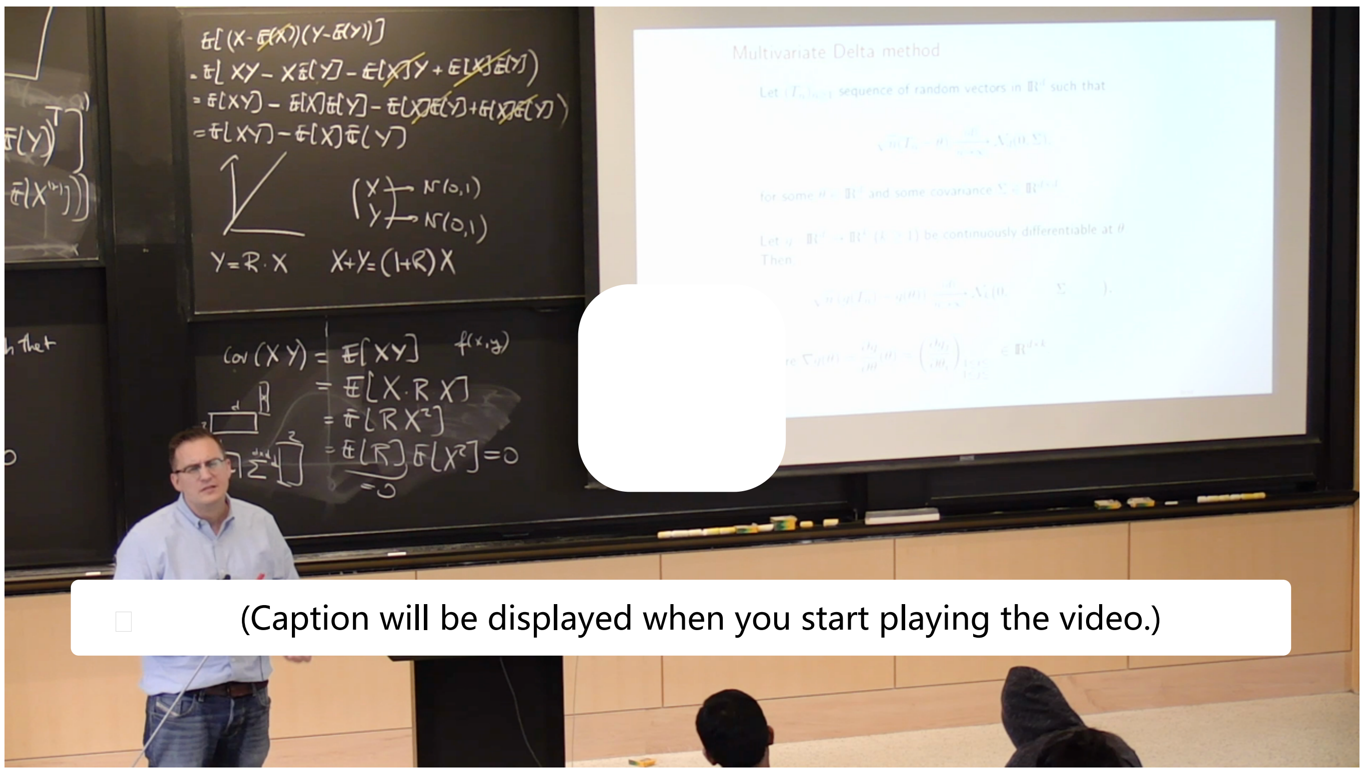
11. Multivariate Delta Method

Multivariate Delta Method



What is the delta method doing for me?

If Xn bar minus mu divided by square root of n

was going to Gaussian, then this would actually

give me a central limit theorem for g

of Xn bar minus g of theta.

Remember this thing?

So it was really when I wanted to take a function of Xn bar

and make it convert to a function of the mean.

And so we have this thing to be Xn bar.

Anything that satisfies this kind of thing will work.

So for us, we just saw that if this is Xn bar and this is mu,

then we actually have this.

But again, this doesn't have to be the case.

In particular, DMLE will not always be of the form Xn bar,

视频

下载视频文件

字幕

下载 SubRip (.srt) file

下载 Text (.txt) file

Gradient Matrix of a Vector Function

4/4 points (graded)

Given a vector-valued function $f: \mathbb{R}^d \to \mathbb{R}^k$, the **gradient** or the **gradient matrix** of f, denoted by ∇f, is the $d \times k$ matrix

$$\nabla f = \begin{pmatrix} | & | & & | \\ \nabla f_1 & \nabla f_2 & \dots & \nabla f_k \\ | & | & & | \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_k}{\partial x_1} \\ \vdots & \dots & \vdots \\ \frac{\partial f_1}{\partial x_d} & \dots & \frac{\partial f_k}{\partial x_d} \end{pmatrix}.$$

This is also the transpose of what is known as the **Jacobian matrix** $\mathbf{J}_f$ of f.

Let $f(x,y,z) = \begin{pmatrix} x^2+y^2+z^2 \\ 2xy \\ y^3+z^3 \\ z^4 \end{pmatrix}$.

How many rows does $\nabla f(x,y,z)$ have?

3

Answer: 3

How many columns does $\nabla f(x, y, z)$ have?

4

Answer: 4

What does column 2 represent in the gradient matrix?

- (selected) Derivative of $2xy$ with respect to x, y, z stacked as a column
- Derivative of the individual functions with respect to y stacked as a column

What is $\nabla f(x,y,z)_{3,2}$?

0

Answer: 0*x

0

Solution:

According to notation developed in the video, the gradient for f is of size 3×4 because it is a function of 3 variables and it outputs 4 values as a column. Column $j \in \{1,2,3,4\}$ of the gradient matrix represents the derivative of the j^{th} function of $f(x, y, z)$ with respect to x, y, z stacked as a column.

$\nabla f(x,y,z)_{3,2}$ is the derivative of function $2xy$ (2nd function) with respect to z (3rd variable). This derivative is equal to 0.

提交 你已经尝试了2次（总共可以尝试2次）

Answers are displayed within the problem

General Statement of the Multivariate Delta Method

The multivariate delta method states that given

- a sequence of random vectors $(\mathbf{T}_n)_{n\geq 1}$ satisfying $\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right) \xrightarrow[n\to\infty]{(d)} \mathbf{T}$,
- a function $\mathbf{g}: \mathbb{R}^d \to \mathbb{R}^k$ that is continuously differentiable at $\vec{\theta}$,

then

$$\sqrt{n}\left(\mathbf{g}(\mathbf{T}_n) - \mathbf{g}(\vec{\theta})\right) \xrightarrow[n\to\infty]{(d)} \nabla \mathbf{g}(\vec{\theta})^T \mathbf{T} \qquad \text{where } \nabla \mathbf{g} = \begin{pmatrix} | & | & \cdots & | \\ \nabla \mathbf{g}_1 & \nabla \mathbf{g}_2 & \cdots & \nabla \mathbf{g}_k \\ | & | & \cdots & | \end{pmatrix}.$$

Common Application

In the lecture and in most applications, $\mathbf{T}_n = \overline{\mathbf{X}}_n$ where $\overline{\mathbf{X}}_n$ is the sample average of $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{iid}{\sim} \mathbf{X}$, and $\vec{\theta} = \mathbb{E}[\mathbf{X}]$. The (multivariate) CLT then gives $\mathbf{T} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{X}})$ where $\Sigma_{\mathbf{X}}$ is the covariance of $\mathbf{X}$. In this case, we have

$$\sqrt{n}\left(\mathbf{g}(\mathbf{T}_n) - \mathbf{g}(\vec{\theta})\right) \xrightarrow[n\to\infty]{(d)} \nabla \mathbf{g}(\vec{\theta})^T \mathbf{T} \sim \mathcal{N}\left(\mu, \nabla \mathbf{g}(\vec{\theta})^T \Sigma_{\mathbf{X}} \nabla \mathbf{g}(\vec{\theta})\right) \qquad (\mathbf{T} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{X}})).$$

(Optional) Proof of Multivariate Delta Method

As in the univariate case, the main idea of the proof of the multivariate delta method is to apply the first order multivariate Taylor theorem (i.e. linear approximation with a remainder term), and then use (multivariate) Slutsky's, and the continuous mapping theorem to establish the required convergence.

Slutsky's theorem and the continuous mapping theorems in higher dimensions are straightforward generalizations of these same theorems in one dimension, i.e. where applicable, scalar random variables are replaced with random vectors.

Proof:
Let $(\mathbf{T}_n)_{n\geq 1}$ be a sequence of random vectors in $\mathbb{R}^d$ such that

$$\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right) \xrightarrow[n\to\infty]{(d)} \mathbf{T},$$

for some $\vec{\theta} \in \mathbb{R}^d$.

Let $\mathbf{g} : \mathbb{R}^d \to \mathbb{R}^k$ be continuously differentiable at $\vec{\theta}$. Then, for any vector $\mathbf{t} \in \mathbb{R}^d$, the first order multivariate Taylor's expansion at $\vec{\theta}$ gives

$$\mathbf{g}(\mathbf{t}) = \mathbf{g}(\vec{\theta}) + \nabla\mathbf{g}(\vec{\theta})^T\left(\mathbf{t} - \vec{\theta}\right) + \left\|\mathbf{t} - \vec{\theta}\right\| \mathbf{u}(\mathbf{t})$$

where $\mathbf{u}(\mathbf{t}) \to \mathbf{0}$ as $\mathbf{t} \to \vec{\theta}$.

Extend the above equation by replacing $\mathbf{t}$ with a random vector $\mathbf{T}_n$, rearrange and multiply both sides by $\sqrt{n}$:

$$\sqrt{n}\left(\mathbf{g}(\mathbf{T}_n) - \mathbf{g}(\vec{\theta})\right) \quad = \quad \nabla\mathbf{g}(\vec{\theta})^T\left(\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right) + \left\|\sqrt{n}(\mathbf{T}_n - \vec{\theta})\right\| \mathbf{u}(\mathbf{T}_n).$$

Let us look at convergence of each term on the right as $n \to \infty$. We will apply the multivariate version of $\boxed{\text{continuous mapping}}$ theorem and $\boxed{\text{Slutsky's theorem}}$ multiple times to our ingredient:

1. $X_n \xrightarrow{d} X \quad \Rightarrow \quad g(X_n) \xrightarrow{d} g(X);$
2. $X_n \xrightarrow{p} X \quad \Rightarrow \quad g(X_n) \xrightarrow{p} g(X);$
3. $X_n \xrightarrow{as} X \quad \Rightarrow \quad g(X_n) \xrightarrow{as} g(X).$

- $X_n + Y_n \xrightarrow{d} X + c;$
- $X_nY_n \xrightarrow{d} cX;$
- $X_n/Y_n \xrightarrow{d} X/c,$ provided that c is invertible,

$$\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right) \xrightarrow[n\to\infty]{(d)} \mathbf{T},$$

which also implies

$$\left(\mathbf{T}_n - \vec{\theta}\right) \xrightarrow[n\to\infty]{(d)/(p)} \mathbf{0}.$$

The first term $\nabla\mathbf{g}(\vec{\theta})^T\left(\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right)$ is a continuous function of $\left(\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right)$, hence

$$\nabla\mathbf{g}(\vec{\theta})^T\left(\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right) \quad \xrightarrow[n\to\infty]{(d)} \left(\nabla\mathbf{g}(\vec{\theta})\right)^T\mathbf{T} \qquad \text{by continuous mapping theorem.}$$

For the second term, the first factor $\left\|\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right\|$ is again a continuous function of $\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)$, and therefore

$$\left\|\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right\| \xrightarrow[n\to\infty]{(d)} \|\mathbf{T}\| \qquad \text{by continuous mapping theorem.}$$

The second factor in the second term is a continuous function of $\mathbf{T}_n$ near $\boldsymbol{\theta}$. Hence

$$\mathbf{u}(\mathbf{T}_n) \xrightarrow[n\to\infty]{(d)/(p)} \mathbf{u}(\vec{\theta}) \; = \; \mathbf{0} \qquad \text{by continuous mapping theorem.}$$

By (multivariate) Slutsky theorem, the entire second term converges to $\mathbf{0}$:

$$\left\|\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right\| \mathbf{u}(\mathbf{T}_n) \xrightarrow[n\to\infty]{(d)/\mathbf{P}} \|\mathbf{T}\|\,(\mathbf{0}) \; = \; \mathbf{0}.$$

Finally, applying the (multivariate) Slutsky theorem to the sum of the two terms gives:

$$\nabla\mathbf{g}(\vec{\theta})^T\left(\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right) + \left\|\sqrt{n}\left(\mathbf{T}_n - \vec{\theta}\right)\right\| \mathbf{u}(\mathbf{T}_n) \xrightarrow[n\to\infty]{(d)} \nabla\mathbf{g}(\vec{\theta})^T\mathbf{T} + \mathbf{0} \; = \; \nabla\mathbf{g}(\vec{\theta})^T\mathbf{T}.$$

This establishes the multivariate delta method.

Hide

讨论

显示讨论

主题： Unit 3 Methods of Estimation:Lecture 10: Consistency of MLE, Covariance Matrices, and Multivariate Statistics / 11. Multivariate Delta Method