УДК 517.518.45 **DOI 10.52928/2070-1624-2022-38-4-6-12**

МЕТОД СТАТИСТИЧЕСКИХ МОМЕНТОВ В ПОЛИНОМИАЛЬНОЙ РЕГРЕССИИ КОРРЕЛЯЦИОННОЙ ЗАВИСИМОСТИ

д-р техн. наук, проф. С. Г. ЕХИЛЕВСКИЙ, канд. физ.-мат. наук, доц. О. В. ГОЛУБЕВА, О. Н. ЗАБЕЛЕНДИК
(Полоцкий государственный университет)

Методами теории вероятностей обоснована лаконичная процедура, позволяющая выразить параметры полиномиальной регрессии условного математического ожидания через смешанные статистические моменты системы случайных величин. Реализованы примеры линейной и квадратичной регрессии. Во втором случае рассмотрение ограничено ситуацией, когда плотность вероятности случайного аргумента является четной функцией. Результат получен без громоздких выкладок, ибо при его получении использованы не начальные статистические моменты, возникающие в методе наименьших квадратов, а смешанные центральные моменты, отражающие вид регрессионной кривой. Показано, что в общем случае учет нелинейности корреляционной зависимости лишь усиливает неравенство, подтверждающее адекватность регрессионного приближения. Обоснована сходимость такой процедуры, если условное математическое ожидание не является полиномом по сути.

Ключевые слова: *условное математическое ожидание, полиномиальная регрессия, смешанные статистические моменты.*

Введение. Обычно регрессию корреляционной зависимости случайных величин осуществляют методом наименьших квадратов [1], не имеющим никакого отношения к теории вероятностей. К тому же его использование в случае нелинейной регрессии приводит к системам линейных алгебраических уравнений высоких порядков и, как следствие, громоздким выкладкам. По этой причине получение регрессионных кривых, как правило, просто не рассматривается [2]. Вместе с тем известно [3; 4], что закон распределения случайной величины (или их системы) можно восстановить, пользуясь соответствующими статистическими моментами. В частности, такая процедура реализована в [5] при моделировании рабочего процесса респиратора на химически связанном кислороде. Это позволило не только выделить асимптотику процесса, но и определить поправки к ней, обусловленные асимметриями и эксцессами высших порядков. Аналогично в [6] метод статистических моментов позволил рассмотреть независимые повторные испытания как асимптотически гауссовский случайный процесс с дискретным временем и проиллюстрировать справедливость локальной теоремы Лапласа лишь в окрестности наиболее вероятного числа успехов.

В настоящей публикации методами теории вероятностей обосновывается процедура, позволяющая выразить параметры линии регрессии условного матожидания через статистические моменты системы случайных величин. Рассматриваются случаи линейной и квадратичной регрессии.

Линейная регрессия корреляционной зависимости. Рассмотрим двумерную случайную величину $\{X,Y\}$ с возможными значениями (x, y). Аппроксимируем условное математическое ожидание Y прямой линией

$$M(Y|\, X = x) = m_Y(x) \approx k\,x + b\,. \qquad (1)$$

Ее параметры k и b выражаются через начальные и центральные моменты $\{X,Y\}$ независимо от их числовых значений. Для упрощения процедуры получения такого выражения для b будем считать, что среднеквадратическое отклонение σ_X мало. В соответствии с неравенством Чебышева $\forall\,\varepsilon > 0$

$$\lim_{\sigma_X \to 0} P\left(\left| x - m_X \right| < \varepsilon\right) \ge \lim_{\sigma_X \to 0}\left(1 - \sigma_X^2/\varepsilon\right) = 1\,.$$

Это значит, что плотность вероятности X превращается в δ-функцию Дирака

$$\lim_{\sigma_X \to 0} f(x) = \delta(x - m_X)\,,$$

поэтому

$$m_Y = \int_{-\infty}^{\infty} m_Y(x) f(x)\,dx = m_Y(m_X)\,,$$

откуда с учетом (1) получим строгое равенство

$$b = m_Y - k\,m_X\,, \tag{2}$$

ибо короткий участок ($\sigma_X \to 0$) линии $m_Y(x)$ аппроксимируется прямой сколь угодно точно.

Подставив (2) в (1), сведем построение регрессионной прямой к отысканию ее углового коэффициента

$$m_Y(x) - m_Y \approx k(x - m_X)\,. \tag{3}$$

По определению

$$m_Y(x) = \int_{-\infty}^{\infty} y\,f(y|\,x)\,d\,y\,, \tag{4}$$

где $f(y|\,x)$ – условная плотность вероятности Y. Умножим условие ее нормировки $1 = \int_{-\infty}^{\infty} f(y|\,x)\,d\,y$ на m_Y и отнимем полученное от (4):

$$m_Y(x) - m_Y = \int_{-\infty}^{\infty} (y - m_Y)\,f(y|\,x)\,d\,y = \frac{1}{f(x)}\int_{-\infty}^{\infty} (y - m_Y)\,f(x,y)\,d\,y\,. \tag{5}$$

Подставив (5) в (3), умножим результат на $(x - m_X)\,f(x)$ и выполним интегрирование по x:

$$\mu_{XY} \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_X)(y - m_Y)\,f(x,y)\,d\,x\,dy = k\int_{-\infty}^{\infty}(x - m_X)^2 f(x)\,d\,x = k\,\sigma_X^2\,. \tag{6}$$

Отсутствие после интегрирования зависимости от x позволяет подбором k обеспечить строгое равенство в (6). Именно такой выбор обеспечивает минимальность среднего значения условной дисперсии Y, полученной в приближении линейной регрессии корреляционной зависимости, в чем можно убедиться[1] добавив к k произвольную поправку Δ :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(y - m_Y - (k+\Delta)(x - m_X)\right)^2 f(x,y)\,d\,x\,dy = \underbrace{\int_{-\infty}^{\infty}(y - m_Y)^2 f(y)\,d\,y}_{=\sigma_Y^2}\;\underbrace{\int_{-\infty}^{\infty} f(x|\,y)\,d\,x}_{=1} -$$

$$-2(k+\Delta)\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - m_X)(y - m_Y)\,f(x,y)\,d\,x\,dy}_{=k\sigma_X^2\ (см.(6))} +$$

$$+(k+\Delta)^2\underbrace{\int_{-\infty}^{\infty}(x - m_X)^2 f(x)\,d\,x}_{=\sigma_X^2}\;\underbrace{\int_{-\infty}^{\infty} f(y|\,x)\,d\,y}_{=1} = \sigma_Y^2 - 2(k+\Delta)k\,\sigma_X^2 + (k+\Delta)^2\sigma_X^2 =$$

$$= \sigma_Y^2 - k^2\sigma_X^2 + \Delta^2\sigma_X^2 > \sigma_Y^2 - k^2\sigma_X^2\,, \tag{7}$$

Данный результат допускает следующую интерпретацию. Правильный выбор значений k и b обеспечивает совпадение точного значения смешанного центрального момента 2-го порядка μ_{XY} с полученным в приближении линейной регрессии корреляционной зависимости. Именно это обстоятельство исключает вклад в среднее значение условной дисперсии Y, обусловленный погрешностью, вносимой

[1] Именно это является обоснованием того, что параметры k и b можно определять методом наименьших квадратов на основе экспериментальных данных по измерению двумерной случайной величины $\{X, Y\}$.

в μ_{XY} процедурой регрессии. При этом оставшаяся погрешность, связанная с процедурой регрессии, обусловлена только смешанными центральными моментами более высоких порядков. Чтобы избавиться от нее нужно добавлять в (1) новые параметры (повышать степень регрессионного полинома).

С учетом соотношения (6) для углового коэффициента k уравнение регрессии (3) примет вид

$$(m_Y(x) - m_Y)/\sigma_Y \approx r_{XY}(x - m_X)/\sigma_X , \tag{8}$$

где

$$r_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{x - m_X}{\sigma_X} \cdot \frac{y - m_Y}{\sigma_Y} f(x, y)\,dx\,dy = \frac{\mu_{XY}}{\sigma_X \sigma_Y} \tag{9}$$

– коэффициент линейной корреляции (центральный смешанный момент 2-го порядка приведенных случайных величин *X* и *Y*).

Согласно (6), (7) и (9) количественной характеристикой разброса $\{X, Y\}$ вокруг регрессионной прямой (8) является интеграл

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(y - m_Y - k(x - m_X)\right)^2 f(x, y)\,dx\,dy = \sigma_Y^2\left(1 - r_{XY}^2\right). \tag{10}$$

Вклад в правую часть (10) дает условная дисперсия $\sigma_Y^2(x)$ и погрешность, возникающая при замене $m_Y(x)$ прямой линией. Чтобы оценить адекватность такого приближения, определим точный вклад среднего значения $\sigma_Y^2(x)$ в дисперсию Y

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - m_Y)^2 f(x, y)\,dx\,dy = \underbrace{\int_{-\infty}^{\infty} (y - m_Y)^2 f(y)\,dy}_{=\sigma_Y^2} \underbrace{\int_{-\infty}^{\infty} f(x|\,y)\,dx}_{=1} = \sigma_Y^2 \cdot 1 =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(y - m_Y(x) + m_Y(x) - m_Y\right)^2 f(x, y)\,dx\,dy = \int_{-\infty}^{\infty} f(x)\,dx \underbrace{\int_{-\infty}^{\infty} \left(y - m_Y(x)\right)^2 f(y|\,x)\,dy}_{\sigma_Y^2(x)} -$$

$$-2\underbrace{\int_{-\infty}^{\infty} \left(m_Y(x) - m_Y\right) f(x)\,dx}_{=0} \underbrace{\int_{-\infty}^{\infty} \left(y - m_Y(x)\right) f(y|\,x)\,dy}_{=0} +$$

$$+\int_{-\infty}^{\infty} \left(m_Y(x) - m_Y\right)^2 f(x)\,dx \underbrace{\int_{-\infty}^{\infty} f(y|\,x)\,dy}_{=1} = M\left(\sigma_Y^2(x)\right) - 2\cdot 0\cdot 0 + D\left(m_Y(x)\right)\cdot 1 , \tag{11}$$

где учтены условия нормировки, а также определения математического ожидания

$$M\left(m_Y(x)\right) = \int_{-\infty}^{\infty} m_Y(x) f(x)\,dx = \int_{-\infty}^{\infty}\underbrace{\int_{-\infty}^{\infty} y f(y|\,x)\,dy}_{=m_Y(x)} f(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x, y)\,dx dy = m_Y ,$$

дисперсии

$$\int_{-\infty}^{\infty} \left(m_Y(x) - m_Y\right)^2 f(x)\,dx = \int_{-\infty}^{\infty} \left(m_Y(x) - M\left(m_Y(x)\right)\right)^2 f(x)\,dx = D\left(m_Y(x)\right), \quad \int_{-\infty}^{\infty} \sigma_Y^2(x) f(x)\,dx = M\left(\sigma_Y^2(x)\right)$$

и равенство нулю центральных моментов первого порядка

$$\int_{-\infty}^{\infty} \left(y - m_Y(x)\right) f(y|\, x)\, dy = \underbrace{\int_{-\infty}^{\infty} y\, f(y|\, x)\, dy}_{=m_Y(x)} - m_Y(x) \underbrace{\int_{-\infty}^{\infty} f(y|\, x)\, dy}_{=1} = 0,$$

$$\int_{-\infty}^{\infty} \left(m_Y(x) - m_Y\right) f(x)\, dx = \underbrace{\int_{-\infty}^{\infty} m_Y(x) f(x)\, dx}_{M\left(m_Y(x)\right)=m_Y} - m_Y \underbrace{\int_{-\infty}^{\infty} f(x)\, dx}_{=1} = 0\,.$$

Согласно (11) и благодаря тому, что разброс $\{X,Y\}$ вокруг условного математического ожидания меньше, чем вокруг регрессионной прямой

$$M\left(\sigma_Y^2(x)\right) = \sigma_Y^2\left(1 - D\left(m_Y(x)\right)/\sigma_Y^2\right) \le \sigma_Y^2\left(1 - r_{XY}^2\right).$$

Иными словами,

$$\eta^2 = D\left(m_Y(x)\right)/\sigma_Y^2 \ge r_{XY}^2\,, \tag{12}$$

причем равенство возможно только в ситуации точного равенства (8), когда корреляция линейна по сути. В остальных случаях адекватность приближения (8) подтверждается силой неравенства

$$\left(\eta^2 - r_{XY}^2\right)/\eta^2 \ll 1\,. \tag{13}$$

Определенную согласно (12) величину η называют корреляционным отношением.

Если условие (13) нарушено, регрессия должна быть нелинейной, чтобы описывающий ее полином лучше «вписывался» в график условного матожидания $m_Y(x)$.

Квадратичная регрессия корреляционной зависимости. Если график $m_Y(x)$ не содержит перегибов, усилить неравенство (13) позволяет квадратичная регрессия

$$m_Y(x) \approx a x^2 + bx + c\,. \tag{14}$$

Выразим ее параметры a, b, c через статистические моменты (в том числе и смешанные) системы случайных величин $\{X,Y\}$.

Для этого умножим (14) на плотность вероятности $f(x)$ и выполним интегрирование по всем x. Точность полученного в результате такого усреднения равенства

$$m_Y = am_{X^2} + bm_X + c \tag{15}$$

при любых a и b обеспечивается соответствующим подбором значения c. Подставив его в (14), получим

$$m_Y(x) - m_Y \approx a\left(x^2 - m_{X^2}\right) + b\left(x - m_X\right). \tag{16}$$

Подставим (5) в (16), умножим результат на $(x - m_X) f(x)$ и выполним интегрирование по x:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_X)(y - m_Y) f(x,y)\, dx\, dy = a\int_{-\infty}^{\infty} \left(x^2 - m_{X^2}\right)(x - m_X) f(x)\, dx + b\underbrace{\int_{-\infty}^{\infty} (x - m_X)^2 f(x)\, dx}_{=\sigma_X^2}. \tag{17}$$

Отсутствие после интегрирования зависимости от x позволяет подбором a и b обеспечить в (17) строгое равенство.

Преобразуем фигурирующий в правой части (17) интеграл

$$\int_{-\infty}^{\infty}\left(x^2-m_{X^2}\right)(x-m_X)f(x)\,dx=\underbrace{\int_{-\infty}^{\infty}x^3 f(x)\,dx}_{=m_{X^3}}-m_{X^2}\underbrace{\int_{-\infty}^{\infty}x f(x)\,dx}_{=m_X}-m_X\underbrace{\int_{-\infty}^{\infty}\left(x^2-m_{X^2}\right)f(x)\,dx}_{=0}, \tag{18}$$

где учтено равенство нулю центрального момента первого порядка. В результате с учетом (17), (18) и (9) получим первое уравнение для определения a и b

$$\sigma_X\sigma_Y r_{XY}=a\left(m_{X^3}-m_{X^2}m_X\right)+b\sigma_X^2\,. \tag{19}$$

Для получения второго уравнения осуществим аналогичную процедуру. После подстановки (5) в (16) умножим результат на $\left(x^2-m_{X^2}\right)f(x)$ и выполним интегрирование по x

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x^2-m_{X^2}\right)\left(y-m_Y\right)f(x,y)\,dx\,dy\equiv\sigma_{X^2}\sigma_Y r_{X^2Y}=$$

$$=a\underbrace{\int_{-\infty}^{\infty}\left(x^2-m_{X^2}\right)^2 f(x)\,dx}_{=\sigma_{X^2}^2}+b\underbrace{\int_{-\infty}^{\infty}\left(x^2-m_{X^2}\right)(x-m_X)f(x)\,dx}_{=m_{X^3}-m_X m_{X^2}}, \tag{20}$$

где последнее равенство записано с учетом (18).

Согласно (20) второе уравнение для определения a и b, записанное в компактной форме, имеет вид

$$\sigma_{X^2}\sigma_Y r_{X^2Y}=a\sigma_{X^2}^2+b\left(m_{X^3}-m_{X^2}m_X\right). \tag{21}$$

Решение системы двух линейных алгебраических уравнений (19), (21) позволяет в общем виде выразить неизвестные a и b через фигурирующие в ней статистические моменты закона распределения $\{X,Y\}$ и завершить процедуру квадратичной регрессии корреляционной зависимости Y от X.

Чтобы упростить анализ полученных результатов, ограничим дальнейшее рассмотрение ситуацией, когда плотность вероятности X является четной функцией ($f(-x)=f(x)$). В этом случае $m_{X^3}=m_X=0$ и согласно (19), (21)

$$b=\sigma_Y r_{XY}/\sigma_X\,,\qquad a=\sigma_Y r_{X^2Y}/\sigma_{X^2}\,, \tag{22}$$

что позволяет записать (16) в аналогичном (8) виде:

$$\frac{m_Y(x)-m_Y}{\sigma_Y}\approx r_{X^2Y}\frac{x^2-m_{X^2}}{\sigma_{X^2}}+r_{XY}\frac{x}{\sigma_X}\,. \tag{23}$$

Приближенное равенство (23) получено без громоздких выкладок, ибо мы отталкивались не от начальных статистических моментов, возникающих в методе наименьших квадратов, а от смешанных центральных моментов, отражающих вид регрессионной кривой.

Покажем, что нелинейность регрессии снижает среднеквадратический разброс Y вокруг ее графика (см. (16) по сравнению с полученным в линейном приближении (см. (10):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(y-m_Y-a\left(x^2-m_{X^2}\right)-b\left(x-m_X\right)\right)^2 f(x,y)\,dx\,dy=\underbrace{\int_{-\infty}^{\infty}\left(y-m_Y\right)^2 f(y)\,dy}_{=\sigma_Y^2}\underbrace{\int_{-\infty}^{\infty}f(x|\,y)\,dx}_{=1}-$$

$$-2\underbrace{\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(y-m_Y)\left(a\left(x^2-m_{X^2}\right)+b(x-m_X)\right)f(x,y)\,dx\,dy}_{=a^2\sigma^2_{X^2}+b^2\sigma^2_X\ (см.(17),(20))}+\underbrace{\int\limits_{-\infty}^{\infty}\left(a\left(x^2-m_{X^2}\right)+b(x-m_X)\right)^2 f(x)\,dx}_{=a^2\sigma^2_{X^2}+b^2\sigma^2_X}\underbrace{\int\limits_{-\infty}^{\infty}f(y|\,x)\,dy}_{=1}=$$

$$=\sigma_Y^2-a^2\sigma^2_{X^2}-b^2\sigma_X^2=\sigma_Y^2\left(1-r^2_{X^2Y}-r^2_{XY}\right)<\sigma_Y^2\left(1-r^2_{XY}\right). \tag{24}$$

При этом неравенство (12) приобретает вид

$$\eta^2=D\left(m_Y(x)\right)\big/\sigma_Y^2\ge r^2_{X^2Y}+r^2_{XY}, \tag{25}$$

причем равенство возможно только в ситуации точного равенства (23), когда корреляция квадратична по сути.

В общем случае учет нелинейности корреляционной зависимости лишь усиливает неравенство, подтверждающее адекватность регрессионного приближения (см. (13)

$$\left(\eta^2-r^2_{X^2Y}-r^2_{XY}\right)\big/\eta^2<\left(\eta^2-r^2_{XY}\right)\big/\eta^2\ll 1. \tag{26}$$

Если и этого недостаточно (например, график условного матожидания содержит точку перегиба), целесообразно заменить в (14) квадратный полином на кубический и т. д. Сходимость такой процедуры (если $m_Y(x)$ не полином по сути) вытекает из того, что правая часть (26) неотрицательна по смыслу.

Разумеется, наличие $f(x,y)$ позволяет непосредственно определить подлинную зависимость $m_Y(x)$. Необходимость в приближении (8) или (23) и, соответственно, критериях их адекватности (13), (26) возникает, когда о корреляции судят на основе экспериментальных данных по измерению $\{X,Y\}$. При этом приближенные числовые значения фигурирующих в (8), (13) и (23), (26) параметров (m_Y, m_X, σ_Y, σ_X, η, r_{XY}, m_{X^2}, σ_{X^2}, r_{X^2Y}) получаются с помощью статистических оценок.

Заключение. Таким образом, методами теории вероятностей обоснована лаконичная процедура, позволяющая выразить параметры полиномиальной регрессии условного математического ожидания через смешанные статистические моменты системы случайных величин $\{X,Y\}$. Реализованы примеры линейной и квадратичной регрессии. Во втором случае рассмотрение ограничено ситуацией, когда плотность вероятности случайного аргумента X является четной функцией. Результат получен без громоздких выкладок, ибо при его получении использованы не начальные статистические моменты, возникающие в методе наименьших квадратов, а смешанные центральные моменты, отражающие вид регрессионной кривой. Показано, что в общем случае учет нелинейности корреляционной зависимости лишь усиливает неравенство, подтверждающее адекватность регрессионного приближения. Обоснована сходимость такой процедуры, если условное математическое ожидание не является полиномом по сути.

ЛИТЕРАТУРА

1. Гмурман, В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – М. : Высш. шк., 1972. – 368 с.
2. Коваленко, И. Н. Теория вероятностей и математическая статистика / И. Н. Коваленко, А. А. Филиппова. – М. : Высш. шк., 1973. – 368 с.
3. Пытьев, Ю. П. Курс теории вероятностей и математической статистики для физиков / Ю. П. Пытьев, И. А. Шишмарев. – М. : Изд-во Моск. ун-та, 1983. – 252 с.
4. Contribution of excesses of the gamma distribution to the asymptotics of factorials with large arguments / S. G. Ekhilevskiy [et al.] // System analysis and information Technology. – 2019. – №1(15) – 2(16). – С. 119–125.
5. Ехилевский, С. Г. Теоретико-вероятностный подход к моделированию респиратора на химически связанном кислороде / С. Г. Ехилевский, О. В. Голубева, Е. П. Потапенко // Безопасность труда в промышленности. – 2020. – № 10. – С. 7–15.
6. Независимые повторные испытания как асимптотически гауссовский случайный процесс / С. Г. Ехилевский [и др.] // Вестн. Полоц. гос. ун-та. Сер. С, Фундам. науки. – 2016. – № 2. – С. 111–116.

REFERENCES

1. Gmurman, V. E. (1972) *Teoriya veroyatnostei i matematicheskaya statistika [Theory of Probability and Mathematical Statistics]*. Moscow: Vysshaya shkola. (In Russ.).
2. Kovalenko, I. N., & Filippova, A. A. (1973). *Teoriya veroyatnostei i matematicheskaya statistika [Theory of Probability and Mathematical Statistics]*. Moscow: Vysshaya shkola. (In Russ.).

3. Pyt'ev, Yu. P., & Shishmarev. I. A. (1983). *Kurs teorii veroyatnostei i matematicheskoi statistiki dlya fizikov [Course of Probability Theory and Mathematical Statistics for Physicists]*. Moscow: Publ. Mosk. un-ta. (In Russ.).
4. Ekhilevskiy, S. G., Golubeva, O. V., Zabelendik, O. N., & Struk, T. S. (2019) Contribution of excesses of the gamma distribution to the asymptotics of factorials with large arguments. *System analysis and information Technology*, *1*(15) – *2*(16), 119–125.
5. Ekhilevskiy, S. G., Golubeva, O. V., & Potapenko, E. P. (2020). Teoretiko-veroyatnostnyi podkhod k modelirovaniyu respiratora na khimicheski svyazannom kislorode. *Bezopasnost' truda v promyshlennosti [Labor safety in industry]*, (10), 7–15. (In Russ.).
6. Ekhilevskiy, S. G., Golubeva, O. V., Potapenko, E. P., & Rud'kova, T. S. (2016) Nezavisimye povtornye ispytaniya kak asimptoticheski gaussovskii sluchainyi protsess [Independent Retests as an Asymptotically Gaussian Stochastic Process]. *Vestnik Polotskogo gosudarstvennogo universiteta. Seriya C, Fundamental'nye nauki [Herald of Polotsk State University. Series C. Fundamental sciences]*, (2), 111–116.

Поступила 17.03.2022

THE METHOD OF STATISTICAL MOMENTS IN A POLYNOMIAL REGRESSION OF CORRELATION DEPENDENCE

S. EKHILEVSKIY, O. GOLUBEVA, O. ZABELENDIK

The methods of probability theory substantiate a brief procedure that allows us to express the parameters of the polynomial regression of conditional mathematical expectation through mixed statistical moments of a system of random variables. Examples of linear and quadratic regression are implemented. In the second case, consideration is limited to the situation when the probability density of a random argument is an even function. The result was obtained without cumbersome calculations, since it was obtained using not the initial statistical moments resulting from the least squares method, but mixed central moments reflecting the type of regression curve. It is shown that in the general case, taking into account the nonlinearity of the correlation dependence only strengthens the inequality, which is the criterion for the adequacy of the regression approximation. The convergence of such a procedure is confirmed if the conditional expectation is not essentially a polynomial.

Keywords: *conditional mathematical expectation, polynomial regression, mixed statistical moments.*