1 Basic convergence concepts and preliminary theorems

P5

1.1 Modes of convergence of a sequence of random variables

Definition 1.1.1 (convergence in probability)**.**

Example 1.1.1. *Bernoulli trials, derive expectations and variance, using chebyshev to derive* $\xrightarrow{p}$

$\boldsymbol{X}_n \xrightarrow{p} \boldsymbol{X}\ if \|\boldsymbol{X}_n - \boldsymbol{X}\| \xrightarrow{p} 0$
$\boldsymbol{X}_n \xrightarrow{p} \boldsymbol{X} \iff component-wise\ convergence$
$X_n = E(X_n) + O_p(\sqrt{\mathrm{Var}(X_n)})$

Definition 1.1.2 (bounded in probability)**.** *for any* ϵ*, there exists a* k*, s.t.*$\sqrt{P(|X_n| > k) \leq \epsilon}$
$X_n = O_p(1)$

Definition 1.1.3 (convergence with probability one)**.**

characterization: $\lim_{n\to\infty} P(|X_m - X| < \epsilon, \forall m \geq n) = 1, every\ \epsilon$

Example 1.1.2. *iid U(0,1),* $X_{(n)} = \max X_1, \ldots, X_n$*, see* $X_{(n)} \xrightarrow{wp1} 1$.

Definition 1.1.4 (convergence in rth mean)**.** $\lim_{n\to\infty} E|X_n - X|^r = 0$

The bigger the r is, the stronger the convergence is. (By Jensen: $EY^{\frac{r}{s}} \geq (EY)^{\frac{r}{s}}$, where $Y = |X_n - X|^s$)

Definition 1.1.5 (convergence in distribution(in law))**.** $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$*, at every continuity point of* F_X

Example 1.1.3. $X_n \sim Uniform(\frac{1}{n}, \ldots, \frac{n-1}{n}, 1)$*, then for any* $t \in [\frac{i}{n}, \frac{i+1}{n}), F_{X_n} = \frac{i}{n}, F_X(t) = t$*, so we have* $\xrightarrow{d}$

Example 1.1.4. $X_n \sim N(0, 1 + n^{-1}), \lim_n F_{X_n}(x) = \phi(x)$

joint convergence in law versus(stronger than)marginal convergence

Example 1.1.5. $X \sim U(0,1)$*, so* $1 - X \sim U(0,1)$*, let* $Y_n = X$ *for* n *odd and* $Y_n = 1 - X$ *for* n *even.*

Suppose X_n, X are integer-valued r.v., we have: $X_n \xrightarrow{d} X \iff P(X_n = k) \to P(X = k)$

1.2 Fundamental results and theorems on convergence

1.2.1 Relationship

Theorem 1.2.1. *the four relationship between the four convergence and the proof respectively.*

Example 1.2.1. $X_i's \overset{i.i.d}{\sim} N(0,1)$*, by the above and Markov ineq. we have:* $X_n \xrightarrow{wp1} 0$

1.2.2 Transformations

P10

Theorem 1.2.2 (Continuous Mapping Theorem)**.** *Note that we need the joint convergence especially when convergence in law!*

Example 1.2.2. *(i) (ii)If* $(X_n, Y_n) \xrightarrow{d} N_2(\mathbf{0}, \boldsymbol{I}_2)$*, then* $\max\{X_n, Y_n\} \xrightarrow{d} \max\{X, Y\}$*, which has the CDF* $[\phi(x)]^2$

Corollary 1.2.1. *linear and quadratic forms of CMT. If we have joint convergence, then* $\boldsymbol{A}\boldsymbol{X}_n \to \boldsymbol{A}\boldsymbol{X}$*, so as the quadratic form. Proof can be conduct using the fact that they are all functions of* $\boldsymbol{X}_n$

Example 1.2.3. *(i)(ii)When we have marginal convergence of wp1 or probability, we have the corresponding convergence of their sum and products. The anagolous statement is wrong in terms of* $\xrightarrow{d}$*, for the joint convergence is needed.*

Example 1.2.4. *(i)For non-degenerate distribution, the function can be non-continous at countable points.* $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \sim N(0,1), then\ \frac{1}{X_n} \xrightarrow{d} Z = \frac{1}{X}$

(ii)if $(X_n, Y_n) \xrightarrow{d} N_2(\mathbf{0}, \boldsymbol{I}_2)$*, then* $\frac{X_n}{Y_n} \xrightarrow{d} standard\ Cauchy$*. The ratio of two independent standard normal rv. and request the joint convergence.*

Example 1.2.5. $\overline{X}_n \xrightarrow{p} \theta \Rightarrow \overline{X}_n^{1/2} \xrightarrow{p} \theta^{1/2}$

Theorem 1.2.3 (Slutsky's Theorem)**.** *Marginal convergence is needed for convergence in law in the case of:* $X_n \xrightarrow{d} X, Y_n \xrightarrow{p} c$*, then we have the corresponding convergence of their sum and products and ratio. Think of the proof.*

$Y_n = X_n + o_p(1)$, then we can investigate Y_n by investigating X_n. It suffixes to show that $Y_n - X_n = o_p(1)$

Example 1.2.6. $X_n \xrightarrow{p} c \iff X_n \xrightarrow{d} c$*, can be proved by definition of* $\xrightarrow{p}$ *and the degenerate distribution of c.*

Example 1.2.7. *Gamma distribution, omited*

Example 1.2.8. *the asymptotic distribution of t-statistic can be derived by CLT, WLLN and Slutsky.*

1.2.3 WLLN and SLLN

Theorem 1.2.4. $X_i's \overset{iid}{\sim} F$

(i)WLLN: $\lim_{x\to\infty} x[1 - F(x) + F(-x)] = 0$

(ii)SLLN: $E[X_1]$ *is finite.*

Example 1.2.9. $X_i's \sim t(2)$*, can use WLLN above.*

Theorem 1.2.5. *$X_1, X_2, \dots$ with finite expectations.*

(i)WLLN: uncorrelated and $\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 = 0$

(ii)SLLN: independent and $\sum_{i=1}^\infty \frac{\sigma_i^2}{c_i^2} < \infty$*,(special: c=i), where c_n ultimately monotone and $c_n \to \infty$, then:*

$$c_n^{-1}\sum_{i=1}^n (X_i - \mu_i) \xrightarrow{wp1} 0.$$

(iii)SLLN with common mean: $\sum_{i=1}^\infty \sigma_i^{-2} = \infty$

Example 1.2.10. *Consider the BLUE of the $X_i \overset{indep}{\sim} (\mu, \sigma_i^2)$, fitted (iii) above.*

Example 1.2.11. *When $(X_i's, Y_i's)$ are iid bivariate samples, then we have: $X_iY_i's$ are iid and are fit for SLLN. We can derive the asymptotic distribution of r_n by this, SLLN and CMT.*

1.2.4 Characterization of convergence in law

Theorem 1.2.6. *(i)(The Portmanteau)*$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \iff E[g(\boldsymbol{X}_n)] \to E[g(\boldsymbol{X})]$

(ii)(Levy-Cramer continuity)Let ϕ denote the CHF, then $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \iff \lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t), \forall \boldsymbol{t}$

(iii)(Cramer-Wold device)$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \iff \boldsymbol{c}'\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{c}'\boldsymbol{X}, \forall \boldsymbol{c}$

(iii) can be proved by (ii). (iii) is often used when calculating the joint distribution

We can get: $\boldsymbol{X}_N \xrightarrow{d} \boldsymbol{X}, \boldsymbol{Y}_n \xrightarrow{d} \boldsymbol{c} \Rightarrow (\boldsymbol{X}_n, \boldsymbol{Y}_n) \xrightarrow{d} (\boldsymbol{X}, \boldsymbol{c})$ using Slutsky and (iii) above.

Example 1.2.12. *Example 1.1.3 revisted. Let $g(x) = x^{10}$ and use (i) above.*

Example 1.2.13. *Bernstein polynomials, using WLLN and (i) above.*

Note the the r.v. of Bin(n,p) can be seen as sum of iid, so we can use LLN or CLT!

Example 1.2.14. *(Proof of LLN and CLT)*

$X_i's$ iid and $T_n = \sum_{i=1}^n X_n$. So we have: $[\frac{\partial \phi_X(t)}{\partial t}]\Big|_{t=0} = \sqrt{-1}EX, [\frac{\partial^2 \phi_X(t)}{\partial t^2}]\Big|_{t=0} = -EX^2,$

(i)We use first-order Taylor, then:

$$\phi_{\frac{T_n}{n}}(t) = [\phi_{X_1}(\frac{t}{n})]^n = [1 + \frac{\sqrt{-1}\mu t}{n} + o(|t|n^{-1})]^n \to exp\{\sqrt{-1}\mu t\} = \phi_\mu(t)$$

(ii)We use first-order Taylor and suppose $\mu = 0$ then:

$$\phi_{\frac{T_n}{\sqrt{n}}}(t) = [\phi_{X_1}(\frac{t}{\sqrt{n}})]^n = [1 - \frac{\sigma^2 t^2}{2n} + o(t^2 n^{-1})]^n \to exp\{-\sigma^2 t^2/2\} = \phi_{N(0,\sigma^2)}(t)$$

(iii)

Theorem 1.2.7. *(i)(Prohorov's)*$X_n \xrightarrow{d} X \Rightarrow X_n = O_p(1)$

(ii)(Polya's)"$F_{X_n} \Rightarrow F_X +$ continuity of F_X" $\Rightarrow \sup_x |F_{X_n} - F_X| \to 0$

Proof of (ii) used cutting of the interval, which can turn a uncountable set into a countable object set. We divide the value range of F_x i.e. [0,1] into k pieces, with the monotonicity of F and the convergence of F_{X_n} at every point in the domain, then let $k \to \infty$

Theorem 1.2.8. *(Scheffe)* $\lim_n f_n(\boldsymbol{x}) = f(\boldsymbol{x}), \forall \boldsymbol{x}, + f$ *and* f_n *is density,* $\Rightarrow \lim_n \int |f_n(\boldsymbol{x}) - f(\boldsymbol{x})| d\boldsymbol{x} = 0$

The sketch of the proof: the integration equals and we let $g_n(\boldsymbol{x}) = [f - f_n] I_{f \geq f_n}$ and use dominated convergence theorem.

Theorem 1.2.9 (Frechet and Shohat). *Note the Carleman condition at P19*

1.2.5 Results on o_p and O_p

$R_n o_p(1) = o_p(R_n), R_n O_p(1) = O_p(R_n)$.

Lemma 1.2.1. *The key of the proof is* $f(\boldsymbol{t}) = \frac{g(\boldsymbol{t})}{\|\boldsymbol{t}\|^r}$

1.3 The central limit theorem

P20

Definition 1.3.1. $\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1)$, *written by* X_n *is* $AN(\mu_n, \sigma_n)$

1.3.1 The CLT for the iid case

Theorem 1.3.1 (Lindeberg-Levy). *iid, finite variance.*

Example 1.3.1 (Confidence intervals). *namely Wald intervals, when* σ *is known*

Example 1.3.2 (Sample variance). *iid and finite fourth moments. We can derive the AS distribution by using:* $Y_n = X_n + o_p(1)$. *We want the AS distribution of* X_n *to be known.*

$$\sqrt{n}(S_n^2 - \sigma^2) = \sqrt{n}(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2) - \sqrt{n}\frac{n}{n-1}(\overline{X}_n - \mu)^2.$$

iid sum, so we can use CLT above.

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4)$$

where $\mu_4 - \sigma^4 = \mathrm{Var}[(X_1 - \mu)^2]$

Example 1.3.3 (Level of the Chi-square test). *For we can take* χ_n^2 *as sum of iid, we can use CLT to derive the AS distribution of it. So simply we have:* $\frac{\chi_{n-1}^2 - (n-1)}{\sqrt{2(n-1)}} \xrightarrow{d} N(0,1)$

Moreover, we have: $\sqrt{n}(\frac{S_n^2}{\sigma^2} - 1) \xrightarrow{d} N(0, \kappa + 2)$, *where* $\kappa = \frac{\mu_4}{\sigma^2} - 3$ *is the kurtosis.*
Simply we derive, $P_{H_0}(\frac{nS_n^2}{\sigma^2} > \chi_{n-1,\alpha}^2) \to 1 - \phi(\frac{z_\alpha\sqrt{2}}{\sqrt{k+2}})$

Theorem 1.3.2 (Multivariate CLT for iid case). *Proved by univariate CLT and Cramer-Wold device.*

An exercise at the P22-.

Example 1.3.4. *We want to find the joint distribution of* $(\overline{X}_n, Z_n)$*. Then* $\mathbf{\Sigma} = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, I\{X_1 = 0\}) \\ * & \mathrm{Var}(I\{X_1 = 0\}) \end{pmatrix}$
Finally we have:

$$\sqrt{n}(\overline{X}_n, Z_n) - (E\overline{X}_n, EZ_n) \xrightarrow{d} \boldsymbol{N}_2(\mathbf{0}, \mathbf{\Sigma})$$

Definition 1.3.2. *Slowly varying at* ∞*, if* $\lim_{n\to\infty} \frac{g(tx)}{g(x)} = 1, \forall t > 0$

Examples: $\log x, \frac{x}{1+x}$

Theorem 1.3.3 (CLT when variance do not exist)**.** *iid,* $v(x) = \int_{-x}^{x} y^2 dF(y)$ *is slowly varying at* ∞

Example 1.3.5. $X_1, \ldots \overset{iid}{\sim} t(2)$

1.3.2 The CLT for the independent not necessarily iid case

Theorem 1.3.4 (Lindeberg-Feller)**.** *independent and finite variance, and the L-F condition:*

$$\frac{1}{s_n^2} \sum_{j=1}^{n} \int_{|x-\mu_j| > \epsilon s_n} (x - \mu_j)^2 dF_j(x) \to 0$$

i.e.

$$\frac{1}{s_n^2} \sum_{j=1}^{n} E[(X_j - \mu_j)^2 I_{|X_j - \mu_j| > \epsilon s_n}]$$

where: $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$

Example 1.3.6. $X_1, \ldots$ *are independent, and* $X_j \sim U(-j, j)$

Theorem 1.3.5 (Liapounov)**.** *independent, finite variance, Liapounov condition:*

$$\frac{1}{s_n^{2+\delta}} \sum_{j=1}^{n} E|X_j - \mu_j|^{2+\delta}$$

Usually choose $\delta = 1$ or 2. If X_n is uniformly bounded and $s_n \to \infty$, then choose $\delta = 1$ and finish.

Example 1.3.7. *independent,* $X_i \sim BIN(p_i, 1)$*. The condition above hold with* $\delta = 1$ *if* $s_n \to \infty$

Theorem 1.3.6 (Hajek-Sidak)**.** *iid, finite variance, condition:*

$$\max_{1 \le i \le n} \frac{c_{ni}^2}{\sum_{j=1}^{n} c_{nj}^2} \to 0$$

Then,

$$\frac{\sum_{i=1}^{n} c_{ni}(X_i - \mu)}{\sigma \sqrt{\sum_{j=1}^{n} c_{nj}^2}} \xrightarrow{d} N(0, 1)$$

Example 1.3.8 (Simplest linear regression)**.** *Essential! Calculate* $\hat{\beta}_1 - \beta_1$ *and use the theorem above.*

Theorem 1.3.7 (Lindeberg-Feller multivariate)**.** *See*

Example 1.3.9 (multivariate regression)**.** *Note that* $\boldsymbol{a}_{ni} \sim n^{-\frac{1}{2}}$

P27

1.3.3 CLT for a random number of summands

Theorem 1.3.8 (Anscombe-Renyi)**.**

Example 1.3.10. *Coupon*

1.3.4 CLT for dependent sequences

Example 1.3.11. *stationary Gaussian sequence, common mean and variance, long-run variance exists.* $\mathrm{Var}(\sqrt{n}(\overline{X}_n - \mu)) = \sigma + \frac{1}{n}\sum_{i\neq j} Cov(X_i Xj) = \sigma^2 + \frac{2}{n}\sum_{i=1}^{n}(n-i)\gamma_i$

Definition 1.3.3 (m-dependent)**.** *for a given fixed m if* $(X_1, \ldots, X_i)$ *and* $X_j, X_{j+1}, \ldots$ *are independent whenever* $j - i > m$

Theorem 1.3.9 (m-dependent sequence)**.** $X_i's$ *is m-dependent with common mean and variance, then*

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \tau^2)$$

where $\tau^2 = \sigma^2 + 2\sum_{i=2}^{m+1} Cov(X_i, X_j)$

See Homework for a proof.

Example 1.3.12. *Suppose* Z_i *are iid with a finite variance* σ^2*, let* $X_i = (Z_i + Z_{i+1})/2$*, we can write:* $\sum_{i=1}^{n} X_i = \sum_{i=2}^{n} Z_i + \frac{Z_1 + Z_{n+1}}{2}$*, that's the* $Y_n = X_n + o_p(1)$

Example 1.3.13. *None*

Theorem 1.3.10. *sample without replacement*

Example 1.3.14. *None*

1.3.5 Accuracy of CLT

Theorem 1.3.11 (Berry-Esseen)**.** *(i)The speed of convergence of CLT is* $n^{-\frac{1}{2}}$ *(ii)*

omited

1.3.6 Edgeworth expansions

p34

2 Transformations of given statistics: The delta method

P41

2.1 Basic result

Theorem 2.1.1 (Delta Theorem)**.** *The core of this is: If*

$\sqrt{n}(T_n - \theta) = O_p(1)$ *and* g *be once differentiable at* θ *with* $g'(\theta) \neq 0$

then we have:

$$\sqrt{n}[g(T_n) - g(\theta)] = \sqrt{n}(T_n - \theta)g'(\theta) + o_p(1)$$

This can be proved by Taylor expansion w.r.t. $g(T_n)$ *at* θ*.*

Special case is when $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$

We can use Delta theorem with Slutsky, for $T_n \xrightarrow{p} \theta$.

Even more generally, we can replace all the $\sqrt{n}$ with any a_n which is a sequence of positive numbers with $a_n \to \infty$

There is an exercise at P43-

Example 2.1.1. *For the iid CLT case with* $g(x) = x_2$*, we should discuss whether* $\mu = 0$*, for Delta Theorem requests* $g'(\mu) \neq 0$*.*

When $\mu = 0$*, we can get the AS distribution by CMT. (*$\sqrt{n}\overline{X}_n/\sigma \xrightarrow{d} N(0,1) \Rightarrow n\overline{X}_n^2/\sigma^2 \sim \chi_1^2$*)*

Example 2.1.2. *Note that r.v. of Bin(n,p) can be seen as sum of iid., then we can use CLT.*

Example 2.1.3. *Suppose* T_n *is the normal case with* $g(x) = |x|$*. We need to discuss whether* $\theta = 0$ *as well, for Delta Theorem requests* g *be once differentiable at* θ*.*

When $\theta = 0$*, we can determine the limit behaviour of* $|T_n|$ *directly.*

$$\begin{aligned} P(\sqrt{n}|T_n| < a) &= P(-a < \sqrt{n}T_n < a) \\ &\to \phi(\frac{a}{\sigma}) - \phi(-\frac{a}{\sigma}) \end{aligned}$$

2.2 Higher-order expansions

Using Taylor expansion, we can easily show that:

Theorem 2.2.1. $\sqrt{n}(T_n - \theta) = O_p(1)$ *and* g *is differentiable* k *at* θ *with* $g^{(k)}(\theta) \neq 0$ *but* $g^{(j)}(\theta) = 0 \ for \ j < k$ *Then:*

$$g(T_n) = g(\theta) + \frac{(T_n - \theta)^k}{k!}g^{(k)}(\theta) + o_p((T_n - \theta)^k)$$

$$(\sqrt{n})^k[g(T_n) - g(\theta)] = \frac{1}{k!}[g^{(k)}(\theta)][\sqrt{n}(T_n - \theta)]^k + o_p(1)$$

Example 2.2.1. *(i)Example 2.1.1 revisited, for* $\mu = 0, n\overline{X}_n^2/\sigma^2 \xrightarrow{d} \frac{1}{2} \cdot 2 \cdot [N(0,1)]^2$

(ii)If $\sqrt{n}\overline{X}_n \xrightarrow{d} N(0,1)$*, then* $-2n(\cos \overline{X}_n - 1) \xrightarrow{d} \chi_1^2$

2.3 Multivariate version of delta theorem

Theorem 2.3.1. *$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta}) \xrightarrow{d} N_k(\mathbf{0}, \boldsymbol{\Sigma(\theta)})$ and g be once differentiable at $\boldsymbol{\theta}$ with $\nabla g(\boldsymbol{\theta})$*

Then:

$$\sqrt{n}(g(\boldsymbol{T}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} N_m(\mathbf{0}, \nabla^T g(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\nabla g(\boldsymbol{\theta}))$$

It can be shown by Cramer-Wold device, which can derive multivariate distribution by turning it into univariate distribution.

Or just make the first-order Taylor expansion:

$$g(\boldsymbol{T}_n) = g(\boldsymbol{\theta}) + \nabla^T g(\boldsymbol{\theta})(\boldsymbol{T}_n - \boldsymbol{\theta}) + o_p(\|\boldsymbol{T}_n - \boldsymbol{\theta}\|).$$

with Corollary 1.2.1 and we can simply obtain the AS distribution.

Example 2.3.1 (Sample variance revisited(Example 1.3.2)). *iid with finite fourth moments. Taking:*

$$\boldsymbol{T}_n = (\overline{X}_n, \overline{X_n^2}), \boldsymbol{\theta} = (EX_1, EX_1^2)^T, \boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_1^2) \\ * & \mathrm{Var}(X_1^2) \end{pmatrix}$$

Then we can use multivariate CLT. Taking $g(u, v) = v - u^2$, we may as well assume $\mu = 0$(or equivalently working with $X_i - \mu$) for the sample variance does not depend on location.

Finally we can see that: $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4)$. After another univariate Delta theorem, we can derive the AS distribution of S_n

Example 2.3.2 (The joint limit distribution). *(i)By using multivariate Delta theorem like the example above, we can get:*

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} \xrightarrow{d} N_x\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{pmatrix}\right)$$

When $\mu_3 = 0$(normal distribution for example), we say they are AS independent.

(ii)After gaining the joint AS distribution of $(\overline{X}_n, S_n^2)$, we can use Multivariate delta th w.r.t. $(\overline{X}_n, S_n^2)$ to obtain the AS distribution of $(S_n^2, \frac{\overline{X}_n}{S_n})$

2.4 Variance-stabilizing transformations(VST)

From the delta theorem, we have: If $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$, then:

$$\sqrt{n}(g(T_n) - g(\theta)) \xrightarrow{d} N(0, [g'(\theta)]^2\sigma^2(\theta))$$

If we want the variance in the asymptotic distribution of $g(T_n)$ to be constant, we set:

$$[g'(\theta)]^2\sigma^2(\theta) = k^2$$

for some constant k. Therefore, we choose the function g as:

$$g(\theta) = k \int \frac{1}{\sigma(\theta)} d\theta$$

Example 2.4.1. *iid from Poisson, choose* $g(\theta) = \int \frac{k}{\sqrt{\theta}} d\theta = 2k\sqrt{\theta} = \sqrt{\theta}$

Then we can calculate the Wald confidence interval for θ

Example 2.4.2 (Sample correlation revisited)**.** *By taking:* $\boldsymbol{T}_n = \left(\overline{X}_n, \overline{Y}_n, \frac{1}{n}\sum_{i=1}^n X_i^2, \frac{1}{n}\sum_{i=1}^n Y_i^2, \frac{1}{n}\sum_{i=1}^n X_iY_i\right)$ *we can derive the AS distribution of the sample correlation* r_n *by multivariate delta theorem. If* (X_i, Y_i) *are iid bivariate Gaussian, then:*

$$\sqrt{n}(r_n - \rho) \xrightarrow{d} N(0, (1-\rho^2)^2)$$

Now we can use VST to calculate the confidence interval of ρ*.*

$$g(\rho) = \int \frac{1}{1-\rho^2} d\rho = \frac{1}{2}\log\frac{1+\rho}{1-\rho} = \operatorname{arctanh}(\rho)$$

which is known as Fisher's z.

2.5 Approximation of moments

The core of this section is Taylor expansion as well.

Theorem 2.5.1. $X_n \xrightarrow{d} X$ *for some* X + $\sup_n E|X_n|^{k+\delta} < \infty$ *for some* $\delta > 0 \Rightarrow E(X_n^r) \to E(X^r), \forall 1 \leq r \leq k$

Theorem 2.5.2 (von Bahr)**.** *iid and finite variance, and for some k,* $E|X_1|^k < \infty$*. Suppose* $Z \sim N(0,1)$*, then:*

$$E(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma})^r = E(Z^r) + O(\frac{1}{\sqrt{n}}), \forall r \leq k$$

Proposition 2.5.1. *iid, finite fourth moment, g has four uniformly bounded derivatives.*

$$(i) E(g(\overline{X}_n)) = g(\mu) + \frac{g''(\mu)\sigma^2}{2n} + O(n^{-2})$$

$$(ii) \operatorname{Var}(g(\overline{X}_n)) = \frac{(g'(\mu))^2\sigma^2}{n} + O(n^{-2})$$

we can simply prove them by Taylor expansion w.r.t. $g(\overline{X}_n)$ *at* μ

Example 2.5.1 (去年考了)**.** *iid, Poisson(*μ*) and wish to estimate* $P(X_1 = 0) = e^{-\mu}$*.*

See the MLE $e^{-\overline{X}_n}$*, we estimate its MSE, so we should estimate its expectation and variance using the proposition above. We have:* $\operatorname{Bias}(e^{-\overline{X}_n}) = \frac{\mu e^{-\mu}}{2n} + O(n^{-2})$*. So* $\operatorname{MSE}(e^{-\overline{X}_n}) = \frac{\mu e^{-2\mu}}{n} + O(n^{-2})$

See the sign statistic $T = \frac{1}{n}\sum_{i=1}^n I(X_i = 0)$*, which is unbiased.*

For:

$$I(X_i = 0)I(X_j = 0) \geq I^2(X_i = 0) = I(X_i = 0)(i \neq j)$$

Then:

$$\mathrm{MSE}(T) = \mathrm{Var}(T) = ET^2 \geq \frac{1}{n^2} \cdot n^2 \cdot E(I(X_1 = 0)) = P(X_1 = 0) = e^{-\mu}$$

Now we can compare these two statistics.

2.6 Multivariate-version Edgeworth expansion

3 The basic sample statistics

P55 empirical cumulative distribution function: ECDF

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leq x\}}$$

3.1 The sample distribution

3.1.1 Basic properties

Proposition 3.1.1. *Note that*

$$nF_n(x) \sim \mathrm{BIN}(F(x), n)$$

then we have: AS $\xrightarrow{wp1}, \xrightarrow{2nd}$ *and AS distribution.*

3.1.2 Kolmogorov-Smirnov distance

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|.$$

can be denoted as $\|F_n(x) - F(x)\|_\infty$

Theorem 3.1.1 (DKW's inequality)**.** *F_n is the ECDF, then exists a positive constant C s.t.*

$$P(D_n > z) \leq Ce^{-2nz^2}, z > 0, \forall n = 1, 2, \ldots,$$

Or:

$$P(\sqrt{n} D_n > z) \leq Ce^{-2z^2}, z > 0, \forall n = 1, 2, \ldots,$$

which clearly demonstrate that:

$$\sqrt{n} D_n = O_p(1)$$

Theorem 3.1.2 (Massart)**.** *If $nz^2 \geq \frac{\log 2}{2}$, then C can equal to 2.*

Corollary 3.1.1. *Let $h_\epsilon = e^{-2\epsilon^2}, P(\sup_{m \geq n} D_m > \epsilon) \leq \frac{C}{1-h_\epsilon} h_\epsilon^n$*

Using: $\cap \leq \sum$

Theorem 3.1.3 (Glivenko-Cantelli)**.**

$$D_n \xrightarrow{wp1} 0$$

Theorem 3.1.4 (Kolmogorov). *Let F be continuous, then:*

$$\lim_{n\to\infty} P(\sqrt{n}D_n \le z) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j+1}e^{-2j^2z^2}, z > 0$$

This is the null AS distribution of K-S statistics! No need to remember.

Proposition 3.1.2. *Let F be continuous, then $\sqrt{n}D_n$ is distribution-free w.r.t. F.*

proof: choose $x_i's$ s.t. $F(x_i) = \frac{i}{n}, i = 1, \ldots, n$. Then $\sup_{x\in(x_{i-1},x_i)} |F_n(x) - F(x)| = \max\{\frac{i}{n} - F(x), F(x) - \frac{i-1}{n}\}$, so we have: $\sqrt{n}d_n \overset{\mathrm{d}}{=} \sqrt{n}\max_{0\le i\le n}\max(\frac{i}{n} - U_{(i)}, U_{(i)} - \frac{i-1}{n})$, where $U_{(1)} \le \ldots \le U_{(n)}$ are order statistics of an indep sample from U(0,1)

A consequence: given $\alpha \in (0, 1)$, there is a well-defined $d = d_{\alpha,n}$ such that, for any continuous CDF F, $P_F(\sqrt{n}D_n > d) = \alpha$

Example 3.1.1 (Kolmogorov-Smirnov confidence intervals). *We choose a $d = d_{\alpha,n}$ like above, then:*

$$\begin{aligned} 1 - \alpha &= P_F(\sqrt{n}D_n \le d) = P_F(\sqrt{n}\|F_n - F\|_\infty \le d) \\ &= P_F(|F_n - F| \le \frac{d}{\sqrt{n}}, \forall x) \\ &= P_F(F_n(x) - \frac{d}{\sqrt{n}} \le F(x) \le F_n(x) + \frac{d}{\sqrt{n}}, \forall x) \end{aligned}$$

3.1.3 Applications: Kolmogorov-Smirnov and other ECDF-based GOF(goodness-of-fit)tests

D_n, C_n, A_n, for Sparse, Dense respectively.

Proposition 3.1.3. *C_n, A_n are both distribution-free*

Power to 1.

consider a CDF F_1, so that there exists η such that $F_1(\eta) \ne F_0(\eta)$. Let $G_n^{-1}(1 - \alpha)$ denote the $(1-\alpha)$th quantile of the distribution of $\sqrt{n}D_n$ under F_0 and by Th 3.1.4 it is an O(1). Note that under F_1, there is $\sqrt{n}(F_n(\eta) - F_1(\eta)) = O_p(1)$ We write the D_n by definition and plug in the $F_1(t)$, then goes to η. We have:

$$P_{F_1}(\sqrt{n}D_n > G_n^{-1}(1 - \alpha)) \ge P_{F_1}(|\sqrt{n}(F_n(\eta) - F_1(\eta)) + \sqrt{n}(F_1(\eta) - F_0(\eta))| > G_n^{-1}(1 - \alpha)) \to 1$$

Example 3.1.2 (The Berk-Jones procedure).

$$H_0 : F = F_0$$

For any given x, we have: $nF_n(x) \sim \mathrm{Bin}(n, F(x))$, if we write p for F(x) and p_0 for $F_0(x)$, that's to test $p = p_0$. The likelihood maximized at $F(x) = F_n(x)$. So the likelihood ratio statistic is:

$$\lambda_n(x) = \frac{F_n(x)^{nF_n(x)}(1 - F_n(x))^{n-nF_n(x)}}{F_0(x)^{nF_0(x)}(1 - F_0(x))^{n-nF_0(x)}}$$

So the Berk-Jones statistic is: $R_n = n^{-1}\sup_x \log\lambda_n(x)$

Example 3.1.3 (The two-sample case)**.** *anagolous to one-sample case, we have* $D_{m,n}, A_{m,n}$

3.1.4 The Chi-square test

Suppose $X_1,\dots,X_n$ are iid from F, and $F = F_0$, F_0 being completely specified. Let S be the support of F_0, and given k, $A_{ki}, i = 1,\dots,k$ be a partition of S. Let $p_{0i} = P_{F_0}(A_{ki})$ and $n_i = \#\{j : x_j \in A_{ki}\}$, i.e. the observed frequency of the partition set A_{ki}. Therefore, under $H_0, E(n_i) = np_{0i}$, with which we compare n_i.

$$K^2 = \sum_{i=1}^{k}\frac{(n_i - np_{0i})^2}{np_{0i}}$$

Theorem 3.1.5 (The asymptotic null distribution)**.** *Under* $H_0, K^2 \xrightarrow{d} \chi^2_{k-1}$

This proof is essential, on P62.

$\boldsymbol{n} = (n_1,\dots,n_k)^T = \sum_{i=1}^n \boldsymbol{Z}_i$ is iid. So by multivariate CLT, there is $\frac{\boldsymbol{n}-n\boldsymbol{p}_0}{\sqrt{n}} \xrightarrow{d} N_k(\boldsymbol{0}, diag(\boldsymbol{p}_0) - \boldsymbol{p}_0\boldsymbol{p}_0^T)$

And $K^2 = \boldsymbol{Y} = diag^{-1}(\mu)\frac{\boldsymbol{n}-n\boldsymbol{p}_0}{\sqrt{n}}$, where $\mu = (\sqrt{p_{01}},\dots,\sqrt{p_{0k}})^T$, and $\boldsymbol{Y} \overset{\mathrm{d}}{=} N_k(\boldsymbol{0},\boldsymbol{\Sigma})$, where $\Sigma = \boldsymbol{I}_k - \boldsymbol{\mu}\boldsymbol{\mu}^T$

The following is an important proof for $K^2 = \boldsymbol{Y}^T\boldsymbol{Y} \xrightarrow{d} \chi^2_{k-1}$.

For $\mathrm{tr}(\boldsymbol{\Sigma}) = k-1$, and $\boldsymbol{\Sigma}$ is symmetric and idempotent, there exists an orthogonal matrix $\boldsymbol{P}$ s.t. $\boldsymbol{P}^T\boldsymbol{\Sigma}\boldsymbol{P} = diag(\lambda_1,\dots,\lambda_k)$, where there is only one 0 in $\lambda_i's$, the others are all 1. Let $\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{X}$, we have $K^2 = \boldsymbol{Y}^T\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{X} = \sum_{i=1}^k \lambda_i w_i$ with $w_i \overset{\mathrm{iid}}{\sim} \chi^2_1$, namely $K^2 \xrightarrow{d} \chi^2_{k-1}$

Example 3.1.4 (The Hellinger statistic)**.** *For* $\sqrt{n}(\boldsymbol{g}(\overline{\boldsymbol{Z}}_n) - \boldsymbol{g}(\boldsymbol{p}_0)) \overset{d}{=} \sqrt{n}\nabla\boldsymbol{g}(\boldsymbol{p}_0)(\overline{\boldsymbol{Z}}_n - \boldsymbol{p}_0)$*, so that in Pearson's* χ^2*, we may replace* $\sqrt{n}(\overline{\boldsymbol{Z}}_n - \boldsymbol{p}_0)$ *by* $\sqrt{n}\nabla^{-1}\boldsymbol{g}(\boldsymbol{p}_0)(\boldsymbol{g}(\overline{\boldsymbol{Z}}_n) - \boldsymbol{g}(\boldsymbol{p}_0))$*. When we take* $\boldsymbol{g}(\boldsymbol{x}) = (\sqrt{x_1},\sqrt{x_k})^T$*, we have the Hellinger statistic.*

See F_1 for Pearson's χ^2.

Proposition 3.1.4. *Under* F_1*, there is* $\frac{n_i}{n} \xrightarrow{p} p_{1i}$*, so according to CMT, we have* $\frac{K^2}{n} \xrightarrow{p} \sum_{i=1}^k \frac{(p_{1i}-p_{0i})^2}{p_{0i}}$

Local alternatives:

$$p_{1i} = p_{0i} + \delta_i n^{-1/2}$$

where:

$$\boldsymbol{1}^T\boldsymbol{\delta} = \sum_i \delta_i = 0$$

Theorem 3.1.6 (The asymptotic alternative distribution)**.** *Under* H_1*, say* $\boldsymbol{p} = \boldsymbol{p}_1 = \boldsymbol{p}_0 + \boldsymbol{\delta}n^{-1/2}$*. Then* $K^2 \xrightarrow{d} \chi^2_{k-1}(\lambda)$*, where* $\lambda = \sum_{i=1}^k \frac{\delta_i^2}{p_{0i}}$ *is the noncentrality parameter.*

See homework for a proof.

3.2 The sample moments

We define kth moment and central moment of F as:

$$\alpha_k = \int_{-\infty}^{\infty} x_k dF(x) = EX_1^k$$
$$\mu_k = \int_{-\infty}^{\infty} (x-\alpha_1)^k dF(x) = E[(X_1-\alpha_1)^k]$$

The sample moments:

$$a_k = \int_{-\infty}^{\infty} x_k dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

$$m_k = \int_{-\infty}^{\infty} (x-\alpha_1)^k dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} (X_i - a_1)^k$$

Proposition 3.2.1. $a_k \xrightarrow{wp1} \alpha_k; E(a_k) = \alpha_k; \mathrm{Var}(a_k) = \frac{\alpha_{2k}-\alpha_k^2}{n}$

Proposition 3.2.2. $\sqrt{n}(a_1-\alpha_1,\ldots,a_k-\alpha_k)^T$ *is* $\mathrm{AN}_k(\mathbf{0},\mathbf{\Sigma})$*, where* $\mathbf{\Sigma} = (\sigma_{ij})_{k\times k}$ *with* $\sigma_{ij} = \alpha_{i+j} - \alpha_i\alpha_j$.

Proposition 3.2.3. *The same as 3.2.2. It's the properties for* $b_k = \frac{1}{n}\sum_{i=1}^{n}(X_i-\alpha_1)^k$

Proposition 3.2.4. *Properties for* μ_k*, we find the connection between* m_k *and* $b_j's$ *using the binomial theorem. And use CMT and multivariate Delta Th.*

3.3 The sample quantiles

Definition of quantile:

$$F^{-1}(p) \equiv \xi_p = \inf\{x : F(x) \geq p\}.$$

$$F_n^{-1}(p) \equiv \hat{\xi}_p = \inf\{x : F_n(x) \geq p\}$$

3.3.1 Basic results

Theorem 3.3.1.

$$P(|\hat{\xi}_p - \xi_p| > \epsilon) \leq 2Ce^{-2n\delta_\epsilon^2}$$

using DKW's inequality to prove it. We can get strong consistency from it as well.

For $P(\hat{\xi}_p \leq t) = P(F_n(t) \geq p)$, and $nF_n(t)$ is binomial, so we can get the distribution of $\hat{\xi}_p$

Theorem 3.3.2. *The AS distribution of* $\sqrt{n}(\hat{\xi}_p - \xi_p)$ *exists when F is continuous at* ξ_p*. When* $F'(\xi_p) > 0$ *exists, we can get the conclusion we derive in the Bahadur.*

proof is omited.

3.3.2 Bahadur's representation

Theorem 3.3.3 (Bahadur's representation)**.** *Let $X_1, \ldots, X_n$ be iid random variables from a CDF F. Suppose that $F'(\xi_p)$ exists and is positive. Then*

$$\hat{\xi}_p = \xi_p + \frac{F(\xi_p) - F_n(\xi_p)}{F'(\xi_p)} + o_p(n^{-1/2})$$

This proof in at P71, representing a pattern for proof. The lemma below is used.

Lemma 3.3.1. *X_n bounded in probability+$\lim_n[P(X_n \le t, Y_n \ge t+\epsilon) + P(X_n \ge t+\epsilon, Y_n \le t)] = 0 \Rightarrow X_n \xrightarrow{p} Y_n$*

plug in $|X_n - Y_n| \le \ge \epsilon$

Corollary 3.3.1.

$$\sqrt{n}[(\hat{\xi}_{p1}, \ldots, \hat{\xi}_{pm}) - (\xi_{p1}, \ldots, \xi_{pm})] \xrightarrow{d} N_m(0, \boldsymbol{D})$$

where,

$$D_{ij} = p_i(1-p_j)/[F'(\xi_{pi})F'(\xi_{pj})]$$

The proof is important, see the homework for the idea of the proof.

Example 3.3.1 (Interquartile range; IQR)**.** *IQR=$\hat{\xi}_{0.75} - \hat{\xi}_{0.25}$, we can derive the AS distribution using the idea of the corollary above. When the sample comes from Gaussian distribution, we can use IQR/1.35 to estimate σ.*

Example 3.3.2 (Gastwirth estimate)**.** *$F(x-\mu), F(-x) = 1 - F(x)$, wish to estimate μ. One idea is to use L-statistics, i.e. a convex combination of order statistics. Gastwirth is: $0.3X_{(n/3)} + 0.4X_{(n/2)} + 0.3X_{(2n/3)}$*

Corollary 3.3.2. $\sqrt{n}(\overline{X}_n - \mu, F_n^{-1}(p) - \xi_p) \xrightarrow{p} N_2(\mathbf{0}, \mathbf{\Sigma})$

Vital. See homework for the proof.

Hint: $-\sqrt{2\pi}\int_{-\infty}^{0} x\phi(x)dx = 1$

3.3.3 Confidence intervals for quantiles

Hint:

$$(F(X_{(1)}, F_{X_{(2)}}), \ldots, F(X_{(n)})) \stackrel{\mathrm{d}}{=} (U_{(1)}, U_{(2)}, \ldots, U_{(n)})$$

We want to derive a distribution-free confidence for ξ_p

We can derive the AS distribution of $\sqrt{n}\hat{\xi}_p$ and then the wald CI of ξ_p, but it is not distribution-free.

One idea is to involve the $X_{(i)}$ in.

Theorem 3.3.4. *Bahadur's condition+integer sequence $k_n + 1 \le k_n \le n$" + "$k_n/n = p + cn^{-1/2} + o(n^{-1/2})$ with a constant c. Then, $\sqrt{n}(X_{(k_n)} - \hat{\xi}_p) \xrightarrow{p} \frac{c}{F'(\xi_p)}$*

Similar proof as Theorem 3.3.3.

Corollary 3.3.3. *k_{1n} and k_{2n} satisfy the condition above with $c{=}{-}z_{\alpha/2}\sqrt{p(1-p)}$ and $z_{\alpha/2}\sqrt{p(1-p)}$ respectively. Then:*

$$P_F(X_{(k_{1n})} \le \xi_p \le X_{(k_{2n})}) = P_F(F(X_{(k_{1n})}) \le p \le F(X_{(k_{2n})})) = P(U_{(k_{1n})} \le p \le U_{(k_{2n})}) \to \alpha$$

3.3.4 Quantile regression

fit along the pth quantile.

4 Asymptotics in parametric inference

4.1 Asymptotic efficient estimation

$$(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \sim AN_k(\mathbf{0}, \boldsymbol{V}_n(\boldsymbol{\theta}))$$

If $\boldsymbol{V}_{1n}(\boldsymbol{\theta}) \le \boldsymbol{V}_{2n}(\boldsymbol{\theta})$, say $\hat{\boldsymbol{\theta}}_{1n}$ is asymptotically more efficient than $\hat{\boldsymbol{\theta}}_{2n}$

Definition 4.1.1. *Assume the Fisher information matrix:*

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = E\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\sum_i \log f_{\boldsymbol{\theta}}(X_i)[\frac{\partial}{\partial\boldsymbol{\theta}}\sum_i \log f_{\boldsymbol{\theta}}(X_i)]^T\right\}$$

is well defined and positive definite for every n. A sequence of estimators $\hat{\boldsymbol{\theta}}_n$ satisfying the one above is sard to be asymptotically efficient iff $\boldsymbol{V}_n(\boldsymbol{\theta}) = [\boldsymbol{I}_n(\boldsymbol{\theta})]^{-1}$

When $\boldsymbol{\beta} = g(\boldsymbol{\theta})$, we deal with it by Delta theorem.

The information inequality is right under regularity conditions, and MLE asymptotically reaches CR bounds. A famous example can be seen at P81 to P82.

4.2 Maximum likelihood estimation

Let $\boldsymbol{X} = \{X_1, \dots, X_n\}$ be iid with distribution $F_{\boldsymbol{\theta}}$ belonging to a family $F = \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^T\}$. Then the likelihood function is:

$$L(\boldsymbol{\theta}; \boldsymbol{X}) = \prod_{i=1}^n f_{\boldsymbol{\theta}}(X_i).$$

And the MLE is given by: $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \log L(\boldsymbol{\theta}; \boldsymbol{X})$. We say MLE must be a root of the likelihood equation(RLE):

$$\frac{\partial \log L}{\partial \theta_i}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$$

But RLE may not be an MLE. We focus on the case of k=1(one-dimension) for simplicity.

Regularity condition at P83

Theorem 4.2.1. *Assume regularity conditions on the family F. Consider iid observations on F_{θ_0}, with $\theta_0 \in \Theta$. Then with probability 1, the likelihood equations admit a sequence of solutions $\{\hat{\theta}_n\}$ satisfying:*

(i) strong consistency: $\hat{\theta}_n \to \theta_0$

(ii)asymptotic normality and efficiency: $\hat{\theta}_n$ is $AN(\theta_0, [nI(\theta_0)]^{-1})$

We first write the important definition and properties.

Definition of score function:

$$s(\boldsymbol{X}, \theta) = \frac{1}{n}\frac{\partial \log L(\theta; \boldsymbol{X})}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f_\theta(X_i)}{\partial \theta}$$

By CLT(it's iid sum), we have:

$$\sqrt{n}s(\boldsymbol{X}, \theta_0) \xrightarrow{d} N(0, I(\theta_0))$$

for

$$E_{\theta_0}[s(\boldsymbol{X}, \theta_0)] = E_{\theta_0}[\frac{\partial \log f_{\theta_0}(X_1)}{\partial \theta_0}] = \int \frac{1}{f_{\theta_0}(x)}\frac{\partial f_{\theta_0}(x)}{\partial \theta_0} f_{\theta_0}(x)dx = 0$$

$$\mathrm{Var}(\frac{\partial \log f_{\theta_0}(X_1)}{\partial \theta_0}) = E_{\theta_0}(\frac{\partial \log f_{\theta_0}(X_1)}{\partial \theta_0})^2 = -E_{\theta_0}(\frac{\partial^2 \log f_{\theta_0}(X_1)}{\partial \theta_0^2}) = -E_{\theta_0}[s'(\boldsymbol{X}, \theta_0)] = I(\theta_0)$$

where we can prove the essential ineq by:

$$E_\theta(\frac{\partial^2 \log f_{\theta_0}(x)}{\partial \theta_0^2}) = E_\theta(\frac{\partial}{\partial \theta}(\frac{1}{f_\theta(x)}\frac{\partial f_\theta(x)}{\partial \theta}))$$

Note that according to regularity condition, we have:

$$|s''(\boldsymbol{X}, \theta)| \leq \frac{1}{n}\sum_{i=1}^{n}|\frac{\partial^3 \log f_\theta(X_i)}{\partial \theta^3}| \leq \frac{1}{n}\sum_{i=1}^{n}|H(X_i)| \equiv \overline{H}(\boldsymbol{X})$$

Consider Taylor's expandion to the score function:

$$s(\boldsymbol{X}, \hat{\theta}_n) = s(\boldsymbol{X}, \hat{\theta}_n) = s(\boldsymbol{X}, \theta_0) + s'(\boldsymbol{X}, \theta_0)(\hat{\theta}_n - \theta_0) + \frac{1}{2}\overline{H}(\boldsymbol{X})\eta^*(\hat{\theta}_n - \theta_0)^2,$$

where $|\eta^*| = \frac{|s''(\boldsymbol{X},\xi)|}{\overline{H}(\boldsymbol{X})} \leq 1$. Thus:

$$\sqrt{n}s(\boldsymbol{X}, \theta_0) = \sqrt{n}(\hat{\theta}_n - \theta_0)\left(-s'(\boldsymbol{X}, \theta_0) - \frac{1}{2}\overline{H}(\boldsymbol{X})\eta^*(\hat{\theta}_n - \theta_0)\right)$$

for $(\hat{\theta}_n - \theta_0) \xrightarrow{wp1} 0$ and the properties above, we can use Slutsky Theorem and get:

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

4.3 Improving the sub-efficient estimates

Method of moments may not be asymptotically efficient, while sometimes $s(\boldsymbol{X}, \theta) = 0$ is difficult to solve. We consider Newton's method using the estimates based on the method of moments or sample quantiles as the initial guess.

Generally, we use:

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - [s'(\boldsymbol{X}, \hat{\theta}^{(k)})]^{-1} s(\boldsymbol{X}, \hat{\theta}^{(k)}), k = 0, 1, 2, \ldots$$

If Fisher information is available, then $s'(\boldsymbol{X}, \hat{\theta}^{(k)}) \xrightarrow{wp1} -I(\theta^0)$, then we can also use:

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + [I(\hat{\theta}^{(k)})]^{-1} s(\boldsymbol{X}, \hat{\theta}^{(k)}), k = 0, 1, 2, \ldots$$

This method is the method of scoring. The scores, $[I(\hat{\theta}^{(k)})]^{-1} s(\boldsymbol{X}, \hat{\theta}^{(k)})$ are increments added to an estimate to improve it.

Example 4.3.1 (Logistic distribution)**.** *Omited*

One-step MLE:

$$\hat{\theta}^{(1)} = \hat{\theta}^{(0)} - [s'(\boldsymbol{X}, \hat{\theta}^{(0)})]^{-1} s(\boldsymbol{X}, \hat{\theta}^{(0)})$$

Theorem 4.3.1. *The conditions in Theorem 4.2.1 hold and that $\hat{\theta}^{(0)}$ is $\sqrt{n}$-consistent for θ. Then:*

(i)The one-step MLE is asymptotically efficient

(ii)The one-step MLE obtained by replacing $s'(\boldsymbol{X}, \hat{\theta}^{(0)})$ with its expected value, $-I(\hat{\theta}^{(0)})$, is asymptotically efficient.

We mainly want to study that $\sqrt{n}(\hat{\theta}^{(1)} - \hat{\theta}_n) \xrightarrow{p} 0$. By definition, we have $(\hat{\theta}^{(1)} - \hat{\theta}_n) = (\hat{\theta}^{(0)} - \hat{\theta}_n) - [s'(\hat{\theta}^{(0)})]^{-1} s(\hat{\theta}^{(0)})$ with $s(\hat{\theta}_n) = 0$. We conduct taylor expansion to $s(\hat{\theta}^{(0)})$ at $\hat{\theta}_n$, we can get a approximation of $(\hat{\theta}^{(0)} - \hat{\theta}_n)$. Prove that $\sqrt{n}(\hat{\theta}^{(0)} - \hat{\theta}_n) = O_p(1)$ and we get the result.

4.4 Hypothesis testing by likelihood method

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \text{ versus } H_1 : \boldsymbol{\theta} \in \Theta_1$$

The likelihood ratio test(LRT) rejects H_0 for small values of: $\Lambda_n = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta}; \boldsymbol{X})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \boldsymbol{X})}$ and $\lambda_n = -2 \log \Lambda_n$ is often used.

Example 4.4.1. *We can test t-test is LRT equivalently.*

If we have r constraints on $\boldsymbol{\theta}$ and only k-r components of it are free to change, which can be denoted as $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_{k-r})$ without loss of generality. We can write that:

$$H_0 : \boldsymbol{\theta} = g(\boldsymbol{\vartheta})$$

Theorem 4.4.1 (Wilks). *Assume the conditions in Theorem 4.2.1 hold, then under $H_0, \lambda_n \xrightarrow{d} \chi_r^2$*

We conduct Taylor expansion to $2\log L(\hat{\boldsymbol{\theta}}_n)$ at $\boldsymbol{\theta}_0$, and using the Taylor expansion in the proof of Theorem 4.2.1 to approximate $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$. The operation before is conduct to the case under H_0 as well.

Finally we get $\lambda_n = n[s(g(\boldsymbol{\vartheta}_0))]^T([\boldsymbol{I}(g(\boldsymbol{\vartheta}))]^{-1} - [\boldsymbol{D}(\boldsymbol{\vartheta})]^T[\tilde{\boldsymbol{I}}(\boldsymbol{\vartheta})]^{-1}\boldsymbol{D}(\boldsymbol{\vartheta}))s(g(\boldsymbol{\vartheta}_0)) + o_p(1)$. Note the $s(\boldsymbol{\theta}_0)$ is asymptotically normal by CLT. So we need only to check the matrix in the middle is idempotent and symmetrix with trace=r. Then by proof of Theorem 3.1.5, we can get the result.

Example 4.4.2. *Under H_1, the λ_n in the last example is to infinity.*

4.5 The Wald and Rao score tests

$$H_0 : R(\boldsymbol{\theta}) = 0$$

The Wald test:

$$W_n = [R(\hat{\boldsymbol{\theta}}_n)]^T \left\{[C(\hat{\boldsymbol{\theta}}_n)]^T[\boldsymbol{I}_n(\hat{\boldsymbol{\theta}}_n)]^{-1}C(\hat{\boldsymbol{\theta}}_n)^{-1}\right\} R(\hat{\boldsymbol{\theta}}_n)$$

where $\hat{\boldsymbol{\theta}}_n$ is an MLE or RLE of θ, $C(\boldsymbol{\theta}) = \frac{\partial R(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$.

The Rao score test:

$$R_n = n[s(\tilde{\boldsymbol{\theta}}_n)]^T[\boldsymbol{I}(\tilde{\boldsymbol{\theta}}_n)]^{-1}s(\tilde{\boldsymbol{\theta}}_n)$$

Theorem 4.5.1. *Assume the conditions in Theorem 4.2.1 hold. Under H_0 we have:*
(i)$W_n \xrightarrow{d} \chi_r^2$
(ii)$R_n \xrightarrow{d} \chi_r^2$

(i) We know the asymptotic distribution of the MLE by theorem 4.2.1, and with Delta theorem and Slutsky theorem, we can simply get the result. Note that under $H_0, R(\boldsymbol{\theta}_0) = 0$.

(ii) By Lagrange multipliers and Taylor expansion on $R(\tilde{\boldsymbol{\theta}}_n)$ at $\boldsymbol{\theta}_0$ and $s(\tilde{\boldsymbol{\theta}}_n)$ at $\boldsymbol{\theta}_0$, where $\tilde{\boldsymbol{\theta}}_n$ is from the Lagrange. see P96 and P96-.

Thus Wald's test, Rao's test and LRT are asymptotically equivalent. Wald needs $\hat{\boldsymbol{\theta}}_n$, Rao needs $\tilde{\boldsymbol{\theta}}_n$, and LRT needs both.

4.6 Confidence sets based on likelihoods

Suppose $A(\boldsymbol{\theta}_0)$ is the acceptance region of a size α test for H_0 and then $C(\boldsymbol{X}) = \{\boldsymbol{\theta} : \boldsymbol{X} \in A(\boldsymbol{\theta})\}$ is a $1-\alpha$ asymptotically correct confidence set for $\boldsymbol{\theta}$ due to $P_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0 \in C(x)) = 1 - \alpha$.

5 Asymptotics in nonparametric inference

5.1 Sign test (Fisher)

5.1.1 Test procedure

$$H_0 : \theta = 0 \text{ versus } H_1 : \theta > 0$$

or

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta > \theta_0$$

The sign test is a test for the median θ . Sign test:

$$S_n = \sum_{i=1}^{n} I(X_i > \theta_0)$$

Large value of S_n leads to reject H_0. And we have:

$$S_n \overset{H_{kjkj0}}{\sim} \mathrm{BIN}(n, \frac{1}{2})$$

And the p-value is $P(\mathrm{BIN}(n, \frac{1}{2}) \geq S_n)$

In the case of $\theta_0 = 0$, we can derive the power of S_n by $\frac{S_n - np_\theta}{[np_\theta(1-p_\theta)]_{1/2}} = \frac{S_n - n(1-F(0))}{[n(1-F(0))(F(0))]^{1/2}}$

5.1.2 Asymptotic Properties

Definition 5.1.1 (Consistency). *Let $\{\phi_n\}$ be a sequence of tests for $H_0 : F \in \Omega_0$ versus $H_1 : F \in \Omega_1$. Then ϕ_n is consistent against the alternatives Ω_1 if*

(i) $E_F(\phi_n) \to \alpha \in (0, 1), \forall \in \Omega_0$

(ii) $E_F(\phi_n) \to 1, \forall F \in \Omega_1$

Example 5.1.1. *$X_1, \ldots, X_n$ is an iid sample from* Cauchy$(\theta, 1)$*, then $\overline{X}_n$ obeys* Cauchy$(\theta, 1)$*. Let k be the αth quantile of the* Cauchy$(0, 1)$*, then we can derive the power of the test for the case $\theta_0 = 0$.*

$$P_\theta(\overline{X}_n > k) = P(C(\theta, 1) > k) = P(\theta + C(0, 1) > k) = P(C(0, 1) > k - \theta)$$

which is a fixed number unrelated to n.

Theorem 5.1.1. *If F is a continuous CDF with unique median θ, then the sign test is consistent for tests on θ.*

Note that $\frac{1}{n}S_n - p_\theta \xrightarrow{p} 0$ under any F. We choose the critical value under H_0 as $k_n = \frac{n}{2} + z_\alpha\sqrt{\frac{n}{4}}$. So under H_1, it follows that $\frac{1}{n}k_n - p_\theta < 0$ for large n ($p_\theta = P_\theta(X_1 > \theta_0) > \frac{1}{2}$)

If $\sqrt{n}(\hat{\theta}_{1n} - \theta) \xrightarrow{d} N(0, \sigma_1^2(\theta))$ and $\sqrt{n}(\hat{\theta}_{2n} - \theta) \xrightarrow{d} N(0, \sigma_2^2(\theta))$, then the asymptotic relative efficiency (ARE) of $\hat{\theta}_{2n}$ with respect to $\hat{\theta}_{1n}$ is $\frac{\sigma_1^2(\theta)}{\sigma_2^2(\theta)}$

The idea of comparing two tests: the threshold sample size to reach the size and power. Suppose we use statistics T s.t. large values of them correspond to rejection of H_0, α, β is the type one error and the power of the test. Suppose $n(\alpha, \beta, \theta, T)$ is the smallest sample size s.t.

$$P_{\theta_0}(T_n \geq c_n) \leq \alpha, P_{T_n \geq c_n} \geq \beta$$

Then two tests based on statistics T_{1n} and T_{2n} can be compared through:

$$e(T_2, T_1) = \frac{n(\alpha, \beta, \theta, T_1)}{n(\alpha, \beta, \theta, T_2)}$$

which is called Pitman ARE, and T_{1n} is preferred if the ratio is less than 1.

Theorem 5.1.2. *Let $-\infty < h < \infty$ and $\theta_n = \theta_0 + \frac{h}{\sqrt{n}}$. Consider the following conditions: (i)exists $\mu(\theta), \sigma(\theta)$,s.t., for all h, $\frac{\sqrt{n}(T_n - \mu(\theta_n))}{\sigma(\theta_n)} \xrightarrow{d} N(0,1)$ (ii)$\mu'(\theta_0) > 0$ (iii)$\sigma(\theta_0) > 0$ Then:*

$$e(T_2, T_1) = \frac{\sigma_1^2(\theta_0)}{\sigma_2^2(\theta_0)}[\frac{\mu_2'(\theta_0)}{\mu_1'(\theta_0)}]^2$$

The proof is possibly to occur in the exam.

证明.

$$\frac{\sqrt{n_1}(T_1 - \mu_1(\theta_n))}{\sigma_1(\theta_n)} \xrightarrow{H_1} N(0,1), \frac{\sqrt{n_1}(T_1 - \mu_1(\theta_0))}{\sigma_1(\theta_0)} \xrightarrow{H_0} N(0,1)$$

We can get the critical value: $c_1 = \mu_1(\theta_0) + z_\alpha \frac{\sigma_1(\theta_0)}{\sqrt{n_1}}$

Then the power under H_1 is:

$$\begin{aligned}
P_{H_1}(T_1 > c_1) &= P_{H_1}(T_1 > \mu_1(\theta_0) + z_\alpha \frac{\sigma_1(\theta_0)}{\sqrt{n_1}}) \\
&= P_{H_1}(\frac{\sqrt{n_1}(T_1 - \mu_1(\theta_n))}{\sigma_1(\theta_n)} > \frac{\sqrt{n_1}(\mu_1(\theta_0) - \mu_1(\theta_n))}{\sigma_1(\theta_n)} + z_\alpha \frac{\sigma_1(\theta_0)}{\sigma_1(\theta_n)}) \\
&\to \phi(-\frac{\sqrt{n_1}(\mu_1(\theta_0) - \mu_1(\theta_n))}{\sigma_1(\theta_n)} - z_\alpha \frac{\sigma_1(\theta_0)}{\sigma_1(\theta_n)}). \text{ (Since the distribution is symmetric.)}
\end{aligned}$$

Since $\sigma_1(\theta_n) \xrightarrow{p} \sigma_1(\theta_0)$, $\mu_1(\theta_n) - \mu_1(\theta_0) \xrightarrow{p} \mu_1'(\theta_0)(\theta_n - \theta_0)$, (By Taylor expansion), then:

$$P_{H_1}(T_1 > c_1) \to \phi(-z_\alpha + \frac{\sqrt{n_1}\mu_1'(\theta_0)(\theta_n - \theta_0)}{\sigma_1(\theta_0)})$$

$$P_{H_1}(T_2 > c_2) \to \phi(-z_\alpha + \frac{\sqrt{n_2}\mu_2'(\theta_0)(\theta_n - \theta_0)}{\sigma_2(\theta_0)})$$

By the definition of Pitman ARE, we let $P_{H_1}(T_1 > c_1) = P_{H_1}(T_2 > c_2)$, deriving $e(T_2, T_1) = \frac{n_1}{n_2}$. Since:

$$\frac{\sqrt{n_1}\mu_1'(\theta_0)(\theta_n - \theta_0)}{\sigma_1(\theta_0)} = \frac{\sqrt{n_2}\mu_2'(\theta_0)(\theta_n - \theta_0)}{\sigma_2(\theta_0)}$$

Then we can solve out: $e(T_2, T_1) = \frac{n_1}{n_2} = \frac{\sigma_1^2(\theta_0)}{\sigma_2^2(\theta_0)}[\frac{\mu_2'(\theta_0)}{\mu_1'(\theta_0)}]^2$ □

Corollary 5.1.1. *In the case of $\theta_0 = 0$, suppose $F(0) = \frac{1}{2}$ and $X_1, \ldots, X_n$ iid from $F(x-\theta)$. For $T_2 = \frac{S_n}{n}$, since $S_n = \sum_{i=1}^n I(X_i > \theta_0)$, then $\mu_2(\theta) = 1 - F(-\theta)$ and $\sigma_2^2(\theta) = \frac{F(-\theta)(1-F(-\theta))}{n}$. Let $\theta = \theta_0 = 0$, we get:*

$$e(S_n, \overline{X}_n) = 4\sigma_F^2 f^2(0)$$

The sign test cannot be arbitrarily bad with respect to the t-test, which is bound by $\frac{1}{3}$, but the t-test can be arbitrarily bad with respect to the sign test.

5.2 Signed rank test (Wilcoxon)

5.2.1 Procedure

$X_1, \ldots, X_n$ are iid observed from some location parameter distribution $F(x-\theta)$ and F is symmetric. Let $\theta = \text{median}(F)$, we want to test: $H_0 : \theta = 0$ versus $H_1 : \theta > 0$.

We start by ranking $|X_i|$ from the smallest to the largest, giving the units ranks $R_1, \ldots, R_n$ and order statistics $|X|_{(1)}, \ldots, |X|_{(n)}$

We define the Wilcoxon signed-rank statistic:

$$T_n = \sum_{i=1}^n R_i I(X_i > 0)$$

where $R_i I(X_i > 0)$ can be called the positive signed rank of X_i . When θ is greater than 0, T_n is expected to be large.

There is no information loss when using the positive signed rank, for positive signed rank+positive signed rank=$\sum_{i=1}^n R_i = \frac{n(n+1)}{2}$

Now we try to derive the null distribution of T_n.

Define:

$$W_i = I(|X|_{(i)} \text{ corresponds to some positive } X_j)$$

Then T_n can be written as:

$$T_n = \sum_{i=1}^n iW_i$$

Proposition 5.2.1. *Under $H_0, W_1, \ldots, W_n$ are i.i.d.* $\text{BIN}(1, \frac{1}{2})$ *variables.*

By the symmetric assumption, $E(W_i) = \frac{1}{2}$.

Then we can simply get $E_{H_0}(T_n) = \frac{n(n+1)}{4}$ and $\text{Var}_{H_0}(T_n) = \frac{n(n+1)(2n+1)}{24}$

Theorem 5.2.1. *Under H_0, $\frac{T_n - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \xrightarrow{d} N(0,1)$*

Now we use U-statistics to derive the asymptotic distribution of T_n under H_1

Proposition 5.2.2. *T_n can also be written as:*

$$T_n = \sum_{i \le j} I\left(\frac{X_i + X_j}{2} > 0\right) = \sum_{i \le j} I\left(X_i + X_j > 0\right)$$

Define the anti-rank to prove this:

$$D_k = \{i : R_i = k, 1 \le i \le n\}$$

$$\begin{aligned}
\sum_{i \le j} I\left(\frac{X_i + X_j}{2} > 0\right) &= \sum_{i=1}^{n} I(X_i > 0) + \sum_{i<j} I(\frac{X_{D_i + D_j}}{2} > 0) \\
&= \sum_{i=1}^{n} I(X_i > 0) + \sum_{i<j} I(X_{D_j} > 0) \\
&= \sum_{j=1}^{n} I(X_j > 0) + \sum_{j=1}^{n} (j-1) I(X_{D_j} > 0) \\
&= \sum_{j=1}^{n} j I(X_{D_j} > 0) = \sum_{i=1}^{n} i W_i
\end{aligned}$$

We define the kernel function $h(x_1, x_2, \ldots, x_r)$ and assume $n \ge r$, where r is called the order. We want to estimate about the parameter $\theta = \theta(F) = E_F h(X_1, X_2, \ldots, X_r)$

A better unbiased estimate than $h(X_1, \ldots, X_r)$ itself is the U-statistics:

$$U = \frac{1}{\binom{n}{r}} \sum_{1 \le i_1 < i_2 < \ldots < i_r \le n} h(X_{i_1}, X_{i_2}, \ldots, X_{i_r})$$

Example 5.2.1. *Let r=1, then sample moments and $F_n(x_0)$ is a U-statistic. Let r=2, $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$, then the sample variance is a U-statistic.*

Example 5.2.2. *Let $h(X_1, X_2) = I(X_1 + X_2 > 0)$, then $U = \frac{1}{\binom{n}{2}} \sum_{i<j} I(X_i + X_j > 0)$ and can be related to Wilcoxon statistic.*

For $k = 1, \ldots, r$, let

$$\begin{aligned}
h_k(x_1, \ldots, x_k) &= E[h(X_1, \ldots, X_r) | X_1 = x_1, \ldots, X_k = x_k] \\
&= E[h(x_1, \ldots, x_k, X_{k+1}, \ldots, X_r)]
\end{aligned}$$

Define $\zeta_k = \mathrm{Var}(h_k(X_1, \ldots, X_k))$

It can be shown that:

$$U_n - \theta = \frac{r}{n} \sum_{i=1}^{n} (h_1(X_i) - \theta) + o_p(n^{-1/2})$$

The iid sum occurs!

Theorem 5.2.2. *Suppose $Eh^2(X_1, \ldots, X_r) < \infty, 0 < \zeta_1 < \infty$. Then:*

$$\frac{U - \theta}{\sqrt{\mathrm{Var}(U)}} \xrightarrow{d} N(0, 1)$$

where $\mathrm{Var}(U) = \frac{1}{n} r^2 \zeta_1 + O(n^{-2})$

Theorem 5.2.3.
$$\frac{T_n - E[T_n]}{\sqrt{\mathrm{Var}(T_n)}} \xrightarrow{d} N(0,1)$$

证明.

$$\begin{aligned}\frac{1}{\binom{n}{2}}T_n &= \frac{1}{\binom{n}{2}}\sum_{i\le j} I(X_i + X_j > 0)\\ &= \frac{1}{\binom{n}{2}}\sum_{i=1}^{n} I(X_i > 0) + \frac{1}{\binom{n}{2}}\sum_{i<j} I(X_i + X_j > 0)\end{aligned}$$

The first is $o_p(1)$ and second is asymptotically normal as Theorem 5.2.2

□

Theorem 5.2.4. *If F is a continous symmetric CDF with unique median θ, then the signed rank test is consistent for tests on θ*

证明. We can derive the critical value by the Theorem above. $t_n = E[T_n] + z_\alpha\sqrt{\mathrm{Var}(T_n)}$, and we can directly derive that: $\frac{t_n}{\binom{n}{2}} \to \frac{1}{2}$. Then the power is:

$$Q_n = P_F(T_n \ge t_n) = P_F\left(\frac{1}{\binom{n}{2}}T_n - p_\theta \ge \frac{1}{\binom{n}{2}}t_n - p_\theta\right)$$

where $p_\theta = P_\theta(X_1 + X_2 > 0) > \frac{1}{2}$ under H_1. According to Theorem 5.2.2, $\frac{1}{\binom{n}{2}}T_n - p_\theta \xrightarrow{p} 0$ under any F. Then $Q_n \to 1$. □

Theorem 5.2.5. *(i) The Pitman ARE is*

$$e(T_n, \overline{X}_n) = 12\sigma_F^2\left(\int_{-\infty}^{\infty} f^2(u)du\right)^2$$

(ii)$\inf_{F\in\mathcal{F}} e(T_n, \overline{X}_n) = \frac{108}{125} \approx 0.864$, *where $\mathcal{F}$ is the family of CDFs satisfying continuous, symmetric and $\sigma_F^2 < \infty$. The equality is attained at F s.t. $f(x) = b(a^2 - x^2), |x| < a$, where $a = \sqrt{5}, b = \frac{3\sqrt{5}}{20}$*

证明. (i) Let $T_{2n} = \frac{1}{\binom{n}{2}}T_n$, since:

$$E[\frac{1}{\binom{n}{2}}T_n] \xrightarrow{p} E[\frac{1}{\binom{n}{2}}\sum_{i<j} I(X_i + X_j > 0)] = E_F(I(X_1 + X_2 > 0)) = P_\theta(X_1 + X_2 > 0)$$

Using conditional expectation to calaculate this:

$$\begin{aligned}E_F(I(X_1 + X_2 > 0)|X_2 = x) &= P(X_1 > -x)\\ &= 1 - F(-x-\theta) = F(x+\theta) \text{ (by symmetric.)}\end{aligned}$$

Then $E_F[I(X_1+X_2>0)|X_2]=F(X_2+\theta)$, then:

$$P_\theta(X_1+X_2>0)=E[E_F[I(X_1+X_2>0)|X_2]]=\int F(x+\theta)f(x-\theta)dx.$$

Then $\mu_n'(\theta)=2\int f(x+\theta)f(x-\theta)dx$ and $\mu_n'(0)=2\int f^2(u)du>0$

According to Theorem 5.2.2., we have: ($h(x_1,x_2)=I(X_1+X_2>0)$ here)

$$\begin{aligned}
\mathrm{Var}(T_{2n}) &= \frac{1}{n}r^2\zeta_1+O(n^{-2})\\
&= \frac{1}{n}2^2\,\mathrm{Var}(h_1(X_1))+O(n^{-2})\\
&= \frac{4}{n}\left\{E[E^2(h(X_1,X_2)|X_1)]-[E[E(h(X_1,X_2)|X_1)]]^2\right\}\\
&= \frac{4}{n}\left\{E[1-F(-X_1)]^2-E^2h(X_1,X_2)\right\}\\
&= \frac{4}{n}\left\{\int[1-F(-x-\theta)]^2f(x-\theta)dx-\left(\int[1-F(-x-\theta)]f(x-\theta)dx\right)^2\right\}\\
&= \frac{4}{n}\left\{\int F^2(x+\theta)f(x-\theta)dx-\left(\int F(x+\theta)f(x-\theta)dx\right)^2\right\}
\end{aligned}$$

So we have: $\sigma_n(0)=\frac{4}{n}\,\mathrm{Var}_F[F(X)]=\frac{4}{n}\frac{1}{12}=\frac{1}{3n}$.

For $T_{1n}=\overline{X}_n$, we have: $\mu_n(\theta)=\theta,\sigma_n^2(\theta)=\frac{\sigma_F^2}{n}$. Then the result follows from Theorem 5.1.2. $\square$