

# Микроэконометрика

## Вступительная лекция

Потанин Богдан Станиславович

старший преподаватель, кандидат экономических наук

2021-2022

Оценка складывается из трех элементов:

Оценка складывается из трех элементов:

- Домашнее задание №1 – 40%

Оценка складывается из трех элементов:

- Домашнее задание №1 – 40%
- Домашнее задание №2 – 40%

Оценка складывается из трех элементов:

- Домашнее задание №1 – 40%
- Домашнее задание №2 – 40%
- Экзамен – 20%

Оценка складывается из трех элементов:

- Домашнее задание №1 – 40%
- Домашнее задание №2 – 40%
- Экзамен – 20%

Экзамен проходит письменно в форме решения задач по материалам лекций.

Существуют два альтернативных способа получить оценку за экзамен:

Оценка складывается из трех элементов:

- Домашнее задание №1 – 40%
- Домашнее задание №2 – 40%
- Экзамен – 20%

Экзамен проходит письменно в форме решения задач по материалам лекций.

Существуют два альтернативных способа получить оценку за экзамен:

- Выступить с презентацией по интересной статье, в которой используются методы, изучаемые в курсе.

Оценка складывается из трех элементов:

- Домашнее задание №1 – 40%
- Домашнее задание №2 – 40%
- Экзамен – 20%

Экзамен проходит письменно в форме решения задач по материалам лекций.

Существуют два альтернативных способа получить оценку за экзамен:

- Выступить с презентацией по интересной статье, в которой используются методы, изучаемые в курсе.
- Принять участие в тестировании онлайн-версии курса.



# Численное дифференцирование

Градиент, Якобиан и Гессиан

Имеется функция  $f(x)$ , где  $x = (x_1, \dots, x_n)^T$  это  $n$ -мерный вектор столбец и  $f : R^n \rightarrow R$ .

# Численное дифференцирование

## Градиент, Якобиан и Гессиан

Имеется функция  $f(x)$ , где  $x = (x_1, \dots, x_n)^T$  это  $n$ -мерный вектор столбец и  $f : R^n \rightarrow R$ .

- Градиент  $\nabla f(x)$  – это вектор столбец частных производных функции по ее аргументам:

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T, \quad \nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}$$

# Численное дифференцирование

## Градиент, Якобиан и Гессиан

Имеется функция  $f(x)$ , где  $x = (x_1, \dots, x_n)^T$  это  $n$ -мерный вектор столбец и  $f : R^n \rightarrow R$ .

- Градиент  $\nabla f(x)$  – это вектор столбец частных производных функции по ее аргументам:

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T, \quad \nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}$$

- Гессиан  $H(f(x))$  – это матрица вторых производных функции по ее аргументам:

$$H(f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

# Численное дифференцирование

## Градиент, Якобиан и Гессиан

Имеется функция  $f(x)$ , где  $x = (x_1, \dots, x_n)^T$  это  $n$ -мерный вектор столбец и  $f : R^n \rightarrow R$ .

- Градиент  $\nabla f(x)$  – это вектор столбец частных производных функции по ее аргументам:

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T, \quad \nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}$$

- Гессиан  $H(f(x))$  – это матрица вторых производных функции по ее аргументам:

$$H(f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Рассмотрим функцию  $g(x)$  такую, что  $g : R^n \rightarrow R^m$ . То есть на вход данная функция принимает  $n$ -мерный вектор, а возвращает  $m$ -мерный вектор. Через  $g(x)_i$  обозначим  $i$ -й элемент возвращаемого  $g(x)$  вектора.

# Численное дифференцирование

## Градиент, Якобиан и Гессиан

Имеется функция  $f(x)$ , где  $x = (x_1, \dots, x_n)^T$  это  $n$ -мерный вектор столбец и  $f : R^n \rightarrow R$ .

- Градиент  $\nabla f(x)$  – это вектор столбец частных производных функции по ее аргументам:

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T, \quad \nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}$$

- Гессиан  $H(f(x))$  – это матрица вторых производных функции по ее аргументам:

$$H(f(x))_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Рассмотрим функцию  $g(x)$  такую, что  $g : R^n \rightarrow R^m$ . То есть на вход данная функция принимает  $n$ -мерный вектор, а возвращает  $m$ -мерный вектор. Через  $g(x)_i$  обозначим  $i$ -й элемент возвращаемого  $g(x)$  вектора.

- Якобиан  $J(g(x))$  – это матрица,  $i$ -я строка которой является градиентом  $g(x)_i$ :

$$J(g(x))_{i, \text{все столбцы}} = \nabla g(x)_i$$

# Численное дифференцирование

## Примеры Градиента, Якобиана и Гессиана

Для функции  $f(x) = x_1^3 + x_1 * e^{x_2}$  получаем:

$$\nabla f(x) = (3x_1^2 + e^{x_2}, x_1 e^{x_2})^T$$

# Численное дифференцирование

## Примеры Градиента, Якобиана и Гессиана

Для функции  $f(x) = x_1^3 + x_1 * e^{x_2}$  получаем:

$$\nabla f(x) = (3x_1^2 + e^{x_2}, x_1 e^{x_2})^T$$

$$H(f(x)) = \begin{bmatrix} 6x_1 & e^{x_2} \\ e^{x_2} & x_1 e^{x_2} \end{bmatrix}$$

# Численное дифференцирование

## Примеры Градиента, Якобиана и Гессиана

Для функции  $f(x) = x_1^3 + x_1 * e^{x_2}$  получаем:

$$\nabla f(x) = (3x_1^2 + e^{x_2}, x_1 e^{x_2})^T$$

$$H(f(x)) = \begin{bmatrix} 6x_1 & e^{x_2} \\ e^{x_2} & x_1 e^{x_2} \end{bmatrix}$$

Для функции  $g(x) = \begin{bmatrix} x_1^3 + x_1 * e^{x_2} \\ x_1 - x_2 \\ x_1 x_2^2 \end{bmatrix}$  имеем:

$$J(g(x)) = \begin{bmatrix} 3x_1^2 + e^{x_2} & x_1 e^{x_2} \\ 1 & -1 \\ x_2^2 & 2x_1 x_2 \end{bmatrix}$$



# Численное дифференцирование

## Формулы численного дифференцирования

Иногда вывести аналитическую формулу для производной бывает достаточно сложно. В таких случаях удобно воспользоваться численным (приблизительным) дифференцированием. Рассмотрим алгоритм дифференцирования  $f(x)$  по  $x_i$ .

# Численное дифференцирование

## Формулы численного дифференцирования

Иногда вывести аналитическую формулу для производной бывает достаточно сложно. В таких случаях удобно воспользоваться численным (приблизительным) дифференцированием. Рассмотрим алгоритм дифференцирования  $f(x)$  по  $x_i$ .

- Подбирается малое приращение  $\varepsilon$  и определяется аргумент с приращением  $x_\varepsilon = (x_1, \dots, x_i + \varepsilon, \dots, x_n)$ .

# Численное дифференцирование

## Формулы численного дифференцирования

Иногда вывести аналитическую формулу для производной бывает достаточно сложно. В таких случаях удобно воспользоваться численным (приблизительным) дифференцированием. Рассмотрим алгоритм дифференцирования  $f(x)$  по  $x_i$ .

- Подбирается малое приращение  $\varepsilon$  и определяется аргумент с приращением  $x_\varepsilon = (x_1, \dots, x_i + \varepsilon, \dots, x_n)$ .
- Осуществляется приблизительный расчет производной по некоторой формуле:

$$\text{Forward difference: } \frac{\partial f(x)}{\partial x_i} \approx \frac{f(x_\varepsilon) - f(x)}{\varepsilon}$$

# Численное дифференцирование

## Формулы численного дифференцирования

Иногда вывести аналитическую формулу для производной бывает достаточно сложно. В таких случаях удобно воспользоваться численным (приблизительным) дифференцированием. Рассмотрим алгоритм дифференцирования  $f(x)$  по  $x_i$ .

- Подбирается малое приращение  $\varepsilon$  и определяется аргумент с приращением  $x_\varepsilon = (x_1, \dots, x_i + \varepsilon, \dots, x_n)$ .
- Осуществляется приблизительный расчет производной по некоторой формуле:

$$\text{Forward difference: } \frac{\partial f(x)}{\partial x_i} \approx \frac{f(x_\varepsilon) - f(x)}{\varepsilon}$$

$$\text{Central difference: } \frac{\partial f(x)}{\partial x_i} \approx \frac{f(x_\varepsilon) - f(x_{-\varepsilon})}{2\varepsilon}$$

# Численное дифференцирование

## Формулы численного дифференцирования

Иногда вывести аналитическую формулу для производной бывает достаточно сложно. В таких случаях удобно воспользоваться численным (приблизительным) дифференцированием. Рассмотрим алгоритм дифференцирования  $f(x)$  по  $x_i$ .

- Подбирается малое приращение  $\varepsilon$  и определяется аргумент с приращением  $x_\varepsilon = (x_1, \dots, x_i + \varepsilon, \dots, x_n)$ .
- Осуществляется приблизительный расчет производной по некоторой формуле:

$$\text{Forward difference: } \frac{\partial f(x)}{\partial x_i} \approx \frac{f(x_\varepsilon) - f(x)}{\varepsilon}$$

$$\text{Central difference: } \frac{\partial f(x)}{\partial x_i} \approx \frac{f(x_\varepsilon) - f(x_{-\varepsilon})}{2\varepsilon}$$

Аналогичные формулы существуют и для вторых производных. Используя их можно приблизительно рассчитать Градиент, Гессиан и Якобиан функций.

# Численная оптимизация

## Мотивация и классификация

Численная оптимизация позволяет находить приблизительный максимум или минимум функции без необходимости искать аналитическое решение.

# Численная оптимизация

## Мотивация и классификация

Численная оптимизация позволяет находить приблизительный максимум или минимум функции без необходимости искать аналитическое решение.

- Методы **локальной** оптимизации (BFGS, градиентный спуск) как правило работают достаточно быстро, но позволяют находить лишь локальные экстремумы. Методы **глобальной** оптимизации (генетический алгоритм, метод отжига – SA) позволяют найти несколько экстремумов, один из которых может оказаться глобальным. Однако, глобальная оптимизация обычно крайне затратна по времени.

# Численная оптимизация

## Мотивация и классификация

Численная оптимизация позволяет находить приблизительный максимум или минимум функции без необходимости искать аналитическое решение.

- Методы **локальной** оптимизации (BFGS, градиентный спуск) как правило работают достаточно быстро, но позволяют находить лишь локальные экстремумы. Методы **глобальной** оптимизации (генетический алгоритм, метод отжига – SA) позволяют найти несколько экстремумов, один из которых может оказаться глобальным. Однако, глобальная оптимизация обычно крайне затратна по времени.
- Методы локальной оптимизации часто опираются на Градиент (градиентный спуск, ADAM) или Гессиан функции (BFGS, BHHH). В последнем случае число итераций алгоритма, как правило, оказывается меньше, но время каждой итерации – больше, особенно, при большом числе оцениваемых параметров.



# Численная оптимизация

## Мотивация и классификация

Численная оптимизация позволяет находить приблизительный максимум или минимум функции без необходимости искать аналитическое решение.

- Методы **локальной** оптимизации (BFGS, градиентный спуск) как правило работают достаточно быстро, но позволяют находить лишь локальные экстремумы. Методы **глобальной** оптимизации (генетический алгоритм, метод отжига – SA) позволяют найти несколько экстремумов, один из которых может оказаться глобальным. Однако, глобальная оптимизация обычно крайне затратна по времени.
- Методы локальной оптимизации часто опираются на Градиент (градиентный спуск, ADAM) или Гессиан функции (BFGS, BHHH). В последнем случае число итераций алгоритма, как правило, оказывается меньше, но время каждой итерации – больше, особенно, при большом числе оцениваемых параметров.

Поскольку число оцениваемых параметров в эконометрических моделях, как правило, относительно невелико (в сравнении с моделями машинного обучения), то чаще используются алгоритмы, использующие информацию о Гессиане (BFGS, BHHH).

# Метод максимального правдоподобия

## Формулировка

- Рассмотрим i.i.d. выборку  $X = (X_1, \dots, X_n)$  из распределения с вектором параметров  $\theta = (\theta_1, \dots, \theta_m)$ . Оценка  $\theta$  может быть получена при помощи метода максимального правдоподобия (ММП оценка) за счет максимизации функции правдоподобия по всем элементам вектора  $\theta$ , то есть по каждому параметру:

$$\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m) = \underset{\theta_1, \dots, \theta_m}{\operatorname{argmax}} L(\theta_1, \dots, \theta_m; X)$$

# Метод максимального правдоподобия

## Формулировка

- Рассмотрим i.i.d. выборку  $X = (X_1, \dots, X_n)$  из распределения с вектором параметров  $\theta = (\theta_1, \dots, \theta_m)$ . Оценка  $\theta$  может быть получена при помощи метода максимального правдоподобия (ММП оценка) за счет максимизации функции правдоподобия по всем элементам вектора  $\theta$ , то есть по каждому параметру:

$$\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m) = \underset{\theta_1, \dots, \theta_m}{\operatorname{argmax}} L(\theta_1, \dots, \theta_m; X)$$

- ММП оценки состоятельные, асимптотически эффективные и асимптотически нормальные:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left([0 \quad \dots \quad 0]^T, i^{-1}(\theta)\right)$$

Где  $i(\theta) = E(-H(\ln L(\theta; X_1)))$  именуется информацией Фишера.

# Метод максимального правдоподобия

## Формулировка

- Рассмотрим i.i.d. выборку  $X = (X_1, \dots, X_n)$  из распределения с вектором параметров  $\theta = (\theta_1, \dots, \theta_m)$ . Оценка  $\theta$  может быть получена при помощи метода максимального правдоподобия (ММП оценка) за счет максимизации функции правдоподобия по всем элементам вектора  $\theta$ , то есть по каждому параметру:

$$\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m) = \underset{\theta_1, \dots, \theta_m}{\operatorname{argmax}} L(\theta_1, \dots, \theta_m; X)$$

- ММП оценки состоятельные, асимптотически эффективные и асимптотически нормальные:

$$\sqrt{n} (\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N} \left( [0 \quad \dots \quad 0]^T, i^{-1}(\theta) \right)$$

Где  $i(\theta) = E(-H(\ln L(\theta; X_1)))$  именуется информацией Фишера.

- Благодаря асимптотической нормальности для того, чтобы строить асимптотические доверительные интервалы и тестировать гипотезы с помощью ММП оценок достаточно найти их асимптотическую ковариационную матрицу:

$$\operatorname{As.Cov}(\hat{\theta}) = (ni(\theta))^{-1}$$

# Метод максимального правдоподобия

## Пример с нормальным распределением

- Имеется i.i.d. выборка  $X = (X_1, \dots, X_n)$  из нормального распределения  $\mathcal{N}(\mu, \sigma^2)$ . Логарифм функции правдоподобия будет иметь вид:

$$\ln L(\mu, \sigma; X) = \sum_{i=1}^n \ln \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i - \theta)^2}{2\sigma^2}} \right) = -0.5n \ln(2\pi) - n \ln(\sigma) - \frac{\sum_{i=1}^n (X_i - \theta)^2}{2\sigma^2}$$

# Метод максимального правдоподобия

## Пример с нормальным распределением

- Имеется i.i.d. выборка  $X = (X_1, \dots, X_n)$  из нормального распределения  $\mathcal{N}(\mu, \sigma^2)$ . Логарифм функции правдоподобия будет иметь вид:

$$\ln L(\mu, \sigma; X) = \sum_{i=1}^n \ln \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i - \theta)^2}{2\sigma^2}} \right) = -0.5n \ln(2\pi) - n \ln(\sigma) - \frac{\sum_{i=1}^n (X_i - \theta)^2}{2\sigma^2}$$

- Максимизируя по  $\mu$  и  $\sigma^2$  получаем ММП оценки соответствующих параметров:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2$$

# Метод максимального правдоподобия

## Пример с нормальным распределением

- Имеется i.i.d. выборка  $X = (X_1, \dots, X_n)$  из нормального распределения  $\mathcal{N}(\mu, \sigma^2)$ . Логарифм функции правдоподобия будет иметь вид:

$$\ln L(\mu, \sigma; X) = \sum_{i=1}^n \ln \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i - \theta)^2}{2\sigma^2}} \right) = -0.5n \ln(2\pi) - n \ln(\sigma) - \frac{\sum_{i=1}^n (X_i - \theta)^2}{2\sigma^2}$$

- Максимизируя по  $\mu$  и  $\sigma^2$  получаем ММП оценки соответствующих параметров:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2$$

- Найдем информацию Фишера и асимптотическую ковариационную матрицу:

$$i([\hat{\mu}, \hat{\sigma}^2]^T) = \begin{bmatrix} \sigma^{-1} & 0 \\ 0 & 0.5\sigma^{-4} \end{bmatrix} \implies As.Cov([\hat{\mu}, \hat{\sigma}^2]^T) = \begin{bmatrix} \sigma & 0 \\ 0 & 2\sigma^4 \end{bmatrix}$$

# Метод максимального правдоподобия

## Пример с нормальным распределением

- Имеется i.i.d. выборка  $X = (X_1, \dots, X_n)$  из нормального распределения  $\mathcal{N}(\mu, \sigma^2)$ . Логарифм функции правдоподобия будет иметь вид:

$$\ln L(\mu, \sigma; X) = \sum_{i=1}^n \ln \left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i - \theta)^2}{2\sigma^2}} \right) = -0.5n \ln(2\pi) - n \ln(\sigma) - \frac{\sum_{i=1}^n (X_i - \theta)^2}{2\sigma^2}$$

- Максимизируя по  $\mu$  и  $\sigma^2$  получаем ММП оценки соответствующих параметров:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2$$

- Найдем информацию Фишера и асимптотическую ковариационную матрицу:

$$i \left( [\hat{\mu}, \hat{\sigma}^2]^T \right) = \begin{bmatrix} \sigma^{-1} & 0 \\ 0 & 0.5\sigma^{-4} \end{bmatrix} \implies As.Cov \left( [\hat{\mu}, \hat{\sigma}^2]^T \right) = \begin{bmatrix} \sigma & 0 \\ 0 & 2\sigma^4 \end{bmatrix}$$

- В данном случае оценку асимптотической ковариационной матрицы можно получить за счет обычной замены  $\sigma^2$  на  $\hat{\sigma}^2$ . Однако, на практике асимптотическую ковариационную матрицу может быть вывести крайне сложно, что мотивирует поиск альтернативных подходов к оцениванию.



# Метод максимального правдоподобия

## Оценивание асимптотической ковариационной матрицы

Существуют три основных подхода к расчету асимптотической ковариационной матрицы:

# Метод максимального правдоподобия

## Оценивание асимптотической ковариационной матрицы

Существуют три основных подхода к расчету асимптотической ковариационной матрицы:

- Обратный Гессиан – стандартный вариант:

$$\widehat{As.Cov}(\hat{\theta}) = H_*, \quad H_* = H^{-1} \left( -\ln L(\hat{\theta}; X) \right)$$

# Метод максимального правдоподобия

## Оценивание асимптотической ковариационной матрицы

Существуют три основных подхода к расчету асимптотической ковариационной матрицы:

- Обратный Гессиан – стандартный вариант:

$$\widehat{As.Cov}(\hat{\theta}) = H_*, \quad H_* = H^{-1} \left( -\ln L(\hat{\theta}; X) \right)$$

- Произведение Якобианов (ВННН) – применяется, когда сложно посчитать Гессиан:

$$\widehat{As.Cov}(\hat{\theta}) = \left( J_*^T J_* \right)^{-1}, \quad J_* = J \left( \left( \ln L(\hat{\theta}; X_1), \dots, \ln L(\hat{\theta}; X_n) \right)^T \right)$$

# Метод максимального правдоподобия

## Оценивание асимптотической ковариационной матрицы

Существуют три основных подхода к расчету асимптотической ковариационной матрицы:

- Обратный Гессиан – стандартный вариант:

$$\widehat{As.Cov}(\hat{\theta}) = H_*, \quad H_* = H^{-1} \left( -\ln L(\hat{\theta}; X) \right)$$

- Произведение Якобианов (ВННН) – применяется, когда сложно посчитать Гессиан:

$$\widehat{As.Cov}(\hat{\theta}) = \left( J_*^T J_* \right)^{-1}, \quad J_* = J \left( \left( \ln L(\hat{\theta}; X_1), \dots, \ln L(\hat{\theta}; X_n) \right)^T \right)$$

- Сэндвич – устойчив к нарушению допущения о распределении (QMLE):

$$\widehat{As.Cov}(\hat{\theta}) = H_*^{-1} J_*^T J_* H_*^{-1}$$

# Метод максимального правдоподобия

## Оценивание асимптотической ковариационной матрицы

Существуют три основных подхода к расчету асимптотической ковариационной матрицы:

- Обратный Гессиан – стандартный вариант:

$$\widehat{As.Cov}(\hat{\theta}) = H_*, \quad H_* = H^{-1} \left( -\ln L(\hat{\theta}; X) \right)$$

- Произведение Якобианов (ВННН) – применяется, когда сложно посчитать Гессиан:

$$\widehat{As.Cov}(\hat{\theta}) = \left( J_*^T J_* \right)^{-1}, \quad J_* = J \left( \left( \ln L(\hat{\theta}; X_1), \dots, \ln L(\hat{\theta}; X_n) \right)^T \right)$$

- Сэндвич – устойчив к нарушению допущения о распределении (QMLE):

$$\widehat{As.Cov}(\hat{\theta}) = H_*^{-1} J_*^T J_* H_*^{-1}$$

- Бутстрап – когда численные производные плохо считаются или оценка  $\hat{\theta}$  получена некоторым необычным методом (не ММП) и ее асимптотическое распределение неизвестно. Достаточно сформировать  $b$  выборок без возвращения из исходной выборки и по каждой из этих выборок оценить вектора параметров  $\hat{\theta}^{(i)}$ , где  $i \in \{1, \dots, b\}$ . Затем асимптотическая ковариационная матрица оценивается как обычная выборочная ковариация по соответствующей выборке  $(\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(b)})$ .

- Рассмотрим функцию  $g(\theta)$  и ММП оценку  $\hat{\theta}$ . В силу дельта метода оценка  $g(\hat{\theta})$  будет асимптотически нормальной, причем:

$$As.E(g(\hat{\theta})) = g(\theta), \quad As.Cov(g(\hat{\theta})) = \nabla g(\theta)^T As.Cov(\hat{\theta}) \nabla g(\theta)$$

- Рассмотрим функцию  $g(\theta)$  и ММП оценку  $\hat{\theta}$ . В силу дельта метода оценка  $g(\hat{\theta})$  будет асимптотически нормальной, причем:

$$As.E(g(\hat{\theta})) = g(\theta), \quad As.Cov(g(\hat{\theta})) = \nabla g(\theta)^T As.Cov(\hat{\theta}) \nabla g(\theta)$$

- Оценка асимптотической ковариационной матрицы будет иметь вид:

$$\widehat{As.Cov}(g(\hat{\theta})) = \nabla g(\hat{\theta})^T \widehat{As.Cov}(\hat{\theta}) \nabla g(\hat{\theta})$$

- Рассмотрим функцию  $g(\theta)$  и ММП оценку  $\hat{\theta}$ . В силу дельта метода оценка  $g(\hat{\theta})$  будет асимптотически нормальной, причем:

$$As.E(g(\hat{\theta})) = g(\theta), \quad As.Cov(g(\hat{\theta})) = \nabla g(\theta)^T As.Cov(\hat{\theta}) \nabla g(\theta)$$

- Оценка асимптотической ковариационной матрицы будет иметь вид:

$$\widehat{As.Cov}(g(\hat{\theta})) = \nabla g(\hat{\theta})^T \widehat{As.Cov}(\hat{\theta}) \nabla g(\hat{\theta})$$

- Для тестирования гипотез и построения доверительных интервалов для  $g(\theta)$  достаточно воспользоваться полученной оценкой асимптотической ковариационной матрицы.



# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тестирование гипотезы о параметре

- Для тестирования гипотезы о параметре  $H_0 : \theta_i = \theta_i^0$  сперва необходимо вычислить тестовую статистику:

$$T(X) = \frac{\hat{\theta}_i - \theta_i^0}{\sqrt{\widehat{As.Var}(\hat{\theta}_i)}}, \quad \widehat{As.Var}(\hat{\theta}_i) = \widehat{As.Cov}(\hat{\theta})_{i,i}$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тестирование гипотезы о параметре

- Для тестирования гипотезы о параметре  $H_0 : \theta_i = \theta_i^0$  сперва необходимо вычислить тестовую статистику:

$$T(X) = \frac{\hat{\theta}_i - \theta_i^0}{\sqrt{\widehat{As.Var}(\hat{\theta}_i)}}, \quad \widehat{As.Var}(\hat{\theta}_i) = \widehat{As.Cov}(\hat{\theta})_{i,i}$$

- Далее в силу асимптотической нормальности ММП оценок и теоремы Slutsky получаем, что:

$$T(X)|H_0 \xrightarrow{d} \mathcal{N}(0, 1)$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тестирование гипотезы о параметре

- Для тестирования гипотезы о параметре  $H_0 : \theta_i = \theta_i^0$  сперва необходимо вычислить тестовую статистику:

$$T(X) = \frac{\hat{\theta}_i - \theta_i^0}{\sqrt{\widehat{As.Var}(\hat{\theta}_i)}}, \quad \widehat{As.Var}(\hat{\theta}_i) = \widehat{As.Cov}(\hat{\theta})_{i,i}$$

- Далее в силу асимптотической нормальности ММП оценок и теоремы Slutsky получаем, что:

$$T(X)|H_0 \xrightarrow{d} \mathcal{N}(0, 1)$$

- Поэтому p-value считается с помощью функции распределения стандартного нормального распределения:

$$\text{p-value} = 2 \min(\Phi(T(X)), 1 - \Phi(T(X)))$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тестирование гипотезы о функции от параметров

- Для тестирования гипотезы о функции от параметров  $H_0 : g(\theta) = g^0$  сперва необходимо вычислить тестовую статистику:

$$T(X) = \frac{g(\hat{\theta}) - g^0}{\sqrt{\widehat{As.Var}(g(\hat{\theta}))}}, \quad \widehat{As.Var}(g(\hat{\theta})) = \widehat{As.Cov}(g(\hat{\theta}))_{i,i}$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тестирование гипотезы о функции от параметров

- Для тестирования гипотезы о функции от параметров  $H_0 : g(\theta) = g^0$  сперва необходимо вычислить тестовую статистику:

$$T(X) = \frac{g(\hat{\theta}) - g^0}{\sqrt{\widehat{As.Var}(g(\hat{\theta}))}}, \quad \widehat{As.Var}(g(\hat{\theta})) = \widehat{As.Cov}(g(\hat{\theta}))_{i,i}$$

- Далее в силу дельта метода и теоремы Слущкого получаем, что:

$$T(X)|H_0 \xrightarrow{d} \mathcal{N}(0, 1)$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тестирование гипотезы о функции от параметров

- Для тестирования гипотезы о функции от параметров  $H_0 : g(\theta) = g^0$  сперва необходимо вычислить тестовую статистику:

$$T(X) = \frac{g(\hat{\theta}) - g^0}{\sqrt{\widehat{As.Var}(g(\hat{\theta}))}}, \quad \widehat{As.Var}(g(\hat{\theta})) = \widehat{As.Cov}(g(\hat{\theta}))_{i,i}$$

- Далее в силу дельта метода и теоремы Slutsky получаем, что:

$$T(X)|H_0 \xrightarrow{d} \mathcal{N}(0, 1)$$

- Поэтому p-value считается с помощью функции распределения стандартного нормального распределения:

$$\text{p-value} = 2 \min(\Phi(T(X)), 1 - \Phi(T(X)))$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест отношения правдоподобий

- Тестируется гипотеза об  $r$  ограничениях на параметры:

$$H_0 : \begin{cases} g_1(\theta) = 0 \\ \dots \\ g_r(\theta) = 0 \end{cases}, \quad g(\theta) = \begin{bmatrix} g_1(\theta) \\ \dots \\ g_r(\theta) \end{bmatrix}$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест отношения правдоподобий

- Тестируется гипотеза об  $r$  ограничениях на параметры:

$$H_0 : \begin{cases} g_1(\theta) = 0 \\ \dots \\ g_r(\theta) = 0 \end{cases}, \quad g(\theta) = \begin{bmatrix} g_1(\theta) \\ \dots \\ g_r(\theta) \end{bmatrix}$$

- Для проведения теста отношения правдоподобий (LR-тест) необходимо сперва найти ММП оценки  $\hat{\theta}$  без учета ограничений, то есть обычным образом.



# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест отношения правдоподобий

- Тестируется гипотеза об  $r$  ограничениях на параметры:

$$H_0 : \begin{cases} g_1(\theta) = 0 \\ \dots \\ g_r(\theta) = 0 \end{cases}, \quad g(\theta) = \begin{bmatrix} g_1(\theta) \\ \dots \\ g_r(\theta) \end{bmatrix}$$

- Для проведения теста отношения правдоподобий (LR-тест) необходимо сперва найти ММП оценки  $\hat{\theta}$  без учета ограничений, то есть обычным образом.
- Затем находятся оценки с учетом ограничений, то есть за счет максимизации функции правдоподобия с учетом ограничений на параметры, накладываемых нулевой гипотезой  $\hat{\theta}^R$ .

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест отношения правдоподобий

- Тестируется гипотеза об  $r$  ограничениях на параметры:

$$H_0 : \begin{cases} g_1(\theta) = 0 \\ \dots \\ g_r(\theta) = 0 \end{cases}, \quad g(\theta) = \begin{bmatrix} g_1(\theta) \\ \dots \\ g_r(\theta) \end{bmatrix}$$

- Для проведения теста отношения правдоподобий (LR-тест) необходимо сперва найти ММП оценки  $\hat{\theta}$  без учета ограничений, то есть обычным образом.
- Затем находятся оценки с учетом ограничений, то есть за счет максимизации функции правдоподобия с учетом ограничений на параметры, накладываемых нулевой гипотезой  $\hat{\theta}^R$ .
- Далее, рассчитываются тестовая статистика и p-value:

$$T(X) = 2 \left( \ln L(\hat{\theta}) - \ln L(\hat{\theta}^R) \right), \quad T(X)|H_0 \xrightarrow{d} \chi^2(r) \implies \text{p-value} = 1 - F_{\chi^2(r)}(T(X))$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест Вальда

- Тестируемая гипотеза, асимптотическое распределение тестовой статистики при верной нулевой гипотезе и способ расчета  $p$ -value аналогичны LR-тесту.

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест Вальда

- Тестируемая гипотеза, асимптотическое распределение тестовой статистики при верной нулевой гипотезе и способ расчета p-value аналогичны LR-тесту.
- Преимущество теста Вальда над LR-тестом заключается в том, что для расчета тестовой статистики достаточно оценить лишь модель без ограничений:

$$T(X) = g(\hat{\theta})^T \left( \nabla g(\hat{\theta})^T \widehat{As.Cov}(\hat{\theta}) \nabla g(\hat{\theta}) \right)^{-1} g(\hat{\theta})$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест Вальда

- Тестируемая гипотеза, асимптотическое распределение тестовой статистики при верной нулевой гипотезе и способ расчета p-value аналогичны LR-тесту.
- Преимущество теста Вальда над LR-тестом заключается в том, что для расчета тестовой статистики достаточно оценить лишь модель без ограничений:

$$T(X) = g(\hat{\theta})^T \left( \nabla g(\hat{\theta})^T \widehat{As.Cov}(\hat{\theta}) \nabla g(\hat{\theta}) \right)^{-1} g(\hat{\theta})$$

- Данный тест удобно применять в случаях, когда на модель накладываются сложные ограничений, максимизация правдоподобия при условии которых является затруднительной технической задачей.

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест множителей Лагранжа

- Тестируемая гипотеза, асимптотическое распределение тестовой статистики при верной нулевой гипотезе и способ расчета  $p$ -value аналогичны LR-тесту.

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест множителей Лагранжа

- Тестируемая гипотеза, асимптотическое распределение тестовой статистики при верной нулевой гипотезе и способ расчета p-value аналогичны LR-тесту.
- Преимущество теста множителей Лагранжа (LM-тест) над LR-тестом и тестом Вальда заключается в том, что для расчета тестовой статистики достаточно оценить лишь модель с ограничениями:

$$T(X) = \nabla \ln L(\hat{\theta}^R; X)^T \left( \widehat{As.Cov}_F \left( \hat{\theta}^R \right) \right)^{-1} \nabla \ln L(\hat{\theta}^R; X)$$

# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест множителей Лагранжа

- Тестируемая гипотеза, асимптотическое распределение тестовой статистики при верной нулевой гипотезе и способ расчета p-value аналогичны LR-тесту.
- Преимущество теста множителей Лагранжа (LM-тест) над LR-тестом и тестом Вальда заключается в том, что для расчета тестовой статистики достаточно оценить лишь модель с ограничениями:

$$T(X) = \nabla \ln L(\hat{\theta}^R; X)^T \left( \widehat{As.Cov}_F \left( \hat{\theta}^R \right) \right)^{-1} \nabla \ln L(\hat{\theta}^R; X)$$

Где под  $\widehat{As.Cov}_F \left( \hat{\theta}^R \right)$  подразумевается, что мы используем любую из формул для оценивания асимптотической ковариационной матриц, но при этом подставляем оценки ограниченной модели в выражение для полной модели.



# Тестирование гипотез с помощью оценок максимального правдоподобия

## Тест множителей Лагранжа

- Тестируемая гипотеза, асимптотическое распределение тестовой статистики при верной нулевой гипотезе и способ расчета p-value аналогичны LR-тесту.
- Преимущество теста множителей Лагранжа (LM-тест) над LR-тестом и тестом Вальда заключается в том, что для расчета тестовой статистики достаточно оценить лишь модель с ограничениями:

$$T(X) = \nabla \ln L(\hat{\theta}^R; X)^T \left( \widehat{As.Cov}_F \left( \hat{\theta}^R \right) \right)^{-1} \nabla \ln L(\hat{\theta}^R; X)$$

Где под  $\widehat{As.Cov}_F \left( \hat{\theta}^R \right)$  подразумевается, что мы используем любую из формул для оценивания асимптотической ковариационной матриц, но при этом подставляем оценки ограниченной модели в выражение для полной модели.

- Данный тест удобно применять в случаях, когда ограниченная модель оценивается проще, чем полная.