Прикладная статистика в машинном обучении 2022

Семинар 5. Проверка гипотез

Золотарев Антон Олегович

3 октября 2022 г.

1 Теория

1.1 Скалярный случай

Необходимо проверить гипотезу:

$$\begin{cases} H_0 : \theta_0 = c, \\ H_A : \theta_0 \neq c \end{cases}$$

$$\mathrm{W} = \frac{(\hat{\theta}_{ML} - \theta_0)^2}{\widehat{Var}(\hat{\theta}_{ML})}$$

$$\mathrm{LR} = 2\left(\max_{\theta_{UR}} l(\theta_{UR}) - \max_{\theta_R} l(\theta_R)\right)$$

$$\mathrm{LM} = \frac{\left(l'(\hat{\theta}_0) - l'(\hat{\theta}_{ML})\right)^2}{\widehat{Var}(l'(\theta_0))} = \left(l'(\hat{\theta}_0) - l'(\hat{\theta}_{ML})\right)^2 \cdot \hat{I}(\theta)^{-1}|_{\theta_0}$$

1.2 Матричный случай

Необходимо проверить гипотезу:

$$\begin{cases} H_0 : \gamma_0 = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \\ \\ H_A : \gamma_0 = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \neq \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \end{cases}$$

γ - подвектор таких θ, которые участвуют в непосредственной проверке гипотезы.

$$\mathrm{W} = (\hat{\gamma}_{ML} - \gamma_0)^T \widehat{Var}(\hat{\gamma}_{ML})^{-1} (\hat{\gamma}_{ML} - \gamma_0)$$

Как найти $\hat{Var}(\hat{\gamma}_{ML})^{-1}$:

- Находим $\hat{I}(\hat{\theta})$
- Находим $\hat{Var}(\hat{\theta}_{ML})$
- Вырезаем подматрицу $\hat{Var}(\hat{\gamma})$
- Обращаем полученное

Для LR ничего не меняется:

$$\mathrm{LR} = 2\left(\max_{\theta_{UR}} l(\theta_{UR}) - \max_{\theta_R} l(\theta_R)\right)$$

$$\mathrm{LM} = s(\gamma_0)^T \hat{I}(\gamma_0)^{-1} s(\gamma_0)$$

где $s(\cdot) = \text{grad } l = \begin{pmatrix} l'_{\theta_1}|_{c_1} \\ ... \\ l'_{\theta_p}|_{c_p} \end{pmatrix}$

2 Почему $W, LR, LM \sim \chi^2(n-p)$?

2.1 Вспомним про χ^2

$$X \sim \chi^2(k) \iff X = \sum_{i=1}^{k} \frac{(Y_i - \mu)^2}{\sigma^2}, Y_i \sim N(\mu, \sigma^2)$$

2.2 Вальд

$$\mathrm{W} = \frac{(\hat{\theta}_{ML} - \theta_0)^2}{\widehat{Var}(\hat{\theta}_{ML})}$$

Помним про асмиптотическую нормальность оценок максимального правдоподобия: $\hat{\theta}_{ML} \sim_{H_0} N(\theta_0, Var(\hat{\theta}_{ML}))$

Получаем, что при выполнении нулевой гипотезы статистика распределена по χ^2, то есть если не отвергаем нулевую гипотезу, то и не отвергаем гипотезу о том, что статистика распределена по χ^2

2.3 Отношение правдоподобия

$$\mathrm{LR} = 2\left(\max_{\theta_{UR}} l(\theta_{UR}) - \max_{\theta_R} l(\theta_R)\right)$$

Покажем частный случай, когда ограничиваем μ при оценке параметров нормального распределения:

$$l(\theta_{UR}) = \ln L(\hat{\theta}) = -\frac{n}{2}\ln(2\pi\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

$$l(\theta_R) = \ln L(\widetilde{\theta}) = -\frac{n}{2}\ln(2\pi\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \mu_0)^2$$

$$\text{LR} = ... = 2\left(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 + \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right) =$$

$$= \frac{1}{\hat{\sigma}^2}\left[\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\hat{\mu})^2\right]\cdot\frac{n}{n} = \frac{n}{\hat{\sigma}^2}\left[\frac{\sum\limits_{i=1}^{n}(x_i-\mu_0)^2}{n} - \frac{\sum\limits_{i=1}^{n}(x_i-\hat{\mu})^2}{n}\right] =$$

$$= \frac{\sum\limits_{i=1}^{n}(x_i-\mu_0)^2}{\hat{\sigma}^2} - n \sim \chi^2$$

В предпоследнем действии в правой части скобки у нас по сути оценка сигмы, которая схлопывается с дробью перед большой скобкой. Остаётся $-n$, который надо как-то использовать, чтобы привести x_i в окончательном $\frac{\sum\limits_{i=1}^{n}(x_i-\mu_0)^2}{\hat{\sigma}^2}$ к $\frac{(\hat{\mu}-\mu_0)^2}{\hat{\sigma}^2}$. Я 40 минут провозился с этим делом, но понял, что одними простыми манипуляциями из матстата без сложной теории здесь не обойтись. Поэтому не буду вводить в заблуждение и просто советую глянуть теорему Вилкса. Ещё немного про тест отношения правдоподобия:

- Откуда берётся 2 в формуле статистики
- Немножко рассуждений про то, почему асимптотическое стремление статистики к χ^2 является не очень-то надёжным фактом

2.4 Множители Лагранжа

Для случая регрессии в целом не сложно показать, почему статистика теста множителей Лагранжа имеет распределение χ^2 (пример, который пока не надо понимать). Показательный пример для тестов на истинные значения параметров в распределениях придумать не получилось, но суть там в том, что скор-функция ($s(\theta) = l'_\theta$) в точке $\theta = \theta_0$ распределена нормально и надо провести "показательство аналогичное случаю с тестом Вальда. Формально надо доказать следующее:

$$l'_\theta|_{\theta_0} \sim N(l'_\theta|_{\hat{\theta}_{ML}}, \hat{I}(\theta_0)) \sim N(0, \hat{I}(\theta_0))$$

Показано здесь