Handout 3: Asymptotics after transformations

Lecturer & author: Georgios Karagiannis georgios.karagiannis@durham.ac.uk
References: [1]

1 Cramer's theorem & the Delta method

Note 1. Cramer theorem is another implication of Slutsky Theorems. Briefly, it refers to convergence in distribution, and it says (more or less) that smooth differential functions of asymptotically Normal variables (or statistics) are asymptotically Normal too. So, Normality can be transmitted under specific conditions.

Theorem 2. *(Cramer Theorem) Let function $g : \mathbb{R}^d \to \mathbb{R}^k$ such that $\dot{g}(x)$ is continuous in a neighborhood of $\mu \in \mathbb{R}^d$. If $X_n \in \mathbb{R}^d$ is a sequence of random vectors such that $n^a(X_n - \mu) \xrightarrow{D} X$, where $a > 0$, then*

$$n^a(g(X_n) - g(\mu)) \xrightarrow{D} \dot{g}(\mu)X$$

Proof. Because $(\frac{1}{n})^a \to 0$ for $a > 0$, and $n^a(X_n - \mu) \xrightarrow{D} X$, then $(X_n - \mu) = \frac{1}{n^a} n^a(X_n - \mu) \xrightarrow{D} 0$, by Slutky theorem. Therefore $(X_n - \mu) \xrightarrow{D} 0 \implies (X_n - \mu) \xrightarrow{P} 0 \implies X_n \xrightarrow{P} \mu$. Hence, I get

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

Because $\dot{g}(x)$ is continuous in a neighborhood $\{x : |x - \mu| < \delta\}$, then by Mean Value Theorem for $\{|x - \mu| < \delta\}$

$$g(x) = g(\mu) + \int_0^1 \dot{g}(\mu + v(x - \mu))\mathrm{d}u\,(x - \mu). \tag{1.1}$$

So for $|X_n - \mu| < \delta$,

$$n^a(g(X_n) - g(\mu)) = \int_0^1 \dot{g}(\mu + v(X_n - \mu))\mathrm{d}u\, n^a(X_n - \mu). \tag{1.2}$$

Because $X_n \xrightarrow{P} \mu$, it is

$$\int_0^1 \dot{g}(\mu + v(X_n - \mu))\mathrm{d}u \xrightarrow{P} \dot{g}(\mu) \tag{1.3}$$

Slutsky Theorems. Then (1.2) becomes

$$n^a(g(X_n) - g(\mu)) \xrightarrow{D} \dot{g}(\mu)X$$

by using Slutsky theorem on (1.2) and because of $n^a(X_n - \mu) \xrightarrow{D} X$ and (1.3). ∎

The same proof with Mann-Wald notation notation is in Appendix

Theorem 3. *(Delta Theorem) Let function $g : \mathbb{R}^d \to \mathbb{R}^k$ such that $\dot{g}(x)$ is continuous in a neighborhood of $\mu \in \mathbb{R}^d$. If $X_n \in \mathbb{R}^d$ is a sequence of random vectors such that $\sqrt{n}(X_n - \mu) \xrightarrow{D} N(0, \Sigma)$ then*

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{D} N(0, \dot{g}(\mu)\Sigma\dot{g}(\mu)^T)$$

Proof. Using Cramer's Theorem for $a = 1/2$, we have $\sqrt{n}(X_n - \mu) \xrightarrow{D} Z$, where $Z \sim \mathrm{N}(0, \Sigma)$, then $\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{D} \dot{g}(\mu)Z$. Because $Z \sim \mathrm{N}(0, \Sigma)$ it is $\dot{g}(\mu)Z \sim \mathrm{N}(0, \dot{g}(\mu)\Sigma\dot{g}(\mu)^T)$. So

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{D} \mathrm{N}(0, \dot{g}(\mu)\Sigma\dot{g}(\mu)^T)$$

□

Example 4. Consider a 2×2 contingency table where $(n_{i,j})$ is the (i, j)th cell count, and π_{ij} is the (i, j)th cell probability.

1. Show that the marginal distribution of the MLE of the odd ratio $\hat{\theta}$ is such that

$$\sqrt{n}(\log(\hat{\theta}) - \log(\theta)) \xrightarrow{D} \mathrm{N}(0, \frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}}).$$

2. Show that

$$\frac{\log(\hat{\theta}) - \log(\theta)}{\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}} \xrightarrow{D} \mathrm{N}(0, 1).$$

Hint: It is $\hat{\theta} = \frac{n_{11}n_{22}}{n_{21}n_{12}} = \frac{p_{11}p_{22}}{p_{21}p_{12}}$, , where $p_{i,j} = n_{i,j}/n$.

Solution.

1.

- In Example 12 (Handout 2), we showed that from the CLT, have

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

  where

$$\mathrm{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T = \begin{bmatrix} (1-\pi_{11})\pi_{11} & -\pi_{11}\pi_{12} & -\pi_{11}\pi_{21} & -\pi_{11}\pi_{22} \\ -\pi_{11}\pi_{12} & (1-\pi_{12})\pi_{12} & -\pi_{12}\pi_{21} & -\pi_{12}\pi_{22} \\ -\pi_{11}\pi_{21} & -\pi_{12}\pi_{21} & (1-\pi_{21})\pi_{21} & -\pi_{21}\pi_{22} \\ -\pi_{22}\pi_{11} & -\pi_{22}\pi_{12} & -\pi_{22}\pi_{21} & (1-\pi_{22})\pi_{22} \end{bmatrix}$$

  for the whole vectorized quantities $\boldsymbol{\pi} = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$, and $\boldsymbol{p} = (p_{11}, ..., p_{22})$.
- It is $\hat{\theta} = \frac{p_{11}p_{22}}{p_{21}p_{12}} \implies \log(\hat{\theta}) = \log(p_{11}) + \log(p_{22}) - \log(p_{12}) - \log(p_{21})$
- So I can specify $g(x) = \log(x_{11}) + \log(x_{22}) - \log(x_{12}) - \log(x_{21})$

- It is
$$\dot{g}(x)=\frac{\mathrm{d}}{\mathrm{d}x}g(x)=(\frac{1}{x_{11}},-\frac{1}{x_{12}},-\frac{1}{x_{21}},\frac{1}{x_{22}})$$
and hence $\dot{g}(x)$ is continuous a.s.
- Because all the assumptions of Delta Method are satisfied, it is
$$\sqrt{n}(\log(\hat{\theta})-\log(\theta))\xrightarrow{D}\mathrm{N}(0,\dot{g}(\pi)(\mathrm{diag}(\pi)-\pi\pi^T)\dot{g}(\pi)^T)$$
with
$$\begin{aligned}\dot{g}(\boldsymbol{\pi})(\mathrm{diag}(\boldsymbol{\pi})-\boldsymbol{\pi}\boldsymbol{\pi}^T)\dot{g}(\boldsymbol{\pi})^T &= \dot{g}(\boldsymbol{\pi})(1,-1,-1,1)^T\\ &=\frac{1}{\pi_{11}}+\frac{1}{\pi_{12}}+\frac{1}{\pi_{21}}+\frac{1}{\pi_{22}}\end{aligned}$$

2. Using Slutsky theorem, and law of large numbers, similar to Example 19 (Handout 2), we find that
$$\frac{\sqrt{\frac{1}{p_{11}}+\frac{1}{p_{12}}+\frac{1}{p_{21}}+\frac{1}{p_{22}}}}{\sqrt{\frac{1}{\pi_{11}}+\frac{1}{\pi_{12}}+\frac{1}{\pi_{21}}+\frac{1}{\pi_{22}}}}\xrightarrow{P}1$$
and by using Slutsky theorem as in Example 19 (Handout 2), we find
$$\frac{\log(\hat{\theta})-\log(\theta)}{\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{21}}+\frac{1}{n_{22}}}}\xrightarrow{D}\mathrm{N}(0,1).$$

Exercise. 20 Consider an M-way contingency table. Consider the quantities obs. cell counts, cell probabilities, cell proportions in their vectorised forms as
$$\boldsymbol{n}=(n_1,...,n_N)^T\;;\qquad \boldsymbol{p}=(p_1,...,p_N)^T\qquad \boldsymbol{\pi}=(\pi_1,...,\pi_N)^T\;;$$
where $n_+=\sum_{j=1}^N n_j$, and $p_j=n_j/n_+$.

1. Consider a constant matrix $\boldsymbol{C}\in\mathbb{R}^{k\times N}$, and show that
$$\sqrt{n}(\boldsymbol{C}\log(\boldsymbol{p})-\boldsymbol{C}\log(\boldsymbol{\pi}))\xrightarrow{D}\mathrm{N}(0,\boldsymbol{C}\mathrm{diag}(\boldsymbol{\pi})^{-1}\boldsymbol{C}^T-\boldsymbol{C}11^T\boldsymbol{C}^T)\tag{1.4}$$
2. Consider a 3×3 contingency table with probabilities in a vectorized form as
$$\boldsymbol{\pi}=(\;\pi_{11}\;,\;\pi_{21}\;,\;\pi_{31}\;,\;\pi_{12}\;,\;\pi_{22}\;,\;\pi_{32}\;,\;\pi_{13}\;,\;\pi_{23}\;,\;\pi_{33}\;)^T$$
Also $\boldsymbol{n}$, $\boldsymbol{p}$ are vectorized likewise. Find the joint asymptotic distribution of the vector of different log odd ratios
$$\log(\boldsymbol{\theta}^{\boldsymbol{C}})=\begin{bmatrix}\log(\frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}})\\ \log(\frac{\pi_{22}\pi_{33}}{\pi_{23}\pi_{32}})\end{bmatrix}$$

Remark 5. Exercise above shows that if we wish to find the joint asymptotic distribution of a number of odds ratio $\boldsymbol{\theta}^{\boldsymbol{C}}$, we can write the vector of odds ratios in a vectorized form as
$$\log(\boldsymbol{\theta}^{\boldsymbol{C}})=\boldsymbol{C}\log(\boldsymbol{\pi})$$

and them use (1.4). Here each row of the matrix $\boldsymbol{C}$ contains zeros except for two $+1$ elements and two -1 elements in the positions multiplied by elements of $\log(\boldsymbol{p})$ to form the given log odds ratio $\log(\boldsymbol{\pi})$.

Exercise sheet

Exercise #19 ; 22 ; 20 ; 21

1.1 Variance stabilizing transformations

Note 6. Variance stabilizing transformations are applied to statistics (data functions) with purpose to define new statistics whose variation does not depend on the unknown nuisance parameters of the sampling distribution (such as the expected value or higher moments of the statistic).

Note 7. Consider we wish to perform inference on the parameter θ based on a statistic $S(\theta, X_{1:n})$; i.e. construct confidence interval for θ. It is often inconvenient when the variance (asymptotic or exact) of the statistic $S(\theta, X_{1:n})$ involves unknown quantities such as moments e.g. $\mu_r = \mathrm{E}(S(\theta, X_{1:n})^r)$.

- One way to address this issue is to plug in the sample analogs of such unknown quantities, e.g. $\overline{X^r}$, and use Slutsky theorems to compute the asymptotic distribution. However, such a treatment may increase the variance of the statistic causing problems, i.e. too wide confidence interval for θ; e.g. recall the Z and T statistics from SC2.

- Another way is to transform the statistic $S(\theta, X_{1:n})$ by creating a new one whose variance (asymptotic or exact) is independent of these unknown (and troublesome) quantities. Variable stabilizing transformations use Theorem 3 as a bases to create transformations whose asymptotic variance is independent of θ. Here, we consider the 1D case.

The idea Let $X_n \in \mathbb{R}$ be a sequence of random vectors such that $\sqrt{n}(X_n - \mu(\theta)) \xrightarrow{D} \mathrm{N}(0, \sigma^2(\theta))$ where $\mu(\theta)$ and $\sigma^2(\theta)$ are functions of the parameter of interest θ. The variable stabilizing transformation $g(\cdot)$ is the solution of the differential equation resulting from the Delta method method such that

$$\dot{g}(\mu(\theta))\sigma(\theta) = \cancel{\text{constant}} = 1 \tag{1.5}$$

Then, by Theorem 3,

$$\sqrt{n}(g(X_n) - g(\mu(\theta))) \xrightarrow{D} \mathrm{N}(0, 1)$$

Example 8. Consider IID random sample from a Poisson distr $X_i \sim \mathrm{Poi}(\lambda)$, $i = 1, ..., n$. Find the variance stabilizing transformation that will allow us to find asymptotic confidence intervals for the parameter λ.

Hint: If $X_i \sim \mathrm{Poi}(\lambda)$, then $\mathrm{E}(X_i) = \mathrm{Var}(X_i) = \lambda$

Solution. By CLT, it is $\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{D} \mathrm{N}(0, \lambda)$. To find the stabilising transofrmation, I need to solve (1.5)

$$\dot{g}(\mu(\theta))\sigma(\theta) = 1 \Longrightarrow \dot{g}(\lambda) = 1/\sqrt{\lambda} \Longrightarrow g(\lambda) = 2\sqrt{\lambda}$$

So, because $\dot{g}(\lambda)^2\lambda = 1$ it is $\sqrt{n}(2\sqrt{\bar{X}_n} - 2\sqrt{\lambda}) \xrightarrow{D} \mathrm{N}(0, 1)$.

The asymptotic $1 - a$ CI for $2\sqrt{\lambda}$ is $\{\, 2\sqrt{\bar{X}_n} \pm z_{1-\frac{a}{2}} \frac{1}{2\sqrt{n}} \,\}$

Example 9. Find the variance stabilizing transformation, for the case that $\sigma^2(\theta) = \theta^b$, $b \neq 0$. Does it ring any bell?

Solution. It is

$$g'(\theta)\sigma(\theta) = 1 \Longleftrightarrow g'(\theta) = \theta^{-b}$$

Then

$$g(\theta) = \int \theta^{-b} \mathrm{d}\theta = \begin{cases} \log(\theta) & \text{, if } b = 1 \\ \frac{2}{2-b}\theta^{\frac{2-b}{2}} & \text{, if } b \neq 1 \end{cases}$$

which is the Box-Cox transformation for $\lambda = \frac{2-b}{2}$ (although a little bit shifted by $\frac{1}{\lambda}$).

1.2 Second order Delta method

Note 10. Implementation of Theorem 3, when $\dot{g}(\mu) = 0$ may be impractical and provide poor approximations because the approximation gets poor. The following theorem shows that we can improve the approximation of Delta method when $\dot{g}(\mu) = 0$, by considering a 2nd order Taylor expansion in (1.1).

Theorem 11. *Let function $g : \mathbb{R} \to \mathbb{R}$ such that $\dot{g}(\mu) = 0$, and $\ddot{g}(x)$ is continuous in an interval of $\mu \in \mathbb{R}$. If $X_n \in \mathbb{R}$ is a sequence of random vectors such that $\sqrt{n}(X_n - \mu) \xrightarrow{D} N(0, \sigma^2)$ then*

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

where $Y \sim \chi^2_1$.

Proof. The derivation is given as an exercise (Exercise 23). □

Example 12. Consider $X, X_1, X_2, ...$ IID from a distribution with $\mu = \mathrm{E}(X)$ and $\sigma^2 = \mathrm{Var}(X) < \infty$.

1. Find the asymptotic distribution of $\bar{X}_n$, $\bar{X}_n^2$ by using the Delta method.
2. Assume $\mu = \mathrm{E}(X) = 0$. Find$\bar{X}_n^2$ by using the 2nd order Delta method.
3. When $\mu = 0$, discuss how/why the 2nd order Delta method provides more accurate than Delta method.

Solution.

1. By using the CPT, it is
$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \mathrm{N}(0, \sigma^2)$$
Let $g(x) = x^2$, and $\dot{g}(x) = 2x$. Then by using Delta method, it is
$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{D} \mathrm{N}(0, 4\mu^2\sigma^2) \tag{1.6}$$
2. Let $g(x) = x^2$, and $\dot{g}(x) = 2x$. I also compute $\ddot{g}(x) = 2$, and $\dot{g}(\mu) = 0$. By using 2nd order Delta method I get
$$n\bar{X}_n^2 \xrightarrow{D} \sigma^2\chi_1^2$$
3. If $\mu = 0$
    - The exact result is
    $$n\frac{\bar{X}_n^2}{\sigma^2} = (\frac{\bar{X}_n - 0}{\sigma/\sqrt{n}})^2 \sim \chi_1^2 \tag{1.7}$$
    as $\bar{X}_n \sim \mathrm{N}(0, \sigma^2/n)$.
    - Delta method gives, $\sqrt{n}(\bar{X}_n - 0) \xrightarrow{D} \mathrm{N}(0, 0)$, aka $\bar{X}_n \xrightarrow{D} 0$, the mass is around a point ... strange
    - 2nd order Delta method gives $n\bar{X}_n^2 \xrightarrow{D} \sigma^2\chi_1^2$, which is a more reasonable result.
    - Of course, that 2nd order Delta method gives the exact result (1.7) here is a coincidence. However it gives an idea that the 2nd order Delta method gives better results than the 1st order Delta method.

Exercise sheet

Exercise #24 ; 23

References

[1] Robert J Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.

A Alternative proof of Theorem 2

Proof. By Taylor expansion of $g(x)$ around μ , we get

$$g(x) = g(\mu) + \mathring{g}(\mu)(x-\mu) + o\left((x-\mu)\right)$$

Because $n^a(X_n - \mu) \xrightarrow{D} X$, then $n^a(X_n - \mu) \xrightarrow{D} O_p(1)$.

So

$$n^a\left(g(X_n) - g(\mu)\right) = \mathring{g}(\mu)\underbrace{[n^a(X_n-\mu)]}_{\xrightarrow{D} X} + \underbrace{o_p\left(\underbrace{n^a(X_n-\mu)}_{\xrightarrow{D}\text{constant}}\right)}_{\xrightarrow{p} 0} \tag{A.1}$$

But for $n \to \infty$, it is

$$n^a(X_n - \mu) \xrightarrow{D} X$$

which means $n^a(X_n - \mu) \xrightarrow{D} O_p(1)$ as well (from Fact 29 in Handout 2).

Also from Fact 29 in Handout 2

$$o_p\left(n^a(X_n-\mu)\right) = o_p\left(O_p(1)\right) = o_p(1)$$

from Fact 29 in Handout 2.

Hence, by Slutsky rule (2.) in (14), Equation A.1 becomes

$$n^a\left(g(X_n) - g(\mu)\right) \xrightarrow{D} \mathring{g}(\mu)X$$

□