Topics in statistics III/IV (MATH3361/4071)	Michaelmas term, 2020-2021
Exercises: Likelihood methods	
Lecturer & Author: Georgios Karagiannis	georgios.karagiannis@durham.ac.uk

1 Handout 1: Basic probability tools in asymptotics

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.

(a) It is

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

(b) It is

$$\begin{aligned}\mathrm{E}_\pi|X-\theta|^2 &= \mathrm{E}_\pi(X-\theta)^T(X-\theta)\\ |\mathrm{E}_\pi(X)-\theta|^2 &= (\mathrm{E}_\pi(X)-\theta)^T(\mathrm{E}_\pi(X)-\theta)\end{aligned}$$

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

$$f(x) = f(0) + \frac{1}{1!}\dot{f}(0)(x-0) + o(x), \text{ as }, \text{ as } x \to 0$$

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

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

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 **??** 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]

[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

So as $n \to \infty$

$$P(X \leq z - 1\epsilon) \leq \lim \inf_{n\to\infty} P(X_n \leq z) \leq \lim \sup_{n\to\infty} P(X_n \leq z) \leq P(X \leq z + 1\epsilon)$$

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.

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

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

---

2 Handout 2: Basic tools for asymptotics in statistics

Exercise 11. 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 $\bar{X}_n \xrightarrow{\text{qm}} \mathrm{E}(X)$

4. Let $\varphi_X(t)=\mathrm{E}(e^{it^TX})$, 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 $\lim_{\epsilon\rightarrow 0}\dot{\varphi}_X(\epsilon)=\dot{\varphi}_X(0)=i\mu^T$ it is

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

Here I used that $\lim_{n\rightarrow\infty}(1+a_n)^n=\exp(\lim_{n\rightarrow\infty}na_n)$ if $\lim_{n\rightarrow\infty}na_n$ 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

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

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

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

(b) It is

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

as the 2nd mode is finite.

Exercise 12. Show

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

Solution.

- Deterministic: If $x_n=O(h_n)$ and $h_n\rightarrow 0$, then $x_n=o(1)$, because we sandwich $|x_n|\leq Kh_n\rightarrow 0$.
- Stochastic: If $x_n=O_P(h_n)$ and $h_n\rightarrow 0$, then $x_n=o_P(1)$. Because $h_n\rightarrow 0$, 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 13. 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

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

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

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

So

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

which implies

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

Exercise 15. ($\star\star$)

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 16. (⋆⋆⋆⋆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

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

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

Solution. It is $\mathrm{E}\left(\left|\underbrace{1\,(X_i\leq x)}_{\xi_i}\right|\right)<\infty$; and $\mathrm{E}\left(\underbrace{1(X\leq x)}_{\xi}\right)=F_X(x)$. So by the strong LLN,

$$\hat{F}_X(x)=\frac{1}{n}\sum_{i=1}^{n}\underbrace{1\,(X_i\leq x)}_{\xi_i}\overset{\text{as}}{\to}\mathrm{E}(\underbrace{1(X\leq x)}_{\xi})=F_X(x)$$

.

Exercise 18. ($\star\star$) 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 Problem Class 2

3 Handout 3: Asymptotics after transformations

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

$$\sqrt{n}(h(\bar{X}, s_x^2) - h(\mu, \sigma^2)) \xrightarrow{D} \mathrm{N}(0, \dot{h}(\mu, \sigma^2)\Sigma_g \dot{h}(\mu, \sigma^2)^T)$$

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

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

$$\bar{\psi} = \frac{1}{n}\sum_{i=1}^{n}\psi_i = \begin{bmatrix} m_1' \\ m_2' \\ m_3' \end{bmatrix}$$

$$\mu_\psi = \mathrm{E}(\psi_i) = \begin{bmatrix} \mathrm{E}(X_i - \mu) \\ \mathrm{E}(X_i - \mu)^2 \\ \mathrm{E}(X_i - \mu)^3 \end{bmatrix} = \begin{bmatrix} \mu \\ \sigma^2 \\ \mu_3 \end{bmatrix};$$

$$\begin{aligned} \Sigma_\psi = \mathrm{Var}(\psi_i) &= \begin{bmatrix} \mathrm{Var}(X_i - \mu) & \mathrm{Cov}((X_i - \mu), (X_i - \mu)^2) & \mathrm{Cov}((X_i - \mu), (X_i - \mu)^3) \\ \mathrm{Cov}((X_i - \mu), (X_i - \mu)^2) & \mathrm{Var}((X_i - \mu)^2) & \mathrm{Cov}((X_i - \mu)^2, (X_i - \mu)^3) \\ \mathrm{Cov}((X_i - \mu), (X_i - \mu)^3) & \mathrm{Cov}((X_i - \mu)^2, (X_i - \mu)^3) & \mathrm{Var}((X_i - \mu)^3) \end{bmatrix}; \\ &= \text{...calculations...} = \begin{bmatrix} \sigma^2 & \mu_3 & \mu_4 \\ \mu_3 & \mu_4 - \sigma^2 & \mu_5 - \sigma^2\mu_3 \\ \mu_4 & \mu_5 - \sigma^2\mu_3 & \mu_6 - \sigma^2\mu_3 \end{bmatrix} \end{aligned}$$

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 $m_3=m_3'-3m_2'm_1'+2(m_1')^2$, 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

  $$q(m_1',m_2',m_3')=m_3'-3m_2'm_1'+2(m_1')^2$$

  $$q(\mu_1',\mu_2',\mu_3')=\mu_3'-3\mu_2'\mu_1'+2(\mu_1')^2=\mu_3$$

  and

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

  and

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

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. ($\star\star\star$) 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

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

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.

1. By using CLT (same as in Example in the CLT section in the Handouts), we get

$$\sqrt{n}(p - \pi) \xrightarrow{D} \mathrm{N}(0, \mathrm{diag}(\pi) - \pi\pi^T)$$

Recall from the example:

Denote the i-th observation by $\xi_i = (\xi_{i,1}, ..., \xi_{i,N})^T$, where

$$\xi_{i,j} = \begin{cases} 1 & \text{, if observation } i \text{ falls in cell } j \\ 0 & \text{, if observation } i \text{ does not fall in cell } j \end{cases}$$

Since its observation falls in only one cell, $\sum_j \xi_{i,j} = 1$ and $\xi_{i,j}\xi_{i,k} = 0$ when $j \neq k$. Therefore p can be considered as the arithmetic mean of $\{\xi_{i,j}\}_{i=1}^n$ IID variables as

$$p = \frac{1}{n}\sum_{i=1}^{n} \xi_i$$

The moments of $\{\xi_i\}$, are equal to

$$\begin{aligned} \mathrm{E}(\xi_i) &= \pi \\ \mathrm{Var}(\xi_i) &= \Sigma \end{aligned}$$

where

$$\begin{aligned} [\Sigma]_{j,j} =& \mathrm{var}(\xi_{i,j}) = \mathrm{E}(\xi_{i,j}^2) - (\mathrm{E}(\xi_{i,j}))^2 = \pi_j(1-\pi_j) \\ [\Sigma]_{j,k} =& \mathrm{cov}(\xi_{i,j}, \xi_{i,k}) = \mathrm{E}(\xi_{i,j}\xi_{i,k}) - \mathrm{E}(\xi_{i,j})\mathrm{E}(\xi_{i,k}) = -\pi_j\pi_k \end{aligned}$$

because

$$\begin{aligned} \mathrm{E}(\xi_{i,j}) =& \mathrm{P}(\xi_{i,j} = 1) = \pi_j \\ \mathrm{E}(\xi_{i,j}^2) =& \mathrm{P}(\xi_{i,j} = 1) = \pi_j \\ \mathrm{E}(\xi_{i,j}\xi_{i,k}) =& 0, \text{ if } j \neq k \end{aligned}$$

Hence

$$\Sigma = \mathrm{diag}(\pi) - \pi\pi^T$$

Therefore, according to the CLT

$$\sqrt{n}(p - \pi) \xrightarrow{D} \mathrm{N}(0, \mathrm{diag}(\pi) - \pi\pi^T) \tag{4}$$

Consider a function

$$g(x) = C\log(x) = C\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$$

Calculate

$$\dot{g}(x) = C\text{diag}(\pi)^{-1} = C \begin{bmatrix} 1/\pi_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1/\pi_N \end{bmatrix}$$

and notice that it is continuous. So Delta method can be used.

Calculate

$$\begin{aligned}
\dot{g}(\mu)\left(\text{diag}(\pi) - \pi\pi^T\right)\dot{g}(\mu)^T &= \dot{g}(\mu)\text{diag}(\pi)\,\dot{g}(\mu)^T - \dot{g}(\mu)\pi\pi^T\dot{g}(\mu)^T \\
&= C\cancelto{I}{\text{diag}(\pi)^{-1}\text{diag}(\pi)}\text{diag}(\pi)^{-1}C^T - C\cancelto{1}{\text{diag}(\pi)^{-1}\pi}\cancelto{1^T}{\pi^T\text{diag}(\pi)^{-1}}C^T \\
&= C\text{diag}(\pi)^{-1}C^T - C11^TC^T
\end{aligned}$$

Hence from Delta method we get

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

(a) Let $\underset{\sim}{\pi} = \begin{bmatrix} \pi_{11} & \pi_{21} & \pi_{31} & \pi_{12} & \pi_{22} & \pi_{32} & \pi_{13} & \pi_{23} & \pi_{33} \end{bmatrix}^T$. In fact, the vector of different log odd ratios

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

can be expressed as $\log(\underset{\sim}{\theta}^C) = C\log(\underset{\sim}{\pi})$ with

$$\begin{aligned}
\log(\underset{\sim}{\theta}^C) &= \begin{bmatrix} \log(\pi_{11}) - \log(\pi_{12}) - \log(\pi_{21}) + \log(\pi_{22}) \\ \log(\pi_{22}) + \log(\pi_{33}) - \log(\pi_{23}) - \log(\pi_{32}) \end{bmatrix} \\
C &= \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \\
\log(\underset{\sim}{\pi}) &= \begin{bmatrix} \log\pi_{11} & \log\pi_{21} & \log\pi_{31} & \log\pi_{12} & \log\pi_{22} & \log\pi_{32} & \log\pi_{13} & \log\pi_{23} & \log\pi_{33} \end{bmatrix}^T
\end{aligned}$$

so

$$\sqrt{n}(\log(\hat{\underset{\sim}{\theta}}^C) - \log(\underset{\sim}{\theta}^C)) \xrightarrow{D} \text{N}(0, \Sigma)$$

where

$$\begin{aligned}
\Sigma &= C\text{diag}(\pi)^{-1}C^T - C11^TC^T = ... = \\
&= \begin{bmatrix} \frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}} & \frac{1}{\pi_{22}} \\ \frac{1}{\pi_{22}} & \frac{1}{\pi_{22}} + \frac{1}{\pi_{32}} + \frac{1}{\pi_{23}} + \frac{1}{\pi_{33}} \end{bmatrix}
\end{aligned}$$

Exercise 21. ($\star\star\star$) 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

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

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

$$\sqrt{n}\frac{S_n^2-\sigma^2}{\sqrt{EX^4-\sigma^4}} \xrightarrow{D} \mathrm{N}(0,1) \tag{5}$$

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 $\log(\sigma^2)$; by performing suitable calculations, so that the boundaries of the confidence interval do not depend on any unknown moments of the real distribution.

Solution. What I wrote in (5) actually means

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

1. I will try to compute the asymptotic distribution of S_n^2 with the Delta Method.

   Consider $g(x) = \log(x)$.

   Because $\frac{\mathrm{d}}{\mathrm{d}x}g(x) = \frac{1}{x}$ and continuous, then the assumptions of Delta method are satisfied, with

   $$\dot{g}(\sigma^2)\Sigma\dot{g}(\sigma^2)^T = EX^4/\sigma^4 - 1$$

   So

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

   (a) If I just use the statistic above as it is in orderr to create the CI, I will get

   $$\{\log(S_n^2) \pm z_{1-\frac{a}{2}}\sqrt{\frac{EX^4/\sigma^4-1}{n}}\}$$

whose bounds depend on the unknown moments $\mathrm{E}(X^4)$, σ^4 .

If we try to find a stabilization transformation to resolve this problem, it will not work. (many students tried that)

So we have to find an alternative way.

Alternatively we can see if we can replace the unknown moments $\mathrm{E}(X^4)$, σ^4 with their sample analogues. Possibly we can justify this by using Slutsky theorem and CLT. I have

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

It is

$$\overline{x^4} \xrightarrow{P} EX^4$$
$$\overline{x} \xrightarrow{P} EX$$
$$\overline{x^2} \xrightarrow{P} EX^2$$

Let $g(a,b)=a-b^2$. Then g is continuous, so by Slusky theorem

$$S_n^2 = g(\overline{x^2},\overline{x}) \xrightarrow{P} g(EX^2,(EX)^2)=\sigma^2$$

Let $f(a,b)=\sqrt{\frac{EX^4/\sigma^4-1}{a/b^2-1}}$. Then because f is continuous apart from a value where the probability is zero, I can use the Slutsky's Theorem and get

$$\sqrt{\frac{EX^4/\sigma^4-1}{\overline{X^4}/S_n^4-1}} \xrightarrow{P} 1$$

By using Slusky theorem we get

$$\underbrace{\sqrt{\frac{EX^4/\sigma^4-1}{\overline{X^4}/S_n^4-1}} \times \sqrt{n}\frac{\log(S_n^2)-\log(\sigma^2)}{\sqrt{EX^4/\sigma^4-1}}}_{=\sqrt{n}\frac{S_n^2-\sigma^2}{\sqrt{\overline{X^4}/S_n^4-1}}} \xrightarrow{D} 1\times \mathrm{N}(0,1)$$

So, an asymptotic CI for $\log(\sigma^2)$ is

$$\{\log(S_n^2) \pm z_{1-\frac{a}{2}}\sqrt{\frac{\overline{X^4}/S_n^4-1}{n}}\}$$

Exercise 23. (⋆⋆⋆) 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

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

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.

We expand $g(X_n)$ by Taylor (2nd degree) around μ. So

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

So

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

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

For the second term, because $\sqrt{n}(X_n - \mu) \xrightarrow{D} \mathrm{N}(0, \sigma^2)$ then $\sqrt{n}(X_n - \mu) = O_p(1)$, then $(\sqrt{n}(X_n - \mu))^2 = O_p(1)$. Hence $o\left((\sqrt{n}(X_n - \mu))^2\right) = o(O_p(1)) = o_p(1)$.

Hence by Slutsky rules:

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

The next exercise is from Homework 3

Exercise 24. (⋆⋆⋆) 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. It is $\mathrm{E}(X_i) = 1p + 0(1-p) = p$, $\mathrm{E}(X_i^2) = 1p + 0(1-p) = p$, and hence $\mathrm{Var}(X_i) = \mathrm{E}(X_i^2) - \mathrm{E}(X_i) = p(1-p)$. Therefor, from the CLT, I have

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{D} \mathrm{N}(0, p(1-p))$$

Therefore, I need a variance stabilization transformation $g(\cdot)$, such that

$$\begin{aligned}\dot{g}(p)^2 p(1-p) = 1 &\Longrightarrow \\ \dot{g}(p)\sqrt{p(1-p)} = 1 &\Longrightarrow \\ \dot{g}(p) = \frac{1}{\sqrt{p(1-p)}} &\Longrightarrow\end{aligned}$$

Then

$$\begin{aligned}g(p) =& \int \frac{1}{\sqrt{p(1-p)}} \mathrm{d}p \\ \overset{\xi=\sqrt{p}}{=} =& 2\underbrace{\int \frac{1}{\sqrt{1-\xi^2}} \mathrm{d}\xi}_{=\arcsin(\xi)} \\ =& 2\arcsin(\xi) \\ =& 2\arcsin(\sqrt{p})\end{aligned}$$

Then the variance stabilization transformation is $g(p) = 2\arcsin(\sqrt{p})$, and hence according to Delta method

$$\sqrt{n}(\arcsin(\hat{p}_n) - \arcsin(p)) \xrightarrow{D} \mathrm{N}(0, \frac{1}{4})$$

---

4 Handout 4: Estimation by the method of Maximum Likelihood

Exercise 25. Consider random sample $x_1, ..., x_n \overset{IID}{\sim} \mathrm{G}(a, b)$, $a > 0$, $b > 0$ with PDF

$$f(x|a, b) = \frac{1}{\Gamma(a)b^a} x^a e^{-x\frac{1}{b}} 1(x > 0)$$

1. Find the moment estimator $\tilde{\theta}$ of $\theta = (a, b)^T$ by using the first raw moment and the first central moment
2. Is the moment estimator $\tilde{\theta}$ consistent and asymptotically Normal?
3. Find the one step estimator by Fisher scoring algorithm.

Hint-1 Digamma function $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log \Gamma(x)$

Hint-2 Trigamma function $\psi_1(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2} \log \Gamma(x)$

Hint-3 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Solution.

1. The first raw moment is the expected value/mean, and the first central moment is the variance.

   The first raw moment is

$$\mathrm{E}(x) = \int_0^1 x \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b} \mathrm{d}x = \int_0^1 \frac{1}{\frac{1}{a}\Gamma(a+1)\frac{1}{b}b^{a+1}} x^{(a+1)-1} e^{-x/b} \mathrm{d}x = ab$$

   and the sample one

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

   The first central moment is

$$\mathrm{var}(x) = \mathrm{E}(x^2) - (\mathrm{E}(x))^2$$

   So

$$\mathrm{E}(x^2) = \int_0^1 x^2 \frac{1}{\Gamma(a)b^a} x^{a-1} e^{-x/b} \mathrm{d}x = \int_0^1 \frac{1}{\frac{1}{a(a+1)}\Gamma(a+2)\frac{1}{b^2}b^{a+2}} x^{(a+2)-1} e^{-x/b} \mathrm{d}x = a(a+1)b^2$$

   and hence

$$\mathrm{var}(x) = \mathrm{E}(x^2) - (\mathrm{E}(x))^2 = ab^2$$

   The sample first central moment is

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

   From the method of moments I get

$$\begin{cases} \mathrm{E}(x|\tilde{a},\tilde{b}) = & \bar{x} \\ \mathrm{var}(x|\tilde{a},\tilde{b}) = & s^2 \end{cases} \Longrightarrow \qquad \begin{cases} \tilde{a} = & \frac{\bar{x}^2}{s^2} \\ \tilde{b} = & \frac{\bar{x}^2}{s^2} \end{cases} \Longrightarrow \qquad \begin{cases} \tilde{a} = & \frac{(\mathrm{E}(x))^2}{\mathrm{var}(x)} = \frac{\bar{x}^2}{s^2} \\ \tilde{b} = & \frac{\mathrm{var}(x)}{\mathrm{E}(x)} = \frac{\bar{x}^2}{s^2} \end{cases}$$

   So the moment estimator is

$$\tilde{\theta} = \begin{bmatrix} \tilde{a} \\ \tilde{b} \end{bmatrix} = \begin{bmatrix} \frac{\bar{x}^2}{s^2} \\ \frac{s^2}{\bar{x}} \end{bmatrix} \tag{6}$$

2. It is consistent because $\tilde{\theta} \xrightarrow{as} \theta$. This is because of the following.

It is

$$\begin{cases} \mathrm{E}(x) = & ab \\ \mathrm{var}(x) = & ab^2 \end{cases} \Longrightarrow \qquad \begin{cases} a = & \frac{(\mathrm{E}(x))^2}{\mathrm{var}(x)} \\ b = & \frac{\mathrm{var}(x)}{\mathrm{E}(x)} \end{cases} \Longrightarrow \qquad \begin{cases} a = & \frac{(\mathrm{E}(x))^2}{\mathrm{var}(x)} \\ b = & \frac{\mathrm{var}(x)}{\mathrm{E}(x)} \end{cases}$$

From SLLN, $\bar{x} \xrightarrow{as} \mathrm{E}(x)$. From SLLN, $\overline{x^2} \xrightarrow{as} \mathrm{E}(x^2)$. From Slutsky Theorem, $s^2 = \overline{x^2} - (\overline{x})^2 \xrightarrow{as} \mathrm{E}(x^2) - \mathrm{E}(x)^2 = \mathrm{var}(x)$

So From Slutsky theorem

$$\tilde{\theta} = \begin{bmatrix} \frac{\bar{x}^2}{s^2} \\ \frac{s^2}{\bar{x}} \end{bmatrix} \xrightarrow{as} \begin{bmatrix} \frac{(\mathrm{E}(x))^2}{\mathrm{var}(x)} \\ \frac{\mathrm{var}(x)}{\mathrm{E}(x)} \end{bmatrix} = \theta$$

It is asymptotically Normal because of the following.

$\bar{x}$ and s^2 are asymptotically Normal by the CLT, as averages of IID quantities. Hence, by Delta method, (16) is asymptotically Normal.

3. Recall the the one-step estimators

$$\text{Newton alg.} \qquad \breve{\theta}_n = \tilde{\theta}_n - \ddot{\ell}_n(\tilde{\theta}_n)^{-1}\dot{\ell}_n(\tilde{\theta}_n) \tag{7}$$

$$\text{Fisher scoring alg.} \qquad \breve{\theta}_n = \tilde{\theta}_n + \frac{1}{n}\mathcal{I}(\tilde{\theta}_n)^{-1}\dot{\ell}_n(\tilde{\theta}_n) \tag{8}$$

For the Fisher algorithm, I need to find $\mathcal{I}(\theta)^{-1}$. It is

$$\begin{aligned}
\log f(x|\theta) &= -\log\Gamma(a) - a\log(b) - \frac{1}{b}x + (a-1)\log(x) \\
\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(x|\theta) &= \begin{bmatrix} -\psi(a) - \log(b) + \log(x) \\ -\frac{a}{b} + \frac{1}{b^2}x \end{bmatrix} \\
\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(x|\theta) &= \begin{bmatrix} -\psi_1(a) & -\frac{1}{b} \\ -\frac{1}{b} & -\frac{2x-ab}{b^3} \end{bmatrix} \\
\mathcal{I}(\theta) &= \begin{bmatrix} \psi_1(a) & \frac{1}{b} \\ \frac{1}{b} & \frac{a}{b^2} \end{bmatrix} \\
\mathcal{I}(\theta)^{-1} &= \frac{1}{a\psi_1(a) - 1}\begin{bmatrix} a & -b \\ -b & b^2\psi_1(a) \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\ell_n(\theta) &= -n\log\Gamma(a) - na\log(b) - \frac{1}{b}\sum_{i=1}^{n} x_i + (a-1)\sum_{i=1}^{n}\log(x_i) \\
\dot{\ell}_n(\theta) &= \begin{bmatrix} -n\psi(a) - n\log(b) + \sum_{i=1}^{n}\log(x_i) \\ -n\frac{a}{b} + n\frac{1}{b^2}\bar{x} \end{bmatrix}
\end{aligned}$$

The Fisher recursion is

$$\breve{\theta}_n = \tilde{\theta}_n + \frac{1}{n}\mathcal{I}(\tilde{\theta}_n)^{-1}\dot{\ell}_n(\tilde{\theta}_n)$$

$$\breve{\theta}_n = \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{n}\frac{1}{\tilde{a}\psi_1(\tilde{a})-1}\begin{bmatrix}\tilde{a} & -\tilde{b}\\ -\tilde{b} & \tilde{b}^2\psi_1(\tilde{a})\end{bmatrix}\begin{bmatrix}-n\psi(\tilde{a}) - n\log(\tilde{b}) + \sum_{i=1}^n \log(x_i)\\ -n\frac{\tilde{a}}{\tilde{b}} + n\frac{1}{\tilde{b}^2}\bar{x}\end{bmatrix}$$

$$= \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{\tilde{a}\psi_1(\tilde{a})-1}\begin{bmatrix}\tilde{a} & -\tilde{b}\\ -\tilde{b} & \tilde{b}^2\psi_1(\tilde{a})\end{bmatrix}\begin{bmatrix}-\psi(\tilde{a}) - \log(\tilde{b}) + \frac{1}{n}\sum_{i=1}^n \log(x_i)\\ -\frac{\tilde{a}}{\tilde{b}} + \frac{1}{\tilde{b}^2}\bar{x}\end{bmatrix}$$

$$= \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{\tilde{a}\psi_1(\tilde{a})-1}\begin{bmatrix}-\tilde{a}\psi(\tilde{a}) - \frac{1}{\tilde{b}}(\bar{x}-\tilde{a}\tilde{b}) - \tilde{a}\log(\tilde{b}) + \frac{\tilde{a}}{n}\sum_{i=1}^n \log(x_i)\\ \tilde{b}\psi(\tilde{a}) - \psi_1(\tilde{a})(\bar{x}-\tilde{a}) + \tilde{b}\log(\tilde{b}) - \frac{\tilde{b}}{n}\sum_{i=1}^n \log(x_i)\end{bmatrix}$$

So bu substituting

$$\breve{\theta}_n = \begin{bmatrix}\frac{\bar{x}^2}{s^2}\\ \frac{s^2}{\bar{x}}\end{bmatrix} + \frac{1}{\frac{\bar{x}^2}{s^2}\psi_1(\frac{\bar{x}^2}{s^2})-1}\begin{bmatrix}-\frac{\bar{x}^2}{s^2}\psi(\frac{\bar{x}^2}{s^2}) - \frac{\bar{x}^2}{s^2}\log(\frac{s^2}{\bar{x}}) + \frac{1}{n}\frac{\bar{x}^2}{s^2}\sum_{i=1}^n \log(x_i)\\ \frac{s^2}{\bar{x}}\psi(\frac{\bar{x}^2}{s^2}) - \psi_1(\frac{\bar{x}^2}{s^2})(\bar{x}-\frac{\bar{x}^2}{s^2}) + \frac{s^2}{\bar{x}}\log(\frac{s^2}{\bar{x}}) - \frac{\tilde{b}}{n}\sum_{i=1}^n \log(x_i)\end{bmatrix}$$

Additionally for the Newton recursion I need

$$\ddot{\ell}_n(\theta) = -n\begin{bmatrix}\psi_1(a) & \frac{1}{b}\\ \frac{1}{b} & \frac{2\bar{x}-ab}{b^3}\end{bmatrix}$$

$$(\ddot{\ell}_n(\theta))^{-1} = -\frac{1}{n}\frac{1}{\psi_1(a)\frac{2\bar{x}-ab}{b}-1}\begin{bmatrix}\frac{2\bar{x}-ab}{b} & -b\\ -b & b^2\psi_1(a)\end{bmatrix}$$

The Newton recursion is

$$\breve{\theta}_n = \tilde{\theta}_n - (\ddot{\ell}_n(\theta))^{-1}\dot{\ell}_n(\tilde{\theta}_n)$$

$$\breve{\theta}_n = \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{n}\frac{1}{\psi_1(\tilde{a})\frac{2\bar{x}-\tilde{a}\tilde{b}}{\tilde{b}}-1}\begin{bmatrix}2\bar{x}-\tilde{a}\tilde{b} & \tilde{b}\\ \tilde{b} & \tilde{b}^2\psi_1(\tilde{a})\end{bmatrix}\begin{bmatrix}-n\psi(\tilde{a}) - n\log(\tilde{b}) + \sum_{i=1}^n \log(x_i)\\ -n\frac{\tilde{a}}{\tilde{b}} + n\frac{1}{\tilde{b}^2}\bar{x}\end{bmatrix}$$

$$= \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{\psi_1(\tilde{a})\frac{2\bar{x}-\tilde{a}\tilde{b}}{\tilde{b}}-1}\begin{bmatrix}2\bar{x}-\tilde{a}\tilde{b} & \tilde{b}\\ \tilde{b} & \tilde{b}^2\psi_1(\tilde{b})\end{bmatrix}\begin{bmatrix}-\psi(\tilde{a}) - \log(\tilde{b}) + \frac{1}{n}\sum_{i=1}^n \log(x_i)\\ -\frac{\tilde{a}}{\tilde{b}} + \frac{1}{\tilde{b}^2}\bar{x}\end{bmatrix}$$

$$= \text{...calculations}$$

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

$$\mathrm{var}_{f_\theta}(\hat{\theta}_n(x)) \geq \frac{1}{n}\dot{g}(\theta)\mathcal{I}(\theta)^{-1}\dot{g}(\theta)^T \tag{9}$$

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

$$\begin{aligned}
\mathrm{E}_{f_\theta}\Psi(X,\theta) =& 0 \quad \text{(you have proved it before)} \\
\dot{g}_n(\theta) =& \frac{\mathrm{d}}{\mathrm{d}\theta}\int \hat{\theta}_n(x) f(x|\theta)\mathrm{d}x = \int \hat{\theta}_n(x)\frac{\frac{\mathrm{d}}{\mathrm{d}\theta}f(x|\theta)}{f(x|\theta)} f(x|\theta)\mathrm{d}x \\
=& \int \hat{\theta}_n(x)\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(x|\theta) f(x|\theta)\mathrm{d}x = \mathrm{E}_{f_\theta}(\hat{\theta}_n(x)(\Psi(x,\theta) - \cancelto{=0}{\mathrm{E}_\theta\Psi(X,\theta)})) \\
=& \mathrm{cov}_{f_\theta}(\hat{\theta}_n(x), \Psi(x,\theta))
\end{aligned} \tag{10}$$

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

Exercise 27. Prove the following statement: Given that the assumptions of Cramer Theorem (for the Normality of MLE) are satisfied, and that $\mathcal{I}(\theta)$ and $\mathcal{J}_n(\theta)$ are continuous on θ, then

$$\sqrt{n}\mathcal{I}(\theta_0)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, I) \tag{11}$$

$$\sqrt{n}\mathcal{I}(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, I) \tag{12}$$

$$\mathcal{J}_n(\hat{\theta}_n)^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, I) \tag{13}$$

where $\hat{\theta}_n$ denotes the MLE, θ_0 denotes the true value of θ, and $A^{1/2}$ denotes the lower triangular matrix of the Cholesky decomposition of A; i.e., $A = A^{1/2}(A^{1/2})^T$.

Solution.

- Eq 11 results from Cramer Theorem, and the properties of covariance matrix.
  - Eq. 12 results by using Cramer Theorem and Slutsly theorems. Precisely, because $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$, Slutsky implies $\mathcal{I}(\hat{\theta}_n) \xrightarrow{a.s.} \mathcal{I}(\theta_0)$ which implies $\mathcal{I}(\hat{\theta}_n)^{1/2}\mathcal{I}(\theta_0)^{-1/2} \xrightarrow{a.s.} I$. Therefore, by Slutsky

    $$\underbrace{\mathcal{I}(\hat{\theta}_n)^{1/2}\mathcal{I}(\theta_0)^{-1/2}\sqrt{n}\mathcal{I}(\theta_0)^{1/2}(\hat{\theta}_n - \theta_0)}_{=\sqrt{n}\mathcal{I}(\hat{\theta}_n)^{1/2}(\hat{\theta}_n-\theta_0)} \xrightarrow{D} \underbrace{I \times \mathrm{N}(0, I)}_{=\mathrm{N}(0,I)}$$

  - Eq. 13 results by using the USLLN and Slutsly theorems. So I just need to show that

    $$\frac{1}{n}\mathcal{J}_n(\hat{\theta}_n) \xrightarrow{a.s.} \mathcal{I}(\theta_0)$$

    Set $U(x, \theta) = -\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log(f(x|\theta))$, and $\mathcal{I}(\theta) = \mathrm{E}(U(x, \theta))$. Then

    $$|\frac{1}{n}\sum_{i=1}^{n}\underbrace{\left(-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log(f(x_i|\hat{\theta}_n))\right)}_{U(x_i,\hat{\theta}_n)} - \mathcal{I}(\theta_0)| \le |\frac{1}{n}\sum_{i=1}^{n}U(x_i, \hat{\theta}_n) - \mathcal{I}(\hat{\theta}_n)| + |\mathcal{I}(\hat{\theta}_n) - \mathcal{I}(\theta_0)| \tag{14}$$

    $$\le \sup_{|\hat{\theta}_n - \theta_0| \le \delta} |\frac{1}{n}\sum_{i=1}^{n}U(x_i, \theta) - \mathcal{I}(\theta)| + |\mathcal{I}(\hat{\theta}_n) - \mathcal{I}(\theta_0)| \tag{15}$$

    The first term converges to zero because the assumptions of the USLLN are satisfied. The second term converges to zero because $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ and hence $\mathcal{I}(\hat{\theta}_n) \xrightarrow{a.s.} \mathcal{I}(\theta_0)$ by using Slutsky theorem.

    So by Slutsy $(\frac{1}{n}\mathcal{J}_n(\hat{\theta}_n))^{1/2}\mathcal{I}(\theta_0)^{-1/2} \xrightarrow{a.s.} I$, and by Slutsky again

    $$\underbrace{(\frac{1}{n}\mathcal{J}_n(\hat{\theta}_n))^{1/2}\mathcal{I}(\theta_0)^{-1/2}I(\theta_0)^{1/2}(\hat{\theta}_n - \theta_0)}_{=\mathcal{J}_n(\hat{\theta}_n)^{1/2}(\hat{\theta}_n-\theta_0)} \xrightarrow{D} \underbrace{I \times \mathrm{N}(0, I)}_{=\mathrm{N}(0,I)}$$

Exercise 28. ($\star\star$) (Shannon-Kolmogorov Information Inequality) Prove the Shannon-Kolmogorov Information Inequality. Let f_0 and f_1 (like $f_0(\cdot) = f(\cdot|\theta_0)$ and $f_1(\cdot) = f(\cdot|\theta_1)$) be PDFs of corresponding distributions with respect to x. Then

$$\mathrm{KL}(f_0, f_1) = \mathrm{E}_0 \log \frac{f_0(X)}{f_1(X)} = \int \log \frac{f_0(X)}{f_1(X)} f_0(X) \mathrm{d}X \geq 0$$

with the equality iff $f_0(x) = f_1(x)$ a.s.

Solution. Function $\log(\cdot)$ is convex, then Jensen's inequality[2] implies

$$-K(f_0, f_1) = \mathrm{E}_0 \log \frac{f_1(X)}{f_0(X)} \;::\; \begin{cases} < \log \mathrm{E}_0 \frac{f_1(X)}{f_0(X)} & \text{, if } f_1(x) \neq f_0(x) \\ = \log \mathrm{E}_0 \frac{f_1(X)}{f_0(X)} & \text{, if } f_1(x) = f_0(x) \end{cases}$$

But

$$\mathrm{E}_0 \frac{f_1(x)}{f_0(x)} = \int \frac{f_1(x)}{f_0(x)} f_0(x) \mathrm{d}x = \int_{S_0} f_1(x) \mathrm{d}x \leq 1$$

at $S_0 = \{x : f_0(x) > 0\}$. Hence,

$$\mathrm{KL}(f_0, f_1) : \begin{cases} > 0 & \text{, if } f_1(x) \neq f_0(x) \\ = 0 & \text{, if } f_1(x) = f_0(x) \end{cases}$$

---

5 Handout 5: Improving sub-efficient estimators

Exercise 29. Consider random sample $x_1, ..., x_n \overset{IID}{\sim} \mathrm{G}(a, b)$, $a > 0$, $b > 0$ with PDF

$$f(x|a, b) = \frac{1}{\Gamma(a) b^a} x^a e^{-x\frac{1}{b}} 1(x > 0)$$

1. Find the moment estimator $\tilde{\theta}$ of $\theta = (a, b)^T$ by using the first raw moment and the first central moment

2. Is the moment estimator $\tilde{\theta}$ consistent and asymptotically Normal?

3. Find the one step estimator by Fisher scoring algorithm.

Hint-1 Digamma function $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log \Gamma(x)$

[2]Jensen's inequality: Consider a function φ, it is

- $\mathrm{E}(\varphi(x)) \leq \varphi(\mathrm{E}(x))$ if $\varphi(\cdot)$ is convex
- $\mathrm{E}(\varphi(x)) \geq \varphi(\mathrm{E}(x))$ if $\varphi(\cdot)$ is concave
- The equality holds if x is constant a.s.

Hint-2 Trigamma function $\psi_1(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\log\Gamma(x)$

Hint-3 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Solution.

1. The first raw moment is the expected value/mean, and the first central moment is the variance.

   The first raw moment is

$$\mathrm{E}(x) = \int_0^1 x\frac{1}{\Gamma(a)b^a}x^{a-1}e^{-x/b}\mathrm{d}x = \int_0^1 \frac{1}{\frac{1}{a}\Gamma(a+1)\frac{1}{b}b^{a+1}}x^{(a+1)-1}e^{-x/b}\mathrm{d}x = ab$$

   and the sample one

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

   The first central moment is

$$\mathrm{var}(x) = \mathrm{E}(x^2) - (\mathrm{E}(x))^2$$

   So

$$\mathrm{E}(x^2) = \int_0^1 x^2\frac{1}{\Gamma(a)b^a}x^{a-1}e^{-x/b}\mathrm{d}x = \int_0^1 \frac{1}{\frac{1}{a(a+1)}\Gamma(a+2)\frac{1}{b^2}b^{a+2}}x^{(a+2)-1}e^{-x/b}\mathrm{d}x = a(a+1)b^2$$

   and hence

$$\mathrm{var}(x) = \mathrm{E}(x^2) - (\mathrm{E}(x))^2 = ab^2$$

   The sample first central moment is

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

   From the method of moments I get

$$\begin{cases} \mathrm{E}(x|\tilde{a},\tilde{b}) = & \bar{x} \\ \mathrm{var}(x|\tilde{a},\tilde{b}) = & s^2 \end{cases} \Longrightarrow \begin{cases} \tilde{a} = & \frac{\bar{x}^2}{s^2} \\ \tilde{b} = & \frac{\bar{x}^2}{s^2} \end{cases} \Longrightarrow \begin{cases} \tilde{a} = & \frac{(\mathrm{E}(x))^2}{\mathrm{var}(x)} = \frac{\bar{x}^2}{s^2} \\ \tilde{b} = & \frac{\mathrm{var}(x)}{\mathrm{E}(x)} = \frac{\bar{x}^2}{s^2} \end{cases}$$

   So the moment estimator is

$$\tilde{\theta} = \begin{bmatrix} \tilde{a} \\ \tilde{b} \end{bmatrix} = \begin{bmatrix} \frac{\bar{x}^2}{s^2} \\ \frac{s^2}{\bar{x}} \end{bmatrix} \tag{16}$$

2. It is consistent because $\tilde{\theta} \xrightarrow{as} \theta$. This is because of the following.

It is

$$
\begin{cases} \mathrm{E}(x) = & ab \\ \mathrm{var}(x) = & ab^2 \end{cases} \Longrightarrow \begin{cases} a = & \frac{(\mathrm{E}(x))^2}{\mathrm{var}(x)} \\ b = & \frac{\mathrm{var}(x)}{\mathrm{E}(x)} \end{cases} \Longrightarrow \begin{cases} a = & \frac{(\mathrm{E}(x))^2}{\mathrm{var}(x)} \\ b = & \frac{\mathrm{var}(x)}{\mathrm{E}(x)} \end{cases}
$$

From SLLN, $\bar{x} \xrightarrow{as} \mathrm{E}(x)$. From SLLN, $\overline{x^2} \xrightarrow{as} \mathrm{E}(x^2)$. From Slutsky Theorem, $s^2 = \overline{x^2} - (\overline{x})^2 \xrightarrow{as} \mathrm{E}(x^2) - \mathrm{E}(x)^2 = \mathrm{var}(x)$

So From Slutsky theorem

$$
\tilde{\theta} = \begin{bmatrix} \frac{\bar{x}^2}{s^2} \\ \frac{s^2}{\bar{x}} \end{bmatrix} \xrightarrow{as} \begin{bmatrix} \frac{(\mathrm{E}(x))^2}{\mathrm{var}(x)} \\ \frac{\mathrm{var}(x)}{\mathrm{E}(x)} \end{bmatrix} = \theta
$$

It is asymptotically Normal because of the following.

$\bar{x}$ and s^2 are asymptotically Normal by the CLT, as averages of IID quantities. Hence, by Delta method, (16) is asymptotically Normal.

3. Recall the the one-step estimators

$$
\text{Newton alg.} \qquad \breve{\theta}_n = \tilde{\theta}_n - \ddot{\ell}_n(\tilde{\theta}_n)^{-1}\dot{\ell}_n(\tilde{\theta}_n) \tag{17}
$$

$$
\text{Fisher scoring alg.} \qquad \breve{\theta}_n = \tilde{\theta}_n + \frac{1}{n}\mathcal{I}(\tilde{\theta}_n)^{-1}\dot{\ell}_n(\tilde{\theta}_n) \tag{18}
$$

For the Fisher algorithm, I need to find $\mathcal{I}(\theta)^{-1}$. It is

$$
\begin{aligned}
\log f(x|\theta) &= -\log\Gamma(a) - a\log(b) - \frac{1}{b}x + (a-1)\log(x) \\
\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(x|\theta) &= \begin{bmatrix} -\psi(a) - \log(b) + \log(x) \\ -\frac{a}{b} + \frac{1}{b^2}x \end{bmatrix} \\
\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(x|\theta) &= \begin{bmatrix} -\psi_1(a) & -\frac{1}{b} \\ -\frac{1}{b} & -\frac{2x-ab}{b^3} \end{bmatrix} \\
\mathcal{I}(\theta) &= \begin{bmatrix} \psi_1(a) & \frac{1}{b} \\ \frac{1}{b} & \frac{a}{b^2} \end{bmatrix} \\
\mathcal{I}(\theta)^{-1} &= \frac{1}{a\psi_1(a) - 1}\begin{bmatrix} a & -b \\ -b & b^2\psi_1(a) \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\ell_n(\theta) &= -n\log\Gamma(a) - na\log(b) - \frac{1}{b}\sum_{i=1}^{n} x_i + (a-1)\sum_{i=1}^{n}\log(x_i) \\
\dot{\ell}_n(\theta) &= \begin{bmatrix} -n\psi(a) - n\log(b) + \sum_{i=1}^{n}\log(x_i) \\ -n\frac{a}{b} + n\frac{1}{b^2}\bar{x} \end{bmatrix}
\end{aligned}
$$

The Fisher recursion is

$$
\begin{aligned}
\breve{\theta}_n &= \tilde{\theta}_n + \frac{1}{n}\mathcal{I}(\tilde{\theta}_n)^{-1}\dot{\ell}_n(\tilde{\theta}_n) \\
\breve{\theta}_n &= \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{n}\frac{1}{\tilde{a}\psi_1(\tilde{a})-1}\begin{bmatrix}\tilde{a} & -\tilde{b}\\ -\tilde{b} & \tilde{b}^2\psi_1(\tilde{a})\end{bmatrix}\begin{bmatrix}-n\psi(\tilde{a}) - n\log(\tilde{b}) + \sum_{i=1}^n \log(x_i)\\ -n\frac{\tilde{a}}{\tilde{b}} + n\frac{1}{\tilde{b}^2}\bar{x}\end{bmatrix} \\
&= \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{\tilde{a}\psi_1(\tilde{a})-1}\begin{bmatrix}\tilde{a} & -\tilde{b}\\ -\tilde{b} & \tilde{b}^2\psi_1(\tilde{a})\end{bmatrix}\begin{bmatrix}-\psi(\tilde{a}) - \log(\tilde{b}) + \frac{1}{n}\sum_{i=1}^n \log(x_i)\\ -\frac{\tilde{a}}{\tilde{b}} + \frac{1}{\tilde{b}^2}\bar{x}\end{bmatrix} \\
&= \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{\tilde{a}\psi_1(\tilde{a})-1}\begin{bmatrix}-\tilde{a}\psi(\tilde{a}) - \frac{1}{\tilde{b}}(\bar{x}-\tilde{a}\tilde{b}) - \tilde{a}\log(\tilde{b}) + \frac{\tilde{a}}{n}\sum_{i=1}^n \log(x_i)\\ \tilde{b}\psi(\tilde{a}) - \psi_1(\tilde{a})(\bar{x}-\tilde{a}) + \tilde{b}\log(\tilde{b}) - \frac{\tilde{b}}{n}\sum_{i=1}^n \log(x_i)\end{bmatrix}
\end{aligned}
$$

So bu substituting

$$
\breve{\theta}_n = \begin{bmatrix}\frac{\bar{x}^2}{s^2}\\ \frac{s^2}{\bar{x}}\end{bmatrix} + \frac{1}{\frac{\bar{x}^2}{s^2}\psi_1(\frac{\bar{x}^2}{s^2})-1}\begin{bmatrix}-\frac{\bar{x}^2}{s^2}\psi(\frac{\bar{x}^2}{s^2}) - \frac{\bar{x}^2}{s^2}\log(\frac{s^2}{\bar{x}}) + \frac{1}{n}\frac{\bar{x}^2}{s^2}\sum_{i=1}^n \log(x_i)\\ \frac{s^2}{\bar{x}}\psi(\frac{\bar{x}^2}{s^2}) - \psi_1(\frac{\bar{x}^2}{s^2})(\bar{x}-\frac{\bar{x}^2}{s^2}) + \frac{s^2}{\bar{x}}\log(\frac{s^2}{\bar{x}}) - \frac{\tilde{b}}{n}\sum_{i=1}^n \log(x_i)\end{bmatrix}
$$

Additionally for the Newton recursion I need

$$
\begin{aligned}
\ddot{\ell}_n(\theta) &= -n\begin{bmatrix}\psi_1(a) & \frac{1}{b}\\ \frac{1}{b} & \frac{2\bar{x}-ab}{b^3}\end{bmatrix} \\
(\ddot{\ell}_n(\theta))^{-1} &= -\frac{1}{n}\frac{1}{\psi_1(a)\frac{2\bar{x}-ab}{b}-1}\begin{bmatrix}\frac{2\bar{x}-ab}{b} & -b\\ -b & b^2\psi_1(a)\end{bmatrix}
\end{aligned}
$$

The Newton recursion is

$$
\begin{aligned}
\breve{\theta}_n &= \tilde{\theta}_n - (\ddot{\ell}_n(\theta))^{-1}\dot{\ell}_n(\tilde{\theta}_n) \\
\breve{\theta}_n &= \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{n}\frac{1}{\psi_1(\tilde{a})\frac{2\bar{x}-\tilde{a}\tilde{b}}{\tilde{b}}-1}\begin{bmatrix}2\bar{x}-\tilde{a}\tilde{b} & \tilde{b}\\ \tilde{b} & \tilde{b}^2\psi_1(\tilde{a})\end{bmatrix}\begin{bmatrix}-n\psi(\tilde{a}) - n\log(\tilde{b}) + \sum_{i=1}^n \log(x_i)\\ -n\frac{\tilde{a}}{\tilde{b}} + n\frac{1}{\tilde{b}^2}\bar{x}\end{bmatrix} \\
&= \begin{bmatrix}\tilde{a}\\ \tilde{b}\end{bmatrix} + \frac{1}{\psi_1(\tilde{a})\frac{2\bar{x}-\tilde{a}\tilde{b}}{\tilde{b}}-1}\begin{bmatrix}2\bar{x}-\tilde{a}\tilde{b} & \tilde{b}\\ \tilde{b} & \tilde{b}^2\psi_1(\tilde{b})\end{bmatrix}\begin{bmatrix}-\psi(\tilde{a}) - \log(\tilde{b}) + \frac{1}{n}\sum_{i=1}^n \log(x_i)\\ -\frac{\tilde{a}}{\tilde{b}} + \frac{1}{\tilde{b}^2}\bar{x}\end{bmatrix} \\
&= \text{...calculations}
\end{aligned}
$$

The next exercise is from Homework 4

Exercise 30. Let $x_1, ..., x_n \overset{IID}{\sim} f_\theta$ with unknown parameter $\theta \in (0,1)$ and PDF

$$
f(x|\theta) = \begin{cases}\theta\exp(-x) + (1-\theta)x\exp(-x) & , x \geq 0\\ 0 & , x < 0\end{cases}
$$

1. Calculate the moment estimator $\tilde{\theta}_n$ of θ, (I give you a bit of freedom here)

2. Calculate the asymptotic distribution of the $\tilde{\theta}_n$

3. Find the 1-step estimator $\breve{\theta}_n$ of θ such that it can be asymptotically efficient.

   Hint: Recall that $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\mathrm{d}x$, and $\Gamma(a) = (a-1)\Gamma(a-1)$

Solution.

1. I will much the 1st central moments (theoretical and sample).. It is

$$\begin{aligned}\mathrm{E}(x|\theta) =& \int_0^\infty x(\theta\exp(-x) + (1-\theta)x\exp(-x))\mathrm{d}x \\ =&\theta\int_0^\infty x^{2-1}\exp(-x)\mathrm{d}x + (1-\theta)\int_0^\infty x^{3-1}\exp(-x))\mathrm{d}x \\ =&\theta\Gamma(2) + (1-\theta)\Gamma(3) = 2-\theta\end{aligned}$$

   and

$$m_1' = \bar{x}$$

   Then

$$\mathrm{E}(x|\tilde{\theta}_n) = m_1' \implies \tilde{\theta}_n = 2 - \bar{x}$$

2. We use use the CLT .

   So

$$\begin{aligned}\mathrm{E}(x^2|\theta) =& \int_0^\infty x^2(\theta\exp(-x) + (1-\theta)x\exp(-x))\mathrm{d}x \\ =&\theta\int_0^\infty x^{3-1}\exp(-x)\mathrm{d}x + (1-\theta)\int_0^\infty x^{4-1}\exp(-x))\mathrm{d}x \\ =&\theta\Gamma(3) + (1-\theta)\Gamma(4) = 6-4\theta\end{aligned}$$

   So

$$\mathrm{var}(x|\theta) = \mathrm{E}(x^2|\theta) - (\mathrm{E}(x|\theta))^2 = 2-\theta^2$$

   So by CLT

$$\sqrt{n}(\bar{x} - (2-\theta)) \xrightarrow{D} \mathrm{N}(0, 2-\theta^2) \tag{19}$$

   Well, for the asympt. distr. of $\tilde{\theta}_n$ I do not really need to use Delta, however I could use it. I can just rearrange (19) as

$$\sqrt{n}(\tilde{\theta}_n - \theta)) = \sqrt{n}((\bar{x}-2) - \theta)) = -\sqrt{n}(\bar{x} - (2-\theta)) \xrightarrow{D} \mathrm{N}(0, 2-\theta^2) \tag{20}$$

3. It is

$$\breve{\theta}_n = \tilde{\theta}_n - \ddot{\ell}_n(\tilde{\theta}_n)^{-1}\dot{\ell}_n(\tilde{\theta}_n)$$

where

$$\begin{aligned}
\ell_n(\theta) &= \sum_{i=1}^{n} \log(\theta \exp(-x_i) + (1-\theta)x_i \exp(-x_i)) \\
&= \sum_{i=1}^{n} (\log(\theta + (1-\theta)x_i) - x_i) \\
\dot{\ell}_n(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\ell_n(\theta) &= \sum_{i=1}^{n} \frac{1-x_i}{\theta + (1-\theta)x_i} \\
&= \sum_{i=1}^{n} \frac{1-x_i}{\theta + (1-\theta)x_i} \\
\ddot{\ell}_n(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\dot{\ell}_n(\theta) &= -\sum_{i=1}^{n} \frac{1-x_i}{(\theta + (1-\theta)x_i)^2}
\end{aligned}$$

So ...

$$\begin{aligned}
\breve{\theta}_n &= \tilde{\theta}_n - \ddot{\ell}_n(\tilde{\theta}_n)^{-1}\dot{\ell}_n(\tilde{\theta}_n) \\
&= \tilde{\theta}_n + \frac{\sum_{i=1}^{n} \frac{1-x_i}{\tilde{\theta}_n + (1-\tilde{\theta}_n)x_i}}{\sum_{i=1}^{n} \frac{1-x_i}{(\tilde{\theta}_n + (1-\tilde{\theta}_n)x_i)^2}} \\
&= 2 - \bar{x} + \frac{\sum_{i=1}^{n} \frac{1-x_i}{2-\bar{x}+(\bar{x}-1)x_i}}{\sum_{i=1}^{n} \frac{1-x_i}{(2-\bar{x}+(\bar{x}-1)x_i)^2}}
\end{aligned}$$

---

6 Handout 6: Confidence intervals and hypothesis tests

Exercise 31. (Log likelihood ratio statistic)

1. Let $x_1, x_2, ..., x_n$ be IID random variables generated from a distribution f_θ labeled by a d-dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$, and admitting PDF $f(\cdot|\theta)$. Assume the conditions from the Cramér Theorem are satisfied, and that θ_0 is the true value. Prove that

$$W_{\mathrm{LR}}(\theta_0) = -2(\ell_n(\theta_0) - \ell_n(\hat{\theta}_n)) \xrightarrow{D} \chi^2_d$$

it is where $\hat{\theta}_n$ is the MLE of θ.

Hint-1 Expand $\ell_n(\theta_0)$ around $\hat{\theta}_n$ by Taylor expansion

Hint-2 Prove that $W_{\text{LR}}(\theta_0) \xrightarrow{a.s} n(\theta_0 - \hat{\theta}_n)^T \mathcal{I}(\theta_0)(\theta_0 - \hat{\theta}_n)$

Hint-3 Prove that $W_{\text{LR}}(\theta_0) \xrightarrow{D} \chi^2_d$

2. Calculate the asymptotic distribution of the statistic

$$\tilde{W}_{\text{LR}}(\theta_0) = -2(\ell_n(\theta_0) - \ell_n(\breve{\theta}_n))$$

where $\breve{\theta}_n$ is the one step estimator produced from the Fisher iterative method using the method of moments estimator as initial step.

Solution.

1. Right, let's expand it,

$$\begin{aligned}
\ell_n(\theta_0) &= \ell_n(\hat{\theta}_n) + \dot{\ell}_n(\hat{\theta}_n)(\theta_0 - \hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)^T \int_0^1 \int_0^1 u\, \ddot{\ell}_n(\hat{\theta}_n + uv(\theta_0 - \hat{\theta}_n))\mathrm{d}u\mathrm{d}v\ (\theta_0 - \hat{\theta}_n) \\
&= \ell_n(\hat{\theta}_n) + \dot{\ell}_n(\hat{\theta}_n)(\theta_0 - \hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)^T\ n \int_0^1 \int_0^1 u\, \frac{1}{n}\ddot{\ell}_n(\hat{\theta}_n + uv(\theta_0 - \hat{\theta}_n))\mathrm{d}u\mathrm{d}v\ (\theta_0 - \hat{\theta}_n)
\end{aligned}$$

So by rearranging the terms

$$-2(\ell_n(\theta_0)-\ell_n(\hat{\theta}_n)) = -\underbrace{\dot{\ell}_n(\hat{\theta}_n)}_{=0}(\theta_0-\hat{\theta}_n)-n(\theta_0-\hat{\theta}_n)^T \underbrace{\int_0^1 \int_0^1 u\, \frac{1}{n}\ddot{\ell}_n(\hat{\theta}_n + uv(\theta_0 - \hat{\theta}_n))\mathrm{d}u\mathrm{d}v}_{\xrightarrow{a.s} -\frac{1}{2}\mathcal{I}(\theta_0)}\ (\theta_0-\hat{\theta}_n)$$

It is

$$\dot{\ell}_n(\hat{\theta}_n) = 0$$

because $\hat{\theta}_n$ is an MLE.

From Cramer' Theorem $\hat{\theta}_n \xrightarrow{a.s} \theta_0$. Then by Slutsky's theorem, $\hat{\theta}_n + uv(\theta_0 - \hat{\theta}_n) \xrightarrow{a.s} \theta_0$. Then by using the USLLN it is

$$\int_0^1 \int_0^1 u\, \frac{1}{n}\ddot{\ell}_n(\hat{\theta}_n + uv(\theta_0 - \hat{\theta}_n))\mathrm{d}u\mathrm{d}v \xrightarrow{a.s} -\frac{1}{2}\mathcal{I}(\theta_0)$$

... in particular:

$$
\begin{aligned}
&\left|\int_0^1\int_0^1 u\,\frac{1}{n}\ddot{\ell}_n(\hat{\theta}_n+uv(\theta_0-\hat{\theta}_n))\mathrm{d}u\mathrm{d}v-\left[-\frac{1}{2}\mathcal{I}(\theta_0)\right]\right|\\
&\qquad<\int_0^1\int_0^1\left|u\,\frac{1}{n}\ddot{\ell}_n(\hat{\theta}_n+uv(\theta_0-\hat{\theta}_n))+u\mathcal{I}(\theta_0)\right|\mathrm{d}u\mathrm{d}v\\
&\qquad<\int_0^1 u\int_0^1\sup_{\theta:|\theta-\theta_0|<\epsilon}\left|\frac{1}{n}\ddot{\ell}_n(\theta)-(-\mathcal{I}(\theta))\right|\mathrm{d}u\mathrm{d}v+\left|\mathcal{I}(\hat{\theta}_n+uv(\theta_0-\hat{\theta}_n))-\mathcal{I}(\theta)\right|\\
&\qquad\to 0
\end{aligned}
$$

So to sum up

$$-2(\ell_n(\theta_0)-\ell_n(\hat{\theta}_n))\xrightarrow{a.s}n(\theta_0-\hat{\theta}_n)^T\mathcal{I}(\theta_0)(\theta_0-\hat{\theta}_n)\tag{21}$$

which implies that

$$\left[-2(\ell_n(\theta_0)-\ell_n(\hat{\theta}_n))\right]-\left[n(\theta_0-\hat{\theta}_n)^T\mathcal{I}(\theta_0)(\theta_0-\hat{\theta}_n)\right]\xrightarrow{p}0\tag{22}$$

From Cramer' Theorem I know that

$$
\begin{aligned}
\sqrt{n}(\hat{\theta}_n-\theta_0)&\xrightarrow{D}\mathrm{N}(0,\mathcal{I}(\theta_0)^{-1})\\
\implies\sqrt{n}\mathcal{I}(\theta_0)^{1/2}(\hat{\theta}_n-\theta_0)&\xrightarrow{D}\mathrm{N}(0,I)\\
\implies n(\hat{\theta}_n-\theta_0)^T\mathcal{I}(\theta_0)(\hat{\theta}_n-\theta_0)&\xrightarrow{D}\chi^2_d
\end{aligned}
$$

But $-2(\ell_n(\theta_0)-\ell_n(\hat{\theta}_n))$ is asymptotic equivalent to $n(\hat{\theta}_n-\theta_0)^T\mathcal{I}(\theta_0)(\hat{\theta}_n-\theta_0)$ from (22). So by the Slutsky's theorem

$$-2(\ell_n(\theta_0)-\ell_n(\hat{\theta}_n))\xrightarrow{D}\chi^2_d$$

2. It is

$$\tilde{W}_{\mathrm{LR}}(\theta_0)=-2(\ell_n(\theta_0)-\ell_n(\breve{\theta}_n))\xrightarrow{D}\chi^2_d$$

because $\breve{\theta}_n$ and $\hat{\theta}_n$ are asymptotic equivalent.

The next exercise is from Homework 4

Exercise 32. Let

$$y_i\overset{\text{ind}}{\sim}\mathrm{Bin}(n,\pi_i)$$

where $i = 1, ..., N$. Consider that the probability of success is modeled such as

$$\text{logit}(\pi_i) = x_i^T \theta \tag{23}$$

where $\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i})$. Here $x_i = (x_{i,1}, ..., x_{i,d})^T$ are known vectors containing the values of the d regressions at the i-th observation, and $\theta \in \mathbb{R}^d$.

1. Show that
$$\pi_i = \frac{e^{x_i^T \theta}}{1 + e^{x_i^T \theta}}$$

2. Assume that the MLE $\hat{\theta}$ of θ is known/calculated. Show that the $(1-a)$ Wald confidence interval for the unknown parameter θ, by using the observed information matrix, is
$$\text{C.I.} : \{\, \theta \in \mathbb{R}^d \, : \, (\hat{\theta} - \theta)^T X^T (\text{diag}_{\forall i}(n \hat{\pi}_i (1 - \hat{\pi}_i))) X (\hat{\theta} - \theta) \le \chi^2_{d, 1-a} \,\}$$
where
$$\hat{\pi}_i = \frac{e^{x_i^T \hat{\theta}}}{1 + e^{x_i^T \hat{\theta}}}$$
X is the so called design matrix from the regression
$$\begin{bmatrix} \text{logit}(\pi_1) \\ \vdots \\ \text{logit}(\pi_N) \end{bmatrix} = \underbrace{\begin{bmatrix} \longleftarrow & x_1^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & x_N^T & \longrightarrow \end{bmatrix}}_{=X} \theta$$
and $\text{diag}_{\forall i}(\heartsuit_i) = \begin{bmatrix} \heartsuit_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \heartsuit_N \end{bmatrix}$.

3. Find the score statistic rejection area for the hypothesis test $\text{H}_0 : \theta = \theta_*$ versus $H_1 : \theta \neq \theta_*$.

Solution.

1. It is
$$\begin{aligned} \text{logit}(\pi_i) =& x_i^T \theta \Longleftrightarrow \\ \log(\frac{\pi_i}{1 - \pi_i}) =& x_i^T \theta \Longleftrightarrow \\ \frac{\pi_i}{1 - \pi_i} =& \exp(x_i^T \theta) \Longleftrightarrow \\ \pi_i =& \frac{e^{x_i^T \theta}}{1 + e^{x_i^T \theta}} \end{aligned}$$

(a) First I need to calculate the observed information matrix $\mathcal{J}_n(\hat{\theta}_n)$. It is

$$
\begin{aligned}
\ell_n(\theta) =& \sum_{i=1}^{N}(\log\binom{n}{y_i} + y_i \log(\frac{e^{x_i^T\theta}}{1+e^{x_i^T\theta}}) + (n-y_i)\log(\frac{e^{x_i^T\theta}}{1+e^{x_i^T\theta}})) \\
\propto& \sum_{i=1}^{N} y_i \log(\frac{e^{x_i^T\theta}}{1+e^{x_i^T\theta}}) + (n-y_i)\log(\frac{e^{x_i^T\theta}}{1+e^{x_i^T\theta}})) \\
=& \sum_{i=1}^{N}(n_i \log(\frac{1}{1+e^{x_i^T\theta}}) + y_i \log(e^{x_i^T\theta}))
\end{aligned}
$$

Then

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\theta_j}\ell_n(\theta) =& \sum_{i=1}^{N}(n\frac{e^{x_i^T\theta}x_{i,j}}{1+e^{x_i^T\theta}} + y_i x_{i,j}) \\
=& \sum_{i=1}^{N}(n\pi_i x_{i,j} + y_i x_{i,j})
\end{aligned}
$$

which in matrix form becomes

$$
\dot{\ell}(\theta) = X^T \mathrm{diag}(y - n\pi)) \tag{24}
$$

It is

$$
\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}\theta_k \mathrm{d}\theta_j}\ell_n(\theta) =& \frac{\mathrm{d}}{\mathrm{d}\theta_k}\sum_{i=1}^{N}(n\pi_i x_{i,j} + y_i x_{i,j}) \\
=& \sum_{i=1}^{N} n \frac{\mathrm{d}}{\mathrm{d}\theta_k}\pi_i \, x_{i,j} \\
=& \sum_{i=1}^{N} n\,(-1)\frac{e^{x_i^T\theta}x_{i,k}}{(1+e^{x_i^T\theta})^2}\, x_{i,j} \\
=& -\sum_{i=1}^{N} n\pi_i(1-\pi_i)x_{i,j}
\end{aligned}
$$

which in matrix form becomes

$$
\ddot{\ell}(\theta) = -X^T(\mathrm{diag}(n\pi_i(1-\pi_i)))X
$$

Hence

$$
\mathcal{J}_n(\hat{\theta}_n) = X^T(\mathrm{diag}(n\hat{\pi}_i(1-\hat{\pi}_i)))X
$$

and the $(1-a)$ Wald's interval is

$$\text{C.I.} : \{ \theta \in \mathbb{R}^d \ : \ (\hat{\theta}_n - \theta)^T X^T (\text{diag}_{\forall i}(n\hat{\pi}_i(1-\hat{\pi}_i))) X (\hat{\theta}_n - \theta) \leq \chi^2_{d,1-a} \ \}$$

(b) The Score statistic is

$$U(\theta) = X^T(y - n\pi))$$

So the rejection area is

$$\{y_{1:n} : U(\theta_*)^T \mathcal{J}_n(\hat{\theta}_n)^{-1} U(\theta_*) \geq \chi^2_{d,1-a}\}$$
$$\{y_{1:n} : (n\pi - y)^T X (X^T(\text{diag}(n_i\hat{\pi}_i(1-\hat{\pi}_i)))X)^{-1} X^T (n\pi - y) \geq \chi^2_{d,1-a}$$

---

7 Handout 7: The Profile likelihood (MLE under the presence of nuisance parameters)

Exercise 33. For $i = 1, ..., k$, let $x_{i,1}, ..., x_{i,n} \overset{\text{IID}}{\sim} \text{Poi}(\theta_i)$. Find the asymptotic likelihood ratio rejection area for teasting the hypothesis

$$\text{H}_0 : \theta_1 = ... = \theta_k$$

Hint: It is

$$f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!} 1(x \in \mathbb{N})$$

Solution. Under H_1, the log-likelihood is

$$\begin{aligned}
\ell_1(\theta) &= \sum_{i=1}^{k} \sum_{j=1}^{n} (-\theta_i + x_{i,j} \log(\theta_i) - \log(x_{i,j}!)) \\
&\propto -n \sum_{i=1}^{k} \theta_i + \sum_{i=1}^{k} \log(\theta_i) \sum_{j=1}^{n} x_{i,j}
\end{aligned}$$

The MLE is

$$0 = \frac{\mathrm{d}}{\mathrm{d}\theta_i}\ell_1(\theta)|_{\theta=\hat{\theta}^{(1)}} = -n + \frac{1}{\hat{\theta}_i^{(1)}}\sum_{j=1}^{n} x_{i,j}$$

$$\implies \hat{\theta}_i^{(1)} = \frac{1}{n}\sum_{j=1}^{n} x_{i,j}$$

$$\implies \hat{\theta}^{(1)} = (\bar{x}_{1,\bullet}, ..., \bar{x}_{k,\bullet})^T$$

and there are $d_1 = k$ free parameters for estimation.

Under H_0, it is the log-likelihood is $\theta_1 = ... = \theta_k = \theta$

$$\ell_0(\theta) = \sum_{i=1}^{k}\sum_{j=1}^{n}(-\theta + x_{i,j}\log(\theta) - \log(x_{i,j}!))$$

$$\propto -nk\theta + \log(\theta)\sum_{i=1}^{k}\sum_{j=1}^{n} x_{i,j}$$

The MLE is

$$0 = \frac{\mathrm{d}}{\mathrm{d}\theta}\ell_0(\theta)|_{\theta=\hat{\theta}^{(0)}} = -nk + \frac{1}{\hat{\theta}^{(0)}}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{i,j}$$

$$\implies \hat{\theta}^{(0)} = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{i,j}$$

$$\hat{\theta}^{(0)} = \bar{x}_{\bullet,\bullet}$$

and there is $d_0 = 1$ free parameter for estimation.

So

$$-2(\ell_n(\hat{\theta}^{(0)}) - \ell_n(\hat{\theta}^{(1)})) \xrightarrow{D} \chi^2_{k-1}$$

where

$$
\begin{aligned}
-2(\ell_n(\hat{\theta}^{(0)}) - \ell_n(\hat{\theta}^{(1)})) =& -2(-nk\hat{\theta}^{(0)} + \log(\hat{\theta}^{(0)}) \sum_{i=1}^{k} \sum_{j=1}^{n} x_{i,j} \\
& - n \sum_{i=1}^{k} \hat{\theta}_i^{(1)} + \sum_{i=1}^{k} \log(\hat{\theta}_i^{(1)}) \sum_{j=1}^{n} x_{i,j}) \\
=& -2(\cancel{-nk\bar{x}_{\bullet,\bullet}} + \log(\bar{x}_{\bullet,\bullet}) \sum_{i=1}^{k} \sum_{j=1}^{n} x_{i,j} \\
& \cancel{+n \sum_{i=1}^{k} \bar{x}_{i,\bullet}} - \sum_{i=1}^{k} \log(\bar{x}_{i,\bullet}) \sum_{j=1}^{n} x_{i,j}) \\
=& 2n \sum_{i=1}^{k} \log(\bar{x}_{i,\bullet}) \bar{x}_{i,\bullet} - 2nk \log(\bar{x}_{\bullet,\bullet}) \bar{x}_{\bullet,\bullet}
\end{aligned}
$$

So the rejection area is

$$
\mathrm{RA} = \{2n \sum_{i=1}^{k} \log(\bar{x}_{i,\bullet}) \bar{x}_{i,\bullet} - 2nk \log(\bar{x}_{\bullet,\bullet}) \bar{x}_{\bullet,\bullet} \geq \chi^2_{k-1,1-a}\}
$$

Exercise 34. Let $x = (x_1, ..., x_c) \sim \mathrm{Mult}(\pi_1, ..., \pi_c)$, with $\pi_i \in (0, \infty)$ and $\sum_{i=1}^{c} \pi_i = 1$. Find the asymptotic likelihood ratio rejection area for teasting the hypothesis

$$
\mathrm{H}_0 : \pi_1 = ... = \pi_c = \frac{1}{c}
$$

Hint: It is

$$
f(x|\theta) = \binom{n}{x_1 ... x_c} \prod_{i=1}^{c} \pi_i^{x_i}
$$

Solution. It is

$$
\ell_n(\pi) = \log \binom{n}{x_1 ... x_c} + \sum_{i=1}^{k} x_i \log(\pi_i)
$$

Lagrandge function is

$$
\mathcal{L}(\pi, \theta) = \log \binom{n}{x_1 ... x_c} + \sum_{i=1}^{k} x_i \log(\pi_i) - \theta(\sum_{i=1}^{c} \pi_i - 1)
$$

Under H_1, the MLE is

$$
\begin{aligned}
0 = \frac{\mathrm{d}}{\mathrm{d}\pi_i}\mathcal{L}(\pi,\theta)|_{\pi=\hat{\pi},\,\theta=\hat{\theta}} &\implies \hat{\pi}_i = \frac{x_i}{\theta} \\
0 = \frac{\mathrm{d}}{\mathrm{d}\theta}\mathcal{L}(\pi,\theta)|_{\pi=\hat{\pi},\,\theta=\hat{\theta}} &\implies \sum_{i=1}^{c}\pi_i = 1 \\
&\implies \hat{\pi}_i = \frac{x_i}{n} \\
&\implies \hat{\pi}^{(1)} = (\frac{x_1}{n},...,\frac{x_c}{n})^T
\end{aligned}
$$

So

$$
\ell(\hat{\pi}^{(1)}) = \log\binom{n}{x_1...x_c} + \sum_{i=1}^{k} x_i\log(\frac{x_i}{n})
$$

with $d_1 = c-1$ free parameters.

Under H_0,

$$
\hat{\pi}^{(0)} = (\frac{1}{c},...,\frac{1}{c})^T
$$

So

$$
\ell(\hat{\pi}^{(0)}) = \log\binom{n}{x_1...x_c} + n\bar{x}\log(\frac{1}{c})
$$

with $d_0 = 0$ free parameters.

So

$$
-2(\ell_n(\hat{\pi}^{(0)}) - \ell_n(\hat{\pi}^{(1)})) \xrightarrow{D} \chi^2_{c-1}
$$

where

$$
\begin{aligned}
-2(\ell_n(\hat{\pi}^{(0)}) - \ell_n(\hat{\pi}^{(1)})) =& -2(n\bar{x}\log(\frac{1}{c}) - \sum_{i=1}^{k} x_i\log(\frac{x_i}{n})) \\
=& 2\sum_{i=1}^{c} x_i\log(\frac{cx_i}{n})
\end{aligned}
$$

So the rejection area is

$$
\mathrm{RA} = \{2\sum_{i=1}^{c} x_i\log(\frac{cx_i}{n}) \geq \chi^2_{c-1,1-a}\}
$$

Exercise 35. [Woodbury matrix identity] Verify that

$$
(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}
$$

if A and C are non-singular.

Solution.

By checking that $(A+UCV)(A+UCV)^{-1}=I$

$$
\begin{aligned}
(A+UCV)\times\left[A^{-1}-A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}\right] \\
&= I+UCVA^{-1}-(U+UCVA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1} \\
&= I+UCVA^{-1}-UC(C^{-1}+VA^{-1}U)(C^{-1}+VA^{-1}U)^{-1}VA^{-1} \\
&= I+UCVA^{-1}-UCVA^{-1}=I.
\end{aligned}
$$

So

$$
(A+UCV)^{-1}=A^{-1}-A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}
$$

---

8 Others

Exercise 36. (Very difficult) Consider a contigency table with N cells. Consider a Multimomial sampling scheme was used to collect n observations. Let $y=(y_1,...,y_N)^T$ be the observed counts, and $\pi=(\pi_1,...,\pi_N)^T$ be the expected probabilities in N cells of a contingency table. Let the total number of observations be $n=\sum_{i=1}^N y_i$. Assume that

$$
y\sim \text{Mult}(n,\pi) \tag{25}
$$

where

$$
f(y|n,\pi)=\binom{n}{y_1...y_N}\prod_{i=1}^n \pi_i^{y_i}
$$

Consider a log-linear model

$$
\pi_i=\pi_i(\theta)=\frac{\exp(x_i^T\theta)}{\sum_{\forall k}\exp(x_i^T\theta)} \tag{26}
$$

$\theta\in\Theta$ is a d-dimensional vector of unknown coefficients, and $x_i=(x_{i,1},...,x_{i,d})^T$ are the values of d regressors.

In a matrix form

$$
\pi=\frac{\exp(X\theta)}{1_d^T\exp(X\theta)}
$$

where

$$
X=\begin{bmatrix}\longleftarrow & x_1^T & \longrightarrow\\ & \vdots & \\ \longleftarrow & x_N^T & \longrightarrow\end{bmatrix}
$$

Assume that Cramer's Theorem conditions are satisfied. Consider that the MLE $\hat{\theta}_n$ of θ is computed/calculated, and that θ_0 is the unknown true value of θ. Then

1. Show that
$$\frac{\mathrm{d}\pi}{\mathrm{d}\theta} = (\mathrm{diag}(\pi) - \pi\pi^T)X$$

2. Show that the likelihood equations to find the MLE $\hat{\theta}$ of θ are such as
$$X^T y = nX^T\pi(\hat{\theta}_n)$$
Does it ring a bell?

3. Consider the j-th single observation $\xi_j = (\xi_{j,1}, ..., \xi_{j,N})^T$ where $\xi_{j,i} = 1$ if it falls in cell i and $\xi_{j,i} = 0$ if it does not fall in cell i. Write the probability distribution $f(\xi_i|...) =?$ in the form of the Multinomial distribution.

4. Calculate the asymptotic distribution of the MLE $\hat{\theta}$ of θ.

   Hint: Use the fact that a single observation falls in only one cell, and use its probability.

5. Calculate the asymptotic distribution of cell probability estimators $\hat{\pi}$ of π.

6. Calculate the Wald's $(1-a)$ CI for θ, that results as an ellipsoid easy to compute or plot in 2D on 3D.

Solution.

1. It is
$$\begin{aligned}
[\frac{\mathrm{d}\pi}{\mathrm{d}\theta}]_{i,j} = \frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_j} =& \frac{\mathrm{d}}{\mathrm{d}\theta_j}\frac{\exp(x_i^T\theta)}{\sum_{\forall k}\exp(x_k^T\theta)} \\
=& \frac{\exp(x_i^T\theta)x_{i,j}\sum_{\forall k}\exp(x_i^T\theta) - \exp(x_i^T\theta)\sum_{\forall k}\exp(x_k^T\theta)x_{k,j}}{[\sum_{\forall k}\exp(x_k^T\theta)]^2} \\
=& \frac{\exp(x_i^T\theta)}{\sum_{\forall k}\exp(x_k^T\theta)}\frac{x_{i,j}\sum_{\forall k}\exp(x_k^T\theta) - \sum_{\forall k}\exp(x_k^T\theta)x_{k,j}}{[\sum_{\forall k}\exp(x_k^T\theta)]^2} \\
=& \frac{\exp(x_i^T\theta)}{\sum_{\forall k}\exp(x_k^T\theta)}\left(x_{i,j}\frac{\sum_{\forall k}\exp(x_k^T\theta)}{\sum_{\forall k}\exp(x_k^T\theta)} - \sum_{\forall k}\frac{\exp(x_k^T\theta)}{\sum_{\forall k}\exp(x_k^T\theta)}x_{k,j}\right) \\
=& \pi_i\left(x_{i,j} - \sum_{\forall k}\pi_k x_{k,j}\right) = \pi_i x_{i,j} - \pi_i\sum_{\forall k}\pi_k x_{k,j}
\end{aligned}$$

So if I write it in a matrix form

$$\begin{aligned}\frac{\mathrm{d}\pi}{\mathrm{d}\theta} &= \mathrm{diag}(\pi)X - \pi\pi^T X \\ &= (\mathrm{diag}(\pi) - \pi\pi^T)X\end{aligned}$$

2. Well,

$$\ell_n(\theta) = \log(\binom{n}{y_1 \ldots y_N}) + \sum_{i=1}^{N} y_i \log(\pi_i(\theta))$$

It is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta_j}\ell_n(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\theta_j}\sum_{i=1}^{N} y_i \log(\pi_i(\theta)) \\ &= \sum_{i=1}^{N} y_i \frac{1}{\pi_i(\theta)} \frac{\mathrm{d}}{\mathrm{d}\theta_j}\pi_i(\theta) \\ &= \sum_{i=1}^{N} y_i \frac{1}{\pi_i(\theta)} \left(\pi_i(\theta)x_{i,j} - \pi_i(\theta)\sum_{\forall k}\pi_k(\theta)x_{k,j}\right) \\ &= \sum_{i=1}^{N} y_i x_{i,j} - \underbrace{\sum_{i=1}^{N} y_i}_{=n} \sum_{\forall k}\pi_k(\theta)x_{k,j} \\ &= \sum_{i=1}^{N} y_i x_{i,j} - n\sum_{\forall k}\pi_k(\theta)x_{k,j}\end{aligned}$$

So

$$\dot{\ell}_n(\theta) = X^T y - nX^T\pi(\theta)$$

Hence

$$0 = \ell_n(\theta)|_{\theta=\hat{\theta}} \quad \Longrightarrow \quad X^T y = nX^T\pi(\hat{\theta})$$

It is the same equation as the one for the log-linear model under the Piosson samp[ling scheme, when $\mu(\theta) = n\pi(\theta)$.

3. Based on the Multinomial sampling scheme, each observation can fall in one only cell. Observation ξ_j can fall in i-cell with probability π_i. So

$$\xi_j \stackrel{\text{IID}}{\sim} \mathrm{Mult}(1, \pi)$$

with

$$f(\xi_j|\pi) = \prod_{i=1}^{N} \pi_i^{\xi_{j,i}}$$

where $\xi_{j,i} \in \{0,1\}^d$ and $\sum_i \xi_{j,i} = 1$.

4. Since Cramer's Theorem conditions are satisfied, I will use Cramer's Theorem. Namely,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}(0, \mathcal{I}(\theta_0)^{-1})$$

where $\mathcal{I}(\theta)$ is the Fisher's information matrix, aka the information matrix for 1 observation, Let's say observation ξ. Therefor, I just need to find $\mathcal{I}(\theta)$ with

$$[\mathcal{I}(\theta)]_{j,k} = \mathrm{E}(\frac{\mathrm{d}}{\mathrm{d}\theta_j}\log(f(\xi|\pi))\frac{\mathrm{d}}{\mathrm{d}\theta_k}\log(f(\xi|\pi)))$$

.

It is

$$\log(f(\xi|\pi)) = \sum_{i=1}^{N}\xi_i \log(\pi_i)$$

It is

$$\frac{\mathrm{d}}{\mathrm{d}\theta_j}\log(f(\xi|\pi)) = \sum_{i=1}^{N}\xi_i \frac{1}{\pi_i}\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_j}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\theta_k}\log(f(\xi|\pi)) = \sum_{i=1}^{N}\xi_i \frac{1}{\pi_i}\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_k}$$

So

$$
\begin{aligned}
[\mathcal{I}(\theta)]_{j,k} =& \mathrm{E}\left(\left(\sum_{i=1}^{N}\xi_i\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_j}\right)\left(\sum_{i'=1}^{N}\xi_{i'}\frac{\mathrm{d}\pi_{i'}}{\mathrm{d}\theta_k}\right)\right)\\
=& \mathrm{E}\left(\sum_{i=1}^{N}\sum_{i'=1}^{N}\xi_i\xi_{i'}\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_j}\frac{\mathrm{d}\pi_{i'}}{\mathrm{d}\theta_k}\right)\\
=& \sum_{i=1}^{N}\sum_{i'=1}^{N}\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_j}\frac{\mathrm{d}\pi_{i'}}{\mathrm{d}\theta_k}\mathrm{E}\left(\xi_i\xi_{i'}\right) = \begin{cases}\mathrm{E}\left(\xi_i^2\right) = 1^2\pi_i + 0^2(1-\pi_i) = \pi_i & , i = i' \\ 0 & , i \neq i'\end{cases}\\
=& \{\text{so we care for those where } \xi_i = \xi_{i'} = 1\}\\
=& \sum_{i=1}^{N}\pi_i\frac{1}{\pi_i}\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_j}\frac{1}{\pi_i}\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_k}\\
=& \sum_{i=1}^{N}\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_j}\frac{1}{\pi_i}\frac{\mathrm{d}\pi_i}{\mathrm{d}\theta_k}\\
=& \sum_{i=1}^{N}\left(\pi_i x_{i,j} - \pi_i\sum_{\forall s}\pi_s x_{s,j}\right)\left(\pi_i x_{i,k} - -\pi_i\sum_{\forall s}\pi_s x_{s,k}\right)\frac{1}{\pi_i}\\
=& \sum_{i=1}^{N}\left(x_{i,j} - (\pi^T X_{:,j})\right)\left(\pi_i x_{i,k} - \pi_i(\pi^T X_{:,k})\right)\\
=& \sum_{i=1}^{N}x_{i,j}\pi_i x_{i,k} - (\sum_{i=1}^{N}x_{i,j}\pi_i)(\pi^T X_{:,k}) - \underbrace{\sum_{i=1}^{N}(\pi^T X_{:,j})(\pi_i x_{i,k}) + (\pi^T X_{:,j})\overbrace{\sum_{i=1}^{N}\pi_i}^{=1}(\pi^T X_{:,k})}_{=0}\\
=& \sum_{i=1}^{N}x_{i,j}\pi_i x_{i,k} - \sum_{i=1}^{N}x_{i,j}\pi_i(\pi^T X_{:,k})\\
=& X_{:,j}^T\mathrm{diag}(\pi)X_{:,k} - (\pi^T X_{:,j})^T(\pi^T X_{:,k})
\end{aligned}
$$

So in a matrix form, it is

$$
\begin{aligned}
\mathcal{I}(\theta) &= X^T\mathrm{diag}(\pi)X - X^T\pi\pi^T X\\
&= X^T(\mathrm{diag}(\pi) - \pi\pi^T)X
\end{aligned}
$$

So

$$
\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \mathrm{N}\left(0, \left(X^T(\mathrm{diag}(\pi_0) - \pi_0\pi_0^T)X\right)^{-1}\right) \tag{27}
$$

where $\pi_0 = \pi(\theta_0)$.

5. Because $\hat{\pi}$ is a continuous function of θ's and because I know that (27), I can use Delta method in order to find the asymptotic distribution of $\hat{\pi}$.

   According to Delta method, it is

$$\sqrt{n}(\hat{\pi} - \pi_0) \xrightarrow{D} \mathrm{N}\left(0, \Sigma_{\pi_0}\right)$$

   where $\pi_0 = \pi(\theta_0)$, and

$$\begin{aligned}
\Sigma_{\pi_0} =& \frac{\mathrm{d}\pi}{\mathrm{d}\theta}|_{\theta=\theta_0} \left(X^T(\mathrm{diag}(\pi_0) - \pi_0\pi_0^T)X\right)^{-1} (\frac{\mathrm{d}\pi}{\mathrm{d}\theta}|_{\theta=\theta_0})^T \\
=& (\mathrm{diag}(\pi_0) - \pi_0\pi_0^T)X \left(X^T(\mathrm{diag}(\pi_0) - \pi_0\pi_0^T)X\right)^{-1} X^T(\mathrm{diag}(\pi_0) - \pi_0\pi_0^T)
\end{aligned}$$

6. Well, the $(1-a)100\%$ confidence interval for θ which is touch to invert is

$$\begin{aligned}
\mathrm{CI}(\theta) =& \left\{\theta \in \mathbb{R}^d : \sqrt{n}(\hat{\theta} - \theta)\mathcal{I}(\theta)(\hat{\theta} - \theta)^T \le \chi^2_{d,1-a}\right\} \\
=& \left\{\theta \in \mathbb{R}^d : \sqrt{n}(\hat{\theta} - \theta) \left(X^T(\mathrm{diag}(\pi(\theta)) - \pi(\theta)\pi(\theta)^T)X\right) (\hat{\theta} - \theta)^T \le \chi^2_{d,1-a}\right\}
\end{aligned}$$

   So probably I would go with the asymptotic equivalent one

$$\begin{aligned}
\mathrm{CI}(\theta) =& \left\{\theta \in \mathbb{R}^d : \sqrt{n}(\hat{\theta} - \theta)\mathcal{I}(\hat{\theta})(\hat{\theta} - \theta)^T \le \chi^2_{d,1-a}\right\} \\
=& \left\{\theta \in \mathbb{R}^d : \sqrt{n}(\hat{\theta} - \theta) \left(X^T(\mathrm{diag}(\hat{\pi}) - \hat{\pi}\hat{\pi}^T)X\right) (\hat{\theta} - \theta)^T \le \chi^2_{d,1-a}\right\}
\end{aligned}$$

   where $\hat{\pi} = \pi(\hat{\theta})$.

   The degrees of freedom of the critical values in the CI are d because there are d free parameters in θ.
