

# Transformations

Patrick Breheny

# Introduction

- It is often the case in statistics that one knows something about the convergence of  $\mathbf{x}_n$ , but then we want to know something about the convergence of some function of the random vector  $g(\mathbf{x}_n)$
- Today, we'll go over three useful tools for drawing these kinds of conclusions
  - The continuous mapping theorem
  - Slutsky's theorem
  - The delta method

# Continuous mapping theorem

- The continuous mapping theorem is a simple, but very useful result
- It says that if  $\mathbf{x}_n \rightarrow \mathbf{x}$  (in any sense), then  $g(\mathbf{x}_n) \rightarrow g(\mathbf{x})$  (in the same sense) if  $g$  is continuous
- **Theorem (continuous mapping):** Let  $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$  be continuous almost everywhere with respect to  $\mathbf{x}$ .
  - If  $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$ , then  $g(\mathbf{x}_n) \xrightarrow{d} g(\mathbf{x})$
  - If  $\mathbf{x}_n \xrightarrow{P} \mathbf{x}$ , then  $g(\mathbf{x}_n) \xrightarrow{P} g(\mathbf{x})$
  - If  $\mathbf{x}_n \xrightarrow{\text{as}} \mathbf{x}$ , then  $g(\mathbf{x}_n) \xrightarrow{\text{as}} g(\mathbf{x})$

# Example #1

- The continuous mapping theorem is extremely useful and allows formal justification of all kinds of things that seem obvious, such as: we should be able to “square both sides” of a convergence statement
- For example,  $Z_n \xrightarrow{d} N(0, 1) \implies Z_n^2 \xrightarrow{d} \chi_1^2$
- If the function *isn't* continuous, however, all bets are off
- For example, suppose  $X_n \xrightarrow{P} 0$ , and

$$g(x) = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0 \end{cases} ;$$

$g(0) = 0$ , but  $g(X_n) \xrightarrow{P} 1$  if  $X_n$  is continuous

## Example #2

- Note that only continuity at 0 was relevant in that last example, since  $P(X = 0) = 1$
- By contrast, if  $X_n \xrightarrow{P} 1$  or  $X_n \xrightarrow{d} N(0, 1)$ , then the CMT would hold since  $g(x)$  would be continuous almost everywhere
- As a multivariate example, we proved the central limit theorem

$$\sqrt{n}(\bar{\mathbf{x}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma});$$

by the continuous mapping theorem, we immediately have the corollary

$$\sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\bar{\mathbf{x}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$$

## Example #3

- The result even extends to matrices (you can imagine stacking the columns of the  $d \times x$  matrix into a giant vector of dimension  $d \cdot k$ )
- Proving these facts is beyond the scope of this course, but:
  - Matrix inversion is a continuous function (unless the matrix is singular)
  - Taking the square root of a positive definite matrix is also continuous
- So for example, if  $\mathbf{X}_n \xrightarrow{P} \mathbf{A}$ , then  $\mathbf{X}_n^{-1} \xrightarrow{P} \mathbf{A}^{-1}$  provided that  $\mathbf{A}$  is not singular

# Asymptotic equivalence

- Another very useful result is Slutsky's theorem, which we will present here in a rather general form
- Before we prove this result, we need to introduce the following lemma concerning asymptotically equivalent sequences
- Two sequences of random vectors  $\mathbf{x}_n$  and  $\mathbf{y}_n$  are said to be *asymptotically equivalent* if  $\mathbf{x}_n - \mathbf{y}_n \xrightarrow{P} \mathbf{0}$ .
- **Lemma (asymptotic equivalence):** If  $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$  and  $\mathbf{x}_n - \mathbf{y}_n \xrightarrow{P} \mathbf{0}$ , then  $\mathbf{y}_n \xrightarrow{d} \mathbf{x}$ .
- In words, the lemma is saying that asymptotically equivalent sequences have the same limiting distributions

# Slutsky's theorem

- This lemma is necessary to prove Slutsky's theorem:
- **Theorem (Slutsky):** If  $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$  and  $\mathbf{y}_n \xrightarrow{P} \mathbf{a}$ , where  $\mathbf{a}$  is a constant, then

$$\begin{bmatrix} \mathbf{x}_n & \mathbf{y}_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathbf{x} \\ \mathbf{a} \end{bmatrix}.$$

- This is perhaps not the form in which you are used to seeing Slutsky's theorem; the name "Slutsky's theorem" is widely used in an inconsistent manner to mean a number of similar results

## Remarks

- This result along with the continuous mapping theorem implies all of the results that people often call “Slutsky's theorem”
- For example, we have the familiar  $X_n + Y_n \xrightarrow{d} X + a$  and  $X_n Y_n \xrightarrow{d} aX$  since addition and multiplication are continuous functions
- But we also have much more complex statements; for example, if  $\mathbf{x}_1, \mathbf{x}_2, \dots$  is an iid sample with mean  $\boldsymbol{\mu}$  and nonsingular variance, then

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu})^\top \mathbf{S}_n^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \xrightarrow{d} \chi_d^2,$$

where  $\bar{\mathbf{x}}$  is the sample mean and  $\mathbf{S}_n$  is the sample variance

# Delta method

- There is one last important type of transformation that Slutsky/CMT do not address: suppose we know the distribution of  $\mathbf{x} - \boldsymbol{\mu}$  and want to know the distribution of  $g(\mathbf{x}) - g(\boldsymbol{\mu})$
- This result is described by a theorem known as the delta method
- **Theorem (Delta method):** Let  $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$  such that  $\nabla g$  is continuous in a neighborhood of  $\boldsymbol{\mu} \in \mathbb{R}^d$  and suppose  $\sqrt{n}(\mathbf{x}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathbf{x}$ . Then

$$\sqrt{n}(g(\mathbf{x}_n) - g(\boldsymbol{\mu})) \xrightarrow{d} \nabla g(\boldsymbol{\mu})^\top \mathbf{x}.$$

## Normal distribution corollary

- This typically comes when dealing with functions of sample moments, which are multivariate normal by the CLT
- **Corollary (Delta method):** Let  $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$  such that  $\nabla g$  is continuous in a neighborhood of  $\boldsymbol{\mu} \in \mathbb{R}^d$  and suppose  $\sqrt{n}(\mathbf{x}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$ . Then

$$\sqrt{n}(g(\mathbf{x}_n) - g(\boldsymbol{\mu})) \xrightarrow{d} N(\mathbf{0}, \nabla g(\boldsymbol{\mu})^\top \boldsymbol{\Sigma} \nabla g(\boldsymbol{\mu})).$$

- Our proof of the delta method illustrates an important point worth noting for the future: applying a Taylor series expansion requires conditions, but these conditions only need to be met with probability tending to 1 in order to establish convergence in distribution (asymptotic equivalence lemma)

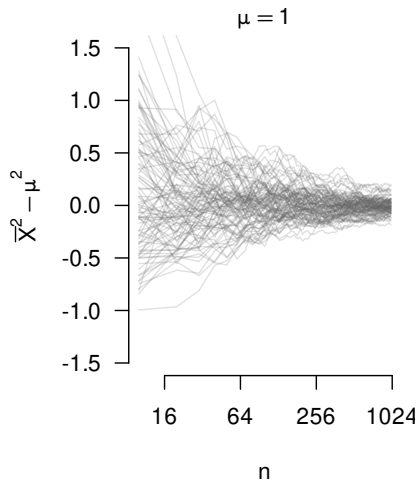
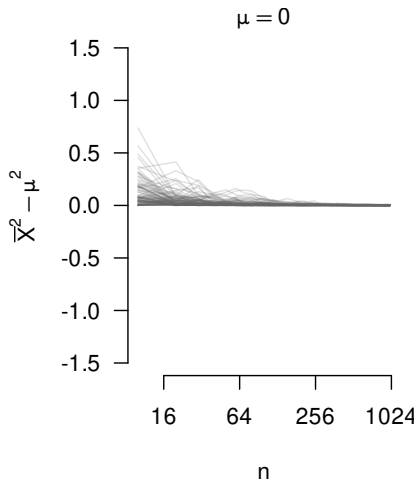
## Example

- While the delta method is certainly a useful result, it should be noted that the rate of convergence can vary widely depending on both  $\mu$  and  $g$
- For example, let's look at the function  $g(\mu) = \mu^2$
- By the delta method and the CLT, we have

$$\sqrt{n}(\bar{x}^2 - \mu^2) \xrightarrow{d} N(0, 4\mu^2\sigma^2)$$

- Note that this isn't even the same rate of convergence for all  $\mu$ :  $\bar{x}^2 - \mu^2$  is  $O_p(1/\sqrt{n})$  in general, but  $O_p(1/n)$  when  $\mu = 0$

## Example (cont'd)



## Remarks

- The relevance to statistical practice is that a common use of the delta method is to derive approximate confidence intervals for unknown parameters
- However, not all transformations and not all values of the unknown parameters converge to normality equally fast
- In practice, this means that some transformations produce much more accurate confidence intervals than others, and it is not always obvious which transformation is best
- Furthermore, a confidence interval procedure can be good for some values of  $\theta$  but poor at other values