1 Сходимость по вероятности

Определение. *Пусть* $(\Omega, \mathcal{F}, \mathbb{P})$ *— вероятностное пространство с определёнными на нём случайными величинами* $\mathbf{X}, \mathbf{X}_n, (n = 1, 2, \dots)$. *Говорят, что* $\{\mathbf{X}_n\}_{n=1}^{\infty}$ ***сходится по вероятности к*** $\mathbf{X}$, *если* $\forall \varepsilon > 0$*:*

$$\lim_{n\to\infty} \mathbb{P}(|\mathbf{X}_n - \mathbf{X}| > \varepsilon) = 0$$

Обозначение: $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}$

Данное свойство означает, что если взять величину $\mathbf{X}_n$ с достаточно большим номером, то вероятность значительного отклонения от предельной величины $\mathbf{X}$ будет небольшой. Однако важно понимать, что если одновременно (т.е. для одного и того же элементарного исхода ω) рассмотреть последовательность $\{X_n(\omega)\}$, то она не обязана сходиться к значению $X(\omega)$, вообще говоря, ни при каком ω. Т.е. сколь угодно далеко могут находиться сильно отклоняющиеся значения, просто их "не очень много поэтому вероятность того, что такое сильное отклонение попадет в заданном эксперименте на конкретно заданный номер n, мала.

В качестве примера рассмотрим вероятностное пространство $\Omega = [0, 1]$, вероятность - мера Лебега (т.е. вероятность любого интервала равна его длине). Случайные величины зададим следующим образом: для первых двух X_1, X_2 разбиваем Ω на два интервала $[0, \frac{1}{2})$ и $(\frac{1}{2}, 1]$ и определяем X_1 равной 1 на первом интервале и 0 на втором, а X_2 - наоборот, 0 на первом интервале и 1 на втором. Далее берем следующие четыре величины X_3, X_4, X_5, X_6, делим Ω на четыре непересекающихся интервала длины $\frac{1}{4}$ и задаем каждую величину равной 1 на своем интервале и 0 на остальных. Затем рассматриваем следующие 8 величин, делим Ω на 8 интервалов и т.д.

В результате для каждого элементарного исхода ω последовательность значений имеет вид:

$$\{X_n(\omega)\} = (\underbrace{1, 0}_{2}, \underbrace{0, 0, 1, 0}_{4}, \underbrace{0, 0, 0, 0, 0, 1, 0, 0}_{8}, \dots)$$

последовательность состоит из серий длин $2, 4, 8, 16, \dots, 2^k, \dots$, причем в каждой серии на каком-либо одном месте (зависящем от выбранного элементарного исхода) стоит значение 1, а на остальных местах - нули.

Случайные величины, входящие в серию с номером k (длины 2^k) принимают значение 1 с вероятностью 2^{-k} и значение 0 с вероятностью $1-2^{-k}$. Из основного определения следует, что данная последовательность сходится по вероятности к случайной величине $X \equiv 0$. При этом ни при одном значении ω последовательность значений X_n не сходится к 0, так как в любой последовательности значений сколь угодно далеко обязательно находятся отстоящие от 0 значения. Однако поскольку длины серий неограниченно возрастают, то вероятность "попасть"на это значение становится сколь угодно малой при выборе элемента последовательности с достаточно большим номером.

Заметим, что вместо значения 1 можно выбрать любое другое (в том числе как угодно быстро возрастающее с ростом n), и тем самым сделать последовательность математических ожиданий $\mathbb{M}X_n$ произвольной (в том числе неограниченной). Данный пример показывает, что сходимость по вероятности не влечет сходимости математических ожиданий (равно как и любых других моментов).

Более сильный вид сходимости, который обеспечивает сходимость последовательностей значений к предельному - сходимость почти всюду.

2 Теорема о сходимости при большой выборке

Если две линейные модели имеют разные наборы признаков и одну и ту же зависимую переменную, то остаточная дисперсия корректной модели (s_n^2) имеет меньшее среднее значение, чем некорректной (t_n^2). Этот раздел показывает, что в довольно общих условиях $s_n^2 < t_n^2$ будет выполняться с вероятностью, близкой к 1, при условии, что размер выборки n достаточно велик. Когда размер выборки невелик, то эксперт, который принимает правило принятия решения о выборе модели с меньшей остаточной дисперсией, может принимать неправильные решения со значительной вероятностью. В настоящем разделе показано, что вероятность принятия «неправильной» модели на основе упомянутого выше правила принятия решений сходится к нулю при увеличении размера выборки. Этот результат кажется очевидным, но доказательство не совсем тривиально.

Мы рассматриваем линейную модель с n наблюдениями и k признаками, которая описывается уравнением

$$y_n = \mathbf{X}_n\beta + \varepsilon_n$$

Предполагается, что ошибка одинаково независимо распределена с нулевым средним и дисперсией σ^2: а так же что матрица $\mathbf{X}_n$ размера $n \times k$ является нестохастической. Стохастическая матрица в теории вероятностей — это неотрицательная матрица, в которой сумма элементов любой строки или любого столбца равна единице. Пусть матрица B_n

размера $k \times n$ будет любой матрицей, такой что $\mathbf{X}_n B_n \mathbf{X}_n = \mathbf{X}_n$ и пусть $\mathbf{X}_n B_n$ будет симметричной. Тогда $\mathbf{X}_n B_n$ является идемпотентной, и ее след равен ее рангу. Определим остаточную дисперсию как

$$s_n^2 = y_n^T (I_n - \mathbf{X}_n B_n) y_n / q$$

где $q = n - rank(\mathbf{X}_n)$.

Предположим, что $y_n = \mathbf{X}_n \beta + \varepsilon_n$ является правильно указанной моделью и что альтернативный набор регрессоров представлен нестохастической матрицей Z_n размера $n \times h$ которая удовлетворяет условию того, что пространство, охватываемое его столбцами, не содержит $X_n \beta$. В противном случае, существует вектор γ, такой что $Z_n \gamma = \mathbf{X}_n \beta$ так что обе спецификации будут правильно описывать $E(y_n)$. Следовательно, спецификация, которая содержит все переменные правильной спецификации, а также некоторые дополнительные нерелевантные переменные, не является неправильной в нашем понимании, хотя и может привести к неэффективной оценке для β. Под спецификацией модели в данном случае будем понимать выбор независимых переменных. Пусть C_n размера $h \times n$ будет матрицей такой, что $Z_n C_n Z_n = Z_n$ и $Z_n C_n$ является симметричной. Тогда остаточная дисперсия «неправильной» модели определяется как

$$t_n^2 = y_n^T (I_n - Z_n C_n) y_n / m$$

где $m = n - rank(Z_n)$. Подставляя $y_n = \mathbf{X}_n \beta + \varepsilon_n$ мы получаем

$$t_n^2 = \beta^T \mathbf{X}_n^T H_n \mathbf{X}_n \beta / m + 2\beta^T \mathbf{X}_n^T H_n \varepsilon_n / m + \varepsilon_n^T H_n \varepsilon_n / m \tag{1}$$

где H_n

$$H_n = I_n - Z_n C_n$$

которая удовлетворяет $H_n^T = H_n$, $H_n^2 = H_n$, $H_n Z_n = 0$. Первый правый член в (1) можно интерпретировать как остаточную дисперсию регрессии $\mathbf{X}_n \beta$ на Z. Предположим положительную нижнюю оценку исходя из нашей теоремы:

$$\theta_n^2 \equiv \beta^T \mathbf{X}_n^T H_n \mathbf{X}_n \beta / m \geq g^2 \tag{2}$$

для определенного числа $g > 0$ и от определенного значения n и далее. Таким образом, мы исключаем возможность того, что θ_n^2 будет сходиться к нулю, что будет означать, что ошибка спецификации исчезнет в долгосрочной перспективе. Очевидно, последовательность $\{\theta_n^2\}_{n=1}^{\infty}$ может сходиться к определенному пределу θ_n^2 или расходиться. В первом случае наш основной результат может быть доказан очень простым способом, см. (5). Во втором случае последовательность может либо колебаться, либо стремиться к бесконечности.

Для дальнейшего повествовния напомним, что в условиях, указанных выше

$$\underset{n\to\infty}{plim}\, s_n^2 = \sigma^2 \tag{3}$$

что было показано в [1] для случая $rank(X_n) = k$ и в [2] для более общего ранга $rank(X_n) \leq k$. Так же можно показать, что

$$\underset{n\to\infty}{plim}\, \varepsilon_n^T H_n \varepsilon_n / m = \sigma^2 \tag{4}$$

Теперь мы можем сформулировать и доказать следующий простой результат. Если $\lim\limits_{n\to\infty} \theta_n^2 = \theta_n^2 \geq g^2$ тогда

$$\underset{n\to\infty}{plim}\, t_n^2 = \sigma^2 + \theta^2 \tag{5}$$

Первый правый член в уравнении (1) сходится к θ^2 по предположению, второй член имеет второй момент равный $4\theta_n^2\sigma^2/m$ так что оно сходится к нулю в среднем квадратичном, а для третьего члена мы имеем (4).

Если последовательность $\{\theta_n^2\}_{n=1}^{\infty}$ расходится, то второй член в (1) может стать существенно отрицательным. Это не всегда верно, что этот второй член сходится к нулю, когда Z и ε независимы. Последовательность $\{\theta_n^2\}_{n=1}^{\infty}$ может стремится к бесконечности. Тогда второе слагаемое, которое имеет нулевое среднее значение и дисперсию $4\theta_n^2\sigma^2/m$, может существенно отличаться от нуля в обоих направлениях. Таким образом, мы должны доказать, что это может быть достаточно компенсировано первым членом. Наш основной результат достаточно общий, чтобы включить как случаи сходимости, так и расхождения.

Теорема. *При предположении* (2) $\lim\limits_{n\to\infty} P[t_n^2 < s_n^2 + \lambda g^2] = 0$ *для каждой* λ *удовлетворяющей* $\lambda < 1$.

Доказательство. Мы ограничимся случаем $0 < \lambda < 1$ во избежание технических осложнений. Имея доказательство для этого случая, обобщение для случая $\lambda \leq 0$ очевидно, поскольку функции распределения монотонно не убывают. Доказательство основано на следующей элементарной теореме в теории вероятностей. Пусть $U_1, U_2, \ldots, U_k$

есть произвольные вещественные случайные величины; тогда

$$P[\sum_{i=1}^{k} U_i < 0] \leq P[U_i < 0 \text{ хотя бы для одного } i] \leq \sum_{i=1}^{k} P[U_i < 0] \tag{6}$$

Неравенство в теореме можно записать в виде

$$t_n^2 - \lambda g^2 - s_n^2 = \sum_{i=1}^{4} U_{in} < 0$$

где

$\mathrm{U}_{1n} = (\theta_n + \phi_1 g)(\theta_n - \phi_2 g) - 2\eta \quad (\theta_n > 0)$
$\mathrm{U}_{2n} = \xi_n + (\phi_2 - \phi_1) g\theta_n$
$\mathrm{U}_{3n} = \varepsilon_n^T H_n \varepsilon_n / m - \sigma^2 + \eta$
$\mathrm{U}_{4n} = -s_n^2 + \sigma^2 + \eta$
$\xi_n = 2\beta^T \mathbf{X}_n^T H_n \varepsilon_n / m.$

Были использованы (1) и (2). Числа η, ϕ_1, ϕ_2 могут быть выбраны по желанию при условии $0 < \phi_1\phi_2 = \eta < 1$, но в контексте нашего доказательства мы ограничим их

$$0 < \phi_1 < \phi_2 < 1, \quad 0 < \eta < \frac{1}{2} g^2 (1 + \phi_1)(1 - \phi_2) \tag{8}$$

Существование η, ϕ_1, ϕ_2 с заданными свойствами может быть просто доказано следующим образом. Выберите любой ϕ_2 такой что $0 < \lambda^{1/2} < \phi_2 < 1$. Тогда $0 < \phi_1 = \lambda/\phi_2 < \lambda^{1/2} < \phi_2$.

Итак, теперь, согласно (6), чтобы доказать теорему, достаточно доказать, что

$$\lim_{n\to\infty} P[U_{in} < 0] = 0, \quad (i = 1, \ldots, 4) \tag{9}$$

Мы докажем эти четыре утверждения в обратном порядке. Во-первых, для i=4 и i=3, (9) сразу же следует из (3) и (4) соответственно. Во-вторых, для i=2:

$$P[U_{2n} < 0] = P[\xi_n < -(\phi_2 - \phi_1) g\theta_n] \leq P[|\xi_n| \geq (\phi_2 - \phi_1) g\theta_n]$$

$\leq \frac{E(\xi_n^2)}{(\phi_2-\phi_1)^2 g^2 \theta_n^2} = \frac{4\sigma^2}{(\phi_2-\phi_1)^2 g^2 m}$

при помощи неравенства Чебышева. В-третьих, так как U_{1n} являются фиксированными числами, (9) сводится к $U_{in} > 0$ для $i = 1$. Итак, для доаказательства того, что

$U_{1n} \geq$ мы начнем с его определения и используем неравенства $\theta_n > g > 0, 0 < \phi_1 < \phi_2 < 1$; см. (2) и (8). Тогда мы получаем:

$\theta_n + \phi_1 g > g(1 + \phi_1) > 0,$
$\theta_n - \phi_2 g > g(1 - \phi_2) > 0$

Объединив эти результаты с (8) мы получим

$$(\theta_n + \phi_1 g)(\theta_n - \phi_2 g) > g^2(1 + \phi_1)(1 - \phi_2) > 2\eta$$

что подразумевает желаемый результат и заканчивает доказательство теоремы. $\square$

Список литературы

[1] Kloek T. Note on consistent estimation of the variance of the disturbances in the linear model. – 1970. – №. 2099-2018-3046.

[2] Drygas H. A note on a paper by T. Kloek concerning the consistency of variance estimation in the linear model //Econometrica (pre-1986). – 1975. – Т. 43. – №. 1. – С. 175.