_m \to Y$ almost surely as $m \in \mathscr{K}_2$ and $m \to \infty$. By construction, it follows that

$$X_m + Y_m \to X + Y \text{ almost surely}$$

as $m \in \mathscr{K}_2$ and $m \to \infty$. Also notice that $\{X_m + Y_m\}_{m \in \mathscr{K}_2}$ is obviously a subsequence of $\{X_{n_k} + Y_{n_k}\}$, therefore, by the sufficiency part of the theorem, we conclude that $X_n + Y_n \to_P X + Y$.

The second statement can be verified by replacing "+" with "·" without changing any other parts. Ultimately, such Cantor diagonal-flavor argument can prove the general

12

continuous mapping theorem: *with the same condition stated in this problem, and suppose that f is a continuous mapping of $\mathbb{R}^2$ into $\mathbb{R}^1$, then $f(X_n, Y_n) \to_P f(X, Y)$, which has extensive applications in asymptotic theories.* $\square$

For the first counterexample, fix X, Y i.i.d. $\sim \mathcal{N}(0, 1)$. Take $X_n \equiv X$ and $Y_n \equiv -X$. Clearly, we have $X_n \Rightarrow X$ and $Y_n \Rightarrow Y$ (where $\Rightarrow$ is an equivalent notation for $\to_d$). But $X_n + Y_n = 0$ doesn't converge in distribution to $X + Y$, which is distributed as $\mathcal{N}(0, 2)$. For the second counter example, fix U, V i.i.d. $\sim \mathcal{U}[-1, 1]$. Take $X_n \equiv U$ and $Y_n \equiv -U$, then $X_n Y_n = -U^2$. Clearly, $X_n \Rightarrow U$ and $Y_n \Rightarrow V$. But $X_n Y_n$ doesn't converge to UV. For if it were true, then it would hold $P[X_n Y_n \le 0] \to P[UV \le 0]$ as $n \to \infty$, since 0 is a continuity point of UV. However, $P[X_n Y_n \le 0] = P[-U^2 \le 0] = P[U^2 \ge 0] = 1$, while $P[UV \le 0] = \int_{-1}^{1} P[V \le 0] \frac{1}{2} du = \frac{1}{2} \times \frac{1}{2} \times 2 = \frac{1}{2}$, contradiction.

Problem 3 *Proof.* Let F_n and F denote the distribution functions of X_n and X, respectively.

Given $\varepsilon > 0$, choose $k \in \mathbb{N}$ sufficiently large such that $\frac{1}{k} < \frac{\varepsilon}{2}$. Since F is continuous on $\mathbb{R}$, by intermediate value theorem, there exist $-\infty = x_0 < x_1 < \cdots < x_{k-1} < x_k = +\infty$ such that $F(x_i) = \frac{i}{k}, i = 0, 1, \ldots, k$. For each $x \in \mathbb{R}$, there exists $i \in \{0, 1, \ldots, k-1\}$ such that $x \in (x_i, x_{i+1}]$ (of course, $x \neq x_k$). By monotonicity of F and F_n, we have

$$
\begin{aligned}
F_n(x) - F(x) \le & F_n(x_{i+1}) - F(x_i) = F_n(x_{i+1}) - F(x_{i+1}) + F(x_{i+1}) - F(x_i) \\
\le & \sup_{0 \le j \le k-1} |F_n(x_j) - F(x_j)| + \frac{1}{k}; \\
F_n(x) - F(x) \ge & F_n(x_i) - F(x_{i+1}) = F_n(x_i) - F(x_i) + F(x_i) - F(x_{i+1}) \\
\ge & -\sup_{0 \le j \le k-1} |F_n(x_j) - F(x_j)| - \frac{1}{k},
\end{aligned}
$$

which is equivalent to $|F_n(x) - F(x)| \le \sup_{0 \le j \le k-1} |F_n(x_j) - F(x_j)| + \frac{1}{k}$. Since this inequality holds for each $x \in \mathbb{R}$, and the upper bound is independent of x, it follows that

$$
\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \le \sup_{0 \le j \le k-1} |F_n(x_j) - F(x_j)| + \frac{1}{k}.
$$

Since $F_n \Rightarrow F$, and F is continuous everywhere, there exists $N \in \mathbb{N}$ such that $\sup_{0 \leq j \leq k-1} |F_n(x_j) - F(x_j)| < \frac{\varepsilon}{2}$ for all $n > N$. Hence for each $n > N$,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \sup_{0 \leq j \leq k-1} |F_n(x_j) - F(x_j)| + \frac{1}{k} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The proof is complete. □

2.1 (a) *Proof.* Given $\varepsilon > 0$, $P[|X_n - 1| \geq \varepsilon] = P[|nY_n| \geq \varepsilon] = P[Y_n \geq \varepsilon/n] = P[Y_n = 1] = \frac{1}{n} \to 0$ as $n \to \infty$, thus $Y_n \to_P 1$. □

(b) *Proof.* Let $Z_n = Y_n - \sum_{i=1}^{n} \frac{1}{i}$ so that $E[Z_n] = 0$ and $\text{Var}(Z_n) = \text{Var}(Y_n) = \sum_{i=1}^{n} \frac{1}{i}$. We shall use the following celebrated asymptotic representation for $\sum_{i=1}^{n} \frac{1}{i}$:

$$\sum_{i=1}^{n} \frac{1}{i} = \log n + \gamma + \varepsilon_n,$$

where γ is a constant, called *Euler constant*, and $\varepsilon_n = o(1)$. Given $\varepsilon > 0$, we have

$$P[|X_n - 1| \geq \varepsilon] = P\left[\left|\frac{Y_n}{\log n} - 1\right| \geq \varepsilon\right] = P[|Y_n - \log n| \geq \varepsilon \log n]$$

$$= P\left[\left|Z_n + \sum_{i=1}^{n} \frac{1}{i} - \left(\sum_{i=1}^{n} \frac{1}{i} - \gamma - \varepsilon_n\right)\right| \geq \varepsilon \log n\right]$$

$$= P[|Z_n + (\gamma + \varepsilon_n)| \geq \varepsilon \log n]$$

$$\leq P[|Z_n| + |\gamma + \varepsilon_n| \geq \varepsilon \log n]$$

$$\leq P\left[|Z_n| \geq \frac{1}{2}\varepsilon \log n\right] + P\left[|\gamma + \varepsilon_n| \geq \frac{1}{2}\varepsilon \log n\right].$$

For sufficiently large n, the second term in the right hand side of the above expressions is zero since $|\gamma + \varepsilon_n| = o(\log n)$. On the other hand, by Chebyshev's inequality,

$$P\left[|Z_n| \geq \frac{1}{2}\varepsilon \log n\right] = P\left[|Z_n - E[Z_n]| \geq \frac{1}{2}\varepsilon \log n\right]$$

$$\leq \frac{4\,\text{Var}(Z_n)}{\varepsilon^2 \log^2 n} = \frac{4\sum_{i=1}^{n} \frac{1}{i}}{\varepsilon^2 \log^2 n} = \frac{4(\log n + \gamma + \varepsilon_n)}{\varepsilon^2 \log^2 n} \to 0$$

14

as $n \to \infty$. Therefore, $X_n \to_P 1$. $\qquad \square$

(c) *Proof.* Since $Y_1^2, \ldots, Y_n^2$ are i.i.d. with $E(Y_1^2) = \mathrm{Var}(Y_1) = 1$, by weak law of large numbers, we have $X_n = \bar{Y}_n \to_P 1$. $\qquad \square$

2.5 (a) *Proof.* Since $E[(X_n - 1)^2] = E[(nY_n)^2] = n^2 E[Y_n^2] = n^2 \times \frac{1}{n} = n \to \infty$ as $n \to \infty$, X_n doesn't converge to 1 in quadratic mean. $\qquad \square$

(b) *Proof.* By Theorem 2.17(a), it suffices to show that $E[X_n] \to 1$ and $\mathrm{Var}(X_n) \to 0$ as $n \to \infty$, which have been shown during the course of proving Exercise 2.1(b). Therefore, $X_n \overset{\mathrm{qm}}{\to} 1$. $\qquad \square$

(c) *Proof.* Since $\sum_{i=1}^n X_i^2 \sim \chi_n^2$, we have $E[Y_n] = 1$ and $\mathrm{Var}(Y_n) = \frac{2}{n} \to 0$ as $n \to \infty$, by Theorem 2.17(a), $X_n \overset{\mathrm{qm}}{\to} 1$. $\qquad \square$

2.8 The statement is true and the proof goes as follows:

Proof. By condition, there exists M such that $P[|X_n| \geq M/2] = 0$ for all n and $|c| < \frac{1}{2}M$. Thus for each $n \in \mathbb{N}$,

$$P[|X_n - c| \geq M] \leq P[|X_n| + |c| \geq M] \leq P[|X_n| \geq M/2] + P[|c| \geq M/2] = 0.$$

Since for each nonnegative random variable Y, we have $E[Y] = \int_0^\infty P[Y \geq t] dt$ (why?), hence

$$E[(X_n - c)^2] = \int_0^\infty P[(X_n - c)^2 \geq t] dt = \int_0^\infty P[|X_n - c| \geq \sqrt{t}] dt$$
$$= \int_0^{M^2} P[|X_n - c| \geq \sqrt{t}] dt$$

in light of $P[|X_n - c| \geq x] \leq P[|X_n - c| \geq M] = 0$ for each $x \geq M$. Now since $X_n \to_P c$, for each fixed $t \in [0, M^2]$, $P[|X_n - c| \geq \sqrt{t}] \to 0$ as $n \to \infty$. The result therefore follows by Lebesgue's dominated convergence theorem, since for every n, $P[|X_n - c| \geq \sqrt{t}]$ is bounded by 1 and $\int_0^{M^2} 1 dt = M^2 < \infty$. $\qquad \square$

2.10 (a) By independence assumption,

$$E[X_i] = c_i E[Y_i] E[2Z_i - 1] = c_i p_i(1-1) = 0;$$

$$E[X_i^2] = c_i^2 E[Y_i^2] E[4Z_i^2 - 4Z_i + 1] = c_i^2 p_i.$$

Hence $\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} E[X_i^2] = \frac{1}{n^2} \sum_{i=1}^{n} c_i^2 p_i$.

(b) *Proof.* Since $E[|X_i|] = c_i E[|Y_i|] E[|2Z_i - 1|] = c_i E[Y_i] = c_i p_i$, by triangle inequality,

$$E[|\overline{X}_n - 0|] = \frac{1}{n} E\left[\left|\sum_{i=1}^{n} X_i\right|\right] \leq \frac{1}{n} \sum_{i=1}^{n} E[|X_i|] = \frac{1}{n} \sum_{i=1}^{n} c_i p_i \to 0$$

as $n \to \infty$ by condition. Since convergence in mean implies convergence in probability, the result follows. $\square$

(c) Take $c_i = i^2$ and $p_i = \frac{1}{i^3}$, $i = 1, 2, \ldots$, then

$$\frac{1}{n} \sum_{i=1}^{n} c_i p_i = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{i} \sim \frac{\log n}{n} \to 0;$$

$$\frac{1}{n^2} \sum_{i=1}^{n} c_i^2 p_i = \frac{1}{n^2} \sum_{i=1}^{n} i = \frac{n(n+1)}{2n^2} \to \frac{1}{2}.$$

Therefore Condition (2.21) holds but $\mathrm{Var}(\overline{X}_n)$ does not converge to zero.

2.14 (a) *Proof.* It is clear that $\hat{\beta}_{0n} = \overline{Y} - \bar{z}_n \hat{\beta}_{1n}$. Direct calculation gives that (notice that $\sum_{i=1}^{n} w_i^{(n)} = 0$)

$$E[\hat{\beta}_{1n}] = \sum_{i=1}^{n} w_i^{(n)}(\beta_0 + \beta_1 z_i + 0) = \beta_1 \sum_{i=1}^{n} \frac{z_i(z_i - \bar{z}_n)}{\sum_{j=1}^{n}(z_j - \bar{z}_n)^2}$$

$$= \beta_1 \sum_{i=1}^{n} \frac{(z_i - \bar{z}_n)(z_i - \bar{z}_n)}{\sum_{j=1}^{n}(z_j - \bar{z}_n)^2} = \beta_1.$$

Thus $\hat{\beta}_{1n}$ is unbiased. Consequently, $E[\hat{\beta}_{0n}] = E[\overline{Y}] - \bar{z}_n E[\hat{\beta}_{1n}] = \beta_0 + \beta_1 \bar{z}_n - \beta_1 \bar{z}_n = \beta_0$, i.e., $\hat{\beta}_{0n}$ is unbiased. $\square$

(b) *Proof.* Using $\sum_{i=1}^{n} w_i^{(n)} = 0$ and $\sum_{i=1}^{n} w_i^{(n)} z_i = 1$, we can rewrite β_1 as $\beta_1 = \sum_{i=1}^{n} w_i^{(n)}(\beta_0 + \beta_1 z_i)$. Thus $\hat{\beta}_{1n} - \beta_1 = \sum_{i=1}^{n} w_i^{(n)}(\beta_0 + \beta_1 z_i + \varepsilon_i) - \sum_{i=1}^{n} w_i^{(n)}(\beta_0 + $

$\beta_1 z_i) = \sum_{i=1}^n w_i^{(n)} \varepsilon_i$. It then follows that

$$\mathrm{Var}(\hat{\beta}_{1n} - \beta_1) = \sum_{i=1}^n w_i^{(n)2} \sigma^2 = \sigma^2 \frac{1}{\sum_{j=1}^n (z_j - \bar{z}_n)^2}.$$

Provided condition 2.18, $\mathrm{Var}(\hat{\beta}_{1n} - \beta_1) \to 0$ as $n \to \infty$, a fortiori, $\hat{\beta}_{1n} - \beta_1 \to_P 0$,

i.e., $\hat{\beta}_{1n}$ is consistent.

Since $\hat{\beta}_{0n} - \beta_0 = \bar{Y} - \bar{z}_n \hat{\beta}_{1n} - \beta_0 = -\bar{z}_n(\hat{\beta}_{1n} - \beta_1) + \bar{\varepsilon}_n$ and $\bar{\varepsilon}_n = o_P(1)$ by WLLN,

it suffices to show that $\mathrm{Var}(\bar{z}_n(\hat{\beta}_{1n} - \beta_1)) \to 0$ as $n \to \infty$, which is exactly the

condition 2.17. The proof is complete. $\qquad\square$

Homework 4

Problem 1 *Proof.* The main tool we shall use for this proof are two well-known inequalities: for any two random variables X and Y which have rth order finite moments, we have

c_r *inequality*

$$E[|X + Y|^r] \leq c_r (E[|X|^r] + E[|Y|^r]),$$

where

$$c_r = \begin{cases} 1 & 0 < r \leq 1, \\ 2^{r-1} & r > 1. \end{cases}$$

Minkowski's inequality

$$E^{\frac{1}{r}}[|X + Y|^r] \leq E^{\frac{1}{r}}[|X|^r] + E^{\frac{1}{r}}[|Y|^r], \quad r \geq 1.$$

Since $X_n \to_{L^r} X$, by definition, $E[|X|^r] < \infty$. If $0 < r \leq 1$, by c_r inequality and $E[|X_n - X|^r] \to 0$ as $n \to \infty$, it follows that for sufficiently large n,

$$E[|X_n|^r] = E[|(X_n - X) + X|^r] \leq E[|X_n - X|^r] + E[|X|^r] < \infty.$$

Interchanging the roles of X_n and X in the above expression, then $E[|X|^r] \leq E[|X_n - X|^r] + E[|X_n|^r]$. Therefore

$$|E[|X_n|^r] - E[|X|^r]| \leq E[|X_n - X|^r] \to 0$$

as $n \to \infty$, i.e., $E[|X_n|^r] \to E[|X|^r]$ as $n \to \infty$.

If $r > 1$, by Minkowski's inequality,

$$E^{\frac{1}{r}}[|X_n|^r] = E^{\frac{1}{r}}[|(X_n - X) + X|^r] \leq E^{\frac{1}{r}}[|X_n - X|^r] + E^{\frac{1}{r}}[|X|^r].$$

Switching the roles of X_n and X above to obtain

$$E^{\frac{1}{r}}[|X|^r] \le E^{\frac{1}{r}}[|X_n - X|^r] + E^{\frac{1}{r}}[|X_n|^r].$$

Thus

$$\left| E^{\frac{1}{r}}[|X_n|^r] - E^{\frac{1}{r}}[|X|^r] \right| \le E^{\frac{1}{r}}[|X_n - X|^r] \to 0$$

as $n \to \infty$. Hence $E^{\frac{1}{r}}[|X_n|^r] \to E^{\frac{1}{r}}[|X|^r]$ as $n \to \infty$, the result then follows by noting that the function $x \mapsto x^r$ is continuous on $(0, +\infty)$. $\qquad \square$

Problem 2 *Proof.*

- $X_n \xrightarrow{L^s} X \Rightarrow X_n \xrightarrow{L^r} X$: Since for $s > r \ge 1$, we have (which may be termed as the monotonicity of the L^p norm):

$$E^{\frac{1}{r}}[|X_n - X|^r] \le E^{\frac{1}{s}}[|X_n - X|^s],$$

  the result follows.

- $X_n \xrightarrow{L^r} X \Rightarrow X_n \xrightarrow{P} X$: This can be easily seen by Markov's inequality:

$$P[|X_n - X| \ge \varepsilon] = P[|X_n - X|^s \ge \varepsilon^s] \le \frac{E[|X_n - X|^s]}{\varepsilon^s}.$$

- $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$: Let F_n and F denote the distribution functions of X_n and F respectively. Given $\varepsilon > 0$, let x be any continuity point of F. If $X_n \le x$, then either $X \le x + \varepsilon$ or $|X_n - X| > \varepsilon$. In other words, $[\omega : X_n(\omega) \le x] \subset [\omega : X(\omega) \le x + \varepsilon] \cup [\omega : |X_n(\omega) - X(\omega)| > \varepsilon]$. Hence

$$F_n(x) = P[X_n \le x] \le P[X \le x + \varepsilon] + P[|X_n - X| > \varepsilon].$$

  Similarly, noting that $[\omega : X(\omega) \le x - \varepsilon] \subset [\omega : X_n(\omega) \le x] \cup [\omega : |X_n(\omega) - X(\omega)| >$

ε], it follows that

$$P[X \le x - \varepsilon] \le P[X_n \le x] + P[|X_n - X| > \varepsilon].$$

Since $X_n \overset{P}{\to} X$, $P[|X_n - X| > \varepsilon] \to 0$ as $n \to \infty$. Thus the above two arrays of inequalities give that

$$F(x - \varepsilon) \le \liminf_{n \to \infty} F_n(x) \le \limsup_{n \to \infty} F_n(x) \le F(x + \varepsilon).$$

Since x is a continuity point of F, by letting $\varepsilon \downarrow 0$, we conclude that $\lim_{n \to \infty} F_n(x) = F(x)$, i.e., $X_n \overset{d}{\to} X$.

- $X_n \overset{a.s.}{\to} X \Rightarrow X_n \overset{P}{\to} X$: It is helpful to express the set $A = [\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)]$ as follows:

$$
\begin{aligned}
&[\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)] \\
=&[\omega : \forall \varepsilon > 0, \exists N_\varepsilon, |X_n(\omega) - X(\omega)| < \varepsilon, \forall n \ge N] \\
=&\bigcap_{\varepsilon > 0} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} [\omega : |X_n(\omega) - X(\omega)| < \varepsilon].
\end{aligned}
$$

By condition we have $P(A) = 1$, which implies that $P[|X_n - X| < \varepsilon] \to 1$ as $n \to \infty$ for as long as ε is fixed, the set $\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} [\omega : |X_n(\omega) - X(\omega)| < \varepsilon]$ contains A.

The counterexamples are given as follows:

- $X_n \overset{d}{\to} X \nRightarrow X_n \overset{P}{\to} X$: Let $X_n \equiv X \sim$ Bernoulli(0.5), $Y \sim$ Bernoulli(0.5) that is independent of X. Clearly, $X_n \overset{d}{\to} Y$, but X_n does not converge to Y in probability, since

$$P[|X_n - Y| \ge 1] = P[|X - Y| \ge 1] = P[X = 1, Y = 0] + P[Y = 1, X = 0] = 0.5,$$

which does not converge to 0 as $n \to \infty$.

20

- $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{a.s.} X$: This counterexample is classical. Let the probability space be $((0,1], \mathscr{B}(0,1], \lambda)$, where $\mathscr{B}(0,1]$ is the Borel σ-field on $(0,1]$ and λ is the Lebesgue measure. Let $X_1 = 1$, $X_2 = I_{(0,1/2]}(\omega)$, $X_3 = I_{(1/2,1]}(\omega)$, $X_4 = I_{(0,1/4]}(\omega)$, $X_5 = I_{(1/4,1/2]}(\omega), \ldots$. In general, if $n = 2^k + m$, where $0 \leq m < 2^k$ and $k \geq 0$, then $X_n = I_{(m2^{-k}, (m+1)2^{-k}]}(\omega)$. By construction, since X_n is moving over $(0,1]$ endlessly, it does not converge at any point $\omega \in (0,1]$ (my first probability teacher described this function as "a shrinking travelling dark cloud"). Therefore, $X_n \xrightarrow{a.s.} 0$. However, the measure (length) of the set it does not converge to 0 tends to 0, since for any given $\varepsilon > 0$, there exists k such that $2^{-k} < \varepsilon$. Then for every $\delta > 0$, for $n \geq 2^k$, we have

$$P[|X_n - 0| \geq \delta] \leq \frac{1}{2^k} < \varepsilon.$$

  That means $X_n \xrightarrow{P} 0$.

- $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{L^1} X$: Set the probability space as above. Define $X_n(\omega) = nI_{(0,n^{-1})}(\omega)$, where I denotes the indicator function. For any fixed $\omega \in (0,1]$, there exists sufficiently large n such that $n^{-1} < x$, hence $X_n \to 0$ everywhere, implying that $X_n \xrightarrow{P} 0$. However, $E[|X_n - 0|] = n \times n^{-1} = 1 \not\to 0$ as $n \to \infty$, i.e., $X_n \not\to_{L^1} 0$.

- $X_n \xrightarrow{L^r} X \not\Rightarrow X_n \xrightarrow{L^s} X$: Set the probability space as above. Define $X_n(\omega) = \sqrt{n}I_{(0,n^{-1})}(\omega)$. It is easy to check that $E[|X_n - 0|] = \frac{1}{\sqrt{n}} \to 0$ as $n \to \infty$ but $E[|X_n - 0|^2] = 1 \not\to 0$ as $n \to \infty$. Hence convergence in mean does not imply convergence in quadratic mean.

- $X_n \xrightarrow{a.s.} X \not\Rightarrow X_n \xrightarrow{L^1} X$: Use the same counterexample as the third item.

- $X_n \xrightarrow{L^1} X \not\Rightarrow X_n \xrightarrow{a.s.} X$: Use the same counterexample as the second item. For $n = 2^k + m$ and all $r > 0$, $E[|X_n - 0|^r] = \frac{1}{2^{kr}} \to 0$ as $k \to \infty$. Hence $X_n \xrightarrow{L^r} 0$ for all $r > 0$, including the case $r = 1$.

$\square$

2.15 (a) *Proof.* Given $\varepsilon > 0$, choose b, c, the partition $\{t_0, \ldots, t_m\}$ and the approximation function h as the hint suggests. Since $X_n \Rightarrow X$ and $t_0, \ldots, t_m$ are all continuity points of F, we have $F_n(t_i) \to F(t_i)$ as $n \to \infty$ for each $i \in \{0, \ldots, m\}$, that is, there exists $N \in \mathbb{N}$, such that $\sup\limits_{0 \le i \le m} |F_n(t_i) - F(t_i)| < \frac{\varepsilon}{2Mm}$ for all $n > N$, where M is such that $|g(x)| \le M$ for all $x \in \mathbb{R}$. It then follows for each $n > N$,

$$
\begin{aligned}
&|E[h(X_n)] - E[h(X)]| \\
=&\left| \sum_{i=1}^{m} g(t_i) P[t_{i-1} < X_n \le t_i] - \sum_{i=1}^{m} g(t_i) P[t_{i-1} < X \le t_i] \right| \\
=&\left| \sum_{i=1}^{m} g(t_i)[(F_n(t_i) - F(t_i)) + (F_n(t_{i-1}) - F(t_{i-1}))] \right| \\
\le& \sum_{i=1}^{m} |g(t_i)||F_n(t_i) - F(t_i)| + \sum_{i=1}^{m} |g(t_i)||F_n(t_{i-1}) - F(t_{i-1})| \\
\le& M \sum_{i=1}^{m} \frac{\varepsilon}{2Mm} + M \sum_{i=1}^{m} \frac{\varepsilon}{2Mm} = \varepsilon.
\end{aligned}
$$

This completes the proof. $\qquad\square$

(b) *Proof.* By the decomposition given in hint and the result of part (a), it remains to show that $E[g(X_n)] - E[h(X_n)]$ can be bounded by arbitrarily small positive ε for all sufficiently large n, and $|E[g(X)] - E[h(X)]|$ can be bounded by small ε. Write

$$
\begin{aligned}
&|E[g(X_n)] - E[h(X_n)]| \\
=&|E[g(X_n)I_{[b \le X_n \le c]}] - E[h(X_n)I_{[b \le X_n \le c]}] + E[g(X_n)I_{[X_n < b] \cup [X_n > c]}] \\
&\quad - E[h(X_n)I_{[X_n < b] \cup [X_n > c]}]| \\
\le&|E[g(X_n)I_{[b \le X_n \le c]}] - E[h(X_n)I_{[b \le X_n \le c]}]| + 4M\varepsilon \\
\le& \sum_{i=1}^{m} \varepsilon |F_n(t_i) - F(t_i)| + \sum_{i=1}^{m} \varepsilon |F_n(t_{i-1}) - F(t_{i-1})| + 4M\varepsilon \\
<& C\varepsilon
\end{aligned}
$$

for some positive constant C whenever n is large enough. Similarly, we can show

the second assertion. This is somewhat a sketch of proof, which is not hard to extend to a rigorous proof so that in the rightmost you will get ε exactly. $\square$

2.16 (a) *Proof.* Given $\varepsilon > 0$, since $t \in C(F)$, there exists $\delta_0 > 0$ such that $|x - t| < \delta_0$ implies that $|F(x) - F(t)| < \varepsilon$. Choose any $\delta \in (0, \delta_0)$, then $0 \le F(t+\delta) - F(t) < \varepsilon$ and $0 \le F(t) - F(t - \delta) < \varepsilon$, which imply $F(t + \delta) < F(t) + \varepsilon$ and $F(t - \delta) > F(t) - \varepsilon$. $\square$

(b) *Proof.* Fix $x \in C(F)$, suppose that $y > x$. Let $f(t)$ be 1 for $t \le x$, 0 for $t \ge y$, and interpolate linearly on $[x, y]$: $f(t) = (y - t)/(y - x)$ for $x \le t \le y$. Since $F_n(x) \le E[f(X_n)]$ and $E[f(X)] \le F(y)$, it follows from the condition that $\limsup_{n \to \infty} F_n(x) \le F(y)$; letting $y \downarrow x$ shows that $\limsup_{n \to \infty} F_n(x) \le F(x)$. Similarly, $F(u) \le \liminf_n F_n(x)$ for $u < x$ and hence $F(x-) \le \liminf_n F_n(x)$. This implies convergence at continuity points. $\square$

2.17 *Proof.* Fix $Z \sim \mathcal{N}(0, 1)$. For each $n \in \mathbb{N}$, define $X_n = Z$, $Y_n = Z + U/n$, where $U \sim \mathcal{U}(-1, 1)$ is taken to be independent of Z.[2] We shall show that:

(i) Every point of $\mathbb{R}^2$ is a continuity point of the distribution function of (Z, Z).

(ii) $(X_n, Y_n) \Rightarrow (Z, Z)$;

(iii) Let $S := \{(x, y) : y = x\}$, i.e., the straight line with slope 1 passing the origin. Then $P[(X_n, Y_n) \in S] \not\to P[(Z, Z) \in S]$. Obviously, S is a connected set in $\mathbb{R}^2$.

To show (i), let's first find the distribution function $F(x, y)$ of (Z, Z). By definition, $F(x, y) = P[(Z, Z) \le (x, y)] = P[Z \le x, Z \le y] = P[Z \le \min(x, y)] = \Phi(\min(x, y))$, where $\Phi(\cdot)$ is the distribution function of Z. Since the mappings $(x, y) \mapsto \min(x, y)$ is continuous on $\mathbb{R}^2$, and $z \mapsto \Phi(z)$ is continuous on $\mathbb{R}^1$, their composition $\Phi(\min(x, y))$ is continuous everywhere on $\mathbb{R}^2$ as well ([5], Theorem 4.7).

To show (ii), because of (i), it suffices to show that as $n \to \infty$, $P[X_n \le x, Y_n \le y] \to \Phi(\min(x, y))$ for every $(x, y) \in \mathbb{R}^2$. Let's first consider the case $x < y$. Then for all

_____
[2]Here is another simpler (yet might be too trivial) counterexample: $X_n = Z, Y_n = Z + 1/n, n = 1, 2, \ldots$.

sufficiently large n, we have $x < y - u/n$ for each $u \in (-1,1)$. It then follows by the independence of Z and U that

$$P[X_n \le x, Y_n \le y] = P[Z \le x, Z + U/n \le y]$$
$$= \int_{-1}^{1} P[Z \le x, Z \le y - u/n] \times \frac{1}{2} \, du = \frac{1}{2} \int_{-1}^{1} \Phi(x) \, du = \Phi(x).$$

The case $x > y$ can be handled similarly. If $x = y$, then

$$P[X_n \le x, Y_n \le x] = P[Z \le x, Z + U/n \le x]$$
$$= \int_{-1}^{1} P[Z \le x, Z \le x - u/n] \times \frac{1}{2} \, du$$
$$= \frac{1}{2} \int_{-1}^{0} P[Z \le x, Z \le x - u/n] \, du + \frac{1}{2} \int_{0}^{1} P[Z \le x, Z \le x - u/n] \, du$$
$$= \frac{1}{2} \int_{-1}^{0} \Phi(x) \, du + \frac{1}{2} \int_{0}^{1} \Phi(x - u/n) \, du$$
$$\rightarrow \Phi(x)$$

as $n \to \infty$. The last step follows from the dominated convergence theorem in view of $\Phi(\cdot)$ is bounded by 1. In summary, $(X_n, Y_n) \Rightarrow (Z, Z)$.

Lastly, let's show (iii). For each fixed $n \in \mathbb{N}$, we have $P[(X_n, Y_n) \in S] = P[X_n = Y_n] = P[Z = Z + U/n] = P[U = 0] = 0$. On the other hand, $P[(Z, Z) \in S] = 1$. Therefore $P[(X_n, Y_n) \in S]$ clearly does not converge to $P[(Z, Z) \in S]$. $\square$

2.21 *Proof.* For simplicity, assume both X_n and Y_n are univariate. Let $X_n \equiv Y_n = Y \sim \mathcal{N}(0,1)$, and $X \sim \mathcal{N}(0,1)$ be independent of Y. Clearly, we have $X_n \Rightarrow X$, $Y_n \rightarrow_P Y$. However, $(X_n, Y_n) = (Y, Y) \not\Rightarrow (X, Y)$, for $P[(Y, Y) \le (0,0)] = P[Y \le 0] = \frac{1}{2} \not\rightarrow P[(X, Y) \le (0,0)] = P[X \le 0]P[Y \le 0] = \frac{1}{4}$. $\square$

2.24 *Proof.*

(a) It is worth pointing out that *the hint is wrong*! As a counterexample, define F as

follows:

$$F(x) = \begin{cases} 0 & x < -1; \\ 1 - \frac{1}{i} & -\frac{1}{i} \leq x < -\frac{1}{i+1}, i = 1, 2, \ldots; \\ 1 & x \geq 0. \end{cases}$$

Since $\lim_{x \to 0} F(x) = F(0) = 1$, 0 is a continuity point of F. However, any neighborhood of 0 contains a (in fact, infinitely many) jump point of F. Therefore, the statement made in the hint is overly strong. Instead of using this wrong statement, notice that the continuity points of F are *dense* in $\mathbb{R}^k$, so we may continue the proof by manipulating the inequalities given in the hint. The proof of the dense statement needs some measure theory, so I will omit it here. For those who are interested in this result, see page 260 of [2].

For an alternative proof for this lemma, see the proof of Theorem 6 in Section 6 of [3], which is one of the optional textbooks listed in syllabus.

(b) By the above lemma, it suffice to show that $(X_n^T, c^T)^T \Rightarrow (X^T, c^T)^T$ and $(0^T, (Y_n - c)^T)^T \to_P (0^T, 0^T)^T$. For clarity, assume $X_n \in \mathbb{R}^n$ and $Y_n \in \mathbb{R}^m$. To show that $(0^T, (Y_n - c)^T)^T \to_P (0^T, 0^T)^T$, we check the definition of convergence in probability. For given $\varepsilon > 0$, since $Y_n \to_P c$, it follows that

$$P[\|(0^T, (Y_n - c)^T)^T - (0, 0)^T\|_{\mathbb{R}^{n+m}} \geq \varepsilon] = P[\|Y_n - c\|_{\mathbb{R}^m} \geq \varepsilon] \to 0$$

as $n \to \infty$, where $\|\cdot\|_{\mathbb{R}^i}$ denotes the Euclidean distance on $\mathbb{R}^i$. Therefore $(0^T, (Y_n - c)^T)^T \to_P (0^T, 0^T)^T$.

We may also verify the weak convergence by original definition. Alternatively, a more compact way is to use the equivalent characterization (Theorem 2.28) for weak convergence. To show $(X_n^T, c^T) \Rightarrow (X^T, c^T)$, by the sufficiency part of Theorem 2.28, we need to show for any given bounded and continuous function g mapping $\mathbb{R}^{n+m}$ into $\mathbb{R}^1$, it holds that $E[g(X_n, c)] \to E[g(X, c)]$ as $n \to \infty$. It is straightforward to verify that given g is bounded and continuous on $\mathbb{R}^{n+m}$, then $f(x) = g(x, c)$ is a bounded

25

and continuous function mapping $\mathbb{R}^n$ to $\mathbb{R}$, thus $E[f(X_n)] \to E[f(X)]$ as $n \to \infty$ for $X_n \Rightarrow X$ and the necessity part of Theorem 2.28. Noting that by construction, $E[f(X_n)] = E[g(X_n, c)]$, $E[f(X)] = E[g(X, c)]$, hence the result follows. $\qquad\square$

Homework 5

Problem 1 (a) *Proof.* First assume $a \in \mathbb{R}$. Given $\varepsilon > 0$, since $a_n \to a$ as $n \to \infty$, there exists $N \in \mathbb{N}$ such that $n > N$ implies that $|a_n - a| < \varepsilon/2$. Let $M = \max_{1 \leq i \leq N} \{|a_i - a|\}$. It follows that for each $n > N \vee 2NM/\varepsilon$ (*remark:* $\vee$ is an extensively used notation in probability theory, for two real numbers x and y, $x \vee y := \max\{x, y\}$),

$$
\begin{aligned}
|A_n - a| &= \left| \frac{1}{n} \sum_{i=1}^{n} a_i - a \right| \\
&= \left| \frac{1}{n} \sum_{i=1}^{n} (a_i - a) \right| \\
&\leq \frac{1}{n} \sum_{i=1}^{N} |a_i - a| + \frac{1}{n} \sum_{i=N+1}^{n} |a_i - a| \\
&\leq \frac{NM}{n} + \frac{n-N}{n} \frac{\varepsilon}{2} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}
$$

This shows that if $a \in \mathbb{R}$, $A_n \to a$ as $n \to \infty$.

This result also holds if $a = +\infty$ or $a = -\infty$. The above proof can be adapted *mutatis mutandis* for these two cases. □

(b) *Proof.* Define $S_n = \sum_{i=1}^{n} a_i$ to be the partial sum sequence for $\{a_n\}$. The convergence of series by definition amounts to that $\lim_{n \to \infty} S_n = a \in \mathbb{R}$. Therefore, for given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $|S_n - a| < \varepsilon$ for all $n > N$. It then follows that for each $n > N + 1$,

$$
|B_n - 0| = |a - S_{n-1}| < \varepsilon.
$$

The proof is thus complete. □

(c) *Proof.* For this problem, we shall apply an important technique called *Abel's summation by parts formula*, which can be found from Theorem 3.41 in [5]. I will also derive it here as one part of our proof.

Let $A_n = a_1 + \cdots + a_n, n = 1, 2, \ldots$. Since $a_1 = A_1, a_n = A_n - A_{n-1}, n \geq 2$, it follows that

$$\sum_{i=1}^{n} a_i b_i = a_1 b_1 + \cdots + a_n b_n$$

$$= A_1 b_1 + (A_2 - A_1) b_2 + (A_3 - A_2) b_3 + \cdots + (A_{n-1} - A_{n-2}) b_{n-1} + (A_n - A_{n-1}) b_n$$

$$= A_1(b_1 - b_2) + A_2(b_2 - b_3) + \cdots + A_{n-1}(b_{n-1} - b_n) + A_n b_n$$

$$= \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) + A_n b_n.$$

Since $A_n \to a$ and $b_n \uparrow +\infty$ as $n \to \infty$, to show that $C_n = \dfrac{1}{b_n} \sum_{i=1}^{n} a_i b_i = \dfrac{1}{b_n} \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) + A_n \to 0$, it is equivalent to show that $\dfrac{1}{b_n} \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) \to -a$ as $n \to \infty$. Given $\varepsilon > 0$, since $A_n \to a$ as $n \to \infty$, there exists $N \in \mathbb{N}$ such that $n > N$ implies that $|A_n - a| < \varepsilon/3$. Denote $\max_{1 \leq i \leq N} \{|A_i - a|\}$ by M. Since $b_n \to +\infty$, there exists $N' \in \mathbb{N}$ such that $n > N'$ implies that $|ab_1/b_n| < \varepsilon/3$ and $M(b_{N+1} - b_1)/b_n < \varepsilon/3$. Hence, by noting the monotonicity of sequence $\{b_n\}$, for each $n > N \vee N' + 1$, we have

$$\left| \frac{1}{b_n} \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) - (-a) \right|$$

$$= \left| \frac{1}{b_n} \sum_{i=1}^{n-1} A_i(b_i - b_{i+1}) + \frac{1}{b_n} \sum_{i=1}^{n-1} a(b_{i+1} - b_i) + \frac{ab_1}{b_n} \right|$$

$$\leq \left| \frac{1}{b_n} \sum_{i=1}^{n-1} (a - A_i)(b_{i+1} - b_i) \right| + \left| \frac{ab_1}{b_n} \right|$$

$$\leq \left| \frac{1}{b_n} \sum_{i=1}^{N} (a - A_i)(b_{i+1} - b_i) \right| + \left| \frac{1}{b_n} \sum_{i=N+1}^{n-1} (a - A_i)(b_{i+1} - b_i) \right| + \left| \frac{ab_1}{b_n} \right|$$

$$\leq \frac{M(b_{N+1} - b_1)}{b_n} + \frac{\varepsilon}{3} \cdot \frac{b_n - b_{N+1}}{b_n} + \frac{\varepsilon}{3}$$

$$< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

This completes the proof. $\qquad \square$

Problem 2 The *Etemadi's maximal inequality* is Theorem 22.5 in [2].

3.3 *Proof.* As the hint suggests and recall what we have done in Homework 1, $[B_n \text{ i.o.}] = \limsup_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k$, it then follows that for each $n \in \mathbb{N}$,

$$P[B_n \text{ i.o.}] = P\left[\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} B_k\right] \leq P\left[\bigcup_{k=n}^{\infty} B_k\right] \leq \sum_{k=n}^{\infty} P(B_k).$$

Let $n \to \infty$ on the right hand side of the above expression and use $\sum_{n=1}^{\infty} P(B_n) < \infty$, we conclude that $P[B_n \text{ i.o.}] = 0$. $\qquad\square$

3.4 (a) *Proof.* Denote $X_1 + X_2 + \cdots + X_n$ by S_n, it is easily seen that

$$S_n^4 = \sum X_\alpha X_\beta X_\gamma X_\delta. \tag{5}$$

Depending on how the indices match up, each term in this sum reduces to one of the following five forms, where in each case the indices are now *distinct*:

$$\begin{cases} X_i^4, \\ X_i^2 X_j^2, \\ X_i^2 X_j X_k, \\ X_i^3 X_j, \\ X_i X_j X_k X_l. \end{cases}$$

By independence assumption, all terms with the latter three forms have 0 expected value. The number of occurrences in the sum (5) of the first form is n. The number of occurrences of the second form is $3n(n-1)$, because there are n choices for the α in (5), three ways to match it with β, γ or δ, and $n-1$ choices for the value common to the remaining two indices. Therefore

$$E[S_n^4] = n\mu_4 + 3n(n-1)\mu_2^2,$$

where $\mu_k = E[X_1^k], k = 2, 4$ (it follows by Schwarz inequality that if $\mu_4 < \infty$, then $\mu_2 < \infty$). $\qquad\square$

(b) *Proof.* By Markov's inequality, for any given $\varepsilon > 0$,

$$P[|\overline{X}_n| \geq \varepsilon] = P[|S_n| \geq n\varepsilon] \leq \frac{E[S_n^4]}{n^4\varepsilon^4} = \frac{n\mu_4 + 3n(n-1)\mu_2^2}{n^4\varepsilon^4} \leq \frac{M}{n^2}$$

for some positive constant M that is independent of n. □

(c) *Proof.* By the result of part (b), we have

$$\sum_{n=1}^{\infty} P[|\overline{X}_n| \geq \varepsilon] \leq M \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

It then follows by Lemma 3.9 that $\overline{X}_n \to_{a.s.} 0$. □

3.7 (a) *Proof.* Without loss of generality, we may assume $\mu = 0$. Let $S_n = X_1 + \cdots + X_n$.

$$\max_{2^{k-1} \leq n < 2^k} |\overline{X}_n| = \max_{2^{k-1} \leq n < 2^k} \frac{|S_n|}{n} \leq \max_{2^{k-1} \leq n < 2^k} \frac{|S_n|}{2^{k-1}} \leq \frac{1}{2^{k-1}} \max_{1 \leq n \leq 2^k} |S_n|$$

Hence by Kolmogorov's inequality,

$$P[Y_k \geq \varepsilon] \leq P\left[\frac{1}{2^{k-1}} \max_{1 \leq n \leq 2^k} |S_n| \geq \varepsilon\right] \leq P\left[\max_{1 \leq n \leq 2^k} |S_n| \geq 2^{k-1}\varepsilon\right] \leq \frac{\sum_{i=1}^{2^k} \mathrm{Var}(X_i)}{2^{2k-2}\varepsilon^2}$$

$$= \frac{4\sum_{i=1}^{2^k} \mathrm{Var}(X_i)}{4^k\varepsilon^2}, \quad k = 1, 2, \ldots.$$

□

(b) *Proof.* By the celebrated geometric series sum formula,

$$\sum_{k=\lceil \log_2 i \rceil}^{\infty} \frac{1}{4^k} = \frac{\frac{1}{4^{\lceil \log_2 i \rceil}}}{1 - \frac{1}{4}} \leq \frac{4}{3} \frac{1}{4^{\log_2 i}} = \frac{4}{3i^2}.$$

By the result of part (a) and Tonelli's theorem (notice that all the summands are nonnegative):

$$\sum_{k=1}^{\infty} P[Y_k \geq \varepsilon]$$

30

$$\leq \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{1}{4^k} \sum_{i=1}^{2^k} \mathrm{Var}(X_i)$$

$$= \frac{4}{\varepsilon^2} \sum_{i=1}^{\infty} \left[\sum_{k=\lceil \log_2 i \rceil}^{\infty} \frac{1}{4^k} \right] \mathrm{Var}(X_i)$$

$$\leq \frac{4}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{4}{3i^2} \mathrm{Var}(X_i)$$

$$= \frac{16}{3\varepsilon^2} \sum_{i=1}^{\infty} \frac{\mathrm{Var}(X_i)}{i^2} < \infty.$$

Hence by Lemma 3.9, $Y_n \to_{a.s.} 0$. Given $\varepsilon > 0$, $Y_n \to_{a.s.} 0$ implies that there exists a set A with probability 1, and for each $\omega \in A$, there exists $N \in \mathbb{N}$, $n > N$ implies that $|Y_n(\omega)| = Y_n(\omega) < \varepsilon$. Now for each $n > 2^N$, there exists $k \in \mathbb{N}$ such that $n \in \{2^{k-1}, \ldots, 2^k - 1\}$, for which case $2^k - 1 \geq n > 2^N$, implying $k > N$, it then follows that

$$|\overline{X}_n(\omega)| \leq \max_{2^{k-1} \leq i < 2^k} |\overline{X}_i(\omega)| = Y_k(\omega) < \varepsilon.$$

This shows that for each $\omega \in A$, $\overline{X}_n(\omega) \to 0$ as $n \to \infty$, hence $\overline{X}_n \to_{a.s.} 0$. $\qquad \square$

Quiz 6

Proof. Denote the distribution function of M_n by F_n. If $x < 0$, then $F_n(x) = 0$. If $x \geq 0$, then

$$F_n(x) = P[M_n \leq x] = P[X_1 \leq x, \ldots, X_n \leq x] = (1 - e^{-\lambda x})^n.$$

Therefore, for each $x \in \mathbb{R}$, $\lim\limits_{n \to \infty} F_n(x) = 0$. Hence F_n *does not* converge in distribution to any *probability* distribution. On the other hand, for each $x \in \mathbb{R}$,

$$P[M_n - \lambda^{-1} \log n \leq x] = F_n(x + \lambda^{-1} \log n) = \left(1 - e^{-(\lambda x + \log n)}\right)^n = \left(1 - \frac{e^{-\lambda x}}{n}\right)^n$$

$$\to e^{-e^{-\lambda x}} =: F(x)$$

as $n \to \infty$. It is straightforward to check $F(\cdot)$ above is nondecreasing, continuous and normalized, hence a valid probability distribution function. This limiting distribution is known as *Gumbel's extreme value distribution.* $\qquad\square$

Homework 6

Problem 1 Recall that the *quantile function* F^- is defined by $F^-(q) = \inf[x : F(x) \geq q], q \in (0, 1)$.

- *Proof.* Suppose that $0 < q_1 < q_2 < 1$, denote $F^-(q_1)$ and $F^-(q_2)$ by ξ_1 and ξ_2 respectively. For each x such that $F(x) \geq q_2$, it follows by $q_2 > q_1$ that $F(x) \geq q_1$, hence $\xi_1 \leq x$. Since this holds for every x such that $F(x) \geq q_2$, it follows that $\xi_1 \leq \inf[x : F(x) \geq q_2] = \xi_2$, i.e., F^- is increasing (though F^- needs not to be *strictly* increasing).

  Fix $q \in (0, 1)$, and let $\{q_n\}$ be a sequence such that $q_n \uparrow q$. By the monotonicity of F^- just proved, the sequence $\{F^-(q_n)\}$ is increasing and bounded above by $F^-(q)$, hence $\{F^-(q_n)\}$ must converge to some limit $a \leq F^-(q)$, by monotone bounded convergence theorem (Theorem 3.14 in [5]). If $a < F^-(q)$, then $F(a) < q$, otherwise $F(a) \geq q$ would imply that $F^-(q) \leq a$, which would be in contradiction to $a < F^-(q)$. On the other hand, since $F^-(q_n) \leq a$, by monotonicity of F, we have $F(F^-(q_n)) \leq F(a)$. By the property which will be shown in the third item, this implies that $q_n \leq F(a)$ for all n, thereby $\limsup_n q_n \leq F(a) < q$, which is in contradiction to $q_n \uparrow q$. Hence $a = F^-(q)$, i.e., $\lim_{q_n \uparrow q} F^-(q_n) = F^-(q)$, that is, F^- is left continuous at q. □

- *Proof.* By definition, $F^-(F(x)) = \inf[y : F(y) \geq F(x)]$. Since $F(x) \geq F(x)$, it follows that $x \in [y : F(y) \geq F(x)]$, hence $x \geq \inf[y : F(y) \geq F(x)] = F^-(F(x))$. □

- *Proof.* For $q \in (0, 1)$, denote $F^-(q)$ by ξ, $[x : F(x) \geq q]$ by E. Since $\xi = \inf E$, for each $n \geq 1$, there exists $x_n \in E$ such that $x_n < \xi + \dfrac{1}{n}$. Since F is monotone and $x_n \in E$, it follows that $F(\xi + 1/n) \geq F(x_n) \geq q$. Let $n \to \infty$ and notice that F is right continuous at ξ, we conclude that $F(\xi) \geq q$, i.e., $F(F^-(q)) \geq q$. □

Remark. By the results proved above, for any distribution function F, we have the

following important relation:

$$\boxed{F(x) \geq q \text{ if and only if } x \geq F^-(q)} \tag{6}$$

This relation has an intuitive interpretation in the sense that $F^- \circ F$ can be viewed as an identity transformation (but it's not a rigorous proof, it just helps you memorize). By considering the contrapositive statements of (6), it also holds that

$$\boxed{x < F^-(q) \text{ if and only if } F(x) < q} \tag{7}$$

Pay close attention that the inequalities in (6) are non-strict, while that in (7) are strict. Do not confuse them with each other!

Problem 2 and Problem 3 below will show you the usefulness of (6) and (7).

Problem 2 [3]

Proof. First suppose that $\{F_n\}$ is a nonrandom sequence of distribution functions such that

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0. \tag{8}$$

Let $\varepsilon > 0$ be given. We wish to show that there exists $N = N(\varepsilon)$ such that for all $n > N$

$$|\operatorname{med}(F_n) - \operatorname{med}(F)| \leq \varepsilon.$$

For simplicity, write $\operatorname{med}(F) = m$. By definition of median and the "well-behaved" condition of F, we can choose $\delta > 0$ such that

$$F(m - \varepsilon) < \frac{1}{2} - \delta < \frac{1}{2} + \delta < F(m + \varepsilon). \tag{9}$$

---

[3]I want to thank Vincent Pisztora for inspiring me to give a simpler and clearer proof, instead of my own proof by contradiction earlier.

Next choose N so that for all $n > N$,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| < \delta.$$

Let $n > N$ and $m_n = \text{med}(F_n) = F_n^-(1/2)$. First note that $|F_n(m+\varepsilon) - F(m+\varepsilon)| < \delta$ implies that

$$F_n(m + \varepsilon) > F(m + \varepsilon) - \delta > \frac{1}{2}.$$

This inequality and (6) further implies $m + \varepsilon \geq F_n^-(1/2) = m_n$. On the other hand, $|F_n(m - \varepsilon) - F(m - \varepsilon)| < \delta$ implies that

$$F_n(m - \varepsilon) < F(m - \varepsilon) + \delta < \frac{1}{2}.$$

This inequality and (7) further implies $m - \varepsilon < F_n^-(1/2) = m_n$. Together, we showed that for all $n > N$,

$$|m_n - m| \leq \varepsilon, \tag{10}$$

which shows $m_n \to m$ as $n \to \infty$. I highlighted the two inequalities above to draw your attention once more that $\geq$ and $\leq$ there cannot be replaced by strict inequalities.

By the Glivenko-Cantelli Theorem, the umbrella condition (8) holds almost surely, with F_n there replaced by $\hat{F}_n$, it therefore follows by (10) that $\text{med}(\hat{F}_n) \to_{a.s.} \text{med}(F)$ as $n \to \infty$. $\qquad \square$

Problem 3 *Proof.* For consistency, I will still use F^- to denote the quantile function. Suppose $0 < q < 1$. Given $\varepsilon > 0$, choose x so that $F^-(q) - \varepsilon < x < F^-(q)$ and $x \in C(F)$. Then (6) implies that $F(x) < q$; $F_n(x) \to F(x)$ now implies that, for n large enough, $F_n(x) < q$ and hence $F^-(q) - \varepsilon < x < F_n^-(q)$. Thus, $\liminf_n F_n^-(q) \geq F^-(q)$.

On the other hand, if $q < q'$ and ε is positive, choose a y for which $F^-(q') < y < F^-(q') + \varepsilon$ and $y \in C(F)$. Now $q < q' \leq F(F^-(q')) \leq F(y)$, and so, for n large enough, $q \leq F_n(y)$ and hence by (6) $F_n^-(q) \leq y < F^-(q') + \varepsilon$. Thus $\limsup_n F_n^-(q) \leq F^-(q')$ if $q < q'$. Therefore, $F_n^-(q) \to F^-(q)$ if F^- is continuous at q.

Since F^- is nondecreasing on $(0,1)$, as showed in the first item of Problem 1, it has at most countably many discontinuities. The proof is complete. □

3.11 *Proof.* Since $|X_n| \geq \inf_{k \geq n} |X_k|$ for each n, it follows that

$$E[|X_n|] \geq E\left[\inf_{k \geq n} |X_k|\right], n = 1, 2, \ldots. \tag{11}$$

Additionally, since the sequence $\{\inf_{k \geq n} |X_k|\}$ is nondecreasing and nonnegative everywhere, by monotone convergence theorem, a passage to limits on both sides of (11) gives that

$$\liminf_n E\left[\inf_{k \geq n} |X_k|\right] = \lim_n E\left[\inf_{k \geq n} |X_k|\right] = E\left[\lim_n \inf_{k \geq n} |X_k|\right]$$
$$= E\left[\liminf_n |X_n|\right] \leq \liminf_n E\left[|X_n|\right].$$

The proof is complete. □

3.12 (a) *Proof.* For any two random variables A and B and given $\alpha > 0$, set $C = |A| \vee |B|$. Since $[|A+B| \geq \alpha] \subset [C \geq \alpha/2]$, it follows that $E[|A+B|I_{[|A+B| \geq \alpha]}] \leq 2E[CI_{[|A+B| \geq \alpha]}] \leq 2E[CI_{[C \geq \alpha/2]}]$. On the other hand, $[C \geq \alpha/2] = [C \geq \alpha/2, C = |A|] \cup [C \geq \alpha/2, C = |B|]$ implies that

$$E[CI_{[|C| \geq \alpha/2]}]$$
$$\leq E[C(I_{[C \geq \alpha/2, C=|A|]} + I_{[C \geq \alpha/2, C=|B|]})]$$
$$\leq E[|A|I_{[|A| \geq \alpha/2]}] + E[|B|I_{[|B| \geq \alpha/2]}]$$

Combining these two inequalities then gives that

$$E[|A+B|I_{[|A+B| \geq \alpha]}] \leq 2E[|A|I_{[|A| \geq \alpha/2]}] + 2E[|B|I_{[|B| \geq \alpha/2]}].$$

Now suppose that $\{A_n\}$ and $\{B_n\}$ are both uniformly integrable. By the inequal-

ity just showed, for each $\alpha > 0$,

$$E[|A_n + B_n|I_{[|A_n+B_n|\geq\alpha]}] \leq 2E[|A_n|I_{[|A_n|\geq\alpha/2]}] + 2E[|B_n|I_{[|B_n|\geq\alpha/2]}], n = 1, 2, \ldots.$$

whence

$$\sup_n E[|A_n + B_n|I_{[|A_n+B_n|\geq\alpha]}] \leq 2\sup_n E[|A_n|I_{[|A_n|\geq\alpha/2]}] + 2\sup_n E[|B_n|I_{[|B_n|\geq\alpha/2]}].$$

Let $\alpha \to \infty$ on both sides of the above inequality, and apply the uniform integrability of $\{A_n\}$ and $\{B_n\}$, it can be seen that $\lim_{\alpha\to\infty} \sup_n E[|A_n + B_n|I_{[|A_n+B_n|\geq\alpha]}] = 0$, that is, $\{A_n + B_n\}$ is uniformly integrable. $\qquad\square$

(b) *Proof.* Since $\{Y_n\}$ is uniformly integrable and $Z_n \overset{d}{=} Y_n$, $\{Z_n\}$ is also uniformly integrable, hence there exists a large α_0, such that $\sup_n E[|Z_n|I_{[|Z_n|\geq\alpha_0]}] < 1$. It then follows that $\sup_n E[|Z_n|] = \sup_n\{E[|Z_n|I_{[|Z_n|<\alpha_0]}] + E[|Z_n|I_{[|Z_n|\geq\alpha_0]}]\} < \alpha_0 + 1 < \infty$. Therefore by Fatou's lemma,

$$E[|Z|] = E[\lim_n |Z_n|] \leq \liminf_n E[|Z_n|] \leq \sup_n E[|Z_n|] < \infty.$$

That is, Z is integrable, implying that as a sequence, $\{-Z\}$ is uniformly integrable. By the result of part (a), $\{Z_n - Z\}$ is uniformly integrable, while this is equivalent to say that $\{X_n\} = \{|Z_n - Z|\}$ is uniformly integrable. $\qquad\square$

(c) *Proof.* Suppose that $\{X_n\}$ is uniformly integrable and $X_n \to_{a.s.} 0$.

Given $\varepsilon > 0$, by the uniform integrability of $\{X_n\}$, there exists $\alpha_0 > 0$ such that $\sup_n E[|X_n|I_{[|X_n|\geq\alpha_0]}] < \varepsilon/2$. On the other hand, it is easily seen that the sequence $\{|X_n|I_{[|X_n|<\alpha_0]}\}$ is dominated by α_0 and converges to 0 almost surely, hence by Lebesgue's dominated convergence theorem, there exists $N \in \mathbb{N}$ such that $n > N$ implies that $E[|X_n|I_{[|X_n|<\alpha_0]}] < \varepsilon/2$. Therefore for each $n > N$,

$$|E[X_n] - 0| \leq E[|X_n|] = E[|X_n|I_{[|X_n|<\alpha_0]}] + E[|X_n|I_{[|X_n|\geq\alpha_0]}] < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

i.e., $E[X_n] \to 0$ as $n \to \infty$. The proof is complete. $\qquad\square$

3.13 *Proof.* Given $\alpha > 0$, $E[|Y_n|^{1+\varepsilon}] = E[|Y_n|^{1+\varepsilon}I_{[|Y_n|<\alpha]}] + E[|Y_n|^{1+\varepsilon}I_{[|X_n|\geq\alpha]}] \geq E[|Y_n| \cdot$
$|Y_n|^{\varepsilon}I_{[|Y_n|\geq\alpha]}] \geq \alpha^{\varepsilon}E[|Y_n|I_{[|Y_n|\geq\alpha]}]$ for each n, which implies that

$$\sup_n E[|Y_n|I_{[|Y_n|\geq\alpha]}] \leq \frac{1}{\alpha^{\varepsilon}}\sup_n E[|Y_n|^{1+\varepsilon}] \to 0$$

as $\alpha \to \infty$, i.e., $\{Y_n\}$ is uniformly integrable. $\square$

3.14 *Proof.* The given hint is not quite relevant. A clearer proof can be made as follows.

Given $\alpha > 0$, it follows by the expectation formula and the law of total probability that

$$
\begin{aligned}
&E[|Y_n|I_{[|Y_n|\geq\alpha]}]\\
&= \int_0^\infty P[|Y_n|I_{[|Y_n|\geq\alpha]} \geq t]\,\mathrm{d}t\\
&= \int_0^\infty \left\{P[|Y_n|I_{[|Y_n|\geq\alpha]} \geq t, |Y_n| \geq \alpha] + P[|Y_n|I_{[|Y_n|\geq\alpha]} \geq t, |Y_n| < \alpha]\right\}\,\mathrm{d}t\\
&= \int_0^\alpha P[|Y_n| \geq \alpha]\,\mathrm{d}t + \int_\alpha^\infty P[|Y_n| \geq t]\,\mathrm{d}t\\
&= \alpha P[|Y_n| \geq \alpha] + \int_\alpha^\infty P[|Y_n| \geq t]\,\mathrm{d}t\\
&\leq \alpha P[|Z| \geq \alpha] + \int_\alpha^\infty P[|Z| \geq t]\,\mathrm{d}t,
\end{aligned}
$$

where the last inequality is due to the provided condition. Same calculation as above shows that the last expression equals to $E[|Z|I_{[|Z|\geq\alpha]}]$. Since this upper bound is independent of n, we have

$$\sup_n E[|Y_n|I_{[|Y_n|\geq\alpha]}] \leq E[|Z|I_{[|Z|\geq\alpha]}].$$

It thus remains to show $E[|Z|I_{[|Z|\geq\alpha]}] \to 0$ as $\alpha \to \infty$, which is a direct consequence of the integrability of Z. Specifically, since $|Z|I_{[|Z|\geq\alpha]}$ is dominated by $|Z|$ and it converges to 0 almost surely as $\alpha \to \infty$, the Lebesgue dominated convergence theorem ensures $E[|Z|I_{[|Z|\geq\alpha]}] \to 0$ as $\alpha \to \infty$. Therefore, $\{Y_n\}$ is uniformly integrable. This completes the proof. $\square$

Homework 7

Problem 1 *Proof.* Since $\int_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty$, the integral $\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\, dt$ exists for every x, for which case it can be shown that the (anti-Fourier) transformation

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t)\, dt$$

defines a probability density function over $\mathbb{R}$. Straightforward calculation shows that

$$f(x) = \frac{1}{\pi} \frac{1 - \cos x}{x^2}, \quad -\infty < x < \infty.$$

This completes proof. $\qquad\square$

Problem 2 (a) Suppose $X \sim \text{Poisson}(\lambda)$, then $P[X = k] = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$. Hence

$$\phi_X(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{\lambda(e^{it}-1)}.$$

(b) Suppose $X \sim \text{Geometric}(p)$, then[4] $P[X = k] = (1-p)^{k-1}p$, $k = 1, 2, \ldots$. Hence

$$\phi_X(t) = \sum_{k=1}^{\infty} e^{itk} (1-p)^{k-1} p = \frac{p}{1-p} \sum_{k=1}^{\infty} [(1-p)e^{it}]^k = \frac{pe^{it}}{1-(1-p)e^{it}}.$$

(c) Suppose $X \sim \text{Exp}(\lambda)$, then $f(x) = \lambda e^{-\lambda x}$, $x > 0$. To evaluate $\phi_X(t) = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x}\, dx$, we need to evaluate

$$C(t) = \int_0^{\infty} \cos(tx) e^{-\lambda x}\, dx \text{ and } S(t) = \int_0^{\infty} \sin(tx) e^{-\lambda x}\, dx$$

separately. By definition, $\phi_X(t) = \lambda C(t) + i\lambda S(t)$, see pp.218-219, p.342 of [2]. Integrating by parts twice, it follows that

$$C(t) = \frac{\lambda}{\lambda^2 + t^2}, \quad S(t) = \frac{t}{\lambda^2 + t^2}.$$

---

[4]Here we use X to denote the number of trials until the first success.

Therefore,

$$\phi_X(t) = \frac{\lambda^2}{\lambda^2 + t^2} + \frac{i\lambda t}{\lambda^2 + t^2} = \frac{\lambda}{\lambda - it}.$$

For a contour integration approach (which needs some complex analysis background), check the following link:

`http://math.stackexchange.com/questions/2172433`

(d) Suppose $X \sim \text{Cauchy}(x_0, \gamma)$, i.e., the density of X is given by

$$f(x) = \frac{1}{\pi\gamma \left[1 + \left(\dfrac{x - x_0}{\gamma} \right)^2 \right]}, \ x \in \mathbb{R}.$$

The characteristic function of X is given by

$$\phi_X(t) = \exp(itx_0 - \gamma|t|).$$

The derivation of this result is somewhat technical. One approach is to use the inversion formula (i.e., inverse Fourier transformation), see p.349 of [2] for details. Another more direct approach is to apply the contour integration technique in complex analysis. For those who are interested, refer to the pdf document I uploaded in the Solution folder.

4.1(b) *Proof.* If $Y \sim \mathcal{N}(0, I_k)$, then $Y_1, \dots, Y_k$ are independent, it follows by part (a) that

$$\phi_Y(t) = E[e^{it^T Y}] = \prod_{j=1}^{k} E[e^{it_j Y_j}] = e^{-\frac{1}{2}\sum_{j=1}^{k} t_j^2} = e^{-\frac{1}{2}t^T t}.$$

Hence if $X = \mu + Q^T Y$ with $\Sigma = QQ^T$, then

$$\phi_X(t) = e^{it^T \mu} E[e^{i(t^T Q)Y}] = e^{it^T \mu} E[e^{i(Q^T t)^T Y}]$$

$$= e^{it^T \mu} e^{-\frac{1}{2}(Q^T t)^T (Q^T t)} = e^{it^T \mu} e^{-\frac{1}{2}t^T QQ^T t} = e^{it^T \mu - \frac{1}{2}t^T \Sigma t}.$$

40

The proof is complete. □

4.4 *Proof.* Define $Y := \Sigma^{-1/2}(X - \mu)$, where $\Sigma^{-1/2} = Q\Lambda^{-1/2}Q^T$. Then $E[Y] = 0$, $\text{Var}(Y) = I$, hence $Y \sim \mathcal{N}(0, I_k)$. Therefore

$$(X - \mu)^T \Sigma^{-1}(X - \mu) = (\Sigma^{-1/2}(X - \mu))^T(\Sigma^{-1/2}(X - \mu)) = Y^T Y = \sum_{j=1}^{k} Y_j^2 \sim \chi_k^2,$$

in view of $Y_1, \ldots, Y_k$ i.i.d. $\sim \mathcal{N}(0, 1)$. □

4.5 (a) *Proof.* Denote $S_n = X_1 + \cdots + X_n$, then $X_1, \ldots, X_n$ i.i.d. $\sim$ Poisson(1) implies that $S_n \sim$ Poisson(n). Noticing that $[Y_n > 0] = [S_n > n]$, hence

$$
\begin{aligned}
&E[Y_n^+] \\
=&E\left[\sqrt{n}\left(\frac{S_n}{n} - 1\right)I[S_n > n]\right] \\
=&\sum_{k=n+1}^{\infty} \sqrt{n}\left(\frac{k}{n} - 1\right)e^{-n}\frac{n^k}{k!} \\
=&\sqrt{n}e^{-n}\sum_{k=n+1}^{\infty}\left[\frac{n^{k-1}}{(k-1)!} - \frac{n^k}{k!}\right] \\
=&n^{n+1/2}\frac{e^{-n}}{n!}
\end{aligned}
$$

The last equality holds because of the obvious telescoping structure and the fact

$$\lim_{N\to\infty}\frac{n^N}{N!} = 0$$

for fixed $n \in \mathbb{N}$. To see this (without applying the Stirling's formula), write

$$\frac{n^N}{N!} = \frac{n}{N} \times \frac{n}{N-1} \times \cdots \times \frac{n}{2n} \times \frac{n^{2n-1}}{(2n-1)!} < \left(\frac{1}{2}\right)^{N-2n+1} \times \frac{n^{2n-1}}{(2n-1)!} \to 0$$

as $N \to \infty$. □

(b) *Proof.* By classical CLT, $Y_n \Rightarrow Z \sim \mathcal{N}(0, 1)$. Since the function $x \mapsto x^+ = x \vee 0$ is continuous in x, it follows by continuous mapping theorem that $Y_n^+ \Rightarrow Z^+$.

Since $E[Z^+] = \dfrac{1}{\sqrt{2\pi}}$, to show the result, it remains to show that $Y_n^+ \Rightarrow Z^+$ implies that $E[Y_n^+] \to E[Z^+]$, which is implied by the uniform integrability of $\{Y_n\}$ (since $Y_n^+ = Y_n \vee 0$ and 0 is clearly uniformly integrable).

Recall that $S_n \sim \text{Poisson}(n)$, thus $E[Y_n^2] = nE[(S_n/n-1)^2] = \frac{1}{n}E[S_n^2] - 2E[S_n] + n = \frac{1}{n}(n+n^2) - 2n + n = 1$ for every n, which implies that $\sup_n E[Y_n^2] = 1 < \infty$. By the result of Exercise 3.13, this implies that $\{Y_n\}$ is uniformly integrable and the proof is complete. $\qquad\square$

4.6 *Proof.* Under the assumption of WLLN, $\mu = E[X_1] < \infty$ exists. For each fixed $t \in \mathbb{R}$, we have

$$\varphi_{\overline{X}_n}(t) = \varphi_{\sum_i X_i}(t/n) = (\varphi_{X_1}(t/n))^n = \left(1 + \frac{it\mu}{n} + o\left(\frac{t}{n}\right)\right)^n \to e^{it\mu}$$

as $n \to \infty$. Notice that $e^{it\mu}$ is exactly the characteristic function of constant μ, thus the WLLN follows from continuity theorem. $\qquad\square$

4.7 *Proof.* By Cramér-Wold device, it suffices to show that for every $c \in \mathbb{R}^k$, it holds that

$$\sqrt{n}(c^T\overline{X}_n - c^T\mu) \Rightarrow \mathcal{N}(0, c^T\Sigma c). \tag{12}$$

But under the given condition, the sequence $c^TX_1, c^TX_2, \ldots$ is a sequence of i.i.d. random variables with mean $c^T\mu$ and variance $c^T\Sigma c$. (12) thus follows directly from the classical CLT. $\qquad\square$

Quiz 7

Problem 1 It seems that this problem missed a crucial condition that the characteristic function $\varphi_X(t) \neq 0, \forall t \in \mathbb{R}^1$, where X is the limiting distribution of X_n. Everyone gets full credits for this problem.

Problem 2 To show this statement by charateristic function, we need the following lemma (see, e.g., [1], p.177):

Lemma. *If the complex numbers c_n have the limit c, then*

$$\lim_{n \to \infty} \left(1 + \frac{c_n}{n}\right)^n = e^c.$$

For a rigorous proof of this lemma, check the most voted answer of the following link[5]:

`http://math.stackexchange.com/questions/374747/`

Since $X_n \sim \text{Bin}(n, p_n)$, the characteristic function of X_n is $\varphi_{X_n}(t) = (1+p_n(e^{it}-1))^n$, rewrite $\varphi_{X_n}(t)$ as

$$\left(1 + \frac{np_n(e^{it} - 1)}{n}\right)^n$$

then applies the lemma with $c_n = np_n(e^{it}-1)$, it follows that $\varphi_{X_n}(t) \to \exp(\lambda(e^{it}-1))$ as $n \to \infty$, which is precisely the characteristic function of the Poisson(λ) distribution. The result follows by the continuity theorem.

---

[5]I pointed this out because the proof for the case that c_n are real numbers (which we are familiar with) is essentially different from that for the case that c_n are complex numbers.

Homework 8

4.12 (a) *Proof.* To show the assertion, it suffices to verify the Lindeberg condition for the triangular array $a_{ni}(X_i - \mu), i = 1, \ldots, n, n = 1, 2, \ldots$. It is easy to see that $s_n^2 = \sum_{i=1}^{n} \text{Var}(a_{ni}(X_i - \mu)) = \sigma^2 \sum_{i=1}^{n} a_{ni}^2$. Given $\varepsilon > 0$,

$$
\frac{1}{s_n^2} \sum_{i=1}^{n} E[a_{ni}^2 (X_i - \mu)^2 I_{[|a_{ni}||X_i - \mu| \geq \varepsilon s_n]}]
$$

$$
\leq \frac{1}{s_n^2} \sum_{i=1}^{n} a_{ni}^2 E[(X_i - \mu)^2 I_{[|X_i - \mu| \geq s_n / \max_{1 \leq i \leq n} |a_{ni}|]}]
$$

$$
= \frac{\sum_{i=1}^{n} a_{ni}^2}{s_n^2} E[(X_1 - \mu)^2 I_{[|X_1 - \mu| \geq s_n / \max_{1 \leq i \leq n} |a_{ni}|]}] \quad \text{since } X_1, \ldots, X_n \text{ are i.i.d.}
$$

$$
= \sigma^{-2} E[(X_1 - \mu)^2 I_{[|X_1 - \mu| \geq s_n / \max_{1 \leq i \leq n} |a_{ni}|]}]
$$

By condition, $s_n / \max_{1 \leq i \leq n} |a_{ni}| \to \infty$ as $n \to \infty$, it thus follows by the integrability of $(X_1 - \mu)^2$ that the right hand side of the above expression tends to 0 as $n \to \infty$. Thus the Lindeberg condition holds. $\square$

(b) *Proof.* By the definitions of $\{w_i^{(n)}\}$ and $\{v_i^{(n)}\}$, it is straightforward to verify that $\hat{\beta}_{1n} - \beta_1 = \sum_{i=1}^{n} w_i^{(n)} \varepsilon_i$ and $\hat{\beta}_{0n} - \beta_0 = \sum_{i=1}^{n} v_i^{(n)} \varepsilon_i$. For notational convenience, denote $\sum_{j=1}^{n} (z_j - \bar{z}_n)^2$ by SS_n.

We will show that one group of sufficient conditions that implies the asymptotic nomalities of $\sqrt{n}(\hat{\beta}_{1n} - \beta_1)$ and $\sqrt{n}(\hat{\beta}_{0n} - \beta_0)$ could be

$$
\boxed{\bar{z}_n \to \mu \in \mathbb{R} \text{ as } n \to \infty}, \tag{13}
$$

$$
\boxed{\frac{SS_n}{n} \to A > 0 \text{ as } n \to \infty}, \tag{14}
$$

$$
\boxed{\frac{\max_{1 \leq i \leq n} (z_i - \bar{z}_n)^2}{SS_n} \to 0 \text{ as } n \to \infty}. \tag{15}
$$

We first show that under conditions (14) and (15), $\sqrt{n}(\hat{\beta}_{1n} - \beta_1)$ is asymptotically

44

normally distributed. Let $\sum_{i=1}^{n} w_i^{(n)} \varepsilon_i$ be T_n in part (a), then we can write

$$\sqrt{n}(\hat{\beta}_{1n} - \beta_1) = \frac{T_n - E[T_n]}{\sqrt{\mathrm{Var}(T_n)}} \times \sqrt{n\,\mathrm{Var}(T_n)}. \tag{16}$$

Since

$$\frac{\max_{1 \le i \le n} (w_i^{(n)})^2}{\sum_{j=1}^{n} (w_j^{(n)})^2} = \frac{\max_{1 \le i \le n} (z_i - \bar{z}_n)^2}{SS_n} \to 0$$

as $n \to \infty$ by (15), the condition in part (a) is satisfied, hence the first factor of the right hand side of (16) converges in distribution to $\mathcal{N}(0,1)$. Additionally, by (14),

$$\sqrt{n\,\mathrm{Var}(T_n)} = \sqrt{n\sigma^2 \frac{1}{SS_n}} \to \sigma\sqrt{A^{-1}}$$

as $n \to \infty$. It follows by (16) and Slutsky's theorem that $\sqrt{n}(\hat{\beta}_{1n} - \beta_1) \Rightarrow \mathcal{N}(0, \sigma^2 A^{-1})$.

Next, we show that under conditions $(13) - (15)$, $\sqrt{n}(\hat{\beta}_{0n} - \beta_0)$ is asymptotically normally distributed. As before, here we let $\sum_{i=1}^{n} v_i^{(n)} \varepsilon_i$ be T_n in part (a) so that we can write

$$\sqrt{n}(\hat{\beta}_{0n} - \beta_0) = \frac{T_n - E[T_n]}{\sqrt{\mathrm{Var}(T_n)}} \times \sqrt{n\,\mathrm{Var}(T_n)}. \tag{17}$$

We have

$$\frac{\max_{1 \le i \le n} (v_i^{(n)})^2}{\sum_{j=1}^{n} (v_j^{(n)})^2}$$

$$= \frac{\max_{1 \le i \le n} \left(\dfrac{1}{n^2} - \dfrac{2\bar{z}_n(z_i - \bar{z}_n)}{nSS_n} + \dfrac{\bar{z}_n^2(z_i - \bar{z}_n)^2}{SS_n^2} \right)}{\dfrac{1}{n} + \dfrac{\bar{z}_n^2}{SS_n}}$$

$$\le \frac{\dfrac{1}{n^2}}{\dfrac{1}{n} + \dfrac{\bar{z}_n^2}{SS_n}} + \frac{\dfrac{2|\bar{z}_n|}{n\sqrt{SS_n}}}{\dfrac{1}{n} + \dfrac{\bar{z}_n^2}{SS_n}} \frac{\max_{1 \le i \le n} |z_i - \bar{z}_n|}{\sqrt{SS_n}} + \frac{\dfrac{\bar{z}_n^2}{SS_n}}{\dfrac{1}{n} + \dfrac{\bar{z}_n^2}{SS_n}} \frac{\max_{1 \le i \le n} (z_i - \bar{z}_n)^2}{SS_n}.$$

45

Under conditions $(13) - (15)$, it is easy to verify that each term in the right hand side of the above inequality is $o(1)$. Therefore condition in part (a) again is satisfied, thus the first factor of the right hand side of (17) converges in distribution to $\mathcal{N}(0, 1)$. Similarly, in this case, by (13), (14):

$$\sqrt{n \operatorname{Var}(T_n)} = \sqrt{n\sigma^2 \left(\frac{1}{n} + \frac{\bar{z}_n^2}{SS_n} \right)} \to \sigma\sqrt{1 + A^{-1}\mu^2}.$$

It then follows by (17) and Slutsky's theorem that

$$\sqrt{n}(\hat{\beta}_{0n} - \beta_0) \Rightarrow \mathcal{N}(0, \sigma^2(1 + A^{-1}\mu^2)).$$

Other sufficient conditions may also work, however, $(13) - (15)$ are standard conditions for dealing with asymptotics under regression settings. □

4.13 *Proof.* For each i, define Z_i by $P[Z_i = 0] = 1 - 1/i^2, P[Z_i = i] = P[Z_i = -i] = 1/(2i^2)$. Clearly, $E[Z_i] = 0, \operatorname{Var}(Z_i) = 1, i = 1, 2, \ldots$. But $\sqrt{n}\bar{Z}_n \to_P 0$ (instead of converging weakly to $\mathcal{N}(0, 1)$). To see this, since $E[|Z_i|] = 1/i, i = 1, \ldots, n$, for any given $\varepsilon > 0$, by triangle inequality and Markov's inequality:

$$P[|\sqrt{n}\bar{Z}_n| \geq \varepsilon]$$
$$= P[|Z_1 + \cdots + Z_n| \geq \sqrt{n}\varepsilon]$$
$$\leq \frac{E[|Z_1 + \cdots + Z_n|]}{\sqrt{n}\varepsilon}$$
$$\leq \frac{\sum\limits_{i=1}^{n} E[|Z_i|]}{\sqrt{n}\varepsilon}$$
$$= \frac{\sum\limits_{i=1}^{n} \frac{1}{i}}{\sqrt{n}\varepsilon}$$
$$= \frac{\log n + O(1)}{\sqrt{n}\varepsilon} \to 0$$

as $n \to \infty$. Therefore $\sqrt{n}\bar{Z}_n \Rightarrow 0$. Since the weak limit is unique, this implies that $\sqrt{n}\bar{Z}_n \not\Rightarrow \mathcal{N}(0, 1)$. □

4.16 (a) *Proof.* For each positive integer n and each positive integer j, we have

$$E[(X_n^j)^+] = 1 \times \frac{1}{2}(1 - 2^{-n}) + \sum_{k=n+1}^{\infty} 2^{kj} \times 2^{-k} = \frac{1}{2} - \frac{1}{2^{n+1}} + \sum_{k=n+1}^{\infty} 2^{(j-1)k} = \infty,$$

$$E[(X_n^j)^-] = 1 \times \frac{1}{2}(1 - 2^{-n}) = \frac{1}{2} - \frac{1}{2^{n+1}} < \infty.$$

Therefore, $E[X_n^j] = E[(X_n^j)^+] - E[(X_n^j)^-] = \infty.$ $\qquad\square$

(b) *Proof.* Define $Y_n = X_n I_{[|X_n|=1]}, n = 1, 2, \ldots$.[6] Obviously, the probability mass function of Y_n is given by

$$P[Y_n = 1] = P[Y_n = -1] = \frac{1}{2}(1 - 2^{-n}), \quad P[Y_n = 0] = 2^{-n}, \quad n = 1, 2, \ldots.$$

Therefore, $E[Y_n] = 0, \mathrm{Var}(Y_n) = E[Y_n^2] = E[Y_n^4] = 1 - 2^{-n}$. Hence $s_n^2 = \sum_{k=1}^{n} \mathrm{Var}(Y_k) = \sum_{k=1}^{n} (1 - 2^{-k}) = n - 1 + (1/2)^n$ and $\sum_{k=1}^{n} E[Y_k^4]/s_n^4 = s_n^2/s_n^4 = 1/s_n^2 \to 0$ as $n \to \infty$. Thus the Lyapunov's condition holds for $\{Y_n\}$, and we can claim the asymptotic normality for $\{Y_n\}$:

$$\frac{1}{s_n}\sum_{k=1}^{n} Y_k = \frac{1}{\sqrt{n - 1 + (1/2)^n}}\sum_{k=1}^{n} Y_k \Rightarrow \mathcal{N}(0, 1).$$

Since $\sqrt{n} \sim \sqrt{n - 1 + (1/2)^n}$, it follows by the Slutsky's theorem that $\sqrt{n}\overline{Y}_n \Rightarrow \mathcal{N}(0, 1)$.

To show $\sqrt{n}\overline{X}_n \Rightarrow \mathcal{N}(0, 1)$, in view of the Slutsky's theorem and $\sqrt{n}\overline{Y}_n \Rightarrow \mathcal{N}(0, 1)$, it suffices to show that $\sqrt{n}(\overline{X}_n - \overline{Y}_n) = o_P(1)$. In fact, we can prove something even stronger, i.e., $\sqrt{n}(\overline{X}_n - \overline{Y}_n) \to_{a.s.} 0$. To see this, let A_n denote the event $[X_n \neq Y_n], n = 1, 2, \ldots$. By definition, $\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P[X_n = 2^k, \text{ for some } k > n] = \sum_{n=1}^{\infty} 2^{-n} = 1 < \infty$, which implies that $P[A_n \text{ i.o.}] = 0$ by the first Borel-Cantelli lemma. This means, there exists an event B with $P(B) = 1$ such that for every $\omega \in B$, there exists a positive integer $N = N(\omega)$,

---

[6]$\{Y_n\}$ is known as the *truncation* of $\{X_n\}$.

such that $X_n(\omega) = Y_n(\omega)$ for all $n > N$. Therefore, for every $n > N$,

$$\sqrt{n}(\overline{X}_n(\omega) - \overline{Y}_n(\omega)) = \frac{1}{\sqrt{n}} \sum_{k=1}^{N} (X_k(\omega) - Y_k(\omega)).$$

The above expression tends to 0 as $n \to \infty$ because we have only finitely many summands. In summary, we showed that for each $\omega \in B$, $\sqrt{n}(\overline{X}_n(\omega) - \overline{Y}_n(\omega)) \to 0$ as $n \to \infty$, i.e., $\sqrt{n}(\overline{X}_n - \overline{Y}_n) \to_{a.s.} 0$. This completes the proof. $\qquad \square$

4.17 (a) *Proof.* It is easy to see that X_{nk} follows geometric distribution with parameter $p_{nk} := (n - k + 1)/n$. Thus $E[X_{nk}] = 1/p_{nk}$, $\mathrm{Var}(X_{nk}) = (1 - p_{nk})/p_{nk}^2$. Also note that $X_{n1}, \dots, X_{nr_n}$ are mutually independent. $\qquad \square$

(b) *Proof.* Following the hint, we check Lyapunov's condition with $\delta = 2$ for $S_n = \sum_{k=1}^{r_n} X_{nk}$. It thus needs to be shown that

$$\frac{1}{\tau_n^4} \sum_{k=1}^{r_n} E\{|X_{nk} - E[X_{nk}]|^4\} \to 0 \qquad (18)$$

as $n \to \infty$. For notational convenience, denote $\sum_{k=1}^{r_n} \frac{1}{p_{nk}^m}$ by $F_n(m)$, $m = 1, 2, 3, 4$. It can be shown that (use the moment generating function of geometric distribution or use Mathematica)

$$\sum_{k=1}^{r_n} E\{|X_{nk} - E[X_{nk}]|^4\}$$

$$= \sum_{k=1}^{r_n} \frac{(p_{nk} - 1)(-p_{nk}^2 + 9p_{nk} - 9)}{p_{nk}^4}$$

$$= - F_n(1) + 10F_n(2) - 18F_n(3) + 9F_n(4)$$

$$\leq 10F_n(2) + 9F_n(4),$$

and

$$\tau_n^4 = \left(\sum_{k=1}^{r_n} \frac{1 - p_{nk}}{p_{nk}^2} \right)^2 = (F_n(2) - F_n(1))^2 = F_n(2)^2 \left(1 - \frac{F_n(1)}{F_n(2)} \right)^2.$$

48

Therefore to show (18), it would be helpful to study the asymptotic behavior of $F_n(m)$. Notice that for $m \geq 1$, the function $t \mapsto (1-t)^{-m}, t \in [0,1)$ is convex and nondecreasing, hence

$$n \int_0^{r_n/n} (1-t)^{-m} \, dt - \left(1 - \frac{r_n}{n}\right)^{-m} + 1 \leq F_n(m) \leq n \int_0^{r_n/n} (1-t)^{-m} \, dt.$$

This inequality and the squeeze principle then imply that (also use $r_n/n \to \rho \in (0,1)$ as $n \to \infty$):

$$\frac{F_n(m)}{n} \to \begin{cases} -\log(1-\rho) & m = 1, \\ \dfrac{1}{m-1}\left[\dfrac{1}{(1-\rho)^{m-1}} - 1\right] & m = 2, 4. \end{cases} \tag{19}$$

Note that all above limits are positive and finite. Collecting above results gives that

$$\frac{1}{\tau_n^4} \sum_{k=1}^{r_n} E\{|X_{nk} - E[X_{nk}]|^4\}$$

$$\leq \frac{1}{F_n(2)^2 \left(1 - \dfrac{F_n(1)}{F_n(2)}\right)^2} [10 F_n(2) + 9 F_n(4)]$$

$$= \frac{10}{n \dfrac{F_n(2)}{n} \left(1 - \dfrac{F_n(1)}{F_n(2)}\right)^2} + \frac{9 \dfrac{F_n(4)}{F_n(2)}}{n \dfrac{F_n(2)}{n} \left(1 - \dfrac{F_n(1)}{F_n(2)}\right)^2}$$

$$\to 0$$

as $n \to \infty$. The above final limit is zero since by (19),

$$\frac{F_n(2)}{n} \to \frac{\rho}{1-\rho}, \frac{F_n(1)}{F_n(2)} \to -\frac{(1-\rho)\log(1-\rho)}{\rho}, \frac{F_n(4)}{F_n(2)} \to \frac{1 - (1-\rho)^3}{3\rho(1-\rho)^2}.$$

And all these limits are bounded away from $0, 1$ and ∞ (the condition $0 < \rho < 1$ is crucial). It may still need some work to show that the second limit above is not 1, but I will omit the details. In summary, Lyapunov's condition is verified

49

thereby the asymptotic normality holds. □

4.18 (a) *Proof.* It is straightforward to verify that $E[X_1] = 2p$, $E[Y_1] = (1-p)^2$, $\text{Var}(X_1) = 2p(1-p)$, $\text{Var}(Y_1) = p(2-p)(1-p)^2$ and $\text{Cov}(X_1, Y_1) = -2p(1-p)^2$. Thus by multivariate CLT,

$$\sqrt{n}\left[(\bar{X}_n, \bar{Y}_n)^T - (2p, (1-p)^2)^T\right] \Rightarrow \mathcal{N}((0,0)^T, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} 2p(1-p) & -2p(1-p)^2 \\ -2p(1-p)^2 & p(2-p)(1-p)^2 \end{pmatrix}.$$

This completes the proof. □

(b) *Proof.* Take $c = (1,1)^T$, then Cramér-Wold device and result of part (a) imply that

$$\sqrt{n}[c^T(\bar{X}_n, \bar{Y}_n)^T - c^T(2p, (1-p)^2)^T] \Rightarrow \mathcal{N}(0, c^T \Sigma c).$$

Simplification gives that

$$\sqrt{n}(\bar{X}_n + \bar{Y}_n - 1 - p^2) \Rightarrow \mathcal{N}(0, p^2(1-p^2)).$$

The proof is complete. □

4.20(a) *Proof.* It is clear that $\{X_{i-1}X_i\}$ is a stationary 1-dependent sequence with finite variance, hence Theorem 4.20 applies. Since $\mu = E[X_0X_1] = E[X_0]E[X_1] = p^2$, $\sigma^2 = \text{Var}(X_0X_1) = E[X_0^2 X_1^2] - p^4 = p^2 - p^4$, $\text{Cov}(X_0X_1, X_1X_2) = E[X_0X_1^2X_2] - E[X_0X_1]E[X_1X_2] = p^3 - p^4$, it follows by Theorem 4.20 that $\sqrt{n}(\bar{X}_n - p^2) \Rightarrow \mathcal{N}(0, p^2 + 2p^3 - p^4)$, or equivalently, in terms of S_n:

$$\frac{S_n - np^2}{\sqrt{n}} \Rightarrow \mathcal{N}(0, p^2 + 2p^3 - p^4).$$

This completes the proof. □

4.21(a) *Proof.* For notational simplicity, denote $I_{[X_i < X_{i-1}, X_i < X_{i+1}]}$ by Y_i, $i = 1, 2, \ldots, n$.

Clearly, $\{Y_i\}$ is a stationary 2-dependent sequence, with

$$\mu = E[Y_1] = P[X_1 < X_0, X_1 < X_2] = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} \int_{x_1}^{\infty} \mathrm{d}F(x_0)\,\mathrm{d}F(x_1)\,\mathrm{d}F(x_2) = \frac{1}{3},$$

$$\sigma^2 = \mathrm{Var}(Y_1) = E[Y_1^2] - (E[Y_1])^2 = E[Y_1](1 - E[Y_1]) = \frac{2}{9},$$

$$\mathrm{Cov}(Y_1, Y_2) = E[Y_1 Y_2] - E[Y_1]E[Y_2] = 0 - \frac{1}{9} = -\frac{1}{9}.$$

The evaluation of $E[Y_1 Y_3]$ is more complicated, as shown below (note the joint the distribution of $(X_0, X_1, X_2, X_3, X_4)$ is $F(x_0)F(x_1)F(x_2)F(x_3)F(x_4)$, $(x_0, x_1, x_2, x_3, x_4) \in \mathbb{R}^5$, then carefully determine the integration region):

$$E[Y_1 Y_3] = P[X_1 < X_0, X_1 < X_2, X_3 < X_2, X_3 < X_4]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_4} \int_{x_3}^{\infty} \int_{-\infty}^{x_2} \int_{x_1}^{\infty} \mathrm{d}F(x_0)\,\mathrm{d}F(x_1)\,\mathrm{d}F(x_2)\,\mathrm{d}F(x_3)\,\mathrm{d}F(x_4)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_4} \int_{x_3}^{\infty} \left[\int_{-\infty}^{x_2} (1 - F(x_1))\,\mathrm{d}F(x_1) \right] \mathrm{d}F(x_2)\,\mathrm{d}F(x_3)\,\mathrm{d}F(x_4)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_4} \int_{x_3}^{\infty} \left[F(x_2) - \frac{1}{2}F(x_2)^2 \right] \mathrm{d}F(x_2)\,\mathrm{d}F(x_3)\,\mathrm{d}F(x_4)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x_4} \left[\frac{1}{3} - \frac{1}{2}F(x_3)^2 + \frac{1}{6}F(x_3)^3 \right] \mathrm{d}F(x_3)\,\mathrm{d}F(x_4)$$

$$= \int_{-\infty}^{\infty} \left[\frac{1}{3}F(x_4) - \frac{1}{6}F(x_4)^3 + \frac{1}{24}F(x_4)^4 \right] \mathrm{d}F(x_4)$$

$$= \frac{2}{15}.$$

Hence $\mathrm{Cov}(Y_1, Y_3) = E[Y_1 Y_3] - E[Y_1]E[Y_3] = 1/45$. By Theorem 4.20, we have

$$\sqrt{n}\left(\bar{Y}_n - \frac{1}{3} \right) \Rightarrow \mathcal{N}(0, 2/9 + 2 \times (-1/9 + 1/45)) = \mathcal{N}(0, 2/45).$$

Or equivalently, in terms of S_n,

$$\frac{S_n - n/3}{\sqrt{n}} \Rightarrow \mathcal{N}\left(0, \frac{2}{45} \right).$$

This completes the proof. □

Homework 9

Problem 1 *Proof.* Carefully evaluating all required population moments, then applying the multivariate CLT gives that

$$\sqrt{n}\left[\begin{pmatrix}\overline{Y}\\ \overline{Y^2}-1\\ \overline{Y^3}-\lambda\end{pmatrix}\right] \Rightarrow \mathcal{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix} 1 & \lambda & \kappa+3\\ \lambda & \kappa+2 & \mu_5/\sigma^5-\lambda\\ \kappa+3 & \mu_5/\sigma^5-\lambda & \mu_6/\sigma^6-\lambda^2 \end{pmatrix}\right) \tag{20}$$

Note that the result is better to be expressed in terms of κ, as the instruction indicates.

Expanding $\sum_{i=1}^{n}(Y_i - \overline{Y})^3$ and using the definitions of $\overline{Y^k}, k = 1, 2, 3$, it can be shown that

$$\ell_n = \boxed{\frac{\overline{Y^3} - 3\overline{Y}\,\overline{Y^2} + 2\overline{Y}^3}{(\overline{Y^2} - \overline{Y}^2)^{3/2}}}. \tag{21}$$

Note do not confuse $\overline{Y^3}$ with $\overline{Y}^3$, etc.

Form (21) of ℓ_n suggests us considering the function

$$\phi(a, b, c) = \frac{c - 3ab + 2a^3}{(b - a^2)^{3/2}},$$

whose derivative at the point $(0, 1, \lambda)$ is $(-3, -3\lambda/2, 1)$, then applying the Delta method. One simple way to express the limiting distribution of $\sqrt{n}(\ell_n - \lambda)$ is to denote by $(T_1, T_2, T_3)^T$ the random vector whose distribution is the trivariate normal given on the right hand side of (20) first, then write

$$\sqrt{n}(\ell_n - \lambda) \Rightarrow \mathcal{N}(0, \operatorname{Var}(-3T_1 - 3\lambda T_2/2 + T_3)).$$

Or more tediously, we can explicitly spell $\operatorname{Var}(-3T_1 - 3\lambda T_2/2 + T_3)$ out as follows:

$$\operatorname{Var}(-3T_1 - 3\lambda T_2/2 + T_3)$$
$$= 9\operatorname{Var}(T_1) + \frac{9}{4}\lambda^2 \operatorname{Var}(T_2) + \operatorname{Var}(T_3) + 9\lambda \operatorname{Cov}(T_1, T_2) - 6\operatorname{Cov}(T_1, T_3) - 3\lambda \operatorname{Cov}(T_2, T_3)$$
$$= 9 + \frac{9}{4}\lambda^2(\kappa + 2) + (\mu_6/\sigma^6 - \lambda^2) + 9\lambda^2 - 6(\kappa + 3) - 3\lambda(\mu_5/\sigma^5 - \lambda)$$

$$= \boxed{-9 - 6\kappa + \frac{9}{4}\lambda^2\kappa + \frac{31}{2}\lambda^2 + \frac{\mu_6}{\sigma^6} - 3\frac{\lambda\mu_5}{\sigma^5}}.$$

This completes the proof. $\qquad\square$

Problem 2 *Proof.* Under the assumption that $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_n\}$ are independent, we have $E[X_1] = \mu_X, E[Y_1] = \mu_Y, E[X_1Y_1] = \mu_X\mu_Y, \operatorname{Var}(X_1) = \sigma_X^2, \operatorname{Var}(Y_1) = \sigma_Y^2, \operatorname{Var}(X_1Y_1) = E[(X_1Y_1)^2] - (E[X_1Y_1])^2 = (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) - \mu_X^2\mu_Y^2 = \sigma_X^2\sigma_Y^2 + \sigma_X^2\mu_Y^2 + \sigma_Y^2\mu_X^2$, and $\operatorname{Cov}(X_1, Y_1) = 0, \operatorname{Cov}(X_1, X_1Y_1) = E[X_1^2Y_1] - E[X_1]E[X_1Y_1] = (\sigma_X^2 + \mu_X^2)\mu_Y - \mu_X^2\mu_Y = \sigma_X^2\mu_Y, \operatorname{Cov}(Y_1, X_1Y_1) = E[X_1Y_1^2] - E[Y_1]E[X_1Y_1] = \mu_X(\sigma_Y^2 + \mu_Y^2) - \mu_X\mu_Y^2 = \mu_X\sigma_Y^2$. It then follows by the multivariate CLT that

$$\sqrt{n}\left[\begin{pmatrix} \overline{X} \\ \overline{Y} \\ \overline{XY} \end{pmatrix} - \begin{pmatrix} \mu_X \\ \mu_Y \\ \mu_X\mu_Y \end{pmatrix}\right] \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & 0 & \sigma_X^2\mu_Y \\ 0 & \sigma_Y^2 & \sigma_Y^2\mu_X \\ \sigma_X^2\mu_Y & \sigma_Y^2\mu_X & \sigma_X^2\sigma_Y^2 + \sigma_X^2\mu_Y^2 + \sigma_Y^2\mu_X^2 \end{pmatrix}\right) \tag{22}$$

Define $g(x, y, z) = \frac{z}{xy}, x \neq 0, y \neq 0$, then $T_n = g(\overline{X}, \overline{Y}, \overline{XY})$ and $g(\mu_X, \mu_Y, \mu_X\mu_Y) = 1$. Additionally, $\nabla g(\mu_X, \mu_Y, \mu_X\mu_Y) = (-1/\mu_X, -1/\mu_Y, 1/(\mu_X\mu_Y))^T$. By the Delta method and (22), it follows that

$$\sqrt{n}(T_n - 1) \Rightarrow \boxed{\mathcal{N}\left(0, \frac{\sigma_X^2\sigma_Y^2}{\mu_X^2\mu_Y^2}\right)}.$$

This completes the proof. $\qquad\square$

5.1 *Proof.* By the classical CLT, we have $\sqrt{n}(X_n/n - 1/2) \Rightarrow \xi \sim \mathcal{N}(0, 1/4)$. Write

$$\delta_n = \frac{X_n}{n}\left(1 - \frac{X_n}{n}\right) = \frac{1}{4} - \left(\frac{X_n}{n} - \frac{1}{2}\right)^2 = \frac{1}{4} - \frac{1}{n}\left[\sqrt{n}\left(\frac{X_n}{n} - \frac{1}{2}\right)\right]^2.$$

Rearrangement and continuous mapping theorem then give that

$$n\left(\delta_n - \frac{1}{4}\right) = -\left[\sqrt{n}\left(\frac{X_n}{n} - \frac{1}{2}\right)\right]^2 \Rightarrow X := -\xi^2.$$

That is, $a_n = n, b_n = 1/4$ and X is defined as above. $\qquad\square$

5.3 (a) *Proof.* By the classical CLT, $\sqrt{n}(X_n/n - p) \Rightarrow \mathcal{N}(0, p(1-p))$. If $p \neq 1/2$, then g

is differentiable at p and $|g'(p)| = 1 \neq 0$, it follows by the Delta method that

$$\sqrt{n}(g(X_n/n) - g(p)) \Rightarrow \mathcal{N}(0, p(1-p)).$$

If $p = 1/2$, then $|g(X_n/n) - g(1/2)| = |X_n/n - 1/2|$. Since the function $x \mapsto |x|$

is continuous on $\mathbb{R}^1$, and $\sqrt{n}(X_n/n - 1/2) \Rightarrow \mathcal{N}(0, 1/4)$. Let $Z \sim \mathcal{N}(0, 1)$, it

follows by the continuous mapping theorem that

$$\sqrt{n}|g(X_n/n) - g(1/2)| \Rightarrow \frac{1}{2}|Z|.$$

Or equivalently, $4n(g(X_n/n) - g(1/2))^2 \Rightarrow \chi_1^2$. $\quad\square$

(b) *Proof.* Note that

$$\frac{\mathrm{d}}{\mathrm{d}x}h(x) = \frac{\mathrm{d}}{\mathrm{d}x}\arcsin(\sqrt{x}) = \frac{1}{2\sqrt{x(1-x)}} \neq 0, 0 < x < 1.$$

By the Delta method and $\sqrt{n}(X_n/n - p) \Rightarrow \mathcal{N}(0, p(1-p))$, we have

$$\sqrt{n}(h(X_n/n) - h(p)) \Rightarrow \mathcal{N}(0, 1/4).$$

Thus $h(x) = \arcsin(\sqrt{x})$ is a variance-stablizing transformation for X_n/n. $\quad\square$

5.4 *Proof.* For notational convenience, denote $\sum_{i=1}^{n} X_i^k/n$ by $\overline{X_n^k}, k = 1, 2$. By condition,
$E[X_1] = \mu, E[X_1^2] = \mu^2 + \sigma^2, \mathrm{Var}(X_1) = \sigma^2, \mathrm{Var}(X_1^2) = 2\sigma^4 + 4\mu^2\sigma^2, \mathrm{Cov}(X_1, X_1^2) = 2\mu\sigma^2$. It then follows by the multivariate CLT that

$$\sqrt{n}\left[\begin{pmatrix}\overline{X}_n \\ \overline{X_n^2}\end{pmatrix} - \begin{pmatrix}\mu \\ \mu^2 + \sigma^2\end{pmatrix}\right] \Rightarrow \mathcal{N}\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\mu^2\sigma^2\end{pmatrix}\right).$$

To facilitate the Delta method, define $g(x, y) := \sqrt{y - x^2}/x, (x, y) \in D = \{(x, y) \in \mathbb{R}^2 : x \neq 0, y \neq x^2\}$. It is easy to verify that g is everywhere differentiable on its

domain, and

$$\nabla g(x, y) = \begin{pmatrix} -\dfrac{y}{x^2\sqrt{y-x^2}} \\ \dfrac{1}{2x\sqrt{y-x^2}} \end{pmatrix}.$$

Under the condition $\mu \neq 0$ (also $\sigma^2 > 0$), $\nabla g(\mu, \mu^2 + \sigma^2)$ is well-defined. Furthermore, $g(\overline{X}_n, \overline{X^2}_n) = \dfrac{S_n}{\overline{X}_n}$, $g(\mu, \mu^2 + \sigma^2) = \dfrac{\sigma}{\mu}$, and

$$\nabla^T g(\mu, \mu^2 + \sigma^2) \begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\mu^2\sigma^2 \end{pmatrix} \nabla g(\mu, \mu^2 + \sigma^2)$$

$$= \frac{1}{\sigma^2\mu^4} \begin{pmatrix} -(\mu^2 + \sigma^2) & \dfrac{1}{2}\mu \end{pmatrix} \begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\mu^2\sigma^2 \end{pmatrix} \begin{pmatrix} -(\mu^2 + \sigma^2) \\ \dfrac{1}{2}\mu \end{pmatrix}$$

$$= \frac{\sigma^4}{\mu^4} + \frac{\sigma^2}{2\mu^2}.$$

Therefore,

$$\sqrt{n}\left(\frac{S_n}{\overline{X}_n} - \frac{\sigma}{\mu} \right) \Rightarrow \boxed{\mathcal{N}\left(0, \frac{\sigma^4}{\mu^4} + \frac{\sigma^2}{2\mu^2} \right)}.$$

This completes the proof. $\square$

5.9 (a) *Proof.* Since $E[X_1] = E[Y_1] = 1/2, \mathrm{Var}(X_1) = \mathrm{Var}(Y_1) = 1/4, \mathrm{Cov}(X_1, Y_1) = \theta - 1/4$, it follows by multivariate CLT that

$$\sqrt{n}\left[\begin{pmatrix} \overline{X}_n \\ \overline{Y}_n \end{pmatrix} - \begin{pmatrix} \dfrac{1}{2} \\ \dfrac{1}{2} \end{pmatrix} \right] \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{4} \begin{pmatrix} 1 & 4\theta - 1 \\ 4\theta - 1 & 1 \end{pmatrix} \right).$$

$\square$

(b) *Proof.* It suffices to find Σ^* and A in Hunter's notes, but we must pay close attention that results such as (5.8) and (5.9) there hold only when $E[X] = E[Y] = 0$, but in this problem, $E[X] = E[Y] = 1/2$. In order to use (5.8) and (5.9), X and Y need to be centered by $X' = X - 1/2, Y' = Y - 1/2$. Also note that both r_n and ρ are invariant under this transformation. For the transformed sample,

direct calculation gives that

$$\Sigma^* = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\theta - \theta^2 \end{pmatrix}, \ A = \begin{pmatrix} -2\rho & -2\rho & 4 \end{pmatrix}.$$

hence

$$A\Sigma^* A^T = \begin{pmatrix} -2\rho & -2\rho & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\theta - \theta^2 \end{pmatrix} \begin{pmatrix} -2\rho \\ -2\rho \\ 4 \end{pmatrix} = 8\theta - 16\theta^2 = 1 - \rho^2.$$

Thus the asymptotic distribution of $\sqrt{n}(r_n - \rho)$ is $\boxed{\mathcal{N}(0, 1 - \rho^2)}$. $\square$

(c) *Proof.* Based on the asymptotic distribution in part (b), a variance-stablizing transformation for r_n should be a function f such that

$$f'(x) = \frac{1}{\sqrt{1 - x^2}}, 0 < x < 1.$$

Clearly, then $f(x) = \arcsin(x), 0 < x < 1$ (or $\sin^{-1}(x)$), which would give us

$$\sqrt{n}(\arcsin(r_n) - \arcsin(\rho)) \Rightarrow \mathcal{N}(0, 1).$$

This completes the proof. $\square$

Homework 10

6.4 *Proof.* Given $x > 0, y > 0$, for sufficiently large n, $(x+y)/n < 1$. By hint, it can be shown that

$$P[nX_{(2)} > x, n(1 - X_{(n-1)}) > y]$$

$$=P[x/n < X_{(2)} < X_{(n-1)} < 1 - y/n]$$

$$=\frac{n!}{0!n!0!}\left(1 - \frac{x}{n} - \frac{y}{n}\right)^n + \frac{n!}{1!(n-1)!0!}\frac{x}{n}\left(1 - \frac{x}{n} - \frac{y}{n}\right)^{n-1} +$$

$$\frac{n!}{0!(n-1)!1!}\left(1 - \frac{x}{n} - \frac{y}{n}\right)^{n-1}\frac{y}{n} + \frac{n!}{1!(n-2)!1!}\frac{x}{n}\left(1 - \frac{x}{n} - \frac{y}{n}\right)^{n-2}\frac{y}{n}$$

$$=\left(1 - \frac{x+y}{n}\right)^n + (x+y)\left(1 - \frac{x+y}{n}\right)^{n-1} + \frac{n(n-1)}{n^2}xy\left(1 - \frac{x+y}{n}\right)^{n-1}$$

$$\to e^{-(x+y)} + (x+y)e^{-(x+y)} + xye^{-(x+y)} = (x+1)e^{-x} \times (y+1)e^{-y}$$

as $n \to \infty$. Note that $(x+1)e^{-x}$ is the tail probability $P[X > x]$ of $X \sim \Gamma(2,1)$. Hence let X, Y be i.i.d. $\Gamma(2,1)$ random variables, then above derivation shows that

$$\begin{pmatrix} nX_{(2)} \\ n(1 - X_{(n-1)}) \end{pmatrix} \Rightarrow \begin{pmatrix} X \\ Y \end{pmatrix}.$$

This completes the proof. $\qquad\square$

6.5 The most convenient way to prove the result is to use the proposition below:

Proposition. *Suppose that $X_1, \ldots, X_n$ i.i.d. $\sim F$, where F is continuous, then*

$$(X_{(1)}, \ldots, X_{(n)}) \overset{d}{=} (F^-(U_{(1)}), \ldots, F^-(U_{(n)})),$$

where F^- is the quantile function [7] of F and $(U_{(1)}, \ldots, U_{(n)})$ is the order statistic of $U_1, \ldots, U_n$ i.i.d. $\sim \mathcal{U}(0,1)$.

The notes applied this result without giving a proof, so I supplement the proof here.

---

[7]See Homework 6 for its definition and basic properties.

Proof of the Proposition. Define $U_i = F(X_i), i = 1, \ldots, n$, then $U_1, \ldots, U_n$ i.i.d. $\sim$ $\mathcal{U}(0,1)$. Therefore, for the $(U_1, \ldots, U_n)$ such defined, and for any $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$P[F^-(U_{(1)}) \leq x_1, \ldots, F^-(U_{(n)}) \leq x_n]$$

$$= P[U_{(1)} \leq F(x_1), \ldots, U_{(n)} \leq F(x_n)] \text{ (by (6) in Homework 6)}$$

$$= P[F(X_{(1)}) \leq F(x_1), \ldots, F(X_{(n)}) \leq F(x_n)] \text{ (by construction)}$$

$$= P[F(X_{(1)}) < F(x_1), \ldots, F(X_{(n)}) < F(x_n)] \text{ (by the continuity of } F)$$

$$= P[X_{(1)} < x_1, \ldots, X_{(n)} < x_n] \text{ (by the monotonicity of } F)$$

$$= P[X_{(1)} \leq x_1, \ldots, X_{(n)} \leq x_n] \text{ (again by the continuity of } F)$$

This completes the proof of the proposition. $\qquad\square$

(a) *Proof.* In this problem, F is continuous, with

$$F^-(u) = \frac{1}{1-u}, \quad 0 < u < 1.$$

Therefore, by the proposition, $(X_{(n-1)}/n, X_{(n)}/n) \overset{d}{=} (1/(n(1-U_{(n-1)})), 1/(n(1-U_{(n)})))$. By the result of Example 6.4, we conclude that

$$\begin{bmatrix} \dfrac{X_{(n-1)}}{n} \\ \dfrac{X_{(n)}}{n} \end{bmatrix} \Rightarrow \begin{bmatrix} \dfrac{1}{Y_1 + Y_2} \\ \dfrac{1}{Y_1} \end{bmatrix},$$

where Y_1, Y_2 are independent standard exponential rando variables. $\qquad\square$

(b) *Proof.* By the continuous mapping theorem, it follows that

$$\frac{X_{(n-1)}}{X_{(n)}} = \frac{X_{(n-1)}/n}{X_{(n)}/n} \Rightarrow \frac{1/(Y_1 + Y_2)}{1/Y_1} = \frac{Y_1}{Y_1 + Y_2} \sim \mathcal{U}(0,1).$$

That $\dfrac{Y_1}{Y_1 + Y_2} \sim \mathcal{U}(0,1)$ can be verified by computing the distribution function of $\dfrac{Y_1}{Y_1 + Y_2}$ directly. $\qquad\square$

59

6.6 *Proof.* Under the assumption of the problem, we can show (one expedient way is to use the representation given in Problem 1) that in fact $X_{(1)}/X_{(2)} \overset{d}{=} \mathcal{U}(0,1)$, for any n, which is stronger than $X_{(1)}/X_{(2)} \Rightarrow \mathcal{U}(0,1)$. The details will be omitted. $\square$

6.7 (a) *Proof.* We shall proceed in a similar way to the Problem 6.5. First it is straight-forward to find the quantile function of F is given by

$$F^-(u) = \theta \log\left(\frac{u}{1-u}\right) = \theta \log u - \theta \log(1-u), \quad 0 < u < 1.$$

Therefore,

$$\theta^{-1}\begin{bmatrix} X_{(n-1)} \\ X_{(n)} \end{bmatrix} \overset{d}{=} \begin{bmatrix} \log(U_{(n-1)}) \\ \log(U_{(n)}) \end{bmatrix} - \begin{bmatrix} \log(1 - U_{(n-1)}) \\ \log(1 - U_{(n)}) \end{bmatrix}.$$

In Example 6.5 we have shown that

$$\begin{bmatrix} -\log(n(1 - U_{(n-1)})) \\ -\log(n(1 - U_{(n)})) \end{bmatrix} \Rightarrow \begin{bmatrix} -\log(Y_1 + Y_2) \\ -\log(Y_1) \end{bmatrix},$$

where Y_1, Y_2 are independent standard exponential random variables.

On the other hand, it can be shown that $U_{(n-1)} \to_P 1$ and $U_{(n)} \to_P 1$ as $n \to \infty$. To see this, given $\varepsilon \in (0,1)$,

$$P[|U_{(n-1)} - 1| \geq \varepsilon] = P[1 - U_{(n-1)} \geq \varepsilon] = P[U_{(n-1)} \leq 1 - \varepsilon]$$

$$= P[\text{At least } n - 1 \text{ of } U_1, \ldots, U_n \leq 1 - \varepsilon]$$

$$= (1 - \varepsilon)^n + \binom{n}{n-1}(1 - \varepsilon)^{n-1}\varepsilon \to 0$$

as $n \to \infty$. The proof of $U_{(n)} \to_P 1$ is similar (and simpler). It then follows that $(\log(U_{(n-1)}), \log(U_{(n)}))' \to_P 0$, in view of the continuous mapping theorem.

Together, by the Slutsky's theorem, we conclude that

$$
\theta^{-1} \begin{bmatrix} X_{(n-1)} - \theta \log n \\ X_{(n)} - \theta \log n \end{bmatrix} \stackrel{d}{=} \begin{bmatrix} -\log(n(1 - U_{(n-1)})) \\ -\log(n(1 - U_{(n)})) \end{bmatrix} + o_P(1) \Rightarrow \begin{bmatrix} -\log(Y_1 + Y_2) \\ -\log(Y_1) \end{bmatrix},
$$

where Y_1, Y_2 are independent standard exponential random variables. Thus the continuous mapping theorem implies that

$$
\theta^{-1}(X_{(n)} - X_{(n-1)}) \Rightarrow \log(Y_1 + Y_2) - \log(Y_1) = -\log\left(\frac{Y_1}{Y_1 + Y_2}\right).
$$

In the Problem 6.5, we mentioned that if Y_1, Y_2 i.i.d. $\sim \text{Exp}(1)$, then $Y_1/(Y_1 + Y_2) \sim \mathcal{U}(0,1)$, which further implies that $-\log(Y_1/(Y_1 + Y_2)) \sim \text{Exp}(1)$. Therefore, we conclude that

$$
X_{(n)} - X_{(n-1)} \Rightarrow \theta \text{Exp}(1). \tag{23}
$$

This completes the proof. $\qquad\square$

(b) The expression (23) and the information given in part (b) imply that the asymptotic 95% confidence interval for θ is

$$
\left[\frac{X_{(n)} - X_{(n-1)}}{3.6889}, \frac{X_{(n)} - X_{(n-1)}}{0.0253} \right] = [0.271(X_{(n)} - X_{(n-1)}), 39.50(X_{(n)} - X_{(n-1)})],
$$

where $X_{(n-1)}$ and $X_{(n)}$ are computed from the observed sample.

6.8 (a) *Proof.* If $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{U}(0, 2\theta)$, then $X_1/(2\theta), \ldots, X_n/(2\theta)$ i.i.d. $\sim \mathcal{U}(0,1)$. In Example 6.3 it has been shown that $n(X_{(1)}/(2\theta), 1 - X_{(n)}/(2\theta))^T \Rightarrow (Y_1, Y_2)$, where Y_1, Y_2 i.i.d. $\sim \text{Exp}(1)$, it then follows by Cramér-Wold device that

$$
n(M - \theta) = (\theta, -\theta)n \begin{pmatrix} \frac{1}{2\theta} X_{(1)} \\ 1 - \frac{1}{2\theta} X_{(n)} \end{pmatrix} \Rightarrow (\theta, -\theta) \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \theta(Y_1 - Y_2).
$$

It can be verified that the pdf of $\theta(Y_1 - Y_2)$ is $f(x) = e^{-|x|/\theta}/(2\theta), x \in \mathbb{R}$, i.e., $\theta(Y_1 - Y_2)$ follows *Laplace distribution* with scale parameter θ. $\qquad\square$

(b) *Proof.* To answer this problem, the term "*Asymptotic Relative Efficiency (ARE)*" needs a rigorous definition. The definition I adopted here is an adaptation of [3], p.91, where this term is only defined for the case that limiting distributions are all normal and the convergence rate are all $n^{-1/2}$. However, in this problem, as we have seen in part (a), the limiting distribution of M is not normal, and its convergence rate is n^{-1} instead of $n^{-1/2}$. To address this issue, let me introduce the following definition:

Definition. A sequence of random variables $\{T_n\}$ is said to have the *asymptotic variance* $v_n(T_n) \equiv v_n$, if there exist a constant sequence $\{\mu_n\}$ and a nondegenerate random variable T with $\mathrm{Var}(T) = 1$ such that

$$\frac{T_n - \mu_n}{\sqrt{v_n}} \Rightarrow T.$$

Given two estimators T_{1n} and T_{2n} of the parameter θ, the *asymptotic relative efficiency* of T_{2n} relative to T_{1n} is defined as

$$\mathrm{ARE}(T_{2n}, T_{1n}) = \lim_{n \to \infty} \frac{v_n(T_{1n})}{v_n(T_{2n})}. \tag{24}$$

Remark: This definition may not have appeared in literature, it is tailored for this particular question by myself. My justification for this definition is that when T_{1n} and T_{2n} do converge to normal distributions with the rate $n^{-1/2}$, my definition agrees with the conventional one such as in [3]. Consequently, your answer could be different from mine (in fact, for this problem, it's quite obvious that M is the best estimator among three candidates in that its approximated variance is smallest). If someone found a mathematically rigorous definition of ARE under general settings, please let me know.

For a systematic introduction to ARE in estimation, see also [7].

Since the variance of a Laplace random variable with scale parameter θ is $2\theta^2$,

by the result of part (a), we have

$$\frac{M - \theta}{\sqrt{2\theta^2/n^2}} \Rightarrow Y_1 - Y_2.$$

Thus $v_n(M) = \boxed{2\theta^2/n^2}$.

By classical CLT, it follows that

$$\frac{\overline{X}_n - \theta}{\sqrt{\theta^2/(3n)}} \Rightarrow Z,$$

where $Z \sim \mathcal{N}(0,1)$. Hence $v_n(\overline{X}_n) = \boxed{\theta^2/(3n)}$.

Finally, by Theorem 6.6, we have

$$\frac{\tilde{X}_n - \theta}{\sqrt{\theta^2/n}} \Rightarrow Z,$$

where $Z \sim \mathcal{N}(0,1)$. Hence $v_n(\tilde{X}_n) = \boxed{\theta^2/n}$.

Therefore, by definition (24), we have

$$\mathrm{ARE}(\overline{X}_n, M) = \lim_{n \to \infty} \frac{2\theta^2/n^2}{\theta^2/(3n)} = 0,$$

$$\mathrm{ARE}(\overline{X}_n, \tilde{X}_n) = \lim_{n \to \infty} \frac{\theta^2/n}{\theta^2/(3n)} = 3,$$

$$\mathrm{ARE}(\tilde{X}_n, M) = \lim_{n \to \infty} \frac{2\theta^2/n^2}{\theta^2/n} = 0.$$

Thus in terms of smaller asymptotic variances, M is the most efficient estimator of θ, $\overline{X}_n$ second, and $\tilde{X}_n$ is the least efficient. $\qquad\square$

6.11 To be clear, here Q_i stands for the $i/4$-th sample quantile, $i = 1, 2, 3$.

(a) *Proof.* For $\mathcal{N}(0,1)$ distribution, since $\varphi(\Phi^-(0.25)) = 0.318$, $\varphi(\Phi^-(0.5)) = 0.399$, $\varphi(\Phi^-(0.75)) = 0.318$, it then follows by Theorem 6.7 (use R to do the computa-

tion) that

$$\sqrt{n}\left(\begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} - \begin{bmatrix} 1/4 \\ 1/2 \\ 3/4 \end{bmatrix}\right) \Rightarrow \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.8568 & 0.9860 & 0.6189 \\ 0.9860 & 1.5708 & 0.9860 \\ 0.6189 & 0.9860 & 1.8568 \end{bmatrix}\right).$$

By Cramér-Wold device (or the Delta-method), we then have

$$\sqrt{n}((Q_3 - Q_2) - (Q_2 - Q_1)) \Rightarrow \mathcal{N}(0, 3.3465).$$

This completes the proof. □

(b) *Proof.* For the standard logistic distribution, with f and F denoting its density function and distribution function, we have (again, use R!)

$$f(F^-(0.25)) = 0.1875, f(F^-(0.5)) = 0.25, f(F^-(0.75)) = 0.1875.$$

Therefore,

$$\sqrt{n}\left(\begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} - \begin{bmatrix} 1/4 \\ 1/2 \\ 3/4 \end{bmatrix}\right) \Rightarrow \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5.3333 & 2.6667 & 1.7778 \\ 2.6667 & 4 & 2.6667 \\ 1.7778 & 2.6667 & 5.3333 \end{bmatrix}\right).$$

The Cramér-Wold device then asserts that

$$\sqrt{n}((Q_3 - Q_2) - (Q_2 - Q_1)) \Rightarrow \mathcal{N}(0, 8.8889).$$

This completes the proof. □

6.12 *Proof.* Denote $\lfloor 0.25n+0.5 \rfloor, \lfloor 0.5n+0.5 \rfloor, \lfloor 0.75n+0.5 \rfloor$ by a_n, b_n, c_n, respectively. Thus

median $= X_{(b_n)} \overset{d}{=} 2\theta U_{(b_n)}$, midquartile $= (X_{(a_n)} + X_{(c_n)})/2 \overset{d}{=} \theta(U_{(a_n)} + U_{(c_n)}), \frac{2}{3}Q_3 =$

$\frac{2}{3}X_{(c_n)} = \frac{4\theta}{3}U_{(c_n)}$, where $U_1, \ldots, U_n$ i.i.d. $\sim \mathcal{U}(0,1)$. By Theorem 6.6:

$$\sqrt{n}\left(\begin{bmatrix} U_{(a_n)} \\ U_{(b_n)} \\ U_{(c_n)} \end{bmatrix} - \begin{bmatrix} 1/4 \\ 1/2 \\ 3/4 \end{bmatrix}\right) \Rightarrow \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3/16 & 1/8 & 1/16 \\ 1/8 & 1/4 & 1/8 \\ 1/16 & 1/8 & 1/16 \end{bmatrix}\right).$$

It thus follows that

$$\sqrt{n}(\text{median} - \theta) \Rightarrow \mathcal{N}(0, \theta^2),$$

$$\sqrt{n}(\text{midquartile} - \theta) \Rightarrow \mathcal{N}(0, \theta^2/2),$$

$$\sqrt{n}(2Q_3/3 - \theta) \Rightarrow \mathcal{N}(0, \theta^2/3).$$

Therefore $2Q_3/3$ has the smallest asymptotic variance, the midquartile second, the median largest. $\qquad\square$

Homework 11

7.3 *Proof.* **(a)** It is easy to verify that the log likelihood function (up to some constant independent of θ) is $\ell(\theta) = \theta^{-1}an - a\theta^{a-1}\sum_{i=1}^{n} x_i^a$. Thus the likelihood equation has a closed form solution $\hat{\theta} = n^{1/a}/(\sum_{i=1}^{n} x_i^a)^{1/a}$. Thus it is unique. Since $\ell''(\theta) = -\theta^{-2}an - a(a-1)\theta^{a-2}\sum_{i=1}^{n} x_i^a$, it follows that $\ell''(\hat{\theta}) = -a^2n/\hat{\theta}^2 < 0$, whence $\hat{\theta}$ maximizes the likelihood function.

(b) Since $\ell(\theta) = n\log\theta - \sum_{i=1}^{n} \log(x_i^2 + \theta^2)$, it follows that $\ell'(\theta) = n\theta^{-1} - \sum_{i=1}^{n} \frac{2\theta}{x_i^2+\theta^2}$. Since $\theta > 0$, the likelihood equation can be equivalently written as $g(\theta) := n - \sum_{i=1}^{n} \frac{2\theta^2}{x_i^2+\theta^2} = 0$. Clearly, $g(\theta)$ is a continuous function of θ in $(0, +\infty)$, with the property $\lim_{\theta\to 0} g(\theta) = n > 0$ and $\lim_{\theta\to+\infty} g(\theta) = -n < 0$. It thus follows by the intermediate value theorem of the continuous function (Theorem 4.23, [5]) that there exists some $\hat{\theta} \in (0, +\infty)$ such that $g(\hat{\theta}) = 0$. In addition, $g'(\theta) = -\sum_{i=1}^{n} \frac{4\theta x_i^2}{(x_i^2+\theta^2)^2} < 0$ for all $\theta > 0$, implying that $g(\theta)$ is strictly decreasing in $(0, +\infty)$, thus $\hat{\theta}$ is also the unique solution to the likelihood equation. The negativity of $g'(\theta)$ also insures that $\hat{\theta}$ is the maximizer.

(c) After recognizing $\ell(\theta) = 2n\log\theta - \sum_{i=1}^{n} \log(x_i^3 + \theta^3)$ and $g(\theta) = 2n - \sum_{i=1}^{n} \frac{3\theta^3}{x_i^3+\theta^3}$, the argument is parallel with part (b). The details are thus omitted here. $\square$

7.4 *Proof.* Since $\ell(\theta) = n\log(1-\theta) + n\log\theta\overline{X}_n$, $\ell'(\theta) = -\frac{n}{1-\theta} + \frac{n\overline{X}_n}{\theta}$, thus the maximum likelihood estimate of θ is $\hat{\theta}_{\text{MLE}} = \frac{\overline{X}_n}{1+\overline{X}_n}$.

By the classical central limit theorem, we have

$$\sqrt{n}\left(\overline{X}_n - \frac{\theta}{1-\theta}\right) \Rightarrow \mathcal{N}(0, \theta/(1-\theta)^2).$$

In which, the mean and variance are easily calculated from f_θ. It then follows by the Delta-method (take $g(x) = x/(1+x)$, $x > 0$) that

$$\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta) \Rightarrow \mathcal{N}(0, \theta(1-\theta)^2).$$

This completes the proof. □

7.6 *Proof.* It is straightforward by definition to calculate that

$$I(\theta) = -E_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X) \right] = \frac{9\theta^4}{\sigma^2}. \tag{25}$$

Personally, the second part of this question is at least unclear — the Fisher information (25) evidently doesn't depend on any particular observation of X, since it is obtained by averaging over the whole probability space (i.e., expectation operator has been exerted already). The author might want to express that *"the observed Fisher information can be zero..."*. By definition, given an observed sample $\{x_1, \ldots, x_n\}$, the *observed (Fisher) information* is defined as

$$J(\theta) = -\sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f_\theta(x_i).$$

By contrast to $I(\theta)$, $J(\theta)$ indeed depends on particular observed values. For this problem, if we happened to have only one observation $x = 5\theta^3/2$, then $J(\theta) = -3(2\theta x - 5\theta^4)/\sigma^2 \big|_{x=5\theta^3/2} = 0$. □

7.7 *Proof.* **(a)** To show $P_{\theta_0}[\hat\theta_n$ is a local maximum$] \to 1$, since ℓ is twice differentiable, it suffices to show that $P_{\theta_0}[\ell''(\hat\theta_n) < 0] \to 1$. By assumption, we may assume $\sup_{\theta \in \Theta} |\ell'''(\theta)| \le nM$ for some $M > 0$. It follows by the mean value theorem that

$$|\ell''(\hat\theta_n) - \ell''(\theta_0)| \le nM|\hat\theta_n - \theta_0|.$$

Combining this with $\hat\theta_n \to_P \theta_0$, it follows that $(\ell''(\hat\theta_n) - \ell''(\theta_0))/n = o_P(1)$. Additionally, we have

$$\frac{1}{n}\ell''(\theta_0) \to_P -I(\theta_0) < 0.$$

Thus by the Slutsky's theorem, $\ell''(\hat\theta_n)/n \to_P -I(\theta_0)$, hence

$$P_{\theta_0}[\ell''(\hat\theta_n) < 0] \ge P_{\theta_0}\left[\left|\frac{1}{n}\ell''(\hat\theta_n) + I(\theta_0)\right| < \frac{1}{2}I(\theta_0)\right] \to 1$$

as $n \to \infty$ (note that the condition $0 < I(\theta_0) < \infty$ is crucial).

(b) If $P_{\theta_0}[\hat{\theta}_{1n} = \hat{\theta}_{2n}] \not\to 1$, then there would exist an $\varepsilon_0 > 0$ and an infinite subsequence $\mathscr{K} \subset \mathbb{N}$ such that $P_{\theta_0}[\hat{\theta}_{1m} \neq \hat{\theta}_{2m}] \geq \varepsilon_0$ for each $m \in \mathscr{K}$. By the result of part (a), $P_{\theta_0}[\hat{\theta}_{km}$ is a local maximum$] \to 1$ as $m \to \infty$, $k = 1, 2$. Denote by E_m the event $[\hat{\theta}_{1m}$ is a local maximum$] \cap [\hat{\theta}_{2m}$ is a local maximum$] \cap [\hat{\theta}_{1m} \neq \hat{\theta}_{2m}]$, $m \in \mathscr{K}$. Then $\omega \in E_m$ and the continuity of ℓ imply that there exists a point between $\hat{\theta}_{1m}(\omega)$ and $\hat{\theta}_{2m}(\omega)$, say $\hat{\theta}_{3m}(\omega)$, such that $\hat{\theta}_{3m}(\omega)$ is a local minimum of ℓ, which further implies that $\ell'(\hat{\theta}_{3m}) = 0$. Additionally, $\hat{\theta}_{km} \to_P \theta_0$, $k = 1, 2$ imply that $\hat{\theta}_{3m} \to_P \theta_0$. Therefore $\{\hat{\theta}_{3m}\}$ is also a sequence of consistent roots of the likelihood equation. Hence by the result of part (a),

$$\limsup_{m \to \infty} P_{\theta_0}[\hat{\theta}_{3m} \text{ is a local minimum}] = 0.$$

On the other hand, $E_m \subset [\hat{\theta}_{3m}$ is a local minimum$]$ and $\liminf_{m \to \infty} P_{\theta_0}(E_m) \geq \varepsilon_0$ imply that

$$\liminf_{m \to \infty} P_{\theta_0}[\hat{\theta}_{3m} \text{ is a local minimum}] \geq \varepsilon_0.$$

This contradiction shows that we must have $P_{\theta_0}[\hat{\theta}_{1n} = \hat{\theta}_{2n}] \to 1$, and the proof is complete. $\qquad\square$

7.11 *Proof.* **(a)** It is straightforward to verify that the second derivative of $\log f_\theta(x)$ is given by $2[(x - \theta)^2 - 1]/[1 + (x - \theta)^2]^2$. In addition, note that if $X \sim f_\theta(x)$, then $X - \theta \sim \text{Cauchy}(0, 1)$, i.e., the standard Cauchy distribution. Therefore,

$$\begin{aligned}
E&\left[\frac{(X - \theta)^2 - 1}{[1 + (X - \theta)^2]^2}\right]\\
&= \int_{-\infty}^{\infty} \frac{x^2 - 1}{(1 + x^2)^2} \cdot \frac{1}{\pi(1 + x^2)} \, dx\\
&= \frac{2}{\pi} \int_0^{\infty} \frac{1}{(1 + x^2)^2} \, dx - \frac{4}{\pi} \int_0^{\infty} \frac{1}{(1 + x^2)^3} \, dx\\
&= \frac{2}{\pi} \int_0^{\pi/2} \cos^2 \theta \, d\theta - \frac{4}{\pi} \int_0^{\pi/2} \cos^4 \theta \, d\theta\\
&= \frac{1}{2} - \frac{3}{4} = -\frac{1}{4}.
\end{aligned}$$

From the second line to the third line of the above calculation, we did the variable substitution $x = \tan\theta$. From the third line to the fourth line, the exponent order is decreased by using the double-angle identity $\cos^2\theta = (\cos(2\theta) - 1)/2$. In conclusion, $I(\theta) = 1/2$.

(b) Since the population median of the Cauchy$(\theta, 1)$ distribution is θ, it follows by Theorem 6.6 that

$$\sqrt{n}(\tilde{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, 1/4f(\theta)^2) = \mathcal{N}(0, \pi^2/4).$$

On the other hand, since we have shown in the text that the one-step estimator $\delta_n^* = \tilde{\theta}_n + 2\ell'(\tilde{\theta}_n)/n$ is asymptotically efficient, it follows that

$$\sqrt{n}(\delta_n^* - \theta) \Rightarrow \mathcal{N}(0, I(\theta)^{-1}) = \mathcal{N}(0, 2),$$

in view of the result obtained in part (a). $\square$

7.14 *Proof.* Since $P[X = x] = P[X_1 = x_1, \ldots, X_k = x_k] = \dfrac{1!}{x_1! \ldots x_k!} p_1^{x_1} \cdots p_k^{x_k}$, the log-likelihood function is

$$\ell(p^*) = x_1 \log p_1 + \cdots + x_k \log p_k.$$

Although written in this way, keep in mind that p_k should be understood as an implicit function of $p^* = (p_1, \ldots, p_{k-1})$. In particular, $\partial p_k/\partial p_j = -1$, $j = 1, \ldots, k - 1$. It is then straightforward to verify that

$$\frac{\partial \ell}{\partial p_j} = \frac{x_j}{p_j} - \frac{x_k}{p_k},$$

$$\frac{\partial^2 \ell}{\partial p_j^2} = -\frac{x_j}{p_j^2} - \frac{x_k}{p_k^2}, \ j = 1, \ldots, k - 1.$$

$$\frac{\partial^2 \ell}{\partial p_j \partial p_{j'}} = -\frac{x_k}{p_k^2}, \ 1 \le j \ne j' \le k - 1.$$

Therefore $E[X_j] = p_j, j = 1, \ldots, k$ implies that

$$I(p^*) = \begin{bmatrix} \dfrac{1}{p_1} + \dfrac{1}{p_k} & \cdots & \dfrac{1}{p_k} \\ \vdots & \ddots & \vdots \\ \dfrac{1}{p_k} & \cdots & \dfrac{1}{p_{k-1}} + \dfrac{1}{p_k} \end{bmatrix} = \Lambda + \mathbf{1}\mathbf{1}^T/p_k,$$

where $\Lambda = diag(p_1^{-1}, \ldots, p_{k-1}^{-1})$, $\mathbf{1} = (1, \ldots, 1)^T$. This completes the proof. $\qquad \square$

7.19 *Proof.* The necessity part is trivial in view of the Cramér-Wold device.

To prove the sufficiency part, assume asymptotic normality holds for $\hat{\boldsymbol{\theta}}_n$, that is

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \Rightarrow \mathcal{N}(0, \Sigma)$$

for some covariance matrix Σ. Since $I^{-1}(\boldsymbol{\theta})$ is the C-R lower bound, we have $\Sigma \geq I^{-1}(\boldsymbol{\theta})$ (here and in the sequel, "$A \geq B$" means the matrix $A - B$ is symmetric *and* non-negative definite). By assumption, the diagonal elements of Σ and $I^{-1}(\boldsymbol{\theta})$ all equal to $\gamma_{ii}, i = 1, \ldots, k$. Thus $\Sigma - I^{-1}(\boldsymbol{\theta})$ takes the form

$$\Delta := \Sigma - I^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} 0 & \Delta_{12} & \cdots & \Delta_{1k} \\ \Delta_{12} & 0 & \cdots & \Delta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{1k} & \Delta_{2k} & \cdots & 0 \end{pmatrix}.$$

Our goal is to show $\Delta \equiv 0$, which can be achieved by noting that Δ is symmetric and non-negative definite. In detail, we shall use the following important characterization of symmetric non-negative definite matrices:

Theorem. *If $\Delta \geq 0$, then all its principal minors are non-negative.*

70

By this theorem, since $\Delta \geq 0$, by considering principal minors

$$\Delta \begin{pmatrix} i & i+1 \\ i & i+1 \end{pmatrix} = \det \left(\begin{pmatrix} 0 & \Delta_{i,i+1} \\ \Delta_{i,i+1} & 0 \end{pmatrix} \right) = -\Delta_{i,i+1}^2 \geq 0, \; i = 1, 2, \ldots, n-1$$

successively, we deduce that $\Delta_{i,i+1} = 0, i = 1, 2, \ldots, n-1$. We then consider principal minors

$$\Delta \begin{pmatrix} i & i+2 \\ i & i+2 \end{pmatrix} = \det \left(\begin{pmatrix} 0 & \Delta_{i,i+2} \\ \Delta_{i,i+2} & 0 \end{pmatrix} \right) = -\Delta_{i,i+2}^2 \geq 0, \; i = 1, 2, \ldots, n-2$$

successively, it follows that $\Delta_{i,i+2} = 0, i = 1, 2, \ldots, n-2$. Proceed in this way, we found that all off-diagonal elements of Δ are zeros, thus $\Delta = 0$, $\Sigma = I^{-1}(\boldsymbol{\theta})$, which completes the proof. $\qquad \square$

Homework 12

Problem 1 *Proof.* See Lemma A–Lemma C in Section 4.4.2 of [6], where the general multivariate cases are treated. $\square$

8.1 *Proof.* Given the sample $\{X_1, \ldots, X_n\}$, it is straightforward to compute

$$\ell(\theta) = n \log \theta + n\theta \log c - \sum_{i=1}^{n} (\theta + 1) \log X_i,$$

$$\ell'(\theta) = \frac{n}{\theta} + n \log c - \sum_{i=1}^{n} \log X_i,$$

$$\ell''(\theta) = -\frac{n}{\theta^2}.$$

Denote $\sum_{i=1}^{n} \log X_i / n$ by $\overline{Y}_n$, then the MLE $\hat{\theta}_n = 1/(\overline{Y}_n - \log c)$, the Fisher information $I(\theta) = 1/\theta^2$. It then follows that

$$W_n = \sqrt{nI(\theta_0)}(\hat{\theta}_n - \theta_0) = \frac{\sqrt{n}}{\theta_0} \left(\frac{1}{\overline{Y}_n - \log c} - \theta_0 \right),$$

$$R_n = \frac{1}{\sqrt{nI(\theta_0)}} \ell'(\theta_0) = \sqrt{n}(1 + \theta_0 \log c - \theta_0 \overline{Y}_n),$$

$$\Delta_n = \ell(\hat{\theta}_n) - \ell(\theta_0) = -n \left[\log \theta_0 + \log(\overline{Y}_n - \log c) - 1 + \theta_0(\overline{Y}_n - \log c) \right].$$

We thus reject $H_0 : \theta = \theta_0$ in favor of $H_1 : \theta \neq \theta_0$ whenever $|W_n| \geq u_{\alpha/2}, |R_n| \geq u_{\alpha/2}$, or $2\Delta_n \geq \chi_1^2(\alpha)$. $\square$

8.2 (a) *Proof.* Since $\ell(\boldsymbol{\theta}) = \sum_{j=1}^{k} X_j \log p_j$ and $E[X_j] = np_j, j = 1, \ldots, k$, it follows that

$$I(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{p_1} + \frac{1}{p_k} & \cdots & \frac{1}{p_k} \\ \vdots & \ddots & \vdots \\ \frac{1}{p_k} & \cdots & \frac{1}{p_{k-1}} + \frac{1}{p_k} \end{pmatrix} = \Lambda + \mathbf{1}\mathbf{1}^T / p_k,$$

where $\Lambda = diag(p_1^{-1}, \ldots, p_{k-1}^{-1}), \mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^{k-1}$. Furthermore, if denoting

$(X_1, \ldots, X_{k-1})^T$ by $\boldsymbol{X}$, then

$$\nabla\ell(\boldsymbol{\theta}) = \left(\frac{X_1}{p_1} - \frac{X_k}{p_k} \quad \cdots \quad \frac{X_{k-1}}{p_{k-1}} - \frac{X_k}{p_k}\right)^T = \Lambda\boldsymbol{X} - \frac{X_k}{p_k}\mathbf{1}.$$

We first show that $W_n = \chi^2 \equiv \sum_{j=1}^{k} \frac{(X_j - np_j)^2}{np_j}$. It follows by $\hat{\boldsymbol{\theta}}_n = \boldsymbol{X}/n$ that

$$\begin{aligned}
W_n &= n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)^T I(\boldsymbol{\theta}^0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \\
&= \frac{1}{n}(\boldsymbol{X} - n\boldsymbol{\theta}^0)^T(\Lambda + \mathbf{1}\mathbf{1}^T/p_k)(\boldsymbol{X} - n\boldsymbol{\theta}^0)^T \\
&= \frac{1}{n}(\boldsymbol{X} - n\boldsymbol{\theta}^0)^T\Lambda(\boldsymbol{X} - n\boldsymbol{\theta}^0)^T + \frac{1}{np_k}(\boldsymbol{X} - n\boldsymbol{\theta}^0)^T\mathbf{1}\mathbf{1}^T(\boldsymbol{X} - n\boldsymbol{\theta}^0)^T \\
&= \sum_{j=1}^{k-1}\frac{(X_j - np_j)^2}{np_j} + \frac{(X_k - np_k)^2}{np_k} \\
&= \sum_{j=1}^{k}\frac{(X_j - np_j)^2}{np_j} = \chi^2.
\end{aligned}$$

We next show that $R_n = \chi^2$, for which we need to evaluate $I(\boldsymbol{\theta})^{-1}$ first. By Sherman-Morrison-Woodbury formula ([4], page 65): $(A + UV)^{-1} = A^{-1} - A^{-1}U(I + VA^{-1}U)^{-1}VA^{-1}$,

$$\begin{aligned}
I(\boldsymbol{\theta})^{-1} &= p_k(p_k\Lambda + \mathbf{1}\mathbf{1}^T)^{-1} \\
&= p_k\left[p_k^{-1}\Lambda^{-1} - p_k^{-1}\Lambda^{-1}\mathbf{1}\left(1 + \mathbf{1}^T p_k^{-1}\Lambda^{-1}\mathbf{1}\right)^{-1}\mathbf{1}^T p_k^{-1}\Lambda^{-1}\right] \\
&= diag(p_1, \ldots, p_{k-1}) - (p_1, \ldots, p_{k-1})^T\left(1 + \sum_{j=1}^{k-1}\frac{p_j}{p_k}\right)^{-1}(p_1, \ldots, p_{k-1})/p_k \\
&= \Lambda^{-1} - \boldsymbol{\theta}\boldsymbol{\theta}^T,
\end{aligned}$$

Noting that $\mathbf{1}^T\Lambda^{-1} = \boldsymbol{\theta}^{0T}$, $\Lambda\boldsymbol{\theta} = \mathbf{1}$, $\boldsymbol{\theta}^{0T}\mathbf{1} = 1 - p_k$, $\boldsymbol{X}^T\mathbf{1} = n - X_k$, we have

$$\begin{aligned}
R_n &= \frac{1}{n}\nabla\ell(\boldsymbol{\theta}^0)^T I^{-1}(\boldsymbol{\theta}^0)\nabla\ell(\boldsymbol{\theta}^0) \\
&= \frac{1}{n}\left(\Lambda\boldsymbol{X} - \frac{X_k}{p_k}\mathbf{1}\right)^T(\Lambda^{-1} - \boldsymbol{\theta}^0\boldsymbol{\theta}^{0T})\left(\Lambda\boldsymbol{X} - \frac{X_k}{p_k}\mathbf{1}\right) \\
&= \frac{1}{n}(\boldsymbol{X} - n\boldsymbol{\theta}^0)\left(\Lambda\boldsymbol{X} - \Lambda n\boldsymbol{\theta}^0 + \Lambda n\boldsymbol{\theta}^0 - \frac{X_k}{p_k}\mathbf{1}\right)
\end{aligned}$$

$$= \frac{1}{n}(\boldsymbol{X} - n\boldsymbol{\theta}^0)^T \Lambda (\boldsymbol{X} - n\boldsymbol{\theta}^0) + (\boldsymbol{X} - n\boldsymbol{\theta}^0)^T \frac{np_k - X_k}{np_k} \mathbf{1}$$

$$= \sum_{j=1}^{k} \frac{(X_j - np_j)^2}{np_j} = \chi^2.$$

In summary, $W_n = R_n = \chi^2$. $\qquad\qquad\Box$

(b) *Proof.* Since $\hat{\boldsymbol{\theta}}_n = \boldsymbol{X}/n$, it follows that

$$2\Delta_n = 2(\ell(\hat{\boldsymbol{\theta}}_n) - \ell(\boldsymbol{\theta}^0)) = 2\sum_{j=1}^{k} X_j \log\left(\frac{X_j}{np_j}\right).$$

This completes the proof. $\qquad\qquad\Box$

8.4 *Proof.* By Theorem 6.7 in Hunter's Notes, we have

$$\sqrt{n}\begin{pmatrix} Q_1 - \theta_0/2 \\ Q_3 - 3\theta_0/2 \end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{\theta_0^2}{4}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\right)$$

It then follows by Cramér-Wold's device that

$$\sqrt{n}(Q_3 - Q_1 - \theta_0) \Rightarrow \mathcal{N}(0, \theta_0^2).$$

For $\overline{X}_n$, by classical CLT,

$$\sqrt{n}(\overline{X}_n - \theta_0) \Rightarrow \mathcal{N}(0, \theta_0^2/3).$$

Based on above two limiting distributions, it is easy to see that

$$A_n = \frac{\theta_0 u_\alpha}{\sqrt{n}}, \quad B_n = \frac{\theta_0 u_\alpha}{\sqrt{3n}}.$$

To find the asymptotic relative efficiency of test A with respect to test B, we need to consider the asymptotic distributions of $\sqrt{n}(Q_3 - Q_1 - \theta_0)/\theta_0$ and $\sqrt{n}(\overline{X}_n - \theta_0)/(\theta_0/\sqrt{3})$ under local alternative hypotheses $H_{an} : \theta_n = \theta_0 + n^{-1/2}\Delta$ for some $\Delta > 0$.

We first show that

$$\frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\theta_0/\sqrt{3}} \Rightarrow_{\theta_n} \mathcal{N}(\sqrt{3}\Delta/\theta_0, 1), \tag{26}$$

where the notation "$\Rightarrow_{\theta_n}$" means weak convergence under probability measure P_{θ_n}. It is easy to check that the triangular array $\{X_{ni}, 1 \le i \le n, n \ge 1\}$ satisfies Lyapunov's condition with $\delta = 1$ (use the fact that X_{ni} is bounded by $2\theta_n$), thereby

$$\frac{\sqrt{n}(\overline{X}_n - \theta_n)}{\theta_n/\sqrt{3}} \Rightarrow_{\theta_n} \mathcal{N}(0, 1).$$

Now write

$$\frac{\sqrt{n}(\overline{X}_n - \theta_0)}{\theta_0/\sqrt{3}} = \frac{\sqrt{n}(\overline{X}_n - \theta_n)}{\theta_n/\sqrt{3}} \times \frac{\theta_n}{\theta_0} + \frac{\sqrt{n}(\theta_n - \theta_0)}{\theta_0/\sqrt{3}}.$$

Hence (26) holds in view of the above two expressions and Slutsky's theorem.

Next we show that

$$\frac{\sqrt{n}(Q_3 - Q_1 - \theta_0)}{\theta_0} \Rightarrow_{\theta_n} \mathcal{N}(\Delta/\theta_0, 1). \tag{27}$$

Denote $Y_{ni} = X_{ni}/(2\theta_n), i = 1, \ldots, n$, by condition, $Y_{n1}, \ldots, Y_{nn}$ i.i.d. $\sim \mathcal{U}(0, 1)$. Clearly, $Q_3 - Q_1 = 2\theta_n(\tilde{Q}_3 - \tilde{Q}_1)$, where $\tilde{Q}_1, \tilde{Q}_3$ are first and third quartiles for Y's. Examining the proof of Theorem 6.1, we found that

$$\sqrt{n}\begin{pmatrix} \tilde{Q}_1 - 1/4 \\ \tilde{Q}_3 - 3/4 \end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{16}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \right)$$

Therefore $\sqrt{n}(\tilde{Q}_3 - \tilde{Q}_1 - 1/2) \Rightarrow \mathcal{N}(0, 1/4)$, implying that $\sqrt{n}(Q_3 - Q_1 - \theta_n)/\theta_n \Rightarrow_{\theta_n} \mathcal{N}(0, 1)$. The same decomposition of the left hand side of (27) and Slutsky's theorem then show that (27) holds.

By (26) and (27), the efficacies of test A and test B are $1/\theta_0$ and $\sqrt{3}/\theta_0$, respectively, hence the asymptotic relative efficiency of test A with respect to test B is

$$\text{ARE}(\text{test A, test B}) = \left(\frac{1/\theta_0}{\sqrt{3}/\theta_0} \right)^2 = \frac{1}{3}.$$

Consequently, test B is more efficient than test A, given the underlying distribution is uniform. □

8.8 (a) *Proof.* Under H_0, it is easy to verify that $E[I_i] = 1/2$, $\text{Var}(I_i) = 1/4$. Since $I_1, I_2, \ldots$ are i.i.d., it follows by classical CLT that

$$\frac{\sum_{i=1}^{n}(I_i - 1/2)}{\sqrt{n/4}} = 2\sum_{i=1}^{n}(I_i - 1/2)/\sqrt{n} \Rightarrow \mathcal{N}(0, 1).$$

That is, the asymptotic distribution of T_n under the null hypothesis is standard normal. □

(b) *Proof.* Under the local alternative hypothesis, $I_{n1}, \ldots, I_{nn}, n \geq 1$ form a triangular array in which $I_{n1}, \ldots, I_{nn}$ are i.i.d. with

$$E[I_{ni}] = P_{\lambda_n, \mu_n}[X_1 > Y_1] = \int_0^\infty P[X_1 > y]f_{Y_1}(y)dy$$

$$= \int_0^\infty e^{-(\lambda + n^{-1/2}\delta)^{-1}y}\lambda^{-1}e^{-\lambda^{-1}y} = \frac{\lambda + n^{-1/2}\delta}{2\lambda + n^{-1/2}\delta} = \rho_n,$$

$$\text{Var}(I_{ni}) = \rho_n(1 - \rho_n),$$

$$s_n^2 = n\rho_n(1 - \rho_n).$$

Therefore,

$$\frac{1}{s_n^3}\sum_{i=1}^{n}E[|I_{ni} - \rho_n|^3] \leq \frac{8n}{n^{3/2}[\rho_n(1 - \rho_n)]^{3/2}} \to 0$$

as $n \to \infty$, i.e., Lyapunov's condition holds with $\delta = 1$ (with abuse of notation). Hence

$$\frac{\sum_{i=1}^{n}(I_i - \rho_n)}{\sqrt{n\rho_n(1 - \rho_n)}} \Rightarrow \mathcal{N}(0, 1).$$

□

(c) *Proof.* Write

$$T_n = \frac{\sum_{i=1}^{n}(I_i - 1/2)}{\sqrt{n}/2} = \frac{\sum_{i=1}^{n}(I_i - \rho_n)}{\sqrt{n\rho_n(1 - \rho_n)}} \times \frac{\sqrt{n\rho_n(1 - \rho_n)}}{\sqrt{n}/2} + \frac{n(\rho_n - 1/2)}{\sqrt{n}/2}. \quad (28)$$

76

Since $\rho_n \to 1/2$ as $n \to \infty$, it follows by Slutsky's theorem and the result of part (b) that the first term of the right hand side of (28) converges weakly to $\mathcal{N}(0,1)$. On the other hand,

$$\frac{n(\rho_n - 1/2)}{\sqrt{n}/2} = 2\sqrt{n}(\rho_n - 1/2) = \frac{\sqrt{n}\delta}{2\sqrt{n}\lambda + \delta} \to \frac{\delta}{2\lambda} \tag{29}$$

as $n \to \infty$. (29) and another application of Slutsky's theorem gives that under local alternative hypothesis specified in (b), $T_n \Rightarrow \mathcal{N}(\delta/(2\lambda), 1)$. $\qquad\square$

9.1(a) *Proof.* Write

$$T_n = (\boldsymbol{V}^{(n)})^T \Sigma^{-1} \boldsymbol{V}^{(n)} + (\boldsymbol{V}^{(n)})^T (S_n^{-1} - \Sigma^{-1}) \boldsymbol{V}^{(n)}. \tag{30}$$

The first term of (30)'s right hand side converges in distribution to χ_k^2 by multivariate CLT and continuous mapping theorem, while its second term is $o_P(1)$, since it can be viewed as $O_P(1) \times o_P(1) \times O_P(1)$. The result then follows from Slutsky's theorem. $\qquad\square$

9.4 *Proof.* First, W^2 can be simplified to

$$W^2 = \sum_{j=1}^k \frac{nm}{Z_j} \left(\frac{X_j}{m} - \frac{Y_j}{n} \right)^2$$
$$= \sum_{j=1}^k \left[\frac{nm/N}{p_j} \left(\frac{X_j}{m} - \frac{Y_j}{n} \right)^2 \right] \times \frac{p_j}{Z_j/N}.$$

Intuitively, since $Np_j/Z_j \to_P 1$ as $N \to \infty$ for every j, it is promising to show our goal by two steps:

Step 1: Show that
$$\tilde{W}^2 := \sum_{j=1}^k \frac{nm/N}{p_j} \left(\frac{X_j}{m} - \frac{Y_j}{n} \right)^2 \Rightarrow \chi_{k-1}^2. \tag{31}$$

Step 2: Provided (31) holds, show that

$$W^2 - \tilde{W}^2 \to_P 0. \tag{32}$$

To show (31), we first show that

$$\sqrt{N\frac{m}{N}\frac{n}{N}}\left(\frac{Y_1}{n}-\frac{X_1}{m},\ldots,\frac{Y_{k-1}}{n}-\frac{X_{k-1}}{m}\right)^T \Rightarrow \mathcal{N}(0,\Sigma), \tag{33}$$

where

$$\Sigma = \begin{pmatrix} p_1(1-p_1) & \cdots & -p_1 p_{k-1} \\ \vdots & \ddots & \vdots \\ -p_{k-1}p_1 & \cdots & p_{k-1}(1-p_{k-1}) \end{pmatrix}.$$

Adopting the same notation in Exercise 8.2, it can be shown that $\Sigma^{-1} = \Lambda + \mathbf{1}\mathbf{1}^T/p_k$ (the canonical forms of Σ and Σ^{-1} should be always kept in mind so that it can be called without calculation).

Under given assumptions, it is routine to show by multivariate CLT that as $N \to \infty$:

$$\sqrt{n}\left(\frac{Y_1}{n}-p_1,\ldots,\frac{Y_{k-1}}{n}-p_{k-1}\right)^T \Rightarrow \mathcal{N}(0,\Sigma),$$

$$\sqrt{m}\left(\frac{X_1}{m}-p_1,\ldots,\frac{X_{k-1}}{m}-p_{k-1}\right)^T \Rightarrow \mathcal{N}(0,\Sigma).$$

Since $n/N \to \alpha \in (0,1)$, the above expressions may be rewritten as

$$\sqrt{N\frac{m}{N}\frac{n}{N}}\left(\frac{Y_1}{n}-p_1,\ldots,\frac{Y_{k-1}}{n}-p_{k-1}\right)^T \Rightarrow \mathcal{N}(0,(1-\alpha)\Sigma),$$

$$\sqrt{N\frac{m}{N}\frac{n}{N}}\left(\frac{X_1}{m}-p_1,\ldots,\frac{X_{k-1}}{m}-p_{k-1}\right)^T \Rightarrow \mathcal{N}(0,\alpha\Sigma).$$

Now the independence between $\boldsymbol{X}$ and $\boldsymbol{Y}$ allows us to conclude (33), which further implies that

$$N\frac{m}{N}\frac{n}{N}\left(\frac{Y_1}{n}-\frac{X_1}{m} \quad \cdots \quad \frac{Y_{k-1}}{n}-\frac{X_{k-1}}{m}\right)\Sigma^{-1}\begin{pmatrix} \frac{Y_1}{n}-\frac{X_1}{m} \\ \vdots \\ \frac{Y_{k-1}}{n}-\frac{X_{k-1}}{m} \end{pmatrix} \Rightarrow \chi^2_{k-1}.$$

Performing similar calculation as in Exercise 8.2 (use the elegant form of Σ^{-1}) shows

that the left hand side of the above expression is precisely $\tilde{W}^2$, hence (31) holds.

Now let's verify (32). Noting that each summand of $\tilde{W}^2$ is non-negative, we thus have

$$
|W^2 - \tilde{W}^2|
$$
$$
= \left| \sum_{j=1}^{k} \left[\frac{nm/N}{p_j} \left(\frac{X_j}{m} - \frac{Y_j}{n} \right)^2 \right] \left(\frac{p_j}{Z_j/N} - 1 \right) \right|
$$
$$
\leq \max_{1 \leq j \leq k} \left| \frac{p_j}{Z_j/N} - 1 \right| \tilde{W}^2
$$
$$
= o_P(1) \times O_P(1) = o_P(1).
$$

Hence (32) holds. The final conclusion then follows from Slutsky's theorem, in light of (31) and (32). $\qquad\square$

References

[1] Kai Lai Chung. *A Course in Probability Theory*. Academic Press, third edition, 2001.

[2] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, Inc., New York, third edition, 1995.

[3] Thomas Ferguson. *A Course in Large Sample Theory*. Springer Science+Business Media, B.V., 1996.

[4] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, fourth edition, 2013.

[5] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, Inc., New York, third edition, 1964.

[6] Robert Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc., New York, 1980.

[7] Robert Serfling. Asymptotic relative efficiency in estimation. *International Encyclopedia of Statistical Science*, 68–72, 2011.