Спецкурс "Регрессия"

Шкляев А.В.

15 января 2019 г.

1 Асимптотические свойства оценок вне нормальной модели

Откажемся от предположения нормальности. В этом случае у нас сохраняется несмещенность оценок, но можем ли мы утверждать, что они состоятельны и асимптотически нормальны?
Оказывается удобным формулировать результаты в терминах $H = X(X^tX)^{-1}X^t$.

- H симметрична
- $h_{i,i} \in [0, 1]$
- $tr(H) = k$
- H имеет k собственных значений равных 1 и остальные n-k — равных 0.
- $\sum_{j=1}^{n} h_{i,j}^2 = h_{i,i}$.

Предпоследнее свойство вытекает из того, что H это оператор проекции на пространство $L = X\vec{a}$, а второе и третье вытекают из него. Последнее свойство вытекает из соотношения $H^2 = H$.

Теорема 1. *1) Пусть $\lambda_{min}(A)$ — минимальное собственное значение матрицы $A = X^tX$. Предположим, что $\lambda_{min}(A) \to \infty$. Тогда оценка МНК $\widehat{a}$ будет состоятельной.*
2) Пусть $h_{i,i} \to 0$, $n \to \infty$. Тогда $\widehat{y}_i$ является состоятельной оценкой для $\mathbf{E}Y_i$.
Соответственно, для состоятельности всех $\widehat{y}_i$ требуется $\max_i h_{i,i} \to 0$, $n \to \infty$.

Доказательство. 1) В силу неравенства Чебышева

$$\mathbf{P}(|\widehat{a}_i - a_i| > \delta) \leq \frac{\mathbf{D}\widehat{a}_i}{\delta^2}.$$

При этом

$$\sum_{i=1}^{k} \mathbf{D}\widehat{a}_i = \mathbf{E}(\widehat{a} - a)^t(\widehat{a} - a) = \mathbf{E}\vec{\varepsilon}^t XA^{-2}X^t\vec{\varepsilon} = \sigma^2 tr(XA^{-2}X^t) = \sigma^2/tr(A).$$

Но $tr(A) = \lambda_1(A) + ... + \lambda_k(A) \geq k\lambda_{min}(A)$. Следовательно, все оценки являются состоятельными.

Обратим внимание, что в действительности нам не требуется одинаковой распределенности и независимости величин ε, а достаточно некоррелированности и ограниченности их дисперсий в совокупности и сходимости к бесконечности $\lambda_{min}(X^tX)$.

2) В силу определения

$$\widehat{y}_i = X\vec{a} + X(X^tX)^{-1}X^t\vec{\varepsilon} = \mathbf{E}Y_i + H\vec{\varepsilon}.$$

Следовательно,

$$\mathbf{P}(|\widehat{y}_i - \mathbf{E}Y_i| > \delta) \leq \frac{\sum_{j=1}^{k} h_{i,j}^2\sigma^2}{\delta^2} = \frac{h_i\sigma^2}{\delta^2} \to 0,$$

что и требовалось доказать. Все оценки $\widehat{y}_i$ будут состоятельными при выполнении условия $\max h_{i,i} \to 0$, $n \to \infty$. □

Теорема 2. *Пусть $\vec{Y} = X\vec{a}+\vec{\varepsilon}$, где ε — н.о.р. с конечной дисперсией σ^2. Предположим, как и прежде, что $\max h_{i,i} \to 0$, $n\to\infty$. Предположим также, что $n^{-1}X^tX$ с ростом n сходится к Σ — положительно определенной матрице. Тогда*

$$\sqrt{n}(\widehat{a}-\vec{a}) \xrightarrow{d} \vec{Z} \sim \mathcal{N}(\vec{0}, \sigma^2\Sigma^{-1}).$$

Доказательство. Воспользуемся тем, что $\widehat{a}-\vec{a} = (X^tX)^{-1}X^t\vec{\varepsilon}$. Тогда

$$\psi_{\widehat{a}-\vec{a}}(\vec{s}) = \psi_{(X^tX)^{-1}X^t\vec{\varepsilon}}(\vec{s}) = \psi_{\vec{\varepsilon}}(X(X^tX)^{-1}\vec{s}) = \exp\left(\sum_{i=1}^{n} \ln\psi_{\varepsilon_i}(X(X^tX)^{-1}\vec{s})_i\right).$$

В случае, если $\sqrt{n}\max_{i\le n}(X(X^tX)^{-1}\vec{s})_i \to 0$ мы можем сказать, что

$$\ln\psi_{\varepsilon_i}(X(X^tX)^{-1}\vec{s})_i) = -\frac{\sigma^2}{2}((X(X^tX)^{-1}\vec{s})_i)^2(1+o(1)),$$

где $o(1)$ равномерно мало по i. Следовательно,

$$\ln\psi_{\sqrt{n}(\widehat{a}-\vec{a})}(\vec{s}) = -(1+o(1))\frac{\sigma^2}{2}\vec{s}^t(X^tX)^{-1}X^tX(X^tX)^{-1}\vec{s} \to -\frac{\sigma^2}{2}\vec{s}^t\Sigma^{-1}\vec{s},$$

что нам и требуется. Остается понять, что $\sqrt{n}\max_{i\le n}(X(X^tX)^{-1}\vec{s})_i \to 0$. Воспользуемся оценкой

$$|\sqrt{n}(X(X^tX)^{-1}\vec{s})_i|^2 = n\left|\sum_{l,j,m,p=1}^{k} X_{i,j}(X^tX)^{-1}_{j,l}s_lX_{i,m}(X^tX)^{-1}_{m,p}s_p\right|$$

и заметим, что если при всех j,m $\max_i X_{i,j}X_{i,m}/n \to 0$, то

$$\frac{1}{n}\max_i\left|\sum_{l,j,m,p=1}^{k} X_{i,j}\Sigma^{-1}_{j,l}s_lX_{i,m}\Sigma^{-1}_{m,p}s_p\right| \le \max_{j,m}|\max_i(X_{i,j}X_{i,m}/n)||r_j||r_m| \to 0,$$

где $r_j = \sum_{m=1}^k \Sigma^{-1}_{j,m}s_m$, откуда

$$\max_i|\sqrt{n}(X(X^tX)^{-1}\vec{s})_i|^2 \to 0.$$

Значит нам достаточно получить условие $\max_i X_{i,j}X_{i,m}/n \to 0$ при любых j, m. Для этого достаточно получить $\max_i X^2_{i,j}/n \to 0$ при всех j (это условие называют условием Нётера). Выведем его из условия $\max_i h_{i,i} \to 0$, $n \to \infty$.

Для этого нам понадобится неравенство

$$\frac{\langle X_{i,\cdot}, \vec{e}\rangle^2}{\vec{e}^tX^tX\vec{e}} \le X^t_{i,\cdot}(X^tX)^{-1}X_{i,\cdot} = h_{i,i},$$

справедливое при любом $\vec{e}\neq\vec{0}$. Действительно, переходя в диагонализирующий X^tX базис, мы переведем $X_{i,\cdot}$ в некий вектор $\vec{z}$, а X^tX в диагональную матрицу D с $d_{i,i}$ на диагонали. Тогда наше отношение равно

$$\left(\frac{\sum_{i=1}^k z_ie_i}{\sqrt{\sum_{i=1}^k d_ie_i^2}}\right)^2 = \left(\frac{\sum_{i=1}^k d_i^{-1/2}z_i\sqrt{d_i}e_i}{\sqrt{\sum_{i=1}^k d_ie_i^2}}\right)^2 = \left(\sum_{i=1}^k d_i^{-1/2}z_ie'_i\right)^2 \le ||(d_1^{-1/2}z_1,...,d_k^{-1}z_k)||^2,$$

где $e' = (\sqrt{d_1}e_1, ..., \sqrt{d_k}e_k)/(\sum_{i=1}^k d_i e_i^2)$ — единичный вектор. Но

$$||(d_1^{-1/2}z_1, ..., d_k^{-1}z_k)||^2 = \sum_{i=1}^k d_i^{-1}z_i^2 = \vec{z}^t D^{-1}\vec{z} = X_{i,\cdot}^t(X^tX)^{-1}X_{i,\cdot},$$

что и требовалось доказать.

Остается заметить, что при подстановке $\vec{e} = (0, ..., 0, 1, 0, ...0)$, где единица находится на j месте,

$$\frac{\langle X_{i,\cdot}, \vec{e}\rangle^2}{\vec{e}^t X^t X\vec{e}} = \frac{x_{i,j}^2}{\sum_{i=1}^n x_{i,j}^2} = \frac{x_{i,j}^2/n}{\sum_{i=1}^n x_{i,j}^2/n},$$

откуда

$$\frac{1}{n}\max_{i\le n} x_{i,j}^2 \le \frac{\sum_{i=1}^n x_{i,j}^2}{n} h_{i,i} \to 0,$$

поскольку первый множитель стремится к $\Sigma_{i,j}$, а второй к нулю. □

1.1 Случай случайных предикторов

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

Теорема 3. *Пусть выполнены следующие условия*
1) $rank(Q) = k$, *где* $Q = \mathbf{E}X^tX$;
2) $\mathbf{E}X_{i,j}\varepsilon_i = 0$, $j = 1, ..., k$;
3) $\mathbf{E}\varepsilon^2 < \infty$.
Тогда $\widehat{a}$ *является состоятельной для* $\vec{a}$.

Доказательство.

$$(X^tX)^{-1}X^t\vec{Y} = \vec{a} + (X^tX)^{-1}X^t\varepsilon = \vec{a} + g\left(\frac{1}{n}X^tX, \frac{1}{n}X^t\varepsilon\right),$$

где $g(A, B) = A^{-1}B$. При этом

$$\frac{1}{n}(X^tX)_{i,j} = \frac{1}{n}\sum_{l=1}^n X_{l,i}X_{l,j} \xrightarrow{P} Q_{i,j}$$

в силу закона больших чисел. Аналогичным образом

$$\frac{1}{n}(X^t\vec{\varepsilon})_i = \frac{1}{n}\sum_{l=1}^n X_{l,i}\varepsilon_l \xrightarrow{P} (\mathbf{E}X^t\vec{\varepsilon})_i = 0$$

Остается заметить, что $g(A, B)$ непрерывное в точке $(Q, 0)$ отображение, поскольку Q невырождена. Действительно,

$$g(A, B)_i = (\det(A))^{-1}\sum_{j=1}^k C_{i,j}B_j,$$

где C — главный минор матрицы A, непрерывно зависящий от ее элементов, $\det(A)$ также непрерывно зависит от элементов матрицы и отделен от нуля в окрестности Q.

Отсюда

$$(X^tX)^{-1}X^t\vec{Y} \xrightarrow{P} \vec{a} + g(\mathbf{E}X^tX, \vec{0}) = \vec{a}.$$

□