Chapter 4. Hypothesis testing

In estimation theory, we have efficiency
In testing, we have power.
QMD is important because
(1) relaxes some condi's to show asymp. normality
(2) will give us in LeCam theory
Local asymptotic normality (LAN)
+ LeCam III $\Longrightarrow$ Power calculation framework!

Framework

$$X_1, \dots, X_n \overset{iid}{\sim} P_\theta \in \{P_{\theta'} : \theta' \in \Theta\}$$

$$H_0 : \theta \in \Theta_0 \subset \Theta$$

against

$$H_1 : \theta \in \boxed{\Theta \setminus \Theta_0} = \Theta_1$$

(1) A test/test-function

$$\phi_n : \mathcal{X}^n \longrightarrow [0, 1]$$

represents a (randomized) test:
(a) For randomized test, ϕ_n outputs the probability to reject H_0
(b) For deterministic tests, $\phi_n \in \{0, 1\}$

not reject

reject

(2) For any given θ, $\pi_n(\theta)$ is defined as:

$$\pi_n(\theta) := \mathbb{E}_\theta[\phi_n(\underbrace{X_1, \dots, X_n}_{\overset{iid}{\sim} P_\theta})]$$

and is called power function of ϕ_n.

(3) Ideally, we wish

$$\sup_{\theta \in \Theta_0} \pi_n(\theta) \leq \alpha$$

But in general, this is nearly impossible because it is a non-asymptotic control.

Instead, we can pick asymptotically valid as a criterion.

(i) Uniformly asymptotic valid

$$\limsup_{n \to \infty} \sup_{\theta \in \Theta_0} \pi_n(\theta) \leq \alpha \quad \text{(strong)}$$

→ usually need to put to non-asymp.

☆ (ii) Pointwisely asymp. valid

21:58

$$\boxed{\sup_{\theta \in \Theta_0} \limsup_{n \to \infty} \pi_n(\theta) \leq \alpha}$$

$$\Updownarrow$$

only need to think limit distr.

$$\boxed{\limsup_{n \to \infty} \pi_n(\theta) \leq \alpha, \ \forall \theta \in \Theta_0.}$$

(weak)

(4) The power/efficiency of a test ϕ_n concerns

$$\pi_n(\theta_1) \text{ for } \theta_1 \in \Theta_1.$$

Three famous tests:

① Wald's test
② Likelihood ratio test
③ Score test

$\rightarrow$ parametric model

Framework for the above 3 tests:

(a)* $\theta = (\psi, \eta)$ nuisance parameter

parameter of interest

$\theta \in \mathbb{R}^d$, $\psi \in \mathbb{R}^m$, $\eta \in \mathbb{R}^{d-m}$

remark: If $m = d$, then $\psi = \theta$.

** $\Theta = T \times N$

where $T \subset \mathbb{R}^m$, $N \subset \mathbb{R}^{d-m}$

*** θ is an interior point of Θ.

(b) $\Theta_0 = \{\theta = (\psi, \eta) : \psi = 0\}$

NOT very restrictive
by reparametrization

* If $m = d$, then $\Theta_0 = \{\theta = 0\}$

and the corresponding H_0 is called a
simple null hypothesis.

** If $m < d$, Θ_0 would have multiple elements,
and the corresponding H_0 is called a
composite null hypothesis.

Wald test:

Goal: $H_0 : \psi = 0$

Strategy:

① estimate ψ based on the MLE

$$\hat{\theta} = (\hat{\psi}, \hat{\eta})$$

② reject H_0 if the "magnitude"
of $\hat{\psi}$ is too large.

Implementation: we need to determine the
threshold of rejection

In other words, get the limiting dist. of $\hat{\psi}$
under Θ_0.

Step 1.

under regularity cond. QMD, Lipschitz, cont.

$$\sqrt{n}(\hat{\theta}-\theta) \Rightarrow N(0, I_\theta^{-1})$$

Step 2.

$$\sqrt{n}(\hat{\psi}-\psi) \Longrightarrow N(0, A_\theta^{-1})$$

linear algebra:

$$I_\theta = \begin{bmatrix} \overset{\mathbb{R}^{m\times m}}{I_{\theta,11}} & I_{\theta,12} \\ \\ \underbrace{I_{\theta,21}}_{= I_{\theta,12}^T} & I_{\theta,22} \end{bmatrix}$$

$$\Longrightarrow \quad I_\theta^{-1} = \begin{bmatrix} A_\theta^{-1} & * \\ \\ * & * \end{bmatrix}$$

where $A_\theta = I_{\theta,11} - I_{\theta,12} I_{\theta,22}^{-1} I_{\theta,21}$

Step 3 Under H_0 / Θ_0, $\psi = 0$, i.e.

$$\underbrace{\sqrt{n}\,\hat{\psi}_n \Longrightarrow N(0, A_\theta^{-1})}_{\text{limiting null dist.}}$$

Step 4. Reject H_0 by noticing

$$\sqrt{n}\, A_\theta^{1/2}\, \hat{\psi}_n \Longrightarrow N(0, I_m)$$

$$n\, \hat{\psi}_n^\top A_\theta\, \hat{\psi}_n \Longrightarrow \chi^2(m)$$

A_θ cont.

$$\hat{\theta} \xrightarrow{P} \theta \Longrightarrow A_{\hat{\theta}} \xrightarrow{P} A_\theta$$

$$n\, \hat{\psi}_n^\top A_{\hat{\theta}}\, \hat{\psi}_n \Longrightarrow \chi^2(m)$$

we reject H_0 if

$$n\, \hat{\psi}_n^\top A_{\hat{\theta}}\, \hat{\psi}_n$$

is larger than $(1-\alpha)\times 100\%$ quantile of $\chi^2(m)$.

Nov/30

Likelihood Ratio Test (LRT)

Strategy

$$D_{KL}(P_\theta, P_{\theta_0})$$

θ_0 is an element in Θ_0

true data generating parameter

is uniquely minimized at $\theta_0 = \theta$

The LRT will reject H_0 if

$$\inf_{\theta_0 \in \Theta_0} D_{KL}(P_\theta, P_{\theta_0}) \text{ is too large.}$$

Define estimator of

$$\underbrace{D_{KL}(P_\theta, P_{\theta_0})}_{\substack{\| \leftarrow \text{please verify} \\ P_\theta[l_\theta - l_{\theta_0}]}}$$

by replacing

① P_θ by P_n

② θ by the MLE $\hat\theta$ ($\leftarrow$ the unrestricted MLE maximizing $P_n l_\theta$ over $\theta \in \Theta$)

This yields

$$P_n[l_{\hat\theta} - l_{\theta_0}]$$

In the end, to estimate

$$\inf_{\theta_0 \in \Theta_0} P_n[l_{\hat\theta} - l_{\theta_0}],$$

it is equivalent to estimating

$$P_n l_{\hat\theta} - \sup_{\theta_0 \in \Theta_0} P_n l_{\theta_0}$$

restricted MLE over Θ_0.

$$= P_n l_{\hat\theta} - P_n l_{\hat\theta_0}$$

where $\hat\theta_0 = (0_m, \hat\eta_0)$

We will reject H_0 if $P_n l_{\hat\theta} - P_n l_{\hat\theta_0}$ is too large.

Implementation

$$L_n := 2n \cdot P_n[\ell_{\hat\theta} - \ell_{\hat\theta_0}]$$

Note: $L_n \geq 0$

To decide the limiting null dist (LND) of L_n:

Step 1. We apply Taylor to L_n:

$$L_n = 2\sum_{i=1}^{n}\left[\ell_{\hat\theta}(X_i) - \ell_{\hat\theta_0}(X_i)\right]$$

$$= -2\sum_{i=1}^{n}\left[\ell_{\hat\theta_0}(X_i) - \ell_{\hat\theta}(X_i)\right]$$

$$= -2\cdot(\hat\theta_0 - \hat\theta)^T \sum_{i=1}^{n} \dot\ell_{\hat\theta}(X_i)$$

$$- \sqrt{n}(\hat\theta_0 - \hat\theta)^T \left[\frac{1}{n}\sum_{i=1}^{n} \ddot\ell_{\tilde\theta_n}(X_i)\right]\sqrt{n}(\hat\theta_0 - \hat\theta)$$

with $\tilde\theta_n$ between $\hat\theta$ and $\hat\theta_0$.

Since $\hat\theta$ is the MLE, we have

$$\sum_{i=1}^{n} \dot\ell_{\hat\theta}(X_i) = 0$$

$\Longrightarrow$ the first order term $= 0$

$$\Longrightarrow L_n = -\sqrt{n}(\hat\theta_0 - \hat\theta)[P_n \ddot\ell_{\tilde\theta_n}]\sqrt{n}(\hat\theta_0 - \hat\theta)$$

Steep2. Under H_0, both $\hat{\theta}_0$ and $\hat{\theta}$ should satisfy

$$\hat{\theta}_0 \xrightarrow{P} \theta, \quad \hat{\theta} \xrightarrow{P} \theta$$

$$\Longrightarrow \quad \tilde{\theta}_n \xrightarrow{P} \theta$$

$$\Longrightarrow \quad \ddot{l}_{\tilde{\theta}_n} \xrightarrow{P} \ddot{l}_\theta$$

under smoothness
continuity
of $\ddot{l}_\theta$

$$\Longrightarrow \quad P_n \ddot{l}_{\tilde{\theta}_n} \xrightarrow{P} P_\theta \ddot{l}_\theta = -I_\theta$$

$$\Longrightarrow \quad L_n = \left[\sqrt{n}(\hat{\theta}_0 - \hat{\theta})^T\right] I_\theta \left[\sqrt{n}(\hat{\theta}_0 - \hat{\theta})\right] + o_p(1)$$

if $\theta \in \Theta_0$.

$$= \left[\sqrt{n}\, I_\theta (\hat{\theta}_0 - \hat{\theta})^T\right] I_\theta^{-1} \left[\sqrt{n}\, I_\theta (\hat{\theta}_0 - \hat{\theta})\right] + o_p(1)$$

Step 3.

$$\hat{\theta} - \theta = I_\theta^{-1} (P_n - P_\theta) \dot{l}_\theta + o_p(1/\sqrt{n})$$

MLE asymp. linear expansion

if $\theta \in \Theta_0$, $\hat{\theta}_0 - \theta = \begin{pmatrix} 0 \\ \hat{\eta}_0 - \eta \end{pmatrix}$

with $\hat{\eta}_0 - \eta = I_{\theta,22}^{-1} (P_n - P_\theta) \dot{l}_{\theta,2} + o_p(1/\sqrt{n})$

$$\Longrightarrow \sqrt{n}\, I_\theta (\hat{\theta}_0 - \hat{\theta})$$

$$= \sqrt{n}\, I_\theta (P_n - P_\theta) \left(\begin{bmatrix} 0 \\ I_{\theta,22}^{-1} \dot{l}_{\theta,2} \end{bmatrix} - I_\theta^{-1} \dot{l}_\theta \right) + o_p(1)$$

$$= \sqrt{n}(\mathbb{P}_n - P_\theta)\begin{bmatrix} -(\dot{\ell}_{\theta,1} - I_{\theta,12} I_{\theta,22}^{-1} \dot{\ell}_{\theta,2}) \\ 0 \end{bmatrix} + o_p(1)$$

$$\Longrightarrow \sqrt{n} I_\theta (\hat{\theta}_0 - \hat{\theta}) \Rightarrow \begin{bmatrix} V \\ 0 \end{bmatrix}$$

where $V \sim N(0, A_\theta)$

$$\Longrightarrow L_n \Rightarrow [V^T, 0]\, I_\theta^{-1} \begin{bmatrix} V \\ 0 \end{bmatrix}$$

$\sim \chi^2(m)$ ← Wilk's Thm.

please verify it.

Score test:

Idea: $P_\theta \dot{\ell}_\theta = 0$ (under QMD)

$\Longrightarrow$ If H_0 is true,

$$P_\theta \dot{\ell}_{(0,\eta)} = 0$$

$$\Longrightarrow \mathbb{P}_n \dot{\ell}_{(0,\eta)} \approx 0$$

$$\Longrightarrow Z_n(\hat{\theta}_0) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_{\hat{\theta}_0}(X_i)$$

Proposal: to use the test statistics

$$S_n := \left[Z_n(\hat\theta_0)\right]^T I_{\hat\theta_0}^{-1} \left[Z_n(\hat\theta_0)\right]$$

with $\hat\theta_0$ is $\underset{\theta_0 \in \Theta_0}{\operatorname{argmax}} \mathbb{P}_n \ell_{\theta_0}$

Limiting dist?

It remains to decide the limiting null dist.

$$\begin{aligned} Z_n(\hat\theta_0) &= Z_n(\hat\theta_0) - \sqrt{n} P_\theta \dot\ell_{\hat\theta_0} \\ &\quad + \sqrt{n} P_\theta \dot\ell_{\hat\theta_0} \\ &= \sqrt{n}(\mathbb{P}_n - P_\theta)\dot\ell_{\hat\theta_0} \\ &\quad + \sqrt{n} P_\theta \dot\ell_{\hat\theta_0} - \sqrt{n} \underbrace{\boxed{P_\theta \dot\ell_\theta}}_{=0} \\ &= \boxed{\sqrt{n}(\mathbb{P}_n - P_\theta)\dot\ell_\theta} \leftarrow \text{CLT} \\ &\quad + \boxed{\sqrt{n}(P_\theta \dot\ell_{\hat\theta_0} - P_\theta \dot\ell_\theta)} \leftarrow \text{Delta method} \\ &\quad + \boxed{\sqrt{n}(\mathbb{P}_n - P_\theta)(\dot\ell_{\hat\theta_0} - \dot\ell_\theta)} \leftarrow \text{Donsker } o_p(1) \end{aligned}$$

$$\Rightarrow Z_n(\hat\theta_0) \overset{H_0}{\Longrightarrow} \begin{bmatrix} V \\ 0 \end{bmatrix} \quad \text{with } V \sim N(0, A_\theta)$$

$$\Rightarrow S_n \Rightarrow \chi^2(m)$$

Remark: (Comparison of the three)

$$\chi^2_{(m)} \overset{H_0}{\Longleftarrow} \begin{cases} W_n = n\hat{\psi}^T A_{\hat{\theta}} \hat{\psi} \\ L_n = 2n P_n [\ell_{\hat{\theta}} - \ell_{\hat{\theta}_0}] \\ S_n = [Z_n(\hat{\theta}_0)^T] I_{\hat{\theta}_0}^{-1} Z_n(\hat{\theta}_0) \end{cases}$$

Under H_0 [Engle's book]

$$W_n - L_n = o_p(1)$$
$$L_n - S_n = o_p(1)$$
$$S_n - W_n = o_p(1)$$

For computation

(i) S_n is only based on $\hat{\theta}_0 \in \Theta_0$ ← constrained MLE

(ii) W_n ~ — — ~ $\hat{\theta}$

Chap 4.2 Local Power analysis

$$\lim_{n\to\infty} \underbrace{\pi_n(\theta)}_{\text{size}} \le \alpha, \quad \forall \theta \in \Theta_0$$

[pointwise asymp. size control]

$$\lim_{n\to\infty} \underbrace{\pi_n(\theta)}_{\text{power}} \text{ to be as large as possible } \forall \theta \in \Theta_1.$$

If we think about the asymp. power

$\lim_{n\to\infty} \pi_n(\theta)$, $\theta \in \Theta_1$.

and if we consider $\theta \in \Theta_1$ above to be fixed, then usually you will get every reasonable (even unreasonable) tests satisfy:

(i) $\forall \underset{\text{fixed}}{\theta} \in \Theta_0$, $\pi_n(\theta) \to 0$ asymp. $n \to \infty$

(ii) $\forall \underset{\text{fixed}}{\theta} \in \Theta_1$, $\pi_n(\theta) \to 1$ $n \to \infty$

<u>Example</u> Let me think about in the previous Wald/LRT/Score test framework.

$$\sqrt{n}(\hat{\psi} - \psi) \Rightarrow N(0, A_\theta^{-1})$$

Let's think about an alternative to Wald:

reject H_0 if $\|\hat{\psi}\| > n^{-1/4}$ ← over conservative

Intuitively too conservative.

However, we can show

(i) $\forall \theta \in \Theta_0$,

$\boxed{\pi_n(\theta)} \xrightarrow{n\to\infty} 0$

Pf: Fix $\forall t$, for n large enough,
we have

$$\pi_n(\theta) = P_\theta(\|\hat{\varphi}\| > n^{-1/4})$$

$$= P_\theta(\sqrt{n}\|\hat{\varphi}\| > n^{1/4})$$

$$\leq P_\theta(\sqrt{n}\|\hat{\varphi}\| > t)$$

By CMT,

$$\sqrt{n}\|\hat{\varphi}\| \overset{H_0}{\Longrightarrow} \|Z\| \text{ with } Z_n \sim N(0, A_\theta^{-1})$$

$$\Rightarrow \limsup \pi_n(\theta) \leq \limsup P_\theta(\sqrt{n}\|\hat{\varphi}\| > t)$$

$$= P(\|Z\| > t)$$

if $t \to \infty$

$$= 0$$

(2) For any fixed $\theta \in \Theta_1$,

$$\pi_n(\theta) \xrightarrow{n \to \infty} 1.$$

Pf. $1 \geq \pi_n(\theta)$

$$= P_\theta[\|\hat{\varphi}\| > n^{-1/4}]$$

$$\geq P_\theta[\|\varphi\| - \|\hat{\varphi} - \varphi\| > n^{-1/4}]$$

$$= P_\theta[\underbrace{n^{1/4}\|\varphi\|}_{①} - \underbrace{n^{1/4}\|\hat{\varphi} - \varphi\|}_{②} > 1]$$

$\forall$ fixed $\theta \in \Theta_1$, $\psi > 0$ so that

(i) $n^{\frac{1}{4}} \|\psi\| \longrightarrow \infty$

(ii) $n^{1/2} \|\hat{\psi} - \psi\| \Rightarrow N(0, \cdots)$

implies that

$$\underbrace{n^{-\frac{1}{4}}}_{o(1)} \underbrace{n^{\frac{1}{2}} \|\hat{\psi} - \psi\|}_{O_p(1)} = o_p(1)$$

$\overset{n\to\infty}{=} 1$.

which means

$$\lim \pi_n(\theta) = 1$$

$\square$

We shouldn't consider previous goal all in asymp.

This motivates the new framework of local power analysis:

$H_0 : \theta \in \Theta_0$

$X_1, X_2, \cdots, X_n \sim P_\theta$

v.s.

$H_{1,n}: \theta_n \in \Theta_1$

local alternative

$X_1, X_2, \ldots, X_n \sim P_{\theta_n}$

$$\boxed{\theta_n = \theta_0 + \frac{h}{\sqrt{n}}}$$

the critical local alternative seq.

Goal: to control the size and maximize the local power

$$\lim_{n\to\infty} \pi_n(\theta_n)$$

[note: under LPA framework,

$$\lim \underbrace{\pi_n(\theta_n)}_{\|\psi\| > n^{-1/4}} = 0$$

power]

Q: To use LPA, we need limiting dist. of $\pi_n(\theta_n)$ with a changing seq. of P_{θ_n}?

We have known a lot of asymp. dist. under fixed P_θ.

Le Cam's change of measure claim

① T under P_{θ_0}

② If we know $\dfrac{dP_{\theta_0+h/\sqrt{n}}}{dP_{\theta_0}}$ $\leftarrow \theta_n$ under P_{θ_0}

③ $\left(T, \dfrac{dP_{\theta+h/\sqrt{n}}}{dP_{\theta_0}}\right)$ under P_{θ_0}

Then, we know the asymp. dist of

T under $P_{\theta_0+h/\sqrt{n}}$

↑ Le Cam's third Lemma

Q: Why Le Cam's third Lemma makes sense?

A: For fixed P and Q, and $Q \ll P$.

then $\forall$ event A,

$$Q(A) = \int \mathbb{1}(z \in A)\, Q(dz)$$

$$= \int \mathbb{1}(z \in A) \overset{②}{\frac{dQ}{dP}} \overset{①}{P(dz)}$$

However, the above huerstic is flawed,

that is, Le Cam's 3rd Lemma is asymp. discuss.!

The sol'n will give us two things.

① Local asymp. normality

$$\frac{dP_{\theta+\frac{h}{\sqrt{n}}}}{dP_{\theta_0}} \xrightarrow{P_{\theta_0}} N(\cdot\, , \cdot)$$

② Le Cam's 1st Lemma.

Logic:

① Goal: to study the power of any test

② Fixed alternative doesn't give us anything.

Instead, we have to study

$\pi(\theta_n)$ with $\theta_n \xrightarrow{n\to\infty} \theta_0 \in \Theta_0$

local alternative seq.

③ To study $\pi(\theta_n)$, it is equiv. to building the limiting dist $\mathcal{J}$ under the local alternative $P_{\theta,n}$.

Le Cam's third Lemma

(a) If P and Q are two prob. meas.,

then

① If we know P

② If we know $\frac{dQ}{dP}$ $(Q \ll P)$
iff

then we know Q.

(b) However, we have to examine/study asymp. version of (a):

(b1) Q1: What is the asymp. version of $\boxed{\text{a.c.}}$

$\{P_n\}_{n\geq 1}$ and $\{Q_n\}_{n\geq 1}$

We say they are contiguity

Def. [Contiguity]

We say $\{Q_n\}$ is contiguous w.r.t. $\{P_n\}$, written as

$$Q_n \lhd P_n \quad \text{if}$$

$\forall$ seq. of events $\{A_n\}$, we have

$$P_n(A_n) \xrightarrow{n\to\infty} 0 \quad \text{must imply}$$

$$Q_n(A_n) \xrightarrow{n\to\infty} 0.$$

The approach to verify contiguity is Le Cam's 1st Lemma.

To prepare for 1st Lemma, some knowledge

1. [Lebesgue decomposition Thm].

$\forall$ measures μ, ν,

$\exists$ unique measures ν^a and $\nu^\perp$ s.t.

① $\nu = \nu^a + \nu^\perp$

② $\nu^a \ll \mu$

③ $\nu^\perp \perp \mu$ ($\nu^\perp(A) > 0 \Leftrightarrow \mu(A) = 0$

$\nu^\perp(A) = 0 \Leftrightarrow \mu(A) > 0$)

2. The following 3 conds. are equiv.

① $Q \ll P$

② $\int L(x) P(dx) = 1$ for

$L: \frac{dQ^a}{dP}$

③ $Q = Q^a$

Lemma (Le Cam 1st)

The following four are equiv:

(a) $Q_n \triangleleft P_n$

(b) $L_n := \frac{dQ_n^a}{dP_n} \xrightarrow{P_n} V$ along a subseq. of $\{1, 2, 3, \cdots\}$

$$\Vert$$

$$\mathbb{E}[V] = 1 .$$

(c) $\dfrac{dP_n^a}{dQ_n} \overset{Q_n}{\Longrightarrow} U$ along a subseq. of $\{1, 2, 3, \dots\}$

$$\Vert$$

$$P(U > 0) = 1 .$$

(d) For any seq. of func.

$$f_n : \mathcal{Z}_n \longmapsto \mathbb{R} ,$$

where $f_n(Z_n) = o_{P_n}(1)$

$$\Vert$$

$$f_n(Z_n) = o_{Q_n}(1)$$

Pf is not required.

Thm. If $L_n := \frac{dQ_n}{dP_n}$ satisfies

$$\log L_n \xrightarrow{P_n} N(\mu, \sigma^2) \text{ and } Q_n \triangleleft P_n$$

then $\mu = -\sigma^2/2$. such ones is called to satisfy local asymp. normality.

[HW will show $Q_n = P_{\theta_0 + \frac{h}{\sqrt{n}}}^{\otimes n}$

$$P_n = P_{\theta_0}^{\otimes n}$$

will give a LAN seq.]

Proof. Based on 1st Lemma (b):

$L_n \xrightarrow{P_n} V$ with

$$V \sim \text{logNormal}(\mu, \sigma^2)$$

and we know

$$\mathbb{E}[V] = \exp(\mu + \sigma^2/2)$$

Then $Q_n \triangleleft P_n$ imply $\mathbb{E}[V]$.

$\therefore \mu = -\sigma^2/2$.

(c) Lemma (Le Cam 3rd Lemma)

Let ① $\{P_n\}, \{Q_n\}$ are prob. meas. s.t.

test

$Q_n \lhd P_n$

② $T_n : Z_n \to \mathbb{R}^d$ be the test statistics.

Then, if

$$\begin{pmatrix} T_n \\ L_n \end{pmatrix} \xrightarrow{P_n} \begin{pmatrix} T \\ V \end{pmatrix}$$

$$L_n = \frac{dQ_n^a}{dP_n}$$

then $\forall$ event $A \subset \mathbb{R}^d$, letting

$$R(A) := \mathbb{E}[\mathbb{1}\{T \in A\} \cdot V]$$

we have

① $R(\cdot)$ is a prob. meas.

② $T_n \xrightarrow{Q_n} R$

Proof. Left to you....

Lemma (Le Cam's 3rd Lemma, user-friendly)

If in the above setting, we know

$$\begin{pmatrix} T \\ \log V \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ -\frac{\theta^2}{2} \end{pmatrix}, \begin{bmatrix} \Sigma & \tau \\ \tau^T & \theta^2 \end{bmatrix} \right)$$

then

$$T_n \xRightarrow{Q_n} \mathcal{N}(\mu+\tau, \Sigma)$$

which should be compared to

$$T_n \xRightarrow{P_n} \mathcal{N}(\mu, \Sigma)$$

<u>Pf.</u> Using version 1, we have

① $R(A) = \mathbb{E}[\mathbb{1}_A(T) \cdot V]$

$$= \mathbb{E}\left[\mathbb{1}_A(T) \cdot \mathbb{E}[V|T]\right]$$

② $\log V | T \sim \mathcal{N}(\cdot, \cdot)$

③ $\mathbb{E}[V|T] = \boxed{\exp(\cdots)}$

④ $R(A) = \int_A \boxed{} \, d\lambda(t)$

$$= \int_A \text{density}(\mathcal{N}(\mu+\tau, \Sigma))\, d\lambda$$

⑤ It thus shows

$R(\cdot)$ is prob. meas. of $N(\mu+\tau, \Sigma)$ $\square$

Next, we will use 3rd Lemma to analyse the local power of Wald test.

① We know $T_n \xRightarrow{P_{\theta_0}}$ a certain dist. $\chi^2(m)$

② LAN gives us $\log \frac{dP_{\theta_0+h/\sqrt{n}}}{dP_{\theta_0}} \Rightarrow N(\cdot,\cdot)$

③ T_n is based on $\hat{\theta}_n$, and "ALE"

$$\hat{\theta}_n = \frac{1}{\sqrt{n}} \sum \cdot + o_{P_{\theta_0}}(1)$$

$$\log \frac{dP_{\theta_0+h/\sqrt{n}}}{dP_{\theta_0}} = \frac{1}{\sqrt{n}} \sum \cdot + o_{P_{\theta_0}}(1)$$

④ $$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n - \theta_0) \\ \log \frac{dP_{\theta_0+h/\sqrt{n}}}{dP_{\theta_0}} \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} \sum \cdot \\ \sum \cdot \end{pmatrix} + o_{P_{\theta_0}}(1)$$

$$\xRightarrow{P_{\theta_0}} N(\cdot, \cdot)$$

⑤ Using Le Cam 3rd,

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \xRightarrow{P_{\theta_0+h/\sqrt{n}}} N(\cdot, \cdot)$$

Thm. [MLE under $P_{\theta_0+h/\sqrt{n}}$]

Assume

① θ is an interior point of Θ;

② $\theta \in \Theta_0$

③ $P_n = P_\theta^{\otimes n}$ ← n IID RVs $\sim P_\theta$

$Q_n = P_{\theta+h/\sqrt{n}}$

④ $\{P_{\tilde\theta} : \tilde\theta \in \Theta\}$ is QMD at θ;

⑤ I_θ is invertible;

⑥ $\tilde\theta \mapsto I_{\tilde\theta}$ is cont. at θ;

⑦ $\sqrt{n}(\underset{\text{MLE}}{\hat\theta_n} - \theta) = \underbrace{\boxed{I_\theta^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_\theta(X_i)}}_{\sqrt{n} I_\theta^{-1} P_n \dot{l}_\theta} + o_{P_n}(1)$

Conclusion:

① $\sqrt{n}(\hat\theta_n - \theta) \overset{Q_n = P_{\theta+h/\sqrt{n}}}{\Longrightarrow} N(h, I_\theta^{-1})$

② $W_n := n \psi_n^T A_{\hat\theta_n} \psi_n \overset{Q_n}{\Longrightarrow} Y$,

where Y is a non-centered χ^2

$\chi^2(m, h_\psi^T A_\theta h_\psi)$, $h = (\underset{\text{first m-dim}}{\boxed{h_\psi}}, h_\eta)$

③ If $h \to \infty$, then

$$\pi_n\left(\theta + \frac{h}{\sqrt{n}}\right) \longrightarrow 1.$$

If $h \to 0$, then

$$\pi_n\left(\theta + \frac{h}{\sqrt{n}}\right) \longrightarrow \alpha$$

<u>Remark</u>. Notice

① $\sqrt{n}(\hat{\theta}_n - \theta) \overset{P_\theta}{\Longrightarrow} N(0, I_\theta^{-1})$

② $\sqrt{n}(\hat{\theta}_n - \theta) \underset{P_{\theta+\frac{h}{\sqrt{n}}}}{\Longrightarrow} N(h, I_\theta^{-1})$

$$\Longleftrightarrow \sqrt{n}\left(\hat{\theta}_n - \left(\theta + \frac{h}{\sqrt{n}}\right)\right) \underset{P_{\theta+\frac{h}{\sqrt{n}}}}{\Longrightarrow} N(0, I_\theta^{-1})$$

In other words, MLE has the property that a shift of DGP $\theta + h/\sqrt{n}$ doesn't change the estimation asymp. dist.

Chap. 4.4 Regular ALE (RALE)

Def. [ALE] μ_n is a generic ALE estimating a certain functional $\mu(\theta) \in \mathbb{R}^m$

if $\exists$ influence func.

$$g_\theta : \mathcal{X} \longmapsto \mathbb{R}^m \quad \text{(imagine Fisher score func.)}$$

s.t. $P_\theta g_\theta = 0$

and $P_\theta[g_\theta g_\theta^T]$ is well-defined

s.t.

$$\underbrace{\sqrt{n}(\mu_n - \mu(\theta)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \overbrace{\underbrace{g_\theta(X_i)}_{\text{MLE, } I_\theta^{-1}\dot{\ell}_\theta}}^{\text{influence func.}} + o_{P_n}(1)}_{\text{a linear term, CLT}}$$

Remark We can apply MCLT

$$\sqrt{n}(\mu_n - \mu(\theta)) \xrightarrow{P_\theta} N(0, P_\theta[g_\theta g_\theta^T])$$

Def. [Regular ALE (RALE)]

① The estimator μ_n is said to be RALE (regular)

if $\forall h \in \mathbb{R}^{|\theta|}$,

$$\sqrt{n}\left(\mu_n - \mu\left(\theta + \frac{h}{\sqrt{n}}\right)\right) \xrightarrow{P_{\theta + h/\sqrt{n}}} Z,$$

where Z doesn't depend on h.

② μ_n is RALE if it is regular and ALE.

<u>Remark</u>. One can show if μ_n is a regular ALE with influence func. g_θ, it must be true that

$$\dot{\mu}(\theta) = P_\theta(\dot{\ell}_\theta g_\theta)$$

$$\left.\frac{\partial}{\partial \theta'}\mu(\theta')\right|_{\theta=\theta'}$$

and in this case, we will say

g_θ to be the gradient of $\mu(\cdot)$ at θ w.r.t. model

$$\{P_{\theta'} : \theta' \in \Theta\}$$

Result 1: $\left[\begin{array}{c}\text{Le Cam 3rd Lemma} \\ + \\ \text{ALE of } \mu \\ + \\ \text{LAN}\end{array}\right]$

$$\Rightarrow \sqrt{n}(\mu_n - \mu(\theta)) \xRightarrow{P_{\theta+\frac{h}{\sqrt{n}}}} N(\cdot, \cdot)$$

Result 2. If μ_n is RALE,

$$\sqrt{n}(\mu_n - \mu(\theta)) \xRightarrow{P_{\theta+\frac{h}{\sqrt{n}}}} \underbrace{N(\cdot\,,\ \cdot)}_{\text{a simpler form}}$$