Topics in statistics III/IV (MATH3361/4071) Michaelmas term, 2020-2021

Exercises: Likelihood methods

Lecturer & Author: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

---

This is out of the scope

Exercise 1. ($\star\star$) From Fatou-Lesbeque Lemma, prove Monotone Convergence theorem. (Hint: Use $Y \equiv 0$, use $\limsup_{n\to\infty} f_n$and $\liminf_{n\to\infty} f_n$)

Solution. Since $0 < X_1 \leq ... \leq \lim_{n\to\infty} X_n = X$ a.s.. Then $\mathrm{E}X_n \leq \mathrm{E}X$ or $\limsup_{n\to\infty} \mathrm{E}X_n \leq \mathrm{E}X$.

From Fatou-Lesbeque Lemma, it is $\liminf_{n\to\infty} \mathrm{E}X_n \geq \mathrm{E}X$. Also the limit $\lim \mathrm{E}X_n$ exists. Then, it is $\lim \mathrm{E}X_n = \mathrm{E}X$

This is out of the scope

Exercise 2. ($\star\star$) From Fatou-Lesbeque Lemma, prove Lesbeque Dominant Convergence theorem. (Hint: Use that $-Y \leq -X_n$ and $-Y \leq X_n$, use $\limsup_{n\to\infty} f_n$and $\liminf_{n\to\infty} f_n$)

Solution.

Since $|X_n| \leq Y$, it is $-Y \leq -X_n$, and because $X_n \xrightarrow{a.s.} X$ it is $\liminf_{n\to\infty} \mathrm{E}(-X_n) \geq \mathrm{E}(-Y) \Longleftrightarrow \limsup_{n\to\infty} \mathrm{E}(X_n) \leq \mathrm{E}(Y)$

Since $|X_n| \leq Y$, it is $-Y \leq X_n$ and because $X_n \xrightarrow{a.s.} X$ it is $\liminf_{n\to\infty} \mathrm{E}(X_n) \geq \mathrm{E}(Y)$

So $\lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(Y)$

Exercise 3. ($\star\star$) Let μ be a constant. Show that $X_n \xrightarrow{\text{qm}} \mu$ if and only if $\mathrm{E}X_n \to \mu$ and $\mathrm{Var}(X_n) \to 0$, both in uni-variate and multivariate case.

Solution. It is $\mathrm{E}(X_n - \mu)^2 = \mathrm{Var}(X_n) + (\mathrm{E}X_n - \mu)^2$. Hence, $\mathrm{E}(X_n - \mu)^2 \to 0$. In the multivariate case, it is $\mathrm{E}(X_n - \mu)^T(X_n - \mu) = \mathrm{E}\sum_{i=1}^{d}(X_{n,i} - \mu_i)^2 \to 0$ by treating each element separately.

Exercise 4. ($\star\star$) Consider a sequence of discrete r.v. $\{X_n\}$ with probability $P(X_n = k) = \frac{1}{n}$, for $k = 1/n, 2/n, ..., n/n$. Show that $X_n \xrightarrow{D} X$ where $X \sim \mathrm{U}(0, 1)$. (Hint: Just use the definition.)

Solution. The probability function is $P(X_n \leq x) = k/n$ for $k/n \leq x \leq (k+1)/n$.

Then because $|k/n - x| < 1/n$, we have $\lim_{n\to\infty} P(X_n \leq x) = x$.

Note that $P(X \leq x) = x$ is the distribution function of the Uniform random variable $X \sim \mathrm{U}(0,1)$. So $X_n \xrightarrow{D} \mathrm{U}(0,1)$.

Exercise 5. $(\star)$

1. Show that
$$\mathrm{E}_\pi(X-\theta)^T(X-\theta) = \mathrm{Var}_\pi(X) + (\mathrm{E}_\pi(X)-\theta)^T(\mathrm{E}_\pi(X)-\theta)$$
, where is a constant point, and X is a random variable $X \sim \mathrm{d}\pi(\cdot)$.

2. Show that
$$\mathrm{E}_\pi|X-\theta|^2 = \mathrm{Var}_\pi(X) + |\mathrm{E}_\pi(X)-\theta|^2$$
, where is a constant point, X is a random variable $X \sim \mathrm{d}\pi(\cdot)$, and $|X| = \sqrt{X_1^2 + ...X_d^2}$ is the Euclidean norm.

Solution.

1. It is
$$\begin{aligned}\mathrm{E}_\pi(X-\theta)^T(X-\theta) &= \mathrm{E}_\pi([X-\mathrm{E}_\pi(X)] + [\mathrm{E}_\pi(X)-\theta])^T([X-\mathrm{E}_\pi(X)] + [\mathrm{E}_\pi(X)-\theta]) = ...\\ &= \mathrm{E}_\pi(X-\theta)^T(X-\theta) + (\mathrm{E}_\pi(X)-\theta)^T(\mathrm{E}_\pi(X)-\theta)\\ &= \mathrm{Var}_\pi(X) + (\mathrm{E}_\pi(X)-\theta)^T(\mathrm{E}_\pi(X)-\theta)\end{aligned}$$

2. It is
$$\mathrm{E}_\pi|X-\theta|^2 = \mathrm{E}_\pi(X-\theta)^T(X-\theta)$$
$$|\mathrm{E}_\pi(X)-\theta|^2 = (\mathrm{E}_\pi(X)-\theta)^T(\mathrm{E}_\pi(X)-\theta)$$
from the definition of the is the Euclidean norm $|X| = \sqrt{X_1^2 + ...X_d^2}$. So the result follows from then previous task.

Exercise 6. Show that
$$\log(1+x) = x + o(x), \text{ as } x \to 0.$$

Solution. Let $f(x) = \log(1+x)$. Then $\dot{f}(x) = \frac{1}{1+x}$. The 1st order Taylor expansion of $f(x)$ around 0 is
$$f(x) = f(0) + \frac{1}{1!}\dot{f}(0)(x-0) + o(x), \text{ as , as } x \to 0$$

where $h = x - 0$.

So

$$\log(1+x) = x + o(x), \text{ as } x \to 0.$$

Exercise 7. Show that

$$\lim_{n\to\infty}(1+\frac{1}{n}a_n)^n = \exp(\lim_{n\to\infty} a_n)$$

provided that $\frac{1}{n}a_n \to 0$, as $n \to \infty$.

Hint: From Taylor expansion, it is

$$\log(1+x) = x + o(x), \text{ as } x \to 0.$$

Solution.

- It is

$$\begin{aligned}(1+\frac{1}{n}a_n)^n =& \exp(n\log(1+\frac{1}{n}a_n))\\ =& \exp(n(\frac{1}{n}a_n + o(\frac{1}{n}a_n)))\\ =& \exp(a_n(1+o(1)))\end{aligned}$$

- Then provided that a_n increases slower than n, aka $\frac{1}{n}a_n \to 0$ it is

$$\lim_{n\to\infty}(1+\frac{1}{n}a_n)^n = \exp(\lim_{n\to\infty} a_n)$$

Exercise 8. It is $X_n \xrightarrow{a.s.} X$ if and only if

$$\text{for every } \epsilon > 0, \quad P(|X_k - X| < \epsilon\,,\, \forall k \geq n) \to 1, \qquad \text{as } n \to \infty,$$

Solution. Let $A_{n,\epsilon} = \{|X_k - X| < \epsilon\,,\, \forall k \geq n\}$. Then

$$P(\lim_{n\to\infty} X_n = X) = P\{\forall\epsilon > 0,\ \exists n > 0, \text{ s.t. } |X_k - X| < \epsilon,\ \forall k \geq n\} = P\{\cap_{\epsilon>0} \cup_{\forall n} A_{n,\epsilon}\}$$

So $X_n \xrightarrow{a.s.} X$ is equivalent to $P\{\cap_{\epsilon>0}\cup_{\forall n} A_{n,\epsilon}\} = 1$. Because sets $\cup_{\forall n} A_{n,\epsilon}$ decrease to $\cap_{\epsilon>0}\cup_{\forall n} A_{n,\epsilon}$ as $\epsilon \to 0$, it is

$$P\{\cap_{\epsilon>0} \cup_{\forall n} A_{n,\epsilon}\} = 1 \iff P\{\cup_{\forall n} A_{n,\epsilon}\} = 1,\ \forall\epsilon > 0$$

Because $A_{n,\epsilon}$ increases to $\cup_{\forall n} A_{n,\epsilon}$ as $n \to \infty$, it is

$$P\{\cup_{\forall n} A_{n,\epsilon}\} = 1 \Longleftrightarrow P\{A_{n,\epsilon}\} = 1, \text{ as } n \to \infty, \; \forall \epsilon > 0$$

Exercise 9. Prove the following relations betweeng different modes of convergence

1. $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X$
2. $X_n \xrightarrow{r} X$, for some $r > 0 \implies X_n \xrightarrow{P} X$
3. $(\star\star\star) X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X$

Solution.

1. For any $\epsilon > 0$, then

$$P(|X_n - X| > \epsilon) \geq P(|X_k - X| < \epsilon \,, \forall k \geq n) \to 1, \text{ as, } n \to \infty$$

from Lemma 6 in the Handout.

2. It is

$$E|X_n - X|^r \geq E(|X_k - X|^r 1(|X_n - X| \geq \epsilon)) \geq \epsilon^r P(|X_n - X| \geq \epsilon) \to 0, \text{ as, } n \to \infty$$

This is Markov inequality (Prob. I)

3. This is difficult and it can be skipped.

For any $\epsilon > 0$, $\{X > z + 1\epsilon\}$ and $|X_n - X| < \epsilon$ imply $\{X_n > z\}$. Hence, $\{X_n > z\} \supseteq \{X > z + \epsilon\} \cap \{|X_n - X| < \epsilon\}$. By taking complements, we get $\{X_n \leq z\} \subseteq \{X \leq z + 1\epsilon\} \cup \{|X_n - X| > \epsilon\}$. So I get $P(X_n \leq z) \leq P(X \leq z + \epsilon) + P(|X_n - X| > \epsilon)$.

In a similar way (by interchanging X and X_n), I get $P(X_n \leq z) \geq P(X \leq z - \epsilon) + P(|X_n - X| > \epsilon)$. [1]

So as $n \to \infty$

$$P(X \leq z - 1\epsilon) \leq \liminf_{n \to \infty} P(X_n \leq z) \leq \limsup_{n \to \infty} P(X_n \leq z) \leq P(X \leq z + 1\epsilon)$$

---

[1] It is:

(a) $\limsup_{n \to \infty} f_n := \lim_{n \to \infty} (\sup_{\forall m \geq n} f_m)$ and $\liminf_{n \to \infty} f_n := \lim_{n \to \infty} (\inf_{\forall m \geq n} f_m)$

(b) It is $\liminf_{n \to \infty} f_n \leq \limsup_{n \to \infty} f_n$ if both exist.

(c) It is $\lim_{n \to \infty} f_n = \liminf_{n \to \infty} f_n = \limsup_{n \to \infty} f_n$ if $\lim_{n \to \infty} f_n$ exists

As $F_X(x) = P(X \leq x)$ is continuous at z, the two ends should converge to $F_X(z) = P(X \leq z)$ as $\epsilon \to 0$, which implies that $\lim_{n\to\infty} F_{X_n}(z) = F_X(z)$

Exercise 10. ($\star\star$) Prove that:

1. if $Z \sim \mathrm{N}(0, I)$ then $\varphi_Z(t) = \exp(-\frac{1}{2}t^T t)$, where $Z \in \mathbb{R}^d$
2. if $X \sim \mathrm{N}(\mu, \Sigma)$ then $\varphi_X(t) = \exp(it^T\mu - \frac{1}{2}t^T\Sigma t)$, where $X \in \mathbb{R}^d$

Hint: Assume as known that if $Z \sim \mathrm{N}(0, 1)$ then $\varphi_Z(t) = \exp(-\frac{1}{2}t^2)$, where $Z \in \mathbb{R}$

Solution 11.

1. It is

$$\begin{aligned}
\varphi_Z(t) &= \mathrm{E}(\exp(it^T Z)) = \mathrm{E}(\exp(i\sum_{j=1}^{d}(t_j Z_j))) = \mathrm{E}(\prod_{j=1}^{d}\exp(it_j Z_j)) = \prod_{j=1}^{d}\mathrm{E}(\exp(it_j Z_j)) \\
&= \prod_{j=1}^{d}\varphi_{Z_j}(t) = \prod_{j=1}^{d}\exp(-\frac{1}{2}t_j^2) = \exp(-\frac{1}{2}\sum_{j=1}^{d}t_j^2) = \exp(-\frac{1}{2}t^T t)
\end{aligned}$$

2. Assume a matrix L such as $\Sigma = LL^T$. It is $X = \mu + LZ$. Then

$$\begin{aligned}
\varphi_X(t) &= \varphi_{\mu+LZ}(t) = e^{it^T\mu}\varphi_Z(L^T t) = e^{it^T\mu}\exp(-\frac{1}{2}(L^T t)^T L^T t) \\
&= e^{it^T\mu}\exp(-\frac{1}{2}t^T LL^T t) = \exp(it^T\mu - \frac{1}{2}t^T\Sigma t)
\end{aligned}$$

Exercise 12. Let $X, X_1, X_2, ...$ be i.i.d. random vectors, and let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

then

1. (Weak law) If $E|X| < \infty$, then $\bar{X}_n \xrightarrow{P} \mathrm{E}(X)$
2. (Strong law) $E|X| < \infty$, iff $\bar{X}_n \xrightarrow{as} \mathrm{E}(X)$
3. (in qm) $E|X|^2 < \infty$, iff $\bar{X}_n \xrightarrow{\mathrm{qm}} \mathrm{E}(X)$
4. Let $\varphi_X(t) = \mathrm{E}(e^{it^T X})$, and $\mu = \mathrm{E}(X)$.

Solution.

1. It is

$$\begin{aligned}\varphi_{\bar{X}_n}(t) =& \varphi_{X_1+\ldots+X_n}(\frac{t}{n}) = \prod_{i=1}^{n}\varphi_{X_j}(\frac{t}{n}) = \left(\varphi_X(\frac{t}{n})\right)^n\\ =& \left(\varphi_X(0) + \left(\int_0^1 \dot{\varphi}_X(u\frac{t}{n})\mathrm{d}u\right)\frac{t}{n}\right)^n\end{aligned}$$

since by the Mean-Value theorem

$$\varphi_X(\frac{t}{n}) = \varphi_X(0) + \left(\int_0^1 \dot{\varphi}_X(u\frac{t}{n})\mathrm{d}u\right)\frac{t}{n}.$$

Because $\varphi_X(0) = 1$, and $\lim_{\epsilon\to 0}\dot{\varphi}_X(\epsilon) = \dot{\varphi}_X(0) = i\mu^T$ it is

$$\lim_{n\to\infty}\varphi_{\bar{X}_n}(t) = \exp\left(\lim_{n\to\infty}\left(\int_0^1 \dot{\varphi}_X(u\frac{t}{n})\mathrm{d}u\right)t\right) = \exp(i\mu^T t) \tag{1}$$

Here I used that $\lim_{n\to\infty}(1+a_n)^n = \exp(\lim_{n\to\infty} na_n)$ if $\lim_{n\to\infty} na_n$ exists (Exercise #7).

So (1) says that the characteristic function of $\bar{X}_n$ converges to a characteristic function of the degenerate random variable μ

$$\varphi_{\bar{X}_n}(t) \to \varphi_\mu(t)$$

From the continuity Theorem 24 it is $\bar{X}_n \xrightarrow{D} \mu$. Then from Theorem 7(3) it is $\bar{X}_n \xrightarrow{P} \mu$ because $\mu = \mathrm{E}(X)$ is just a constant point.

2. Proof is out of the scope; for more details see in[**?**].

3. It is

$$\begin{aligned}\mathrm{E}|\bar{X}_n - \mu|^2 =& \mathrm{E}(\bar{X}_n-\mu)^T(\bar{X}_n-\mu)\\ =& \frac{1}{n^2}\sum_i\sum_j \mathrm{E}(X_i-\mu)^T(X_j-\mu)\\ \overset{\text{simplify}}{=}& \frac{1}{n^2}\sum_i \mathrm{E}(X_i-\mu)^T(X_i-\mu) \overset{\text{iid}}{=} \frac{1}{n^2}n\mathrm{E}(X-\mu)^T(X-\mu)\\ =& \frac{1}{n}\mathrm{Var}(X) \to 0\end{aligned}$$

as the 2nd mode is finite.

Exercise 13. Show

$$\text{If} \quad h_n \to 0, \quad \text{and} \quad X_n = O_P(h_n) \quad \text{then} \quad X_n = o_P(1).$$

Solution.

- Deterministic: If $x_n = O(h_n)$ and $h_n \to 0$, then $x_n = o(1)$, because we sandwich $|x_n| \leq Kh_n \to 0$.
- Stochastic: If $x_n = O_P(h_n)$ and $h_n \to 0$, then $x_n = o_P(1)$. Because $h_n \to 0$, for sufficiently large $n > 0$ $Kh_n \leq \delta$. Also as $x_n = O_P(h_n)$ for any $\epsilon > 0$ I can find a $K > 0$ such that $P(|x_n| \leq Kh_n) \geq 1 - \epsilon$. Putting both together, for any $\epsilon > 0$ and any $\delta > 0$, I can get K such that, for sufficiently large $n > 0$, I can get

$$P(|x_n| \leq \delta) \geq P(|x_n| \leq Kh_n) \geq 1 - \epsilon$$

Exercise 14. Let $X_1, X_2, ...$ IID random vectors $X_i \in \mathbb{R}^d$ with mean $\mathrm{E}(X_i) = \mu$ and finite covariance matrix $\mathrm{Var}(X_i) < \infty$ for all $i = 1, ...,$ Then

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \mathrm{N}(0, \Sigma)$$

Solution. We 'll gonna use again the characteristic function, and its property with the IID variables. It is

$$\sqrt{n}(\bar{X}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j - \mu)$$

Hence, for any $t \in \mathbb{R}^d$

$$\begin{aligned}
\varphi_{\sqrt{n}(\bar{X}_n - \mu)}(t) &= \varphi_{\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j - \mu)}(t) = \varphi_{\sum_{j=1}^{n} (X_j - \mu)}(\frac{t}{\sqrt{n}}) \\
&= \prod_{j=1}^{n} \varphi_{(X_j - \mu)}(\frac{t}{\sqrt{n}}) \\
&= \left(\varphi_{(X_j - \mu)}(\frac{t}{\sqrt{n}}) \right)^n = \left(\varphi_{(X - \mu)}(\frac{t}{\sqrt{n}}) \right)^n
\end{aligned}$$

Here, let $\varphi(t) := \varphi_{(X_j - \mu)}(t)$ for notation convenience, as $X_1, X_2, ...$ are IID and hence have the same moments. We use Taylor expansion around 0 as

$$\varphi_{(X-\mu)}(\frac{t}{\sqrt{n}}) = \cancelto{1}{\varphi_{(X-\mu)}(0)} + \cancelto{0}{\dot{\varphi}_{(X-\mu)}(0)} \frac{t}{\sqrt{n}} + t^T \left(\int_0^1 \int_0^1 v \ddot{\varphi}_{(X-\mu)}(0 + vu\frac{t}{n}) \mathrm{d}u \mathrm{d}v \right) \frac{t}{n}$$

because $\ddot{\varphi}_X(t)$ is obviously continuous. So

$$\begin{aligned}\varphi_{\sqrt{n}(\bar{X}_n-\mu)}(t) &= \left(\varphi_{(X-\mu)}(\frac{t}{\sqrt{n}})\right)^n \\ &= \left(1 + t^T\left(\int_0^1\int_0^1 v\ddot{\varphi}_{(X-\mu)}(vu\frac{t}{n})\mathrm{d}u\mathrm{d}v\right)\frac{t}{n}\right)^n\end{aligned}$$

Because $\lim_{n\to\infty}(1+a_n)^n = \exp(\lim_{n\to\infty} na_n)$ if $\lim_{n\to\infty} na_n$ exists (Exercise #7), it is

$$\begin{aligned}\lim_{n\to\infty}\varphi_{\sqrt{n}(\bar{X}_n-\mu)}(t) &= \lim_{n\to\infty}\left(1 + t^T\left(\int_0^1\int_0^1 v\ddot{\varphi}_{(X-\mu)}(vu\frac{t}{n})\mathrm{d}u\mathrm{d}v\right)\frac{t}{n}\right)^n \\ &= \exp\left(\lim_{n\to\infty} t^T\left(\int_0^1\int_0^1 v\ddot{\varphi}_{(X-\mu)}(vu\frac{t}{n})\mathrm{d}u\mathrm{d}v\right)t\right) \\ &= \exp\left(t^T\left(\int_0^1\int_0^1 v(-\Sigma)\mathrm{d}u\mathrm{d}v\right)t\right) \\ &= \exp(-\frac{1}{2}t^T\Sigma\, t)\end{aligned} \tag{2}$$

This is because $\ddot{\varphi}_{(X-\mu)}(\cdot)$ is continuous so $\lim_{n\to\infty}\ddot{\varphi}_{(X-\mu)}(u\frac{t}{n}) = \ddot{\varphi}_{(X-\mu)}(0) = -\mathrm{E}((X-\mu)^T(X-\mu)) = -\Sigma$.

Since $\lim_{n\to\infty}\varphi_{\sqrt{n}(\bar{X}_n-\mu)}(t) = \exp(-\frac{1}{2}t^T\Sigma\, t)$, aka $\varphi_{\sqrt{n}(\bar{X}_n-\mu)}(t) \to \varphi_Z(t)$ where $Z \sim \mathrm{N}(0,\Sigma)$, it is $\sqrt{n}(\bar{X}_n-\mu) \xrightarrow{D} \mathrm{N}(0,\Sigma)$.

Exercise 15. ($\star\star$) Consider that $\sqrt{n}(X_n-\mu) \underset{D}{\rightarrow} Z$, where $Z \sim \mathrm{N}(0,\Sigma)$ for $\Sigma > 0$ (positive definite). Show that $X_n \xrightarrow{P} \mu$. (Hint: Use the concept 'bounded in probability)'

Solution. I show this result by using 2 ways.

First way: It is $\sqrt{n}(X_n-\mu) \underset{D}{\rightarrow} Z$, so

$$\sqrt{n}(X_n-\mu) = O_P(1) \implies (X_n-\mu) = O_P(1)O_P(1/\sqrt{n}) = O_P(1)o_P(1) = o_P(1)$$

So $X_n \xrightarrow{P} \mu$.

Second way: I observe that

$$(X_n-\mu) = \underbrace{\frac{1}{\sqrt{n}}}_{=A_n}\underbrace{\sqrt{n}(X_n-\mu)}_{=B_n}$$

It is $A_n = \sqrt{n}(X_n-\mu) \underset{D}{\rightarrow} Z$, and $B_n = \frac{1}{\sqrt{n}} \to 0$. By Slutsky theorem it is

$$(X_n-\mu) = A_nB_n \underset{D}{\rightarrow} 0Z = 0$$

So

$$(X_n - \mu) \underset{D}{\longrightarrow} 0$$

which implies

$$X_n \xrightarrow{P} \mu$$

Exercise 16. ($\star\star$)

1. If $X_1, X_2, ...$ are IID in $\mathbb{R}^2$ with distribution giving probability

$$\mathrm{P}(X = x) = \begin{cases} \theta_1 & , \text{if } x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \theta_2 & , \text{if } x = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ \theta_1 + \theta_2 & , \text{if } x = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{cases}$$

there $\theta_1 + \theta_2 \leq 1$.. What is the asymptotic distribution of $\bar{X}_n$ given the CLT?

2. If $X_1, X_2, ...$ are IID from a Poisson distribution $\mathrm{Poi}(\theta)$ distribution as

$$\mathrm{P}(x|\theta) = \frac{e^{-\theta}\theta^x}{x!} 1(x \in \{0, 1, 2, ...\})$$

Let Z_n be the proportion of zeros observed $Z_n = \frac{1}{n}\sum_{j=1}^{n} 1(X_j = 0)$. What is the joint asymptotic distribution of $(\bar{X}_n, Z_n)$

Solution.

1. It is $\mu = \mathrm{E}(X) = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$, $\mathrm{E}(X^T X) = \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix}$, so $\mathrm{Var}(X) = \mathrm{E}(X - \mathrm{E}(X))^T (X - \mathrm{E}(X)) = \mathrm{E}(X^T X) - \mu^T \mu = \begin{bmatrix} \theta_1(1-\theta_1) & -\theta_1\theta_2 \\ -\theta_1\theta_2 & \theta_2(1-\theta_2) \end{bmatrix}$ The CLT says $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \mathrm{N}(0, \Sigma)$

2. It is $E(X) = \theta$, $E(1(X = 0)) = \exp(\theta)$, $\mathrm{Var}(X) = \theta$, $\mathrm{Var}(1(X = 0)) = \exp(-\theta)(1 - \exp(-\theta))$ and $E(X1(X = 0)) = 0$, so $\mathrm{cov}(X, 1(X = 0)) = -\theta \exp(-\theta)$. So $\Sigma = \begin{bmatrix} \theta & -\theta\exp(-\theta) \\ -\theta\exp(-\theta) & \exp(-\theta)(1-\exp(-\theta)) \end{bmatrix}$. The CLT says $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \mathrm{N}(0, \Sigma)$

Exercise 17. ($\star\star\star\star$Super difficult) (The autoregressive model) Consider that $\{\epsilon_n\}$ are IID, with mean $\mathrm{E}(\epsilon_n) = \mu$, and variance $\mathrm{Var}(\epsilon_n) = \sigma^2$, $\forall n$. A time series $\{X_n\}_{n\geq 1}$ is modeled as $X_n \sim \mathrm{AR}(\beta)$ where $\beta \in (-1, 1)$ if

$$\begin{aligned} X_n =& \beta X_{n-1} + \epsilon_n; \text{ for } n \geq 2 \\ X_1 =& \epsilon_1 \end{aligned}$$

Show that $\bar{X}_n \xrightarrow{\text{qm}} \mu/(1-\beta)$

1. Show that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} \epsilon_j(1-\beta^{n-j+1})/(1-\beta)$
2. Find $\lim_{n\to\infty} \mathrm{E}(\bar{X}_n) =?$
3. Show that $\lim_{n\to\infty} \mathrm{Var}(\bar{X}_n) = 0$
4. Show that $\bar{X}_n \xrightarrow{\text{qm}} \mu/(1-\beta)$

[Hint] (1.) Show that $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} \epsilon_j(1-\beta^{n-j+1})/(1-\beta)$ (2) Find $\lim_{n\to\infty} \mathrm{E}(\bar{X}_n) = \mu/(1-\beta)$; (3) Show that $\lim_{n\to\infty} \mathrm{Var}(\bar{X}_n) = 0$, (4.) ...

Solution.

1. It is $X_i = \sum_{j=1}^{i} \epsilon_j \beta^{n-j}$. So

$$\bar{X}_n = \frac{1}{n}\frac{\mu}{1-\beta}\sum_{i=1}^{n}\sum_{j=1}^{i}\epsilon_j\beta^{n-j} = \frac{1}{n}\sum_{j=1}^{n}\epsilon_j\frac{1-\beta^{n-j+1}}{1-\beta}$$

2. It is

$$\begin{aligned} \mathrm{E}\bar{X}_n &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\epsilon_j)\frac{1-\beta^{n-j+1}}{1-\beta} = \frac{1}{n}\frac{\mu}{1-\beta}\sum_{j=1}^{n}(1-\beta^{n-j+1}) \\ &= \frac{1}{n}\frac{\mu}{1-\beta}\sum_{j=1}^{n}(1-\beta^{j}) \\ &= \frac{1}{n}\frac{\mu}{1-\beta}(n - \frac{\beta(1-\beta^n)}{1-\beta}) \\ &= \frac{\mu}{1-\beta} - \frac{\beta\mu}{n}\frac{(1-\beta^n)}{(1-\beta)^2} \end{aligned}$$

So $\lim \mathrm{E}\bar{X}_n = \frac{\mu}{1-\beta}$

3. It is

$$\mathrm{Var}(\bar{X}_n) = \sum_{i=1}^{n} \mathrm{Var}(\epsilon_j)(\frac{1}{n}\frac{1-\beta^{n-j+1}}{1-\beta})^2 = \sigma^2 \frac{1}{n^2}\sum_{i=1}^{n} \frac{(1-\beta^{n-j+1})^2}{(1-\beta)^2}$$
$$\leq \sigma^2 \frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{(1-\beta)^2} \;\leq\; \sigma^2\frac{1}{n}$$

as $\beta \in (0,1)$. So $\lim \mathrm{Var}(\bar{X}_n) = 0$

4. It is

$$\begin{aligned}\lim(\mathrm{E}\bar{X}_n\text{-}\frac{\mu}{1-\beta})^2 &= \lim(\mathrm{Var}(\bar{X}_n) + (\mathrm{E}\bar{X}_n - \frac{\mu}{1-\beta})^2)\\ &= \lim\mathrm{Var}(\bar{X}_n) + (\lim\mathrm{E}\bar{X}_n - \frac{\mu}{1-\beta})^2\\ &= 0\end{aligned}$$

So $\bar{X}_n \xrightarrow{\text{qm}} \mu/(1-\beta)$

Exercise 18. $(\star\star)$ Let $X_i \overset{\text{IID}}{\sim} F_X$ for $i = 1, ..., n$, and $F_X = P(X \leq x)$. Show that the empirical distribution function $\hat{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} 1(x \in [x_i, \infty))$ is a strongly consistent estimator of F_X.

Solution. It is $\mathrm{E}(\hat{F}_X(x)) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(1(x \in (-\infty, x_i])) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{P}(x \in (-\infty, x_i]) \leq \frac{1}{n}\sum_{i=1}^{n} 1 < \infty$ So the strong LLN applies.

The next exercise is from Problem Class 2

Exercise 19. Consider random variables $X, X_1, X_2, ...$, where $\mu_n = \mathrm{E}(X-\mu)^n$, and $\mu = \mathrm{E}(X)$

1. Show that,

$$\sqrt{n}(\begin{bmatrix}\bar{X}\\ s_x^2\end{bmatrix} - \begin{bmatrix}\mu\\ \sigma^2\end{bmatrix}) \xrightarrow{D} \mathrm{N}(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2 & \mu_3\\ \mu_3 & \mu_4-\sigma^4\end{bmatrix})$$

2. Show that the asymptotic distribution of the coefficient of variation $\mathrm{cv} = \frac{s_x}{\bar{X}}$, is

$$\sqrt{n}(\frac{s_x}{\bar{X}} - \frac{\sigma}{\mu}) \xrightarrow{D} \mathrm{N}(0, \frac{\mu_4-\sigma^4}{4\mu^2\sigma^2} - \frac{\mu_3}{\mu^3} + \frac{\sigma^4}{\mu^4})$$

3. Show that the asymptotic distribution of the 3rd central moment $m_3 = \frac{1}{n}\sum_{i=1}^{n}(X_j - \bar{X})^3$ is

$$\sqrt{n}(m_3 - \mu_3) \xrightarrow{D} \mathrm{N}(0, \mu_6 - \mu_3^2 - 6\sigma^2\mu_4 + 9\sigma^6)$$

Solution.

1.

- I observe that

$$\begin{aligned}
\begin{bmatrix} \bar{X} \\ s_x^2 \end{bmatrix} - \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix} &= \begin{bmatrix} \bar{X} - \mu \\ \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 - (\bar{X} - \mu)^2 \end{bmatrix} - \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu))^2 \end{bmatrix} - \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix} \\
&= \begin{bmatrix} m_1' \\ m_2' - (m_1')^2 \end{bmatrix} - \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix}
\end{aligned}$$

  where $m_1' = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$ and $m_2' = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$.

- I will try to find the joint asymptotic distribution of $(m_1', m_2')^T$ by CLT, and then the asymptotic distribution of $(\bar{X}, s_x^2)^T$ by Delta method.

- Let

$$\xi_i = \begin{bmatrix} X_i - \mu \\ (X_i - \mu)^2 \end{bmatrix}$$

  IID random vectors. It is

$$\mu_\xi = \mathrm{E}(\xi_i) = \begin{bmatrix} \mathrm{E}(X_i - \mu) \\ \mathrm{E}(X_i - \mu)^2 \end{bmatrix} = \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix}$$

  and

$$\begin{aligned}
\Sigma_\xi = \mathrm{Var}(\xi_i) &= \begin{bmatrix} \mathrm{Var}(X_i - \mu) & \mathrm{Cov}((X_i - \mu), (X_i - \mu)^2) \\ \mathrm{Cov}((X_i - \mu), (X_i - \mu)^2) & \mathrm{Var}(X_i - \mu)^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix}
\end{aligned}$$

  since

$$\begin{aligned}
\mathrm{Cov}((X_i - \mu), (X_i - \mu)^2) &= \mathrm{E}(((X_i - \mu) - \mathrm{E}(X_i - \mu))((X_i - \mu)^2 - \mathrm{E}(X_i - \mu)^2)) \\
&= \mathrm{E}(((X_i - \mu) - \mu_1)((X_i - \mu)^2 - \mu_2)) \\
&= \mathrm{E}(\ (X_i - \mu)^3 - (X_i - \mu)\mu_2 - \mu_1(X_i - \mu)^2 + \mu_1\mu_2\ ) \\
&= \mathrm{E}(X_i - \mu)^3 - \cancelto{0}{\mathrm{E}(X_i - \mu)}\mu_2 - \mu_1\cancelto{\mu_2}{\mathrm{E}(X_i - \mu)^2} + \mu_1\mu_2 \\
&= \mathrm{E}(X_i - \mu)^3 = \mu_3
\end{aligned}$$

It is
$$\bar{\xi} = \begin{bmatrix} m_1' \\ m_2' \end{bmatrix}$$

So by CLT, I have,

$$\sqrt{n}(\begin{bmatrix} m_1' \\ m_2' \end{bmatrix} - \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix}) \xrightarrow{D} \mathrm{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix})$$

- Now, I will calculate the asymptotic distribution of $(\bar{X}, s_x^2)^T$ by Delta method.
  Let,
$$g(x, y) = \begin{bmatrix} x \\ y - x^2 \end{bmatrix}$$
  with
$$\dot{g}(x, y) = \frac{\mathrm{d}g(x, y)}{\mathrm{d}(x, y)} = \begin{bmatrix} -1 & 0 \\ -2x & 1 \end{bmatrix}$$
  So
$$g(\underbrace{m_1', m_2'}_{=\bar{\xi}}) = \begin{bmatrix} m_1' \\ m_2' - (m_1')^2 \end{bmatrix} = \begin{bmatrix} \bar{X} - \mu \\ s_x^2 \end{bmatrix}; \qquad g(\underbrace{0, \sigma^2}_{=\mu_\xi}) = \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix}$$
$$\dot{g}(\underbrace{0, \sigma^2}_{=\mu_\xi}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \qquad \Sigma_g = \dot{g}(\underbrace{0, \sigma^2}_{=\mu_\xi})\Sigma_\xi \dot{g}(\underbrace{0, \sigma^2}_{=\mu_\xi})^T = \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix}$$
  Then, according to Delta theorem
$$\sqrt{n}(g(\bar{\xi}) - g(\mu_\xi)) \xrightarrow{D} \mathrm{N}(0, \dot{g}(\mu_\xi)\Sigma_\xi \dot{g}(\mu_\xi)^T)$$
$$\sqrt{n}(\begin{bmatrix} \bar{X} \\ s_x^2 \end{bmatrix} - \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}) \xrightarrow{D} \mathrm{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{bmatrix})$$

2. Since I have the asymptotic distribution of $(\bar{X}, s_x^2)^T$, I can use the Delta method.
   - Let $h(a, b) = \sqrt{b}/a$, with $\dot{h}(a, b) = (-\frac{\sqrt{b}}{a^2}, \frac{1}{2a\sqrt{b}})$.
   - Then
$$h(\bar{X}, s_x^2) = \frac{s_x}{\bar{X}}; \qquad h(\mu, \sigma^2) = \frac{\sigma}{\mu}$$
$$\dot{h}(\mu, \sigma^2) = \begin{bmatrix} -\frac{\sigma}{\mu^2}, & \frac{1}{2\mu\sigma} \end{bmatrix};$$

$$
\begin{aligned}
\Sigma_h &= \dot{h}(\mu,\sigma^2)\Sigma_g \dot{h}(\mu,\sigma^2)^T \\
&= \frac{\mu_4-\sigma^4}{4\mu^2\sigma^2} - \frac{\mu_3}{\mu^3} + \frac{\sigma^4}{\mu^4}
\end{aligned}
$$

- Then, according to Delta theorem

$$
\begin{aligned}
\sqrt{n}(h(\bar{X}, s_x^2) - h(\mu,\sigma^2)) &\xrightarrow{D} \mathrm{N}(0, \dot{h}(\mu,\sigma^2)\Sigma_g \dot{h}(\mu,\sigma^2)^T) \\
\sqrt{n}(\frac{s_x}{\bar{X}} - \frac{\sigma}{\mu}) &\xrightarrow{D} \mathrm{N}(0, \frac{\mu_4-\sigma^4}{4\mu^2\sigma^2} - \frac{\mu_3}{\mu^3} + \frac{\sigma^4}{\mu^4})
\end{aligned}
$$

3. I observe that

$$
\begin{aligned}
m_3 &= \frac{1}{n}\sum_{i=1}^{n}((\underbrace{X_i-\mu}_{=Z_i}) - (\underbrace{\bar{X}-\mu}_{=\bar{Z}}))^3 = \\
&= \frac{1}{n}\sum_{i=1}^{n} Z_i^3 - 3\frac{1}{n}\sum_{i=1}^{n} Z_i^2\bar{Z} + 2\bar{Z} \\
&= m_3' - 3m_2'm_1' + 2(m_1')^2
\end{aligned}
$$

where $m_j' = \frac{1}{n}\sum_{i=1}^{n} Z_i^j = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^j$, since $Z_i = X_i - \mu$.

- I will use the CLT to calculate the joint asymptotic distribution of $(m_1', m_2', m_3')^T$ and then I will use Delta method to calculate that of m_3.

  I specify

$$
\psi_i = \begin{bmatrix} Z_i \\ Z_i^2 \\ Z_i^3 \end{bmatrix} = \begin{bmatrix} X_i-\mu \\ (X_i-\mu)^2 \\ (X_i-\mu)^3 \end{bmatrix};
$$

  which are IID, with

$$
\begin{aligned}
\bar{\psi} &= \frac{1}{n}\sum_{i=1}^{n}\psi_i = \begin{bmatrix} m_1' \\ m_2' \\ m_3' \end{bmatrix} \\
\mu_\psi = \mathrm{E}(\psi_i) &= \begin{bmatrix} \mathrm{E}(X_i-\mu) \\ \mathrm{E}(X_i-\mu)^2 \\ \mathrm{E}(X_i-\mu)^3 \end{bmatrix} = \begin{bmatrix} \mu \\ \sigma^2 \\ \mu_3 \end{bmatrix};
\end{aligned}
$$

$$\Sigma_\psi = \mathrm{Var}(\psi_i) = \begin{bmatrix} \mathrm{Var}(X_i-\mu) & \mathrm{Cov}((X_i-\mu),(X_i-\mu)^2) & \mathrm{Cov}((X_i-\mu),(X_i-\mu)^3) \\ \mathrm{Cov}((X_i-\mu),(X_i-\mu)^2) & \mathrm{Var}((X_i-\mu)^2) & \mathrm{Cov}((X_i-\mu)^2,(X_i-\mu)^3) \\ \mathrm{Cov}((X_i-\mu),(X_i-\mu)^3) & \mathrm{Cov}((X_i-\mu)^2,(X_i-\mu)^3) & \mathrm{Var}((X_i-\mu)^3) \end{bmatrix};$$

$$= \text{...calculations...} = \begin{bmatrix} \sigma^2 & \mu_3 & \mu_4 \\ \mu_3 & \mu_4-\sigma^2 & \mu_5-\sigma^2\mu_3 \\ \mu_4 & \mu_5-\sigma^2\mu_3 & \mu_6-\sigma^2\mu_3 \end{bmatrix}$$

For instance, you can compute the covariance terms as

$$\begin{aligned}
\mathrm{Cov}((X_i-\mu)^2,(X_i-\mu)^3) &= \mathrm{E}(\,((X_i-\mu)^2-\mathrm{E}(X_i-\mu)^2)\,((X_i-\mu)^3-\mathrm{E}(X_i-\mu)^3)\,) \\
&= \mathrm{E}(\,((X_i-\mu)^2-\mu_2)\,((X_i-\mu)^3-\mu_3)\,) \\
&= \mathrm{E}(\,(X_i-\mu)^5-\mathrm{E}(X_i-\mu)^2\mu_3-\mu_2(X_i-\mu)^3+\mu_2\mu_3\,) \\
&= \mu_5-\mu_2\mu_3
\end{aligned}$$

So by CLT

$$\sqrt{n}(\underbrace{\begin{bmatrix} m'_1 \\ m'_2 \\ m'_3 \end{bmatrix}}_{=\bar{\psi}} - \underbrace{\begin{bmatrix} \mu \\ \sigma^2 \\ \mu_3 \end{bmatrix}}_{=\mu_\psi}) \xrightarrow{D} \mathrm{N}(0\underbrace{\begin{bmatrix} \sigma^2 & \mu_3 & \mu_4 \\ \mu_3 & \mu_4-\sigma^2 & \mu_5-\sigma^2\mu_3 \\ \mu_4 & \mu_5-\sigma^2\mu_3 & \mu_6-\sigma^2\mu_3 \end{bmatrix}}_{=\Sigma_\psi})$$

- Now, in order to find the asymptotic distribution of $m_3 = m'_3 - 3m'_2m'_1 + 2(m'_1)^2$, I will use Delta method

  Let

  $$q(a,b,c) = c - 3ab + 2a^3$$

  then

  $$\dot{q}(a,b,c) = \frac{\mathrm{d}}{\mathrm{d}(a,b,c)}q(a,b,c) = \begin{bmatrix} -3b+6a^2, & -3a, & 1 \end{bmatrix}$$

  So

  $$q(m'_1,m'_2,m'_3) = m'_3 - 3m'_2m'_1 + 2(m'_1)^2$$

  $$q(\mu'_1,\mu'_2,\mu'_3) = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^2 = \mu_3$$

  and

  $$\dot{q}(\mu,\sigma^2,\mu_3) = \begin{bmatrix} -3\sigma^2, & \mu, & 1 \end{bmatrix}$$

and

$$\dot{q}(\mu_\psi)\Sigma_\psi\dot{q}(\mu_\psi)^T = \begin{bmatrix}-3\sigma^2, & \mu, & 1\end{bmatrix}\begin{bmatrix}\sigma^2 & \mu_3 & \mu_4 \\ \mu_3 & \mu_4-\sigma^2 & \mu_5-\sigma^2\mu_3 \\ \mu_4 & \mu_5-\sigma^2\mu_3 & \mu_6-\sigma^2\mu_3\end{bmatrix}\begin{bmatrix}-3\sigma^2, & \mu, & 1\end{bmatrix}^T$$
$$= \mu_6-\mu_3^2-6\sigma^2\mu_4+9\sigma^6$$

So the asymptotic distribution of m_3 is such that

$$\sqrt{n}(q(\bar{\psi})-q(\mu_\psi)) \xrightarrow{D} \mathrm{N}(0, \dot{q}(\mu_\psi)\Sigma_\psi\dot{q}(\mu_\psi)^T)$$
$$\sqrt{n}(m_3-\mu_3) \xrightarrow{D} \mathrm{N}(0, \mu_6-\mu_3^2-6\sigma^2\mu_4+9\sigma^6)$$

Exercise 20. ($\star\star$) Assume X_1,X_2,X_3 independent from Uniform distribution $\mathrm{U}(0,1)$. Compare the exact, Normal approximation, and Edgeworth approximation.

Hint: The exact result is $P(X_1+X_2+x_3 \le 2) = 0.8333$

Solution.

It is $\mu = 1/2$, $\sigma^2 = 1/12$, $\kappa_3 = 0$. Also, $\mathrm{E}(X-1/2)^4 = \int_0^1 (x-1/2)^4 \mathrm{d}x = 1/80$. So $\kappa_4 = \mathrm{E}(X-1/2)^4/\sigma^4 - 3 = -1.2$.

So

Normal approx. $P(X_1+X_2+x_3 \le 2) = P(\sqrt{3}(\bar{X}_3-\mu)^2/\sigma \le (\frac{2}{3}-\frac{1}{2})\sqrt{12}\sqrt{3}) \approx \Phi(1) = 0.8413$

Edgeworth Expansion. $P(X_1+X_2+x_3 \le 2) \approx \Phi(1) + 0 - 1.2(1-3)/(24\times 3)\phi(1) = 0.8332$

The next exercise is from Homework 3

Exercise 21. ($\star\star\star$) Consider an M-way contingency table and consider the quantities obs. cell counts, cell probabilities, cell proportions in their vectorised forms as

$$\underset{\sim}{n} = (n_1, ..., n_N)^T \; ; \qquad \underset{\sim}{\pi} = (\pi_1, ..., \pi_N)^T \; ; \qquad \underset{\sim}{p} = (p_1, ..., p_N)^T$$

where $n = \sum_{j=1}^N n_j$, and $p_j = n_j/n$.

1. Consider a constant matrix $C \in \mathbb{R}^{k\times N}$, and show that

$$\sqrt{n}(C\log(\underset{\sim}{p}) - C\log(\underset{\sim}{\pi})) \xrightarrow{D} \mathrm{N}(0, C\mathrm{diag}(\pi)^{-1}C^T - C11^TC^T) \tag{3}$$

2. Consider a 3×3 contingency table with probabilities $(\pi_{i,j})$. Find the joint asymptotic distribution of the vector of different log odd ratios

$$\log(\underset{\sim}{\theta}^C) = \begin{bmatrix} \log(\frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}}) \\ \log(\frac{\pi_{22}\pi_{33}}{\pi_{23}\pi_{32}}) \end{bmatrix}$$

Solution.

Exercise 22. ($\star\star\star$) Consider a random sample $X, X_1, X_2, ...$ an IID sample with finite moments $E(X) = 0$, and $\mathrm{E}(X^4) < \infty$.

1. Show that if $m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ then

$$\sqrt{n}(\begin{bmatrix} m_1 \\ m_2 \end{bmatrix} - \begin{bmatrix} 0 \\ \sigma^2 \end{bmatrix}) \xrightarrow{D} \mathrm{N}(0, \Sigma)$$

   where $\Sigma = \begin{bmatrix} Var(X) & Cov(X^2, X) \\ Cov(X^2, X) & Var(X^2) \end{bmatrix}$

2. Find an $(1 - a)\%$ asymptotic confidence interval for S_n^2.

Solution.

1. Consider $\xi_i = \begin{bmatrix} X_i \\ X_i^2 \end{bmatrix}$, and $\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$ then $\bar{\xi} = (m_1, m_2)^T$. So from the CLT, we get

$$\sqrt{n}(\bar{\xi} - E\begin{bmatrix} X_i \\ X_i^2 \end{bmatrix}) \xrightarrow{D} \mathrm{N}(0, \begin{bmatrix} Var(X) & Cov(X^2, X) \\ Cov(X^2, X) & Var(X^2) \end{bmatrix})$$

   which is what I want to show

2. I will try to compute the asymptotic distribution of S_n^2 with the Delta Method, and then I ll compute the asymptotic confidence interval.

   - Because $S_n^2 = m_2 - (m_1)^2$, I consider $g((x, y)) = y - x^2$.
   - Because $\frac{\mathrm{d}}{\mathrm{d}(x,y)} g((x, y)) = (-2x, 1)$ and continuous, then the assumptions of Delta method are satisfied, with

$$\begin{aligned} \dot{g}((0, \sigma^2))\Sigma\dot{g}((0, \sigma^2))^T =& Var(X^2) = E((X^2)^2) - (E(X^2))^2 \\ =& EX^4 - (E(X^2) - (\overset{0}{\cancel{EX}})^2)^2 \\ =& EX^4 - (Var(X))^2 = EX^4 - \sigma^4 \end{aligned}$$

- So

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{D} \mathrm{N}(0, EX^4 - \sigma^4)$$

or

$$\sqrt{n}\frac{S_n^2 - \sigma^2}{\sqrt{EX^4 - \sigma^4}} \xrightarrow{D} \mathrm{N}(0,1)$$

- By using slusky theorem it is $\frac{EX^4-\sigma^4}{\overline{X^4}-S^4} \xrightarrow{D} 1$
- and again by using slusky theorem it is

$$\sqrt{n}\frac{S_n^2 - \sigma^2}{\sqrt{\overline{X^4} - S^4}} \xrightarrow{D} \mathrm{N}(0,1)$$

- Hence

$$\{S_n^2 \pm z_{1-\frac{a}{2}}\sqrt{\frac{\overline{X^4} - S^4}{n}}\}$$

The next exercise is from Homework 3

Exercise 23. ($\star\star\star$) Consider an IID sample $X, X_1, X_2, ...$ with $EX = 0$, $EX^4 < \infty$. Consider that

$$\sqrt{n}\frac{S_n^2 - \sigma^2}{\sqrt{EX^4 - \sigma^4}} \xrightarrow{D} \mathrm{N}(0,1) \tag{4}$$

where $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

1. Find the asymptotic distribution of $\log(S_n^2)$.
2. Produce the $1 - a$ asymptotic confidence interval for $\log(\sigma_n^2)$; by performing suitable calculations, so that the boundaries of the confidence interval do not depend on any unknown moments of the real distribution.

Solution.

Exercise 24. ($\star\star\star$) Let function $g : \mathbb{R} \to \mathbb{R}$ such that $\dot{g}(x)$ and $\ddot{g}(x)$ are continuous in a neighborhood of $\mu \in \mathbb{R}$, and $\dot{g}(\mu) = 0$. Prove the following statement:

- If $X_n \in \mathbb{R}$ is a sequence of random vectors such that $\sqrt{n}(X_n - \mu) \xrightarrow{D} \mathrm{N}(0, \sigma^2)$ then

$$n(g(X_n) - g(\mu)) \xrightarrow{D} \frac{\sigma^2 \ddot{g}(\mu)}{2}\chi_1^2$$

Hint-1. Use Taylor expansion of 2nd order.

Hint-2. The Taylor expansion of function $f : \mathbb{R} \to \mathbb{R}$ around point x_0 is:

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!}(x - x_0) f^{(k)}(x_0) + R_n(x)$$

where $R_n(x) = \frac{1}{n!} \int_{x_0}^{x} (x - t)^n f^{(n)}(x_0) = o((x - x_0)^n)$ as $x \to x_0$, provided that the n-th derivative $f^{(n)}(x)$ exists in some interval containing x_0.

Solution.

1. It is $\sqrt{n}(X_n - \mu) \xrightarrow{D} \mathrm{N}(0, \sigma^2)$. We expand $g(X_n)$ by Taylor (2nd degree) around μ. So

$$\begin{aligned} g(x) =& g(\mu) + \cancelto{0}{\dot{g}(\mu)}(x - \mu) + \frac{\ddot{g}(\mu)}{2}(x - \mu)^2 + o((x - \mu)^2) \\ \approx& \frac{\ddot{g}(\mu)}{2}(x - \mu)^2 \end{aligned}$$

Or

$$n(g(X_n) - g(\mu)) \approx \frac{\sigma^2 \ddot{g}(\mu)}{2} (\sqrt{n} \frac{X_n - \mu}{\sigma})^2$$

Because $\sqrt{n}(X_n - \mu) \xrightarrow{D} \mathrm{N}(0, \sigma^2)$, it is $\sqrt{n}\frac{X_n - \mu}{\sigma} \xrightarrow{D} \mathrm{N}(0, 1)$.

So by Slutsky theorem, $(\frac{X_n - \mu}{\sigma})^2 \xrightarrow{D} \chi^2_1$. Then

$$n(g(X_n) - g(\mu)) \approx \frac{\sigma^2 \ddot{g}(\mu)}{2} \chi^2_1$$

The next exercise is from Homework 3

Exercise 25. ($\star\star\star$) Consider random sample $X, X_1, X_2, ...$ IID from a Bernoulli distribution with probability of success p. Find the variance stabilization transformation for the estimator average $\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Solution.

Exercise 26. Prove the Information inequality theorem:

Let $x \in \mathbb{R}^d$ random vector following distribution $\mathrm{d}f_\theta(\cdot)$ labeled by an parameter $\theta \in \Theta \subset \mathbb{R}^r$ and admitting PDF $f(\cdot|\theta)$. Consider an estimator $\hat{\theta}_n := \hat{\theta}_n(x) \in \Theta \subset \mathbb{R}^r$ such that $g(\theta) = \mathrm{E}_{f_\theta}(\hat{\theta}_n)$

exists on Θ. Assume that, $\frac{\mathrm{d}}{\mathrm{d}\theta}f(x|\theta)$ exists ; $\frac{\mathrm{d}}{\mathrm{d}\theta}$ can pass under the integral sign in $\int f(x|\theta)\mathrm{d}x$ and $\int \hat{\theta}_n(x)f(x|\theta)\mathrm{d}x$. Then

$$\mathrm{var}_{f_\theta}(\hat{\theta}_n(x)) \geq \frac{1}{n}\dot{g}(\theta)\mathcal{I}(\theta)^{-1}\dot{g}(\theta)^T \tag{5}$$

where $\mathcal{I}(\theta)$ is the Fisher's information matrix.

- The quantity $\frac{1}{n}\dot{g}(\theta)\mathcal{I}(\theta)^{-1}\dot{g}(\theta)^T$ is called Cramer-Rao lower bound (CRLB).

Hint-1: Use $0 \leq \mathrm{var}_{f_\theta}(\hat{\theta}_n - \dot{g}(\theta)\mathcal{I}(\theta)^{-1}\Psi(x,\theta)) = ...$

Hint-2: Use $\mathrm{var}_{f_\theta}(A+B) = \mathrm{var}_{f_\theta}(A) + \mathrm{var}_{f_\theta}(B) + 2\mathrm{cov}_{f_\theta}(A,B)$

Solution. Let $\Psi(x,\theta) = (\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(x|\theta))^T$.

It is

$$\begin{aligned}
\mathrm{E}_{f_\theta}\Psi(X,\theta) =& 0 \quad \text{(you have proved it before)} \\
\dot{g}_n(\theta) =& \frac{\mathrm{d}}{\mathrm{d}\theta}\int \hat{\theta}_n(x)f(x|\theta)\mathrm{d}x = \int \hat{\theta}_n(x)\frac{\frac{\mathrm{d}}{\mathrm{d}\theta}f(x|\theta)}{f(x|\theta)}f(x|\theta)\mathrm{d}x \\
=& \int \hat{\theta}_n(x)\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(x|\theta)f(x|\theta)\mathrm{d}x = \mathrm{E}_{f_\theta}(\hat{\theta}_n(x)(\Psi(x,\theta) - \cancelto{=0}{\mathrm{E}_\theta\Psi(X,\theta)})) \\
=& \mathrm{cov}_{f_\theta}(\hat{\theta}_n(x), \Psi(x,\theta))
\end{aligned} \tag{6}$$

So

$$\begin{aligned}
0 \leq& \mathrm{var}_{f_\theta}(\hat{\theta}_n - \dot{g}(\theta)\mathcal{I}(\theta)^{-1}\Psi(x,\theta)) \\
=& \mathrm{var}_{f_\theta}(\hat{\theta}_n) - 2\mathrm{cov}_{f_\theta}(\hat{\theta}_n, \dot{g}(\theta)\mathcal{I}(\theta)^{-1}\Psi(x,\theta)) + \mathrm{var}_{f_\theta}(\dot{g}(\theta)\mathcal{I}(\theta)^{-1}\Psi(x,\theta)) \\
=& \mathrm{var}_{f_\theta}(\hat{\theta}_n) - 2\frac{1}{n}\dot{g}(\theta)\mathcal{I}(\theta)^{-1}\dot{g}(\theta)^T + \frac{1}{n}\dot{g}(\theta)\mathcal{I}(\theta)^{-1}\mathcal{I}(\theta)\mathcal{I}(\theta)^{-1}\dot{g}(\theta)^T \\
=& \mathrm{var}_{f_\theta}(\hat{\theta}_n) - \frac{1}{n}\dot{g}(\theta)\mathcal{I}(\theta)^{-1}\dot{g}(\theta)^T
\end{aligned}$$

and the proof is done