$$\times f_{bb}^{d}(-z-y_1)dx_1dzdy_1 + \iiint\limits_{\Pi^3} f_{aa}^{X}(x_1)f_{bb}^{X}(z-x_1)f_{ab}^{d}(y_1)f_{ba}^{d}(-z-y_1)\times$$

$$\times \exp(i\tau(2y_1-z))dx_1dzdy_1 + \iiiint\limits_{\Pi^4} f_{ab}^{X}(x_1)f_{ba}^{X}(z-x_1)f_{abab}^{d}(y_1,-z-y_1,y_3)\times$$

$$\times \exp(i\tau(2x_1-z+y_1+y_3))dx_1dzdy_1dy_3 + \iiint\limits_{\Pi^3} f_{ab}^{X}(x_1)f_{ba}^{X}(z-x_1)f_{aa}^{d}(y_1)f_{bb}^{d}(-z-y_1)$$

$$\times \exp\big(i\tau(2x_1-z)\big)dx_1dzdy_1 + \iiint\limits_{\Pi^3} f_{ab}^{X}(x_1)f_{ba}^{X}(z-x_1)f_{ab}^{d}(y_1)f_{ba}^{d}(-z-y_1)\times$$

$$\times \exp\big(i\tau(2x_1+2y_1-g)\big)dx_1dzdy_1\Big].$$

Доказательство очевидным образом вытекает из свойств ядра Фейра и условий теоремы.

Литература

1. *Труш Н Н.* Асимптотические методы статистического анализа временых рядов.// Мн.: БГУ, 1999.
2. *Marshall R. J.* Autocorrelation estimation of time series with randomly missing observations // Biometrika. 67. 3. P 567–570.

О РИСКЕ ПРОГНОЗИРОВАНИЯ ДЛЯ МОДЕЛИ ЛИНЕЙНОЙ РЕГРЕССИИ С «ПРОПУСКАМИ»

Е. В. Ковалевский

Введение

Проблема обработки данных с пропусками возникает в самых разнообразных приложениях статистического анализа. Многие исследователи стремятся избавиться от пропусков с тем, чтобы впоследствии провести обработку полных данных стандартными средствами [1]. Такой подход подразумевает либо исключение некомплектных наблюдений, либо их заполнение некоторыми значениями. Оба подхода могут оказаться удовлетворительными в том случае, если доля пропусков мала. В противном случае оба метода могут привести к сильному различию статистических выводов, сделанных при наличии полных данных и тех же данных с пропусками.

В практических приложениях часто оказывается полезной модель множественной линейной регрессии

$$y_t = \theta' x_t + \xi_t, t = 1,\dots,T, \qquad (1)$$

где T – количество наблюдений, $y_t \in R, t = 1,\dots,T$ – выходные переменные, $x_t \in R^n, t = 1,\dots,T$ – векторы входных переменных, $\theta \in R^n$ –

вектор параметров модели, $\xi_t \in R, t=1,...,T$ – независимые, одинаково распределенные случайные величины, $E\{\xi_t\}=0, D\{\xi_t\}=\sigma^2, t=1,...,T$.

Предположим также, что ошибки имеют нормальное распределение:

$$L(\xi_t)=N(0,\sigma^2), t=1,...,T.$$

В обозначениях

$$Y=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \in R^T, X=\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_T \end{pmatrix} \in R_{T\times n}, \xi=\begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_T \end{pmatrix} \in R^T, L(\xi)=N_T\left(0_{T\times 1}, \sigma^2 I_{T\times T}\right)$$

модель (1) принимает вид

$$Y=X\theta+\xi. \tag{2}$$

В предположении, что X – матрица полного ранга, хорошо известны оценки параметров модели

$$\widehat{\theta}=\left(X^T X\right)^{-1} X^T Y, s^2=\frac{1}{T-1}\left(Y-X\widehat{\theta}\right)^T\left(Y-X\widehat{\theta}\right). \tag{3}$$

Тогда прогноз выходной переменной $\widehat{y}_{T+\tau}$ в некоторый будущий момент времени $T+\tau$ определяется по формуле

$$\widehat{y}_{T+\tau}=\widehat{\theta}' x_{T+\tau}, \tau \in N. \tag{4}$$

Риск прогнозирования (4) определяется по формуле

$$r\{\widehat{y}_{T+\tau}\}:=E\left\{\left(\widehat{y}_{T+\tau}-y_{T+\tau}\right)^2\right\}=\sigma^2\left(1+x'_{T+\tau}\left(X^T X\right)^{-1} x_{T+\tau}\right). \tag{5}$$

В статье рассматривается зависимость риска прогнозирования (5) в случаях наличия одного пропуска в векторе $x_{T+\tau}$ или в матрице X.

1. Случай пропуска в векторе входных переменных

Пусть имеет место модель (2). Предположим, что в векторе $x_{T+\tau}$ пропущена i_0-ая координата $x_{T+\tau,i_0}$, пропуски в X и Y отсутствуют. В этих условиях оценки вектора θ и дисперсии σ^2 имеют классический вид (3). Рассмотрим зависимость риска прогнозирования (5) от значения переменной $x_{T+\tau,i_0}$:

$$r\left\{x_{T+\tau,i_0}\right\}=Ax^2_{T+\tau,i_0}+Bx_{T+\tau,i_0}+C,$$

$$A = \sigma^2 \left(\left(X^T X \right)^{-1} \right)_{i_0 i_0},$$

$$B = 2\sigma^2 \sum_{j \neq i_0} x_{T+\tau,j} \left(\left(X^T X \right)^{-1} \right)_{i_0 j},$$

$$C = \sigma^2 \left(1 + \sum_{\substack{i \neq i_0 \\ j \neq i_0}} x_{T+\tau,i} \left(\left(X^T X \right)^{-1} \right)_{ij} x_{T+\tau,j} \right).$$

Вычислим верхнюю и нижнюю границы риска в предположении, что переменная $x_{T+\tau,i_0}$ может принимать значения из некоторого отрезка $[x_{i_0}^-, x_{i_0}^+]$.

$$0 = \frac{\partial r\left(x_{T+\tau,i_0} \right)}{\partial x_{T+\tau,i_0}} = 2Ax_{T+\tau,i_0} + B,$$

$$\widetilde{x}^*{}_{T+\tau,i_0} := \frac{-B}{2A} = \frac{-\sum\limits_{j \neq i_0} x_{T+\tau,j} \left(\left(X^T X \right)^{-1} \right)_{i_0 j}}{\left(\left(X^T X \right)^{-1} \right)_{i_0 i_0}},$$

$$x^*{}_{T+\tau,i_0} = \begin{cases} \widetilde{x}^*{}_{T+\tau,i_0}, \widetilde{x}^*{}_{T+\tau,i_0} \in [x_{i_0}^-, x_{i_0}^+], \\ x_{i_0}^-, \widetilde{x}^*{}_{T+\tau,i_0} < x_{i_0}^-, \\ x_{i_0}^+, \widetilde{x}^*{}_{T+\tau,i_0} > x_{i_0}^+. \end{cases}$$

В точке $x^*{}_{T+\tau,i_0}$ достигается минимум риска, и он равен

$$\sigma^2 \left(1 + \left(x_{T+\tau} \mid x^*_{T+\tau,i_0} \right)' \left(X^T X \right)^{-1} \left(x_{T+\tau} \mid x^*_{T+\tau,i_0} \right) \right),$$

где $\left(x_{T+\tau} \mid x^*_{T+\tau,i_0} \right)$ – вектор входных переменных, в котором i_0-я переменная приняла значение $x^*{}_{T+\tau,i_0}$.

Верхняя граница риска

$$r^{\max} = r\left\{ x^{\max}{}_{T+\tau,i_0} \right\},$$

где $x^{\max}{}_{T+\tau,i_0} = \arg\max \left\{ r\left(x_{T+\tau} \mid x_{i_0}^- \right), r\left(x_{T+\tau} \mid x_{i_0}^+ \right) \right\}$.

2. Случай пропуска в матрице данных

Пусть имеет место модель (2). Предположим, что в матрице X пропущен $t_0 i_0$-й элемент $x_{t_0 i_0}$, пропуски в $x_{T+\tau}$ и Y отсутствуют. В отличие от предыдущего случая (пропуски в $x_{T+\tau}$) здесь не могут быть вычислены классические оценки θ и σ^2, поскольку в формулах (3) присутствует матрица X, а следовательно, и пропущенное значение $x_{t_0 i_0}$. Однако мы можем считать $x_{t_0 i_0}$ параметром, как и в случае с $x_{T+\tau, i_0}$, и исследовать зависимость риска прогнозирования $y_{T+\tau}$ от значения $x_{t_0 i_0}$.

Теорема. В условиях модели (2) в случае, когда пропущено одно значение переменной $x_{t_0 i_0}$ в матрице X, стационарные точки функционала риска (5) $r\{x_{t_0 i_0}\}$ прогнозирования переменной $y_{T+\tau}$ являются решением следующего уравнения:

$$0 = \sigma^2 \sum_{i,j=1}^{m} x_{T+\tau,i} x_{T+\tau,j} \times$$

$$\times \left(\left(\sum_{\substack{k=1 \\ k \neq j}}^{m} (-1)^k x_{t_0 k} A_{(i,i_0;j,k)}\left(X^T X\right) + \sum_{\substack{k=1 \\ k \neq i}}^{m} (-1)^{i_0} x_{t_0 k} A_{(i,k;j,i_0)}\left(X^T X\right) \right) \left| X^T X \right| - \right.$$

$$\left. - A_{(j;i)}\left(X^T X\right) \sum_{k=1}^{m} \left((-1)^k + (-1)^{i_0} \right) x_{t_0 k} A_{(i_0;k)}\left(X^T X\right) \right),$$

где $A_{(i;j)}(B)$ – алгебраическое дополнение матрицы B к ij-му ее элементу.

Доказательство теоремы не приводится из-за его громоздкости.

3. Приближенный алгоритм поиска минимума риска

В случае наличия многих пропусков не удалось получить аналитические результаты, однако реализован алгоритм построения гистограммы значений риска (5) на равномерной сетке путем заполнения пропусков и вычисления значения риска.

Ниже приведен пример, иллюстрирующий данный алгоритм:

$$X = \begin{pmatrix} 105.6 & 104.9 \\ 0 & 107.6 \\ 102 & 106 \\ 101.5 & 104.1 \\ 100.6 & 104.5 \\ 102.3 & 105.6 \\ 102 & 0 \\ 101.3 & 103.7 \\ 101.8 & 104.5 \\ 101.3 & 104.8 \end{pmatrix}; x_{T+\tau} = \begin{pmatrix} 103.9 \\ 101.3 \end{pmatrix};$$

$$\sigma^2 = 1.$$

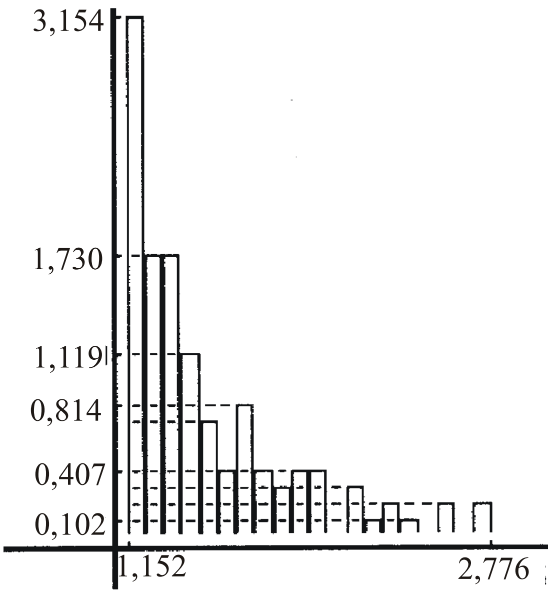

Рис. Гистограмма значений риска

Здесь нулевые значения в матрице X означают пропуск. Объем выборки рисков равен 121, что обеспечивается одиннадцатью возможными значениями по обоим пропущенным значениям из отрезка [90;110]. По гистограмме можно оценить нижнюю и верхнюю границы риска, а также вероятность попадания риска прогнозирования в заданный интервал.

Литература

1. *Литтл Р. Дж. А., Рубин Д. Б.* Статистический анализ данных с пропусками / Пер. с англ. А. М. Никифорова. М.: Финансы и статистика, 1990. 336 с.

К ОЦЕНКЕ СНИЗУ ЧИСЛА НЕЦЕЛОЧИСЛЕННЫХ ВЕРШИН МНОГОГРАННИКА ТРЕХИНДЕКСНОЙ АКСИАЛЬНОЙ ЗАДАЧИ О НАЗНАЧЕНИЯХ

В. М. Кравцов

Известно [1], что многогранник $M(3,n) = \{x = \| x_{ijt} \|_n : x_{ijt} \geq 0 \ \forall (i,j,t)$

$$\in N_n^3, \ \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ijt} = 1 \ \forall t \in N_n, \ \sum_{i=1}^{n} \sum_{t=1}^{n} x_{ijt} = 1 \ \forall j \in N_n,$$

$$\sum_{j=1}^{n} \sum_{t=1}^{n} x_{ijt} = 1 \ \forall i \in N_n\},$$

где $N_n = \{1, 2, ..., n\}$, $N_n^3 = N_n \times N_n \times N_n$, $n \geq 2$, трехиндексной аксиальной задачи о назначениях не является целочисленным, т. е. имеет еще и вершины с дробными компонентами.