УДК 519.85

В.Н. КУЗЬМЕНКО, Э.И. НЕНАХОВ

О ВЗАИМОЗАМЕНЯЕМОСТИ НЕКОТОРЫХ ФУНКЦИЙ ОШИБКИ ПРИ РЕГРЕССИОННОМ АНАЛИЗЕ

Рассматривается возможность сведения вычисления суперквантильной ошибки к вычислению ошибки «Рокафеллара» для случайной величины с дискретным равномерным распределением. Приводятся два варианта сведения и исследуется вопрос эквивалентности использования исходной ошибки и предложенных вариантов в регрессионном анализе.

Введение. В работе [1] строится теория неких наборов функций, определенных для случайной величины и связанных с понятием риска (Risk Quadrangle – риск-квадрат). В каждый набор входит четыре функции – риск, отклонение, ошибка и потери. Функции связаны между собой общими для любого набора соотношениями. Выбор того или иного набора позволяет оценивать различные характеристики случайной величины. Мы рассматриваем наборы, связанные с суперквантильной ошибкой [2] и ошибкой, вычисляемой как линейная комбинация квантильных ошибок (ошибкой «Рокафеллара») [1]. При определенных условиях первый набор может быть замене вторым в задачах регрессионного анализа.

Рассматривается случайная величина X с функцией распределения $F_X(x)=P(X\le x)$.

Для некоторого $\alpha\in(0,1)$ будем рассматривать такие характеристики: α-квантиль – $q_\alpha(X)$; α-суперквантиль – $Q_\alpha(X)$.

α-квантиль определяется как обратная к $F_X(x)$ функция, т. е. $q_\alpha(X)=\{x\,|\,F_X(x)=\alpha\}$. В общем случае такое определение может давать неоднозначное значение или нужное значение x может не существовать. Поэтому, рассматривается два крайних значения – верхнее $q^+_\alpha(X)=\inf\{x\,|\,F_X(x)>\alpha\}$ и нижнее $q^-_\alpha(X)=\sup\{x\,|\,F_X(x)<\alpha\}$, а α-квантиль для целей статьи определяется как интервал $q_\alpha(X)=[q^-_\alpha(X),q^+_\alpha(X)]$, хотя стандартно он определяется как $q^-_\alpha(X)$.

α -квантиль обозначают также как $VaR_{\alpha}(X)$ (value-at-risk).

α -суперквантиль определяется как среднее значение X для значений не меньших, чем $q_{\alpha}^{-}(X)$, или среднее значение на «правом α -хвосте» распределения: $Q_{\alpha}(X)=\int_{\alpha}^{1} q_{\beta}^{-}(X)d\beta$. Если для некоторых β $q_{\beta}(X)$ является интервалом, то выбор значения в пределах интервала не влияет на значение интеграла. α -суперквантиль обозначают также как $CVaR_{\alpha}(X)$ (conditional value-at-risk).

В статистике и регрессионном анализе используются различные виды функций ошибок. Мы будем рассматривать три вида.

Квантильная ошибка $\varepsilon_{\alpha}(X)=\dfrac{1}{1-\alpha}\int_{0}^{1}\max\{0,q_{\beta}^{-}(X)\}d\beta-EX$, где E – символ среднего значения.

Ошибка, основанная на смеси квантильных ошибок, (ошибка Рокафеллара)

$$\varepsilon_{R}(X)=\min_{B_1,\dots,B_K}\left\{\sum_{k=1}^{K}\lambda_k\varepsilon_{\alpha_k}(X-B_k)\mid\sum_{k=1}^{K}\lambda_k B_k=0\right\}-EX,$$

где $\alpha_k\in(0,1)$, $\lambda_k>0$, $\sum_{k=1}^{K}\lambda_k=1$.

Суперквантильная ошибка $\overline{\varepsilon}_{\alpha}(X)=\dfrac{1}{1-\alpha}\int_{0}^{1}\max\{0,Q_{\beta}(X)\}d\beta-EX$.

Перечисленные ошибки согласно [1] входят в «регулярные квадраты» связанных функций – error, regret, deviation, risk. Согласно свойств «регулярных квадратов» величина $D(X)=\min\limits_{C}\varepsilon(X-C)$ – это отклонение (deviation), соответствующее ошибке, а $S(X)=\arg\min\limits_{C}\varepsilon(X-C)$ – статистика.

Для перечисленных ошибок отклонения равны $D_{\alpha}(X)=\min\limits_{C}\varepsilon_{\alpha}(X-C)=$ $=Q_{\alpha}(X)-EX$; $D_R(X)=\min\limits_{C}\varepsilon_R(X-C)=\sum_{k=1}^{K}\lambda_k Q_{\alpha_k}(X)-EX$; $\overline{D}_{\alpha}(X)=$ $=\min\limits_{C}\overline{\varepsilon}_{\alpha}(X-C)=\dfrac{1}{1-\alpha}\int_{\alpha}^{1}Q_{\beta}(X)d\beta-EX$, а соответствующие статистики – $s_{\alpha}(X)=q_{\alpha}(X)$; $s_R(X)=\sum_{k=1}^{K}\lambda_k q_{\alpha_k}(X)$; $\overline{S}_{\alpha}(X)=Q_{\alpha}(X)$.

В регрессионном анализе функции ошибок используются для построения моделей, оцениваемой величины, в зависимости от независимых факторов и оценки характеристик распределения случайной величины. Для этого коэффициенты модели находятся путем минимизации функции ошибки.

В случае линейной регрессии или модели, в которую входит линейный коэффициент с постоянным множителем (intercept) минимизация функции ошибки может быть заменена минимизацией отклонения. Пусть X будет случайной величиной, задаваемой результатами наблюдений x_i, $\overline{f}_i$ и моделью M: $X=\{x_i-M(\overline{c},\overline{f}_i)-C\mid i=1,\dots,n\}$, где n – количество наблюдений, x_i – значения зависимой величины, $\bar{f}_i$ – вектора значений независимых факторов, $\bar{c}$ – вектор неизвестных коэффициентов, C – intercept, все вероятности равны, $p_i = 1/n$.

Введем случайную величину $Y = \{x_i - M(\bar{c}, \bar{f}_i) \mid i = 1,...,n\}$.

Тогда $\min\limits_{\bar{c},C} \varepsilon(X(\bar{c}, C)) = \min\limits_{\bar{c},C} \varepsilon(Y(\bar{c}) - C) = \min\limits_{\bar{c}} D(Y(\bar{c}))$, а оптимальный коэффициент C^* становится зависимым от остальных коэффициентов и равным оцениваемой величине $C^*(Y(\bar{c})) = \arg\min\limits_{C} \varepsilon(Y(\bar{c}) - C) = S(Y(\bar{c}))$.

Далее будем рассматривать линейную модель $X = \{x_i - -(\bar{c}, \bar{f}_i) - C \mid i = 1,...,n\}$, где $(\bar{c}, \bar{f}_i)$ – скалярное произведение, $\bar{\varepsilon}_\alpha$, ε_R – ошибки, $\bar{D}_\alpha$, D_R – соответствующие отклонения. Отклонение $\bar{D}_\alpha$ необходимо вычислять интеграл от суперквантиля. В случае использования D_R достаточно вычислять сумму суперквантилей. Предполагается, что вычислять такую сумму проще и быстрее, что может быть существенным выигрышем при оптимизации.

Первый вопрос ставится следующим образом. Можно ли заменить интеграл такой суммой суперквантилей, чтобы отклонения $\bar{D}_\alpha(Y(\bar{c}))$, $D_R(Y(\bar{c}))$ были равны при различных значениях вектора $\bar{c}$? То есть речь идет о нахождении параметров λ_k, α_k и их количества K.

Введем ряд обозначений и определений:

$\beta_i = \dfrac{i}{n}$, $\sigma_i = 1 - \beta_i$, $i = 0,...,n$, $\delta = \beta_i - \beta_{i-1} = \dfrac{1}{n}$;

i_α – такой индекс $i = 1,...,n$, что $\beta_{i_\alpha - 1} < \alpha \le \beta_{i_\alpha}$; $\delta_\alpha = \beta_{i_\alpha} - \alpha < \delta$; (заметим, что при $\alpha \in (0,1)$ $i_\alpha \in \{1,...,n\}$);

$p'_{i_\alpha} = \delta_\alpha/(1-\alpha)$, $p'_i = \delta/(1-\alpha)$ для $i_\alpha < i \le n$;

$\gamma_{i_\alpha} = 1 - \delta_\alpha / \ln(1 + \delta_\alpha/\sigma_i)$ для $i_\alpha < n$; $\gamma_{i_\alpha} = 1 - \delta_\alpha/2$ для $i_\alpha = n$;

$\gamma_i = 1 - \delta/\ln(1 + \delta/\sigma_i)$ для $i_\alpha < i < n$; $\gamma_i = 1 - \delta/2$ для $i = n > i_\alpha$;

$\bar{p}_i = \sigma_i/(\delta(1-\alpha))(\delta_\alpha + \sigma_{i+1}\ln(\sigma_{i+1}/(1-\alpha)))$ для $i = i_\alpha - 1$;

$\bar{p}_i = \sigma_i/(\delta(1-\alpha))(\delta - \delta_\alpha + \sigma_{i+1}\ln(\sigma_{i+1}/\sigma_i) - \sigma_{i-1}\ln(\sigma_i/(1-\alpha)))$ для $i = i_\alpha < n$;

$\bar{p}_i = \sigma_i/(\delta(1-\alpha))(\sigma_{i+1}\ln(\sigma_{i+1}/\sigma_i) - \sigma_{i-1}\ln(\sigma_i/\sigma_{i-1}))$ для $i_\alpha < i < n$.

(Чтобы сделать изложение короче, выражение $\sigma\ln\sigma$ полагается равным 0 для $\sigma = 0$. Это результат предельного перехода $\lim\limits_{\sigma\to 0} \sigma\ln\sigma = 0$).

Теорема 1. В случае конечного дискретного равномерного распределения для любого вектора $\bar{c}$ справедливы равенства

$$\frac{1}{1-\alpha}\int_\alpha^1 Q_\beta(Y(\bar{c}))d\beta = \sum_{i=i_\alpha}^{n} p'_i Q_{\gamma_i}(Y(\bar{c})) = \sum_{i=i_\alpha-1}^{n-1} \bar{p}_i Q_{\beta_i}(Y(\bar{c})). \qquad (1)$$

Доказательство. В силу громоздкости используемых формул приведем лишь основные соображения.

Для любого интервала $[\beta_{i-1},\beta_i]$ существует $\gamma_i \in (\beta_{i-1},\beta_i)$ такое, что $\int_{\beta_{i-1}}^{\beta_i} Q_\beta d\beta = (\beta_i - \beta_{i-1})Q_{\gamma_i}$. В силу равномерности распределения значения β_i, $i = 1,...,n,$ не зависят от $\bar{c}$. Они являются точками разрыва $q_\alpha^-(Y(\bar{c}))$ как функции от α, а на интервалах (β_{i-1},β_i) функция $q_\alpha^-(Y(\bar{c}))$ постоянна. Поэтому коэффициенты γ_i также не зависят от $\bar{c}$.

Для второй суммы существуют коэффициенты ρ_{1i}, ρ_{2i} такие, что $\int_{\beta_{i-1}}^{\beta_i} Q_\beta d\beta = \rho_{1i}Q_{\beta_i} + \rho_{2i}Q_{\beta_{i-1}}$. В силу равномерности распределения ρ_{1i}, ρ_{2i} и $\bar{p}_i = (\rho_{1i} + \rho_{2i+1})/(1-\alpha)$ также не зависят от $\bar{c}$.

Непосредственно проверяется, что определенные выше коэффициенты $\gamma_i, \bar{p}_i$ удовлетворяют указанным требованиям. Следовательно, равенства (1) справедливы при любом $\bar{c}$. ☻

Назовем первую сумму в (1) суммой типа I и соответствующее отклонение обозначим $D_R^I(Y(\bar{c}))$, вторую – типа II и отклонение обозначим $D_R^{II}(Y(\bar{c}))$.

Следствием теоремы является положительный ответ на первый вопрос: для построенных сумм суперквантилей (сумм первого и второго типов) $\overline{D}_\alpha(Y(\bar{c})) = D_R^I(Y(\bar{c})) = D_R^{II}(Y(\bar{c}))$ независимо от $\bar{c}$ (суммы и интеграл отличаются от отклонений на $E(Y(\bar{c}))$).

Следствием теоремы есть такие свойства и соотношения: $\sum_{i=i_\alpha}^{n} p_i' = \sum_{i=i_\alpha-1}^{n-1} \bar{p}_i = 1$, что непосредственно проверяется; $\gamma_i \in (\beta_{i-1},\beta_i)$, что говорит о том, что в точках γ_i $q_{\gamma_i}(Y(\bar{c}))$ определяется однозначно.

Второй вопрос, который мы ставим, – как соотносятся статистики $\overline{S}_\alpha(Y(\bar{c})),\ s_R^I(Y(\bar{c})), s_R^{II}(Y(\bar{c}))$?

(Далее будем опускать $\bar{c},$ так как результаты от него не зависят).

Согласно определения [1] $\overline{S}_\alpha(Y) = Q_\alpha(Y)$.

Для $D_R^I(Y)$ согласно определения из [1], а также однозначности $q_{\gamma_i}(Y)$ в точках γ_i и согласно определения $Q_\alpha(Y)$ ($CVaR_\alpha(Y)$) из [3] $s_R^I(Y) = \sum_{i=i_\alpha}^{n} p_i' s_{\gamma_i}(Y) = \sum_{i=i_\alpha}^{n} p_i' q_{\gamma_i}(Y) = Q_\alpha(Y)$.

Видим, что $\overline{S}_\alpha(Y) = s_R^I(Y)$.

Для $D_R^{II}(Y)$ согласно определения [1] и неоднозначности $q_\alpha(Y)$ в точках β_i статистика является «суммой интервалов». Для случая $i_\alpha > 1$

$$s_R^{II}(Y) = \sum_{i=i_\alpha-1}^{n-1} \overline{p}_i s_{\beta_i}(Y) = \sum_{i=i_\alpha-1}^{n-1} \overline{p}_i q_{\beta_i}(Y) =$$

$$= \left[\sum_{i=i_\alpha-1}^{n-1} \overline{p}_i q_{\beta_i}^-(Y), \sum_{i=i_\alpha-1}^{n-1} \overline{p}_i q_{\beta_i}^+(Y) \right] = \left[\sum_{i=i_\alpha-1}^{n-1} \overline{p}_i q_{\gamma_i}(Y), \sum_{i=i_\alpha-1}^{n-1} \overline{p}_i q_{\gamma_{i+1}}(Y) \right],$$

где значение γ_i для $i = i_\alpha - 1$ определено как $\gamma_i = (\beta_i + \beta_{i-1})/2$.

В случае $i_\alpha = 1$ в сумму входит бесконечный интервал $s_{\beta_0}(Y) = (-\infty, q_{\gamma_1}(Y)]$, тогда $s_R^{II}(Y) = \left(-\infty, \sum_{i=0}^{n-1} \overline{p}_i q_{\gamma_{i+1}}(Y) \right]$.

Для ответа на вопрос о соотношении $\overline{S}_\alpha(Y)$ и $s_R^{II}(Y)$ исследуем некоторые свойства коэффициентов $\overline{p}_i$, p_i'.

Лемма 1. Последовательность $(\overline{p}_i)$ монотонно возрастающая для $i = i_\alpha + 1, \ldots, n-1$.

Доказательство. Рассмотрим общую формулу $\overline{p}_i$ для $i_\alpha < i < n$, как функцию непрерывных параметров σ_i и δ, опуская индекс i,

$$\overline{p}(\sigma,\delta) = \frac{\sigma}{\delta(1-\alpha)} \left((\sigma-\delta)\ln\left(\frac{\sigma-\delta}{\sigma}\right) - (\sigma+\delta)\ln\left(\frac{\sigma}{\sigma+\delta}\right) \right).$$

Выполнив замену $\delta/\sigma = \rho$, получаем

$$\overline{p}(\rho,\delta) = \frac{\delta}{\rho(1-\alpha)} \big((1-\rho)\ln(1-\rho) + (1+\rho)\ln(1+\rho) \big).$$

Увеличение индекса i для $\overline{p}_i$ эквивалентно увеличению ρ для $\overline{p}(\rho,\delta)$. δ для $i_\alpha < i < n$ остается неизменным.

Производную от $\overline{p}(\sigma,\delta)$ по ρ при $\rho \le 1$ оценим, используя ряд Тейлора,

$$\overline{p}_\rho'(\rho,\delta) = \frac{\delta\rho}{(1-\alpha)} \left(\frac{1}{3} + \frac{4}{15}\rho^2 + \frac{3}{14}\rho^4 + \ldots \right) > 0.$$

Следовательно, $(\overline{p}_i)_{i=i_\alpha+1}^{n-1}$ есть монотонно возрастающей. ☻

Лемма 2. При $i = i_\alpha + 1, \ldots, n-1$ справедливо $\overline{p}_i > \delta/(1-\alpha) = p_{i+1}'$.

Доказательство. Вычислим предел отношения $\overline{p}_i$ и p_{i+1}', перейдя, как и выше, к непрерывной функции $\overline{p}(\sigma,\delta)$,

$$\lim_{\delta\to 0}\frac{\bar{p}(\sigma,\delta)}{\delta/(1-\alpha)}=\lim_{\delta\to 0}\frac{1}{\delta}\left(\frac{\sigma^2}{\delta}\ln\left(1-\frac{\delta^2}{\sigma^2}\right)+\sigma\ln\left(1+\frac{2\delta}{\sigma-\delta}\right)\right)=$$

$$=\lim_{\delta\to 0}\left[\left(-1-\frac{\delta^2}{2\sigma^2}-\ldots\right)+\frac{\sigma}{\delta}\left(\frac{2\delta}{\sigma-\delta}-\frac{1}{2}\left(\frac{2\delta}{\sigma-\delta}\right)^2+\ldots\right)\right]=-1+2=1.$$

То есть в пределе при $\delta\to 0$ $\bar{p}(\sigma,\delta)=\delta/(1-\alpha)$.

Теперь рассмотрим производные по δ при $0<\delta\le\sigma$:

$$\bar{p}'_{\delta}(\sigma,\delta)=\frac{1}{(1-\alpha)}\left[-\frac{\sigma^2}{\delta^2}\ln\left(1-\frac{\delta^2}{\sigma^2}\right)\right]>\frac{1}{(1-\alpha)}\ \text{ и }\ p'_{i+1}(\delta)'_{\delta}=\frac{1}{(1-\alpha)},\ i=i_\alpha+1,\ldots,n-1.$$

Следовательно, $\bar{p}(\sigma,\delta)$ растет быстрее при увеличении δ в пределах $0<\delta\le\sigma$ при фиксированном σ, т. е. $\bar{p}_i>p'_{i+1}$ для $i=i_\alpha+1,\ldots,n-1$. ☻

Лемма 3. Справедливо соотношение $\bar{p}_{i_\alpha-1}<p'_{i_\alpha}$ при $0<\delta_\alpha\le\delta$.

Доказательство. При $\delta_\alpha=0$ $\sigma_{i_\alpha}=1-\alpha$ следовательно $p'_{i_\alpha}=\delta_\alpha/(1-\alpha)=0$,

$$\bar{p}_{i_\alpha-1}=\frac{\sigma_{i_\alpha-1}}{\delta(1-\alpha)}\left(\delta_\alpha+\sigma_{i_\alpha}\ln\left(\frac{\sigma_{i_\alpha}}{1-\alpha}\right)\right)=\frac{\sigma_{i_\alpha-1}}{\delta(1-\alpha)}\left(0+\sigma_{i_\alpha}\ln\left(\frac{\sigma_{i_\alpha}}{\sigma_{i_\alpha}}\right)\right)=0.$$

Далее рассмотрим рост p'_{i_α} и $\bar{p}_{i_\alpha-1}$ при увеличении δ_α, оценив производные $\left((1-\alpha)p'_{i_\alpha}(\delta_\alpha)\right)'_{\delta_\alpha}=1$,

$$\left((1-\alpha)\bar{p}_{i_\alpha-1}(\delta_\alpha)\right)'_{\delta_\alpha}=\frac{\sigma_{i_\alpha-1}}{\delta}\left(\frac{\delta_\alpha}{\sigma_{i_\alpha}+\delta_\alpha}\right)=\frac{\delta_\alpha}{\delta}\left(\frac{\sigma_{i_\alpha}+\delta}{\sigma_{i_\alpha}+\delta_\alpha}\right)\le 1.$$

Следовательно, $\bar{p}_{i_\alpha-1}<p'_{i_\alpha}$ для любого δ_α при $0<\delta_\alpha\le\delta$. ☻

Лемма 4. $\bar{p}_{i_\alpha-1}+\bar{p}_{i_\alpha}>p'_{i_\alpha}$.

Доказательство. Утверждение проверяется непосредственным сложением

$$\bar{p}_{i_\alpha-1}+\bar{p}_{i_\alpha}=\frac{1}{\delta(1-\alpha)}\left(\sigma_{i_\alpha}\delta+\delta\delta_\alpha+\sigma_{i_\alpha}\sigma_{i_\alpha+1}\ln\left(\frac{\sigma_{i_\alpha+1}}{\sigma_{i_\alpha}}\right)\right)>\frac{\delta_\alpha}{(1-\alpha)}=p'_{i_\alpha}.$$ ☻

Лемма 5. Рассмотрим три последовательности: первая $x_1\le x_2\le\cdots\le x_n$, вторая $(p_i)_{i=1}^n, p_i>0,\sum_{i=1}^n p_i=1$, третья $(h_i)_{i=1}^n, h_i>0,\sum_{i=1}^n h_i=1$. Пусть для любого $k\in\{1,\ldots,n\}$ выполняется $\sum_{i=k}^n(h_i-p_i)\ge 0$. Тогда $\sum_{i=1}^n p_i x_i\le\sum_{i=1}^n h_i x_i$.

Доказательство. Утверждение является общим и известным. ☻

Чтобы получить ответ на вопрос о соотношении статистик, рассмотрим три суммы, соответствующие статистике $s_R^I(Y)$ и границам статистики $s_R^{II}(Y)$. Для упрощения записей обозначим $q_{\gamma_i}(Y)$ как x_i. x_0 будет обозначать $-\infty$. Заметим, что $x_{i_\alpha-1} \le x_{i_\alpha} \le \ldots \le x_n$.

$$S_1 = s_R^I(Y) = \sum_{i=i_\alpha}^{n} p'_i q_{\gamma_i}(Y) = \sum_{i=i_\alpha}^{n} p'_i x_i ,$$

$$S_2 = \sum_{i=i_\alpha-1}^{n-1} \bar{p}_i q_{\gamma_i}(Y) = \sum_{i=i_\alpha}^{n} \bar{p}_{i-1} x_{i-1} \text{ и } S_3 = \sum_{i=i_\alpha-1}^{n-1} \bar{p}_i q_{\gamma_{i+1}}(Y) = \sum_{i=i_\alpha}^{n} \bar{p}_{i-1} x_i .$$

Используя соотношения между коэффициентами p'_i и $\bar{p}_{i-1}$ из лемм 2 и 3 и лемму 5, получаем, что $S_1 \le S_3$.

При $i_\alpha = 1$ в S_2 входит $x_0 = -\infty$, следовательно, $S_2 < S_1$.

Для $i_\alpha > 1$ введем обозначения $\Delta p_i = \bar{p}_{i-1} - p'_i$, $\Delta x_i = x_i - x_{i-1}$, тогда

$$S_2 = \sum_{i=i_\alpha}^{n} \bar{p}_{i-1} x_{i-1} = \sum_{i=i_\alpha}^{n} (p'_i + \Delta p_i)(x_i - \Delta x_i) = S_1 - \sum_{i=i_\alpha}^{n} p'_i \Delta x_i + \sum_{i=i_\alpha}^{n} \Delta p_i x_{i-1} .$$

Увеличим $x_{i_\alpha-1}$, положив $x_{i_\alpha-1} = x_{i_\alpha}$. Тогда $\Delta x_{i_\alpha} = 0$ и, учитывая, что $p'_i = \delta/(1-\alpha)$ для $i \in \{i_\alpha+1, \ldots, n\}$, получаем

$$S_2 \le S_1 - \delta/(1-\alpha)(x_n - x_{i_\alpha}) + \sum_{i=i_\alpha}^{n} \Delta p_i x_{i-1} . \tag{2}$$

Оценим сумму $\Delta p_{i_\alpha+1}$ и Δp_{i_α}, учитывая лемму 4,

$$\Delta p_{i_\alpha} + \Delta p_{i_\alpha+1} = \bar{p}_{i_\alpha-1} - p'_{i_\alpha} + \bar{p}_{i_\alpha} - p'_{i_\alpha+1} = \bar{p}_{i_\alpha-1} + \bar{p}_{i_\alpha} - p'_{i_\alpha} - \delta/(1-\alpha) > -\delta/(1-\alpha) .$$

Оценим сверху сумму в (2) с учетом того, что $x_{i_\alpha-1} = x_{i_\alpha}$, $\sum_{i=i_\alpha}^{n} \Delta p_i = 0$, и полученной оценки: $\sum_{i=i_\alpha}^{n} \Delta p_i x_{i-1} = \left(\Delta p_{i_\alpha} + \Delta p_{i_\alpha+1}\right) x_{i_\alpha} + \sum_{i=i_\alpha+2}^{n} \Delta p_i x_{i-1} \le$

$$\le \left(\Delta p_{i_\alpha} + \Delta p_{i_\alpha+1}\right) x_{i_\alpha} + x_{n-1} \sum_{i=i_\alpha+2}^{n} \Delta p_i = -\left(\Delta p_{i_\alpha} + \Delta p_{i_\alpha+1}\right)\left(x_{n-1} - x_{i_\alpha}\right) \le \delta(1-\alpha)\left(x_{n-1} - x_{i_\alpha}\right).$$

Получаем

$$S_2 \le S_1 - \delta/(1-\alpha)(x_n - x_{i_\alpha}) + \delta/(1-\alpha)(x_{n-1} - x_{i_\alpha}) = S_1 - \delta/(1-\alpha)(x_n - x_{n-1}) \le S_1 .$$

Таким образом, доказана такая теорема.

Теорема. Статистика $s_R^I(Y(\bar{c}))$ – это элемент интервала, определяемый статистикой $s_R^{II}(Y(\bar{c}))$.☻

Выводы. В случае оценивания значения суперквантиля с помощью регрессионной модели, которая включает постоянный коэффициент (intercept), наряду с суперквантильной ошибки может быть использована ошибка «Рокафеллара». Существует несколько вариантов задания коэффициетов для ошибки «Рокафеллара», при которых минимальные значения суперквантильной ошибки и ошибки «Рокафеллара» совпадают. Однако выбор коэффициентов может влиять на получаемое значение оцениваемого суперквантиля (статистики). Приведенные результаты показывают, что в одном варианте статистика определяется однозначно, а в другом – она может принадлежать интервалу, в который попадает и первая оценка.

В.М. Кузьменко, Е.І. Ненахов

ПРО ВЗАЄМОЗАМІНЮВАНІСТЬ ФУНКЦІЙ ПОМИЛКИ ПРИ РЕГРЕСІЙНОМУ АНАЛІЗІ

Розглядається можливість зведення обчислення суперквантильної похибки до обчислення похибки «Рокафеллара» для випадкової величини із дискретним рівномірним розподілом. Наводяться два варіанти зведення та досліджується питання еквівалентності використання першої похибки та запропонованих варіантів у регресійному аналізі.

V.M. Kuzmenko, E.I. Nenakhov

A REDUCTION OF SUPERQUANTILE ERROR TO ROCKAFELLAR ERROR
IN REGRESSION ANALISYS

An opportunity of reduction of superquantile error to Rockafellar error for uniform discrete distribution is considered. Two variants of reduction are shown and a question of equivalence of using superquantile error and reductions is analyzed.

1. *Rockafellar R.T., Uryasev S.* The Fundamental Risk Quadrangle in Risk Management, Optimization, and Statistical Estimation // Surveys in Operations Research and Management Science. – 2013. – **18**. – P. 33 – 53.
2. *Rockafellar R.T., Royset J.O., Miranda S.I.* Superquantile regression with applications to buffered reliability, uncertainty quantification and conditional value-at-risk // European J. Operations Research. – 2014. – **234**. – P. 140 – 154.
3. *Rockafellar R.T., Uryasev S.* Conditional value-at-risk for general loss distributions // Journal of Banking and Finance. – 2002. – **26**. – P. 1443 – 1471.

Получено 18.04.2016