Микроэконометрика

Вступительная лекция

Потанин Богдан Станиславович

доцент, научный сотрудник, кандидат экономических наук

2022-2023

Оценка складывается из четырех элементов:

О курсе

Организационная информация

Оценка складывается из четырех элементов:

- Домашнее задание №1 – 30%

О курсе

Организационная информация

Оценка складывается из четырех элементов:

- Домашнее задание №1 – 30%
- Домашнее задание №2 – 30%

О курсе
Организационная информация

Оценка складывается из четырех элементов:

- Домашнее задание №1 – 30%
- Домашнее задание №2 – 30%
- Презентация – 20%

Оценка складывается из четырех элементов:

- Домашнее задание №1 – 30%
- Домашнее задание №2 – 30%
- Презентация – 20%
- Экзамен – 20%

Презентацию можно делать либо индивидуально, либо в паре с другим студентом.

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

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

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

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

Численное дифференцирование

Градиент, Якобиан и Гессиан

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

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

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

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

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

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

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

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

- 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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:

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

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.

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 $\varepsilon > 0$ и определяется аргумент с приращением $x_\varepsilon = (x_1, ..., x_i + \varepsilon, ..., x_n)$.

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 $\varepsilon > 0$ и определяется аргумент с приращением $x_\varepsilon = (x_1, ..., x_i + \varepsilon, ..., x_n)$.
- Осуществляется приблизительный расчет производной по некоторой формуле:

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

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 $\varepsilon > 0$ и определяется аргумент с приращением $x_\varepsilon = (x_1, ..., x_i + \varepsilon, ..., x_n)$.
- Осуществляется приблизительный расчет производной по некоторой формуле:

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

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

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 $\varepsilon > 0$ и определяется аргумент с приращением $x_\varepsilon = (x_1, ..., x_i + \varepsilon, ..., x_n)$.
- Осуществляется приблизительный расчет производной по некоторой формуле:

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

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

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

Продифференцируем функцию $f(x) = x_1^2 x_2$ в точке $x_1 = 3$, $x_2 = 5$ двумя способами.

Продифференцируем функцию $f(x) = x_1^2 x_2$ в точке $x_1 = 3$, $x_2 = 5$ двумя способами.

- Аналитическое дифференцирование:

$$\frac{\partial f(x)}{\partial x_1} = 2x_1x_2 \implies \frac{\partial f(3,5)}{\partial x_1} = 2 \times 3 \times 5 = 30$$

Численное дифференцирование

Пример

Продифференцируем функцию $f(x) = x_1^2 x_2$ в точке $x_1 = 3$, $x_2 = 5$ двумя способами.

- Аналитическое дифференцирование:

$$\frac{\partial f(x)}{\partial x_1} = 2x_1x_2 \implies \frac{\partial f(3,5)}{\partial x_1} = 2 \times 3 \times 5 = 30$$

$$\frac{\partial f(x)}{\partial x_2} = x_1^2 \implies \frac{\partial f(3,5)}{\partial x_2} = 3^2 = 9$$

Продифференцируем функцию $f(x) = x_1^2 x_2$ в точке $x_1 = 3$, $x_2 = 5$ двумя способами.

- Аналитическое дифференцирование:

$$\frac{\partial f(x)}{\partial x_1} = 2x_1x_2 \implies \frac{\partial f(3,5)}{\partial x_1} = 2 \times 3 \times 5 = 30$$

$$\frac{\partial f(x)}{\partial x_2} = x_1^2 \implies \frac{\partial f(3,5)}{\partial x_2} = 3^2 = 9$$

- Численное дифференцирование методом forward difference с приращением $\varepsilon = 0.001$:

$$\frac{\partial f(3,5)}{\partial x_1} \approx \frac{f(3+0.001,5) - f(3,5)}{0.001} = \frac{(3+0.001)^2 \times 5 - 3^2 \times 5}{0.001} = 30.005$$

Численное дифференцирование

Пример

Продифференцируем функцию $f(x) = x_1^2 x_2$ в точке $x_1 = 3$, $x_2 = 5$ двумя способами.

- Аналитическое дифференцирование:

$$\frac{\partial f(x)}{\partial x_1} = 2x_1x_2 \implies \frac{\partial f(3,5)}{\partial x_1} = 2 \times 3 \times 5 = 30$$
$$\frac{\partial f(x)}{\partial x_2} = x_1^2 \implies \frac{\partial f(3,5)}{\partial x_2} = 3^2 = 9$$

- Численное дифференцирование методом forward difference с приращением $\varepsilon = 0.001$:

$$\frac{\partial f(3,5)}{\partial x_1} \approx \frac{f(3+0.001,5) - f(3,5)}{0.001} = \frac{(3+0.001)^2 \times 5 - 3^2 \times 5}{0.001} = 30.005$$
$$\frac{\partial f(3,5)}{\partial x_2} \approx \frac{f(3,5+0.001) - f(3,5)}{0.001} = \frac{3^2 \times (5+0.001) - 3^2 \times 5}{0.001} = 9$$

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

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 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 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).

Алгоритм **градиентного спуска** является одним из простейших численных методов нахождения минимума функции.

- Выбираем произвольную начальную точку x_0.

Алгоритм **градиентного спуска** является одним из простейших численных методов нахождения минимума функции.

- Выбираем произвольную начальную точку x_0.
- Считаем градиент функции в этой точке $\nabla f(x_0)$.

Алгоритм **градиентного спуска** является одним из простейших численных методов нахождения минимума функции.

- Выбираем произвольную начальную точку x_0.
- Считаем градиент функции в этой точке $\nabla f(x_0)$.
- Переходим в новую точку $x_1 = x_0 - \alpha\nabla f(x_0)$, где α – малая положительная константа.

Алгоритм **градиентного спуска** является одним из простейших численных методов нахождения минимума функции.

- Выбираем произвольную начальную точку x_0.
- Считаем градиент функции в этой точке $\nabla f(x_0)$.
- Переходим в новую точку $x_1 = x_0 - \alpha\nabla f(x_0)$, где α – малая положительная константа.
- Повторяем процедуру до тех пор, пока не будут соблюдены условия остановки, например, о том, что $|\nabla f(x_0)| < \varepsilon$, где ε – маленькое положительное число.

Алгоритм **градиентного спуска** является одним из простейших численных методов нахождения минимума функции.

- Выбираем произвольную начальную точку x_0.
- Считаем градиент функции в этой точке $\nabla f(x_0)$.
- Переходим в новую точку $x_1 = x_0 - \alpha \nabla f(x_0)$, где α – малая положительная константа.
- Повторяем процедуру до тех пор, пока не будут соблюдены условия остановки, например, о том, что $|\nabla f(x_0)| < \varepsilon$, где ε – маленькое положительное число.

Нетрудно показать аналитически, что функция $f(x) = x^2 - 2x$ достигает минимума в точке $x^* = 1$. В качестве альтернативы аналитическому решению попробуем приблизиться к минимуму с помощью 10 итераций описанного алгоритма, произвольным образом полагая $x_0 = 3$ и $\alpha = 0.2$.

i	0	1	2	3	4	5	6	7	8	9	10
x_i	3	2.20	1.72	1.43	1.26	1.16	1.09	1.06	1.03	1.02	**1.01**
$\nabla f(x_i)$	4	2.40	1.44	0.86	0.52	0.31	0.19	0.11	0.07	0.04	0.02
$f(x_i)$	3	0.44	-0.48	-0.81	-0.93	-0.98	-0.99	-1.00	-1.00	-1.00	-1.00

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

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

- Рассмотрим i.i.d. выборку $X = (X_1, ..., X_n)$ из распределения с вектором параметров $\theta = (\theta_1, ..., \theta_m)$. Оценка θ может быть получена при помощи метода максимального правдоподобия (ММП оценка) за счет максимизации функции правдоподобия по всем элементам вектора θ, то есть по каждому параметру:
$$\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_m) = \underset{\theta_1, ..., \theta_m}{\operatorname{argmax}}\, L(\theta_1, ..., \theta_m; X)$$
- ММП оценки состоятельные, асимптотически эффективные и асимптотически нормальные:
$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0 & ... & 0\end{bmatrix}^T, i^{-1}(\theta)\right)$$
Где $i(\theta) = E\left(-H\left(\ln L(\theta; X_1)\right)\right)$ именуется информацией Фишера.

Метод максимального правдоподобия

Формулировка

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

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

- ММП оценки состоятельные, асимптотически эффективные и асимптотически нормальные:

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0 & \dots & 0\end{bmatrix}^T, i^{-1}(\theta)\right)$$

Где $i(\theta) = E\left(-H\left(\ln L(\theta; X_1)\right)\right)$ 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:

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

Метод максимального правдоподобия

Пример с нормальным распределением

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

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

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

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

- Максимизируя по μ и σ^2 получаем ММП оценки соответствующих параметров:

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

Метод максимального правдоподобия

Пример с нормальным распределением

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

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

- Максимизируя по μ и σ^2 получаем ММП оценки соответствующих параметров:

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

- Найдем информацию Фишера и асимптотическую ковариационную матрицу:

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

Метод максимального правдоподобия

Пример с нормальным распределением

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

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

- Максимизируя по μ и σ^2 получаем ММП оценки соответствующих параметров:

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

- Найдем информацию Фишера и асимптотическую ковариационную матрицу:

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

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

Существуют три основных подхода к расчету асимптотической ковариационной матрицы:

Существуют три основных подхода к расчету асимптотической ковариационной матрицы:

- Обратный Гессиан – стандартный вариант:

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

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 (BHHH) – применяется, когда сложно посчитать Гессиан:

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

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 (BHHH) – применяется, когда сложно посчитать Гессиан:

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

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

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

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 (BHHH) – применяется, когда сложно посчитать Гессиан:

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

- 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 $i \in \{1, ..., b\}$. Затем асимптотическая ковариационная матрица оценивается как обычная выборочная ковариация по соответствующей выборке $(\hat{\theta}^{(1)}, ..., \hat{\theta}^{(b)})$.

Вернемся к примеру с выборкой из нормального распределения и для удобства запишем функцию правдоподобия для одного наблюдения:

$$\ln L\left(\mu, \sigma^2; X_i\right) = -0.5\ln(2\pi) - \ln(\sigma) - \frac{(X_i - \mu)^2}{2\sigma^2}$$

Вернемся к примеру с выборкой из нормального распределения и для удобства запишем функцию правдоподобия для одного наблюдения:

$$\ln L\left(\mu, \sigma^2; X_i\right) = -0.5\ln(2\pi) - \ln(\sigma) - \frac{(X_i-\mu)^2}{2\sigma^2}$$

Посчитаем частные производные:

$$\frac{\partial \ln L\left(\mu, \sigma^2; X_i\right)}{\partial \mu} = \frac{X_i-\mu}{\sigma^2}, \quad \frac{\partial \ln L\left(\mu, \sigma^2; X_i\right)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{(X_i-\mu)^2}{2\sigma^4}$$

Вернемся к примеру с выборкой из нормального распределения и для удобства запишем функцию правдоподобия для одного наблюдения:

$$\ln L\left(\mu, \sigma^2; X_i\right) = -0.5\ln(2\pi) - \ln(\sigma) - \frac{(X_i-\mu)^2}{2\sigma^2}$$

Посчитаем частные производные:

$$\frac{\partial \ln L\left(\mu, \sigma^2; X_i\right)}{\partial \mu} = \frac{X_i-\mu}{\sigma^2}, \quad \frac{\partial \ln L\left(\mu, \sigma^2; X_i\right)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{(X_i-\mu)^2}{2\sigma^4}$$

Выпишем оценку Якобиана, заменяя истинные значения параметров их оценками:

$$J_* = \begin{bmatrix} \frac{X_1-\hat{\mu}}{\hat{\sigma}^2} & \frac{(X_1-\hat{\mu})^2}{2\hat{\sigma}^4} - \frac{1}{2\hat{\sigma}^2} \\ \vdots & \\ \frac{X_n-\hat{\mu}}{\hat{\sigma}^2} & \frac{(X_n-\hat{\mu})^2}{2\hat{\sigma}^4} - \frac{1}{2\hat{\sigma}^2} \end{bmatrix}$$

Вернемся к примеру с выборкой из нормального распределения и для удобства запишем функцию правдоподобия для одного наблюдения:

$$\ln L\left(\mu,\sigma^2;X_i\right)=-0.5\ln(2\pi)-\ln(\sigma)-\frac{(X_i-\mu)^2}{2\sigma^2}$$

Вернемся к примеру с выборкой из нормального распределения и для удобства запишем функцию правдоподобия для одного наблюдения:

$$\ln L\left(\mu,\sigma^2;X_i\right) = -0.5\ln(2\pi) - \ln(\sigma) - \frac{(X_i-\mu)^2}{2\sigma^2}$$

Посчитаем вторые производные:

$$\frac{\partial^2 \ln L\left(\mu,\sigma^2;X_i\right)}{\partial^2\mu} = -\frac{1}{\sigma^2},\quad \frac{\partial^2 \ln L\left(\mu,\sigma^2;X_i\right)}{\partial^2\sigma^2} = \frac{1}{2\sigma^4} - \frac{(X_i-\mu)^2}{\sigma^6},\quad \frac{\partial^2 \ln L\left(\mu,\sigma^2;X_i\right)}{\partial\mu\partial\sigma^2} = \frac{\mu - X_i}{\sigma^4}$$

Вернемся к примеру с выборкой из нормального распределения и для удобства запишем функцию правдоподобия для одного наблюдения:

$$\ln L\left(\mu, \sigma^2; X_i\right) = -0.5\ln(2\pi) - \ln(\sigma) - \frac{(X_i-\mu)^2}{2\sigma^2}$$

Посчитаем вторые производные:

$$\frac{\partial^2 \ln L\left(\mu, \sigma^2; X_i\right)}{\partial^2\mu} = -\frac{1}{\sigma^2}, \quad \frac{\partial^2 \ln L\left(\mu, \sigma^2; X_i\right)}{\partial^2\sigma^2} = \frac{1}{2\sigma^4} - \frac{(X_i-\mu)^2}{\sigma^6}, \quad \frac{\partial^2 \ln L\left(\mu, \sigma^2; X_i\right)}{\partial\mu\partial\sigma^2} = \frac{\mu - X_i}{\sigma^4}$$

Выпишем оценку (скорректированного на знак) обратного Гессиана, заменяя истинные значения параметров и складывая вторые производные, посчитанные по отдельным наблюдениям:

$$H_* = -\begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & \frac{\sum\limits_{i=1}^{n}(\hat{\mu}-X_i)}{\hat{\sigma}^4} \\ \frac{\sum\limits_{i=1}^{n}(\hat{\mu}-X_i)}{\hat{\sigma}^4} & \frac{n}{2\hat{\sigma}^4} - \frac{\sum\limits_{i=1}^{n}(X_i-\hat{\mu})^2}{\hat{\sigma}^6} \end{bmatrix}^{-1} = -\begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{2n}{\hat{\sigma}^4} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\hat{\sigma}^2}{n} & 0 \\ 0 & \frac{\hat{\sigma}^2}{2n} \end{bmatrix}$$

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

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

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

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

- Оценка асимптотической ковариационной матрицы будет иметь вид:

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

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

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

- Оценка асимптотической ковариационной матрицы будет иметь вид:

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

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

- Вернемся к примеру с выборкой из нормального распределения и рассмотрим функцию:

$$g(\mu,\sigma^2)=\mu e^{2\sigma^2}$$

- Вернемся к примеру с выборкой из нормального распределения и рассмотрим функцию:

$$g(\mu,\sigma^2)=\mu e^{2\sigma^2}$$

- Найдем градиент данной функции:

$$\nabla^T g(\mu,\sigma^2)=\left(e^{2\sigma^2},2\mu e^{2\sigma^2}\right)^T$$

- Вернемся к примеру с выборкой из нормального распределения и рассмотрим функцию:

$$g(\mu, \sigma^2) = \mu e^{2\sigma^2}$$

- Найдем градиент данной функции:

$$\nabla^T g(\mu, \sigma^2) = \left(e^{2\sigma^2}, 2\mu e^{2\sigma^2}\right)^T$$

- Выпишем выражение для оценки асимптотической ковариационной матрицы:

$$\widehat{As.Cov}\left(g(\hat{\mu}, \hat{\sigma}^2)\right) = \left(e^{2\hat{\sigma}^2}, 2\hat{\mu}e^{2\hat{\sigma}^2}\right) \widehat{As.Cov}(\hat{\mu}, \hat{\sigma}^2) \left(e^{2\hat{\sigma}^2}, 2\hat{\mu}e^{2\hat{\sigma}^2}\right)^T$$

- Вернемся к примеру с выборкой из нормального распределения и рассмотрим функцию:

$$g(\mu, \sigma^2) = \mu e^{2\sigma^2}$$

- Найдем градиент данной функции:

$$\nabla^T g(\mu, \sigma^2) = \left(e^{2\sigma^2}, 2\mu e^{2\sigma^2}\right)^T$$

- Выпишем выражение для оценки асимптотической ковариационной матрицы:

$$\widehat{As.Cov}\left(g(\hat{\mu}, \hat{\sigma}^2)\right) = \left(e^{2\hat{\sigma}^2}, 2\hat{\mu} e^{2\hat{\sigma}^2}\right) \widehat{As.Cov}(\hat{\mu}, \hat{\sigma}^2) \left(e^{2\hat{\sigma}^2}, 2\hat{\mu} e^{2\hat{\sigma}^2}\right)^T$$

- Если, например, асимптотическая ковариационной матрица была оценена методом обратного Гессиана, то получаем:

$$\widehat{As.Cov}\left(g(\hat{\theta})\right) = \left(e^{2\hat{\sigma}^2}, 2\hat{\mu} e^{2\hat{\sigma}^2}\right) \begin{bmatrix} \frac{\hat{\sigma}^2}{n} & 0 \\ 0 & \frac{\hat{\sigma}^2}{2n} \end{bmatrix} \left(e^{2\hat{\sigma}^2}, 2\hat{\mu} e^{2\hat{\sigma}^2}\right)^T$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- Поэтому p-value считается с помощью функции распределения стандартного нормального распределения.
- Например, в случае двухсторонней альтернативы $H_1 : \theta_i \neq \theta_i^0$ получаем:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- Поэтому p-value считается с помощью функции распределения стандартного нормального распределения.
- Например, в случае двухсторонней альтернативы $H_1 : \theta_i \neq \theta_i^0$ получаем:

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

Пример, часть 1

- Логарифм зарплаты в обществе имеет нормальное распределение с математическим ожиданием μ и дисперсией σ^2. Посчитанные по выборке из 200 человек выборочное среднее и выборочная дисперсия логарифмов зарплат оказались равны 10 и 4 соответственно. Протестируем гипотезу о том, что дисперсия зарплаты случайно взятого индивида равняется 4.2 против альтернативы о том, что не равняется.

Тестирование гипотез с помощью оценок максимального правдоподобия

Пример, часть 1

- Логарифм зарплаты в обществе имеет нормальное распределение с математическим ожиданием μ и дисперсией σ^2. Посчитанные по выборке из 200 человек выборочное среднее и выборочная дисперсия логарифмов зарплат оказались равны 10 и 4 соответственно. Протестируем гипотезу о том, что дисперсия зарплаты случайно взятого индивида равняется 4.2 против альтернативы о том, что не равняется.
- Из условия получаем, что $\hat{\mu} = 10$ и $\hat{\sigma}^2 = 4$, откуда, методом обратного Гессиана, находим оценку асимптотической ковариационной матрицы, а из нее, асимптотической дисперсии:

$$\widehat{As.Cov}\left(\left[\hat{\mu}, \hat{\sigma}^2\right]^T\right) = \begin{bmatrix} \frac{4}{200} & 0 \\ 0 & \frac{4}{2\times 200} \end{bmatrix} = \begin{bmatrix} 0.02 & 0 \\ 0 & 0.01 \end{bmatrix} \implies \widehat{As.Var}\left(\hat{\sigma}^2\right) = 0.01$$

- Логарифм зарплаты в обществе имеет нормальное распределение с математическим ожиданием μ и дисперсией σ^2. Посчитанные по выборке из 200 человек выборочное среднее и выборочная дисперсия логарифмов зарплат оказались равны 10 и 4 соответственно. Протестируем гипотезу о том, что дисперсия зарплаты случайно взятого индивида равняется 4.2 против альтернативы о том, что не равняется.
- Из условия получаем, что $\hat{\mu} = 10$ и $\hat{\sigma}^2 = 4$, откуда, методом обратного Гессиана, находим оценку асимптотической ковариационной матрицы, а из нее, асимптотической дисперсии:

$$\widehat{As.Cov}\left(\left[\hat{\mu}, \hat{\sigma}^2\right]^T\right) = \begin{bmatrix} \frac{4}{200} & 0 \\ 0 & \frac{4}{2\times 200} \end{bmatrix} = \begin{bmatrix} 0.02 & 0 \\ 0 & 0.01 \end{bmatrix} \implies \widehat{As.Var}\left(\hat{\sigma}^2\right) = 0.01$$

Тестируется гипотеза $H_0 : \sigma^2 = 4.2$, против альтернативы $H_1 : \sigma^2 \neq 4.2$.

$$T(X) = \frac{4 - 4.2}{\sqrt{0.01}} = -2 \implies \text{p-value} = 2\min\left(\Phi\left(-2\right), 1 - \Phi\left(-2\right)\right) \approx 0.046$$

Тестирование гипотез с помощью оценок максимального правдоподобия

Пример, часть 1

- Логарифм зарплаты в обществе имеет нормальное распределение с математическим ожиданием μ и дисперсией σ^2. Посчитанные по выборке из 200 человек выборочное среднее и выборочная дисперсия логарифмов зарплат оказались равны 10 и 4 соответственно. Протестируем гипотезу о том, что дисперсия зарплаты случайно взятого индивида равняется 4.2 против альтернативы о том, что не равняется.
- Из условия получаем, что $\hat{\mu} = 10$ и $\hat{\sigma}^2 = 4$, откуда, методом обратного Гессиана, находим оценку асимптотической ковариационной матрицы, а из нее, асимптотической дисперсии:

$$\widehat{As.Cov}\left(\left[\hat{\mu}, \hat{\sigma}^2\right]^T\right) = \begin{bmatrix} \frac{4}{200} & 0 \\ 0 & \frac{4}{2\times 200} \end{bmatrix} = \begin{bmatrix} 0.02 & 0 \\ 0 & 0.01 \end{bmatrix} \implies \widehat{As.Var}\left(\hat{\sigma}^2\right) = 0.01$$

Тестируется гипотеза $H_0 : \sigma^2 = 4.2$, против альтернативы $H_1 : \sigma^2 \neq 4.2$.

$$T(X) = \frac{4 - 4.2}{\sqrt{0.01}} = -2 \implies \text{p-value} = 2\min\left(\Phi\left(-2\right), 1 - \Phi\left(-2\right)\right) \approx 0.046$$

Поскольку p-value < 0.05, то нулевая гипотеза отвергается на 5%-м уровне значимости.

Тестирование гипотез с помощью оценок максимального правдоподобия

Пример, часть 2

- В предыдущем примере протестируем гипотезу о том, что вероятность того, что логарифм зарплаты случайно взятого индивида меньше 12 равняется 0.8, против альтернативы о том, что не равняется.

Тестирование гипотез с помощью оценок максимального правдоподобия

Пример, часть 2

- В предыдущем примере протестируем гипотезу о том, что вероятность того, что логарифм зарплаты случайно взятого индивида меньше 12 равняется 0.8, против альтернативы о том, что не равняется.
- Выразим вероятность, о которой тестируется гипотеза, как функцию от параметров:

$$P\left(X_i < 12\right) = \Phi\left(\frac{12-\mu}{\sigma}\right) = g(\mu, \sigma^2) \implies g(4, 10) \approx 0.84$$

- В предыдущем примере протестируем гипотезу о том, что вероятность того, что логарифм зарплаты случайно взятого индивида меньше 12 равняется 0.8, против альтернативы о том, что не равняется.
- Выразим вероятность, о которой тестируется гипотеза, как функцию от параметров:

$$P\left(X_i < 12\right) = \Phi\left(\frac{12-\mu}{\sigma}\right) = g(\mu, \sigma^2) \implies g(4, 10) \approx 0.84$$

Найдем градиент данной вероятности по параметрам:

$$\nabla g(\mu, \sigma^2) = \left(-\frac{\phi\left(\frac{12-\mu}{\sigma}\right)}{\sigma}, \frac{(\mu-12)\,\phi\left(\frac{12-\mu}{\sigma}\right)}{2\sigma^3}\right)^T \implies \nabla g(10, 4) \approx (-0.12, -0.03)^T$$

- В предыдущем примере протестируем гипотезу о том, что вероятность того, что логарифм зарплаты случайно взятого индивида меньше 12 равняется 0.8, против альтернативы о том, что не равняется.
- Выразим вероятность, о которой тестируется гипотеза, как функцию от параметров:

$$P\left(X_i < 12\right) = \Phi\left(\frac{12-\mu}{\sigma}\right) = g(\mu, \sigma^2) \implies g(4, 10) \approx 0.84$$

Найдем градиент данной вероятности по параметрам:

$$\nabla g(\mu, \sigma^2) = \left(-\frac{\phi\left(\frac{12-\mu}{\sigma}\right)}{\sigma}, \frac{(\mu-12)\,\phi\left(\frac{12-\mu}{\sigma}\right)}{2\sigma^3}\right)^T \implies \nabla g(10, 4) \approx (-0.12, -0.03)^T$$

Применим дельта-метод:

$$\widehat{As.Var}\left(g(\hat{\mu}, \hat{\sigma}^2)\right) = (-0.12, -0.03)\begin{bmatrix}0.02 & 0\\ 0 & 0.01\end{bmatrix}(-0.12, -0.03)^T \approx 0.000297$$

- В предыдущем примере протестируем гипотезу о том, что вероятность того, что логарифм зарплаты случайно взятого индивида меньше 12 равняется 0.8, против альтернативы о том, что не равняется.
- Выразим вероятность, о которой тестируется гипотеза, как функцию от параметров:

$$P\left(X_i < 12\right) = \Phi\left(\frac{12-\mu}{\sigma}\right) = g(\mu,\sigma^2) \implies g(4,10) \approx 0.84$$

Найдем градиент данной вероятности по параметрам:

$$\nabla g(\mu,\sigma^2) = \left(-\frac{\phi\left(\frac{12-\mu}{\sigma}\right)}{\sigma}, \frac{(\mu-12)\,\phi\left(\frac{12-\mu}{\sigma}\right)}{2\sigma^3}\right)^T \implies \nabla g(10,4) \approx (-0.12,-0.03)^T$$

Применим дельта-метод:

$$\widehat{As.Var}\left(g(\hat{\mu},\hat{\sigma}^2)\right) = (-0.12,-0.03)\begin{bmatrix}0.02 & 0\\ 0 & 0.01\end{bmatrix}(-0.12,-0.03)^T \approx 0.000297$$

Протестируем гипотезу $H_0 : P\left(X_i < 12\right) = 0.8$, против альтернативы $H_1 : P\left(X_i < 12\right) \neq 0.8$.

$$T(X) = \frac{0.84-0.8}{\sqrt{0.000297}} = 2.32 \implies \text{p-value} = 2\min\left(1-\Phi\left(2.32\right), \Phi\left(2.32\right)\right) \approx 0.02$$

- Тестируется гипотеза об r ограничениях на параметры:

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

- Тестируется гипотеза об r ограничениях на параметры:

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

- 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 (LR-тест)

- Тестируется гипотеза об r ограничениях на параметры:

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

- 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 $\hat{\theta}^R$.

- Тестируется гипотеза об r ограничениях на параметры:

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

- 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:

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

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

- 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:

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

- 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:

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

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

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

- 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:

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

- 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:

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

Где под L_F и $\widehat{As.Cov}_F$ подразумевается, что мы рассчитываем логарифм правдоподобия и считаем выражение для оценки асимптотической ковариационной матрицы в точках, соответствующих оценкам ограниченной модели $\hat{\theta}^R$, но с использованием правдоподобия полной модели.

- 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:

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

  Где под L_F и $\widehat{As.Cov}_F$ подразумевается, что мы рассчитываем логарифм правдоподобия и считаем выражение для оценки асимптотической ковариационной матрицы в точках, соответствующих оценкам ограниченной модели $\hat{\theta}^R$, но с использованием правдоподобия полной модели.
- Данный тест удобно применять в случаях, когда ограниченная модель оценивается проще, чем полная.