# Спецкурс "Регрессия"

Шкляев А.В.

15 января 2019 г.

# 1 Асимптотические свойства оценок вне нормальной модели

## 1.1 Случай случайных предикторов

Удобную форму достаточных условий можно получить в предположении, что предикторы выбираются случайным образом.
Предположим, что векторы $(X_{i,1},...,X_{i,k},\varepsilon_k)$ н.о.р. При этом мы, вообще говоря, не требуем чтобы $\mathbf{E}(\varepsilon_k|X_{k,\cdot}) = 0$.

**Теорема 1.** *Пусть выполнены условия 1)-3) и дополнительно потребуем* $\mathbf{E}\varepsilon_i^4$, $\mathbf{E}X_{1,i}^4 < \infty$ *Тогда* $\widehat{a}$ *является асимптотически нормальной для* $\vec{a}$ *с асимптотической ковариационной матрицей* $S = Q^{-1}RQ^{-1}$, *где* $R_{i,j} = \mathbf{E}X_{1,i}X_{1,j}\varepsilon_1^2$.

*Доказательство.* Мы должны доказать, что

$$\sqrt{n}(\widehat{a} - \vec{a}) \xrightarrow{d} Z \sim \mathcal{N}(0, S).$$

При этом

$$\sqrt{n}(\widehat{a} - \vec{a}) = \left(\frac{1}{n}(X^tX)_{i,j}\right)^{-1} \frac{1}{\sqrt{n}}X^t\vec{\varepsilon}.$$

При этом

$$\left(\frac{1}{n}(X^tX)_{i,j}\right)^{-1} \xrightarrow{P} Q^{-1},\ \frac{1}{\sqrt{n}}X^t\vec{\varepsilon} \xrightarrow{d} Z \sim \mathcal{N}(0, R),$$

откуда и вытекает требуемое. □

При условии $\mathbf{E}(\varepsilon|x) = 0$ матрица $S$ упрощается до $\sigma^2 Q^{-1}$.

## 1.2 Оценка ковариационной матрицы

Для построения асимптотического доверительного интервала для неизвестного параметра нам требуется избавиться от зависимости матрицы $S$ от параметров. Рассмотрим случай случайных предикторов. Воспользуемся тем, что

$$Q_{i,j} = \mathbf{E}X_{1,i}X_{1,j}$$

есть предел $\widehat{Q}_{i,j} = n^{-1}\sum_{l=1}^n X_{l,i}X_{l,j}$ по вероятности. Менее очевидна состоятельность оценки для

$$R_{i,j} = \mathbf{E}X_{1,i}X_{1,j}\varepsilon_1^2.$$

Естественно использовать оценку $\widehat{R}_{i,j} = n^{-1}\sum_{l=1}^n X_{l,i}X_{l,j}(Y_l - \langle X_{l,\cdot}, \widehat{a}\rangle)$. Иначе говоря,

$$\widehat{R} = \frac{1}{n}X^t\widehat{D}X,$$

где $\widehat{D}$ — диагональная матрица с $\widehat{\varepsilon}_l = Y_l - \langle X_{l,\cdot}, \widehat{a}\rangle$ на диагонали. Оказывается, что в условиях теоремы 3 это так:

$$\widehat{R} = n^{-1}\sum_{l=1}^{n} X_{l,i}X_{l,j}\varepsilon_l^2 - 2n^{-1}\sum_{l=1}^{n}\sum_{m=1}^{n} X_{l,i}X_{l,j}X_{l,m}(\widehat{a}_m - a_m)\varepsilon_l + \sum_{l=1}^{n}\sum_{m=1}^{k}\sum_{r=1}^{k} X_{l,i}X_{l,j}X_{l,m}(\widehat{a}_m - a_m)X_{l,r}(\widehat{a}_r - a_r). \tag{1}$$

При этом

$$\frac{1}{n}\left|\sum_{m=1}^{n}\sum_{l=1}^{n} X_{l,i}X_{l,j}X_{l,m}(\widehat{a}_m - a_m)\varepsilon_l\right| \le \frac{1}{n}\left|\left|\sum_{l=1}^{n} |X_{l,i}X_{l,j}X_{l,\cdot}|\varepsilon_l\right|\right| \, ||\widehat{a} - \vec{a}|| \le \frac{1}{n}\sum_{l=1}^{n} ||X_{l,i}X_{l,j}X_{l,\cdot}\varepsilon_l|| \, ||\widehat{a} - \vec{a}||$$

Здесь первый множитель сходится к $\mathbf{E}X_{1,i}X_{1,j}||X_{1,\cdot}||\varepsilon_1$, а второй к нулю, откуда второе слагаемое (1) стремится к нулю. Здесь мы пользуемся, тем, что первое математическое ожидание мажорируется

$$\mathbf{E}||X_{1,\cdot}||^3\varepsilon_1 \le (\mathbf{E}||X_{1,\cdot}^4||)^{3/4}(\mathbf{E}\varepsilon_1^4)^{1/4} < \infty.$$

Аналогичным образом

$$\sum_{l=1}^{n}\sum_{m=1}^{k}\sum_{r=1}^{k} X_{l,i}X_{l,j}X_{l,m}X_{l,r}(\widehat{a}_m - a_m)(\widehat{a}_r - a_r)^2 \le \frac{1}{n}\sum_{l=1}^{n} |X_{l,i}||X_{l,j}|||X_{l,\cdot}||||X_{l,\cdot}||||\widehat{a} - \vec{a}||^2$$

сходится к 0, поскольку $||\widehat{a} - \vec{a}||^2 \to 0$, а в

$$\frac{1}{n}\sum_{l=1}^{n} |X_{l,i}||X_{l,j}|||X_{l,\cdot}||||X_{l,\cdot}|| \to \mathbf{E}|X_{1,i}||X_{1,j}|||X_{1,\cdot}||^2.$$

## 2 Взвешенная оценка МНК

Предположим, что мы рассматриваем модель $\vec{Y} = X\vec{a} + \vec{\varepsilon}$, где $\mathbf{D}\varepsilon_i = \sigma_i^2$ — различны. При этом $\sigma_i^2$ различны и известны, а $\varepsilon_i$ независимы. Обозначим $W$ матрицу с $1/\sigma_i$ на диагонали. Естественно свести задачу к предыдущей, перейдя к соотношениям

$$\widetilde{Y} = \widetilde{X}\vec{a} + \widetilde{\varepsilon},\ \widetilde{\varepsilon} = W\varepsilon_i,\ \widetilde{X} = WX,\ \widetilde{Y} = W\vec{Y}.$$

Применяя развитую нами теорию, получаем оценку МНК

$$\widehat{a}_W = (\widetilde{X}^t\widetilde{X})^{-1}\widetilde{X}^t\widetilde{Y} = (X^tW^2X)^{-1}X^tW^2\vec{Y}.$$

Такая оценка называется взвешенной оценкой МНК. При этом оценкой для $\sigma^2$ будет $WRSS_n/n$, где

$$WRSS_n = ||\widetilde{Y} - \widetilde{X}\widehat{a}_W||^2 = (W\vec{Y} - WX\widehat{a})^t(W\vec{Y} - WX\widehat{a}) =$$
$$\vec{Y}^t(E - WV^2X(X^tW^2X)^{-1}X^t)W^2(XE - (X^tW^2X)^{-1}X^tW^2)\vec{Y} = \vec{Y}^tW^2\vec{Y} - \vec{Y}^tW^2X(X^tW^2X)^{-1}X^tW^2\vec{Y}.$$

Предыдущие результаты дают нам результаты для этой модели — оценки $\widehat{a}_W$ и $WRSS/(n-k)$ несмещенные, при нормальности $\varepsilon$ они будут иметь нормальное распределение, $WRSS/\sigma^2 \sim \chi^2_{n-k}$. Аналогичным образом можно в нормальном случае можно проверять гипотезу снижения размерности с помощью статистики Фишера

$$F = \frac{(WRSS_{L_0} - WRSS_{L_1})/(k-m)}{WRSS_{L_1}/(n-k)},$$

которая имеет распределение $F_{k-m,n-k}$ при выполнении гипотезы. По существу в этом случае мы рассматриваем ту же задачу ортогонализации в базисе, матрица Грамма которого есть $W^2$.