ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

А.С. Шведов

НЕЧЕТКО-СЛУЧАЙНЫЕ МЕТОДЫ В ОБОЛОЧЕЧНОМ АНАЛИЗЕ ДАННЫХ

Препринт WP2/2014/05

Серия WP2
Количественный анализ в экономике

Москва 2014

Редактор серии WP2
«Количественный анализ в экономике»
В.А. Бессонов

Шведов, А. С.
Нечетко-случайные методы в оболочечном анализе данных [Электронный ресурс] : препринт WP2/2014/05 / А. С. Шведов ; Нац. исслед. ун-т «Высшая школа экономики». – Электрон. текст. дан. (240 Кб). – М. : Изд. дом Высшей школы экономики, 2014. – 28 с. – (Серия WP2 «Количественный анализ в экономике»).

В рамках нечеткого подхода предлагается новый алгоритм расчета эффективности организации, приводятся результаты расчетов. Дается описание замены вероятностного ограничения на детерминированное ограничение в задаче стохастического линейного программирования. Показана возможность сведения нечетко-случайной задачи линейного программирования оболочечного анализа данных к детерминированной нелинейной задаче математического программирования.

УДК 519.856
ББК 22.172

Классификация JEL: C67

Ключевые слова: нечетко-случайная оптимизация, анализ эффктивности

Shvedov, A. S.
Fuzzy random approach to data envelopment analysis [Electronic resource] : Working paper WP2/2014/05 / A. S. Shvedov ; National Research University Higher School of Economics. – Electronic text data (240 Kb). – Moscow : Publishing House of the Higher School of Economics, 2014. – 28 p. – (Series WP2 "Quantitative Analysis of Russian Economy"). (in Russian).

A new technique for measuring the relative efficiencies of a set of entities is proposed in a fuzzy environment. Results of computations are presented. Further, conversion of a chance constrained stochastic linear program into a deterministic is described. A fuzzy random DEA model is transformed to a crisp model.

JEL Classification: 67

Key words: fuzzy random optimization, efficiency analysis

Препринты Национального исследовательского университета «Высшая школа экономики» размещаются по адресу: http://www.hse.ru/org/hse/wp

1. Введение

Оболочечный анализ данных для оценки технической эффективности организации (data envelopment analysis, DEA) предложен в работе [4] и с тех пор приобрел широкое распространение. Основная идея – сопоставление с другими схожими организациями.

Рассматриваются организации, каждая из которых характеризуется затратами (или входными переменными) $x = (x_1, \ldots, x_p)$ и выпуском (или выходными переменными) $y = (y_1, \ldots, y_q)$. Все входные переменные $x_1, \ldots, x_p$ и выходные переменные $y_1, \ldots, y_q$ – неотрицательные действительные числа.

Техническая эффективность понимается как максимизация выпуска при минимизации затрат. И для самого определения технической эффективности, и для расчета эффективности каждой организации требуется математическая формализация задачи. При анализе технической эффективности не используются цены, соответствующие входным переменным $x_1, \ldots, x_p$ и выходным переменным $y_1, \ldots, y_q$, (если эти цены определены). На следующем этапе технически эффективные единицы сравниваются между собой уже с учетом цен.

Оболочечный анализ данных применяется в очень многих областях человеческой деятельности, и слово "организация" надо понимать здесь в самом широком смысле. Это могут быть и промышленные предприятия, и военные формирования, часто этот подход применяется для оценки эффективности образовательных и медицинских учреждений, в сельском хозяйстве, используется он и для сравнения готовящихся проектов (см., например, [8]).

Пусть сравниваются n организаций; $x_{1\,i}, \ldots, x_{p\,i}$ – затраты для i-й организации; $y_{1\,i}, \ldots, y_{q\,i}$ – выпуск для i-й организации; $i = 1, \ldots, n$. Техническую эффективность i-й организации можно определить как

$$\frac{\sum\limits_{k=1}^{q} v_k y_{k\,i}}{\sum\limits_{j=1}^{p} u_j x_{j\,i}}\,,$$

где $u_1, \ldots, u_p, v_1, \ldots, v_q$ – неотрицательные действительные числа. Остается вопрос, какими должны быть веса u_j и v_k.

Пусть r – номер организации, техническую эффективность которой нужно определить, $1 \leq r \leq n$. Идея оболочечного анализа данных состоит в решении задачи математического программирования

$$\frac{\sum\limits_{k=1}^{q} v_k y_{k\,r}}{\sum\limits_{j=1}^{p} u_j x_{j\,r}} \to \max$$

при условии (1)

$$\frac{\sum\limits_{k=1}^{q} v_k y_{k\,i}}{\sum\limits_{j=1}^{p} u_j x_{j\,i}} \leq 1, \qquad i = 1, \ldots, n,$$

относительно неизвестных $u_1, \ldots, u_p, v_1, \ldots, v_q$.

Полученное значение целевой функции и объявляется эффективностью r-й организации. Ясно, что это число всегда лежит между 0 и 1.

Идея оболочечного анализа данных, предложенная в работе [4], заключается в том, чтобы “позволить” каждой организации подобрать веса u_j и v_k наиболее выгодным для себя образом и произвести расчет эффективности с использованием именно этих весов.

Привлекательной стороной такого подхода является еще и то, что этот подход непараметрический. Никакие предположения о форме зависимости выпуска $y_1, \ldots, y_q$ от затрат $x_1, \ldots, x_p$ не используются.

Нетрудно увидеть, что эта задача математического программирования может быть заменена на задачу линейного программирования

$$\sum_{k=1}^{q} v_k y_{k\,r} \to \max$$

при условиях (2)

$$\sum_{j=1}^{p} u_j x_{j\,r} = 1,$$

$$\sum_{j=1}^{p} u_j x_{j\,i} \geq \sum_{k=1}^{q} v_k y_{k\,i}, \qquad i = 1, \ldots, n.$$

Эти условия можно также записать в виде

$$\sum_{j=1}^{p} u_j x_{j\,r} \leq 1, \qquad -\sum_{j=1}^{p} u_j x_{j\,r} \leq -1,$$

$$-\sum_{j=1}^{p} u_j x_{j\,i} + \sum_{k=1}^{q} v_k y_{k\,i} \leq 0, \qquad i = 1, \ldots, n.$$

Тогда двойственная задача линейного программирования имеет вид

$$\theta_1 - \theta_2 \to \min$$

при условиях

$$x_{j\,r}\theta_1 - x_{j\,r}\theta_2 - \sum_{i=1}^{n} x_{j\,i}\lambda_i \geq 0, \qquad j = 1, \ldots, p,$$

$$\sum_{i=1}^{n} y_{k\,i}\lambda_i \geq y_{k\,r}, \qquad k = 1, \ldots, q,$$

где $\theta_1, \theta_2, \lambda_1, \ldots, \lambda_n$ – неотрицательные действительные числа.

Замечание, что θ_1 и θ_2 входят в уравнения только в виде разности $\theta_1 - \theta_2$, дает окончательную задачу линейного программирования

$$\theta \to \min$$

при условиях (3)

$$x_{j\,r}\theta \geq \sum_{i=1}^{n} x_{j\,i}\lambda_i, \qquad j = 1, \ldots, p,$$

$$\sum_{i=1}^{n} y_{k\,i}\lambda_i \geq y_{k\,r}, \qquad k = 1, \ldots, q.$$

Отрицательные значения θ можно исключить из рассмотрения, при них не могут выполняться первые p условий. (Если не рассматривать возможность $x_{1\,r} = \ldots = x_{p\,r} = 0$, что неявно было сделано при постановке исходной задачи математического программирования.) С другой стороны, $\theta = 1$, $\lambda_r = 1$, $\lambda_i = 0$ при $i \neq r$ – допустимое решение задачи линейного программирования. Поэтому для технической эффективности должны выполняться ограничения $0 \leq \theta \leq 1$. Чем ближе к 1 значение целевой функции θ, тем выше техническая эффективность организации.

Задачи математического программирования (1), (2), (3) являются лишь базовой, исходной моделью оболочечного анализа данных. В различных приложениях используются многочисленные модификации и усовершенствования этой базовой модели.

Оболочечный анализ данных тесно связан с анализом технической эффективности на основе понятия множества производственных возможностей, а также с предшествующей работой [9]. Подробнее об этом см., например, [3].

Большая часть публикаций по оболочечному анализу данных относится к случаю, когда значения входных и выходных переменных не определены точно. Это может быть связано и с отсутствием достоверной информации, и с неточностью измерений, и с самой природой информации. (Иногда в качестве причины указывают высокую чувствительность результатов оболочечного анализа данных к небольшим изменениям в данных.) Одной из основополагающих в этой области является работа [7].

Применение методов нечеткой математики для передачи неопределенности входных и выходных переменных в оболочечном анализе данных начато в работах [16], [11], [10]. Применение аппарата теории вероятностей дается, например, в работах [13], [6], [5].

Интересным подходом к передаче неопределенности является соединение нечетких методов и вероятностных методов. Для этого используется понятие нечетко-случайной величины. Применение такого математического аппарата в оболочечном анализе данных впервые производится, по-видимому, в работе [15]. Из более поздних работ назовем [14].

В работе [1] предлагается новое определение нечетко-случайных величин, обсуждается связь этого определения с другими существующими подходами. В настоящей работе это определение используется при оболочечном анализе данных.

В параграфе 2 настоящей работы предлагается новый алгоритм расчета технической эффективности организации, когда для передачи неопределенности используются нечеткие числа (а не нечетко-случайные величины). Приводятся результаты расчетов. В основе этого алгоритма лежат те же идеи, что и в работе [11], но применяются эти идеи к разным задачам математического программирования. В работе [11] берется задача математического программирования сходная с (1), в настоящей работе – задача линейного программирования (3). В параграфе 3 дается описание замены вероятностного ограничения в задаче стохастического линейного программирования на детерминированное ограничение, несколько обобщается подход из работ [12] и [2]. В параграфе 4 соединяются методы, изложенные в параграфах 2 и 3, показана возможность сведения нечетко-случайной задачи линейного программирования оболочечного анализа данных к детерминированной нелинейной задаче математического программирования.

2. Оболочечный анализ данных при нечетких затратах и выпуске

В этом параграфе рассматривается задача линейного программирования (3), когда $x_{j\,i}$, $y_{k\,i}$ – нечеткие числа. Существуют различные определения нечетких чисел, здесь используется то же определение, что и в [1]. Хотя отношение $\geq$ и может быть перенесено на нечеткие числа, в настоящей работе используется другой подход.

Обозначим через $X_{j\,i}(\eta)$ η-срез нечеткого числа $x_{j\,i}$, через $Y_{k\,i}(\eta)$ – η-срез нечеткого числа $y_{k\,i}$; $i = 1, \ldots, n$, $j = 1, \ldots, p$, $k = 1, \ldots, q$, $0 \leq \eta \leq 1$.

Левый и правый индексы нечеткого числа $x_{j\,i}$ обозначаются, соответственно, $X^L_{j\,i}(\eta)$ и $X^U_{j\,i}(\eta)$. Левый и правый индексы нечеткого числа $y_{k\,i}$ обозначаются, соответственно, $Y^L_{k\,i}(\eta)$ и $Y^U_{k\,i}(\eta)$. Таким образом,

$$X_{j\,i}(\eta) = [X^L_{j\,i}(\eta), X^U_{j\,i}(\eta)], \qquad Y_{k\,i}(\eta) = [Y^L_{k\,i}(\eta), Y^U_{k\,i}(\eta)].$$

Напомним, что при $\eta' < \eta''$ имеют место включения

$$X_{j\,i}(\eta'') \subseteq X_{j\,i}(\eta'), \qquad Y_{k\,i}(\eta'') \subseteq Y_{k\,i}(\eta').$$

Задачу линейного программирования (3) можно рассматривать как построение функции θ от $n(p+q)$ действительных аргументов $x_{j\,i}$, $y_{k\,i}$. Обозначим данную функцию

$$\theta(x_{j\,i}, y_{k\,i}; i = 1, \ldots, n; j = 1, \ldots, p\,; k = 1, \ldots, q)$$

или, сокращенно,

$$\theta(x_{j\,i}, y_{k\,i}).$$

Если аргументы $x_{j\,i}$, $y_{k\,i}$ – нечеткие числа, то и значение функции θ – нечеткое число. В соответствии с определением функции от нечетких аргументов получаем выражения для индексов нечеткого числа θ (в оставшихся формулах этого параграфа, до примера, $x_{j\,i}$, $y_{k\,i}$ – действительные числа)

$$\theta_1(\eta) = \inf_{\substack{x_{j\,i} \in X_{j\,i}(\eta),\, y_{k\,i} \in Y_{k\,i}(\eta) \\ i=1,\ldots,n;\, j=1,\ldots,p\,;\, k=1,\ldots,q}} \theta(x_{j\,i}, y_{k\,i}),$$

$$\theta_2(\eta) = \sup_{\substack{x_{j\,i} \in X_{j\,i}(\eta),\, y_{k\,i} \in Y_{k\,i}(\eta) \\ i=1,\ldots,n;\, j=1,\ldots,p\,;\, k=1,\ldots,q}} \theta(x_{j\,i}, y_{k\,i}).$$

Задачу линейного программирования (3) можно записать в виде

$$\theta \to \min$$

при условиях

$$x_{j\,r}(\theta - \lambda_r) \geq \sum_{\substack{i=1 \\ i \neq r}}^{n} x_{j\,i}\lambda_i, \qquad j = 1, \ldots, p,$$

$$\sum_{\substack{i=1 \\ i \neq r}}^{n} y_{k\,i}\lambda_i \geq y_{k\,r}(1 - \lambda_r), \qquad k = 1, \ldots, q.$$

Для простоты исключим возможность нулевых значений $x_{j\,r}$ при любом j. Тогда задачу линейного программирования можно решать в области $\theta \geq \lambda_r$.

Чем шире область изменения аргументов $\theta, \lambda_1, \ldots, \lambda_n$, тем меньшее значение целевой функции может быть достигнуто.

При увеличении коэффициентов $x_{j\,r}$, $y_{k\,i}$ в левых частях неравенств область изменения аргументов расширяется. При уменьшении коэффициентов $x_{j\,i}$, $y_{k\,r}$ в правых частях неравенств область изменения аргументов также расширяется.

Следовательно, для нахождения $\theta_1(\eta)$ надо решать задачу линейного программирования

$$\theta \to \min$$

при условиях (4)

$$X^U_{j\,r}(\eta)(\theta - \lambda_r) \geq \sum_{\substack{i=1 \\ i \neq r}}^{n} X^L_{j\,i}(\eta)\lambda_i, \qquad j = 1, \ldots, p,$$

$$\sum_{\substack{i=1 \\ i \neq r}}^{n} Y^U_{k\,i}(\eta)\lambda_i \geq Y^L_{k\,r}(\eta)(1 - \lambda_r), \qquad k = 1, \ldots, q.$$

Для нахождения $\theta_2(\eta)$ надо решать задачу линейного программирования

$$\theta \to \min$$

при условиях (5)

$$X^L_{j\,r}(\eta)(\theta - \lambda_r) \geq \sum_{\substack{i=1\\ i\neq r}}^{n} X^U_{j\,i}(\eta)\lambda_i, \qquad j = 1,\ldots,p,$$

$$\sum_{\substack{i=1\\ i\neq r}}^{n} Y^L_{k\,i}(\eta)\lambda_i \geq Y^U_{k\,r}(\eta)(1-\lambda_r), \qquad k = 1,\ldots,q.$$

В следующем примере приводятся результаты расчетов, полученные путем решения задач линейного программирования (4) и (5).

Пример. Рассмотрим те же данные, что и в работе [11]. Сравниваются 4 организации, $p = 1$, $q = 1$. В таблице 1 приведены данные, показывающие затраты и выпуск для каждой из 4 организаций.

i	$x_{1\,i}$	$y_{1\,i}$
1	(11 , 12 , 12 , 14)	(10 , 10 , 10 , 10)
2	(30 , 30 , 30 , 30)	(12 , 13 , 14 , 16)
3	(40 , 40 , 40 , 40)	(11 , 11 , 11 , 11)
4	(45 , 47 , 52 , 55)	(12 , 15 , 19 , 22)

Табл. 1. Затраты и выпуск для 4 организаций для расчетов №1 и №2.

Нечеткое число представлено с помощью 4 действительных чисел. Нечеткое число $x_{j\,i}$ представляется при помощи чисел $X^L_{j\,i}(0)$, $X^L_{j\,i}(1)$, $X^U_{j\,i}(1)$, $X^U_{j\,i}(0)$. Нечеткое число $y_{k\,i}$ представляется при помощи чисел $Y^L_{k\,i}(0)$, $Y^L_{k\,i}(1)$, $Y^U_{k\,i}(1)$, $Y^U_{k\,i}(0)$.

η	$\theta_1(\eta)$	$\theta_2(\eta)$
0.0	1.0000	1.0000
0.1	1.0000	1.0000
0.2	1.0000	1.0000
0.3	1.0000	1.0000
0.4	1.0000	1.0000
0.5	1.0000	1.0000
0.6	1.0000	1.0000
0.7	1.0000	1.0000
0.8	1.0000	1.0000
0.9	1.0000	1.0000
1.0	1.0000	1.0000

Табл. 2. Эффективность θ первой организации в расчете №1; $\theta_1(\eta)$ и $\theta_2(\eta)$ – индексы нечеткого числа θ.

η	$\theta_1(\eta)$	$\theta_2(\eta)$
0.0	0.4400	0.7467
0.1	0.4477	0.7268
0.2	0.4555	0.7072
0.3	0.4633	0.6879
0.4	0.4712	0.6688
0.5	0.4792	0.6500
0.6	0.4872	0.6315
0.7	0.4953	0.6132
0.8	0.5035	0.5952
0.9	0.5117	0.5775
1.0	0.5200	0.5600

Табл. 3. Эффективность θ второй организации в расчете №1; $\theta_1(\eta)$ и $\theta_2(\eta)$ – индексы нечеткого числа θ.

η	$\theta_1(\eta)$	$\theta_2(\eta)$
0.0	0.3025	0.3850
0.1	0.3053	0.3795
0.2	0.3080	0.3740
0.3	0.3107	0.3685
0.4	0.3135	0.3630
0.5	0.3163	0.3575
0.6	0.3190	0.3520
0.7	0.3218	0.3465
0.8	0.3245	0.3410
0.9	0.3273	0.3355
1.0	0.3300	0.3300

Табл. 4. Эффективность θ третьей организации в расчете №1; $\theta_1(\eta)$ и $\theta_2(\eta)$ – индексы нечеткого числа θ.

η	$\theta_1(\eta)$	$\theta_2(\eta)$
0.0	0.2400	0.6844
0.1	0.2496	0.6625
0.2	0.2594	0.6411
0.3	0.2694	0.6200
0.4	0.2797	0.5995
0.5	0.2902	0.5793
0.6	0.3009	0.5597
0.7	0.3119	0.5404
0.8	0.3230	0.5215
0.9	0.3345	0.5031
1.0	0.3462	0.4851

Табл. 5. Эффективность θ четвертой организации в расчете №1; $\theta_1(\eta)$ и $\theta_2(\eta)$ – индексы нечеткого числа θ.

Все нечеткие числа считаются трапецоидальными, то есть функции $X_{j\,i}^L(\eta)$, $X_{j\,i}^U(\eta)$, $Y_{k\,i}^L(\eta)$, $Y_{k\,i}^U(\eta)$ – линейные функции аргумента η.

Из данных видно, что первая организация является явным лидером. Третья организация однозначно проигрывает второй организации.

Расчет №1 проведен с данными из таблицы 1. Результаты расчета представлены в таблицах 2 – 5.

Для первой организации эффективность оказалась равной 1 и в приводимом расчете, и в расчете из работы [11]. Для остальных трех организаций результаты приводимых расчетов весьма сильно отличаются от результатов расчетов из работы [11]. Это может быть объяснено тем, что в работе [11] в целевой функции в задаче (1) число использумых весов $p+q+1$, а не $p+q$, как в настоящей работе.

η	$\theta_1(\eta)$	$\theta_2(\eta)$
0.0	0.8182	1.0000
0.1	0.8401	1.0000
0.2	0.8627	1.0000
0.3	0.8861	1.0000
0.4	0.9101	1.0000
0.5	0.9350	1.0000
0.6	0.9606	1.0000
0.7	0.9871	1.0000
0.8	1.0000	1.0000
0.9	1.0000	1.0000
1.0	1.0000	1.0000

Табл. 6. Эффективность θ второй организации в расчете №2; $\theta_1(\eta)$ и $\theta_2(\eta)$ – индексы нечеткого числа θ.

η	$\theta_1(\eta)$	$\theta_2(\eta)$
0.0	0.5156	0.6875
0.1	0.5222	0.6818
0.2	0.5288	0.6762
0.3	0.5357	0.6707
0.4	0.5428	0.6653
0.5	0.5500	0.6600
0.6	0.5574	0.6548
0.7	0.5651	0.6496
0.8	0.5729	0.6445
0.9	0.5810	0.6395
1.0	0.5893	0.6346

Табл. 7. Эффективность θ третьей организации в расчете №2; $\theta_1(\eta)$ и $\theta_2(\eta)$ – индексы нечеткого числа θ.

η	$\theta_1(\eta)$	$\theta_2(\eta)$
0.0	0.4091	1.0000
0.1	0.4270	1.0000
0.2	0.4454	1.0000
0.3	0.4645	1.0000
0.4	0.4842	1.0000
0.5	0.5047	1.0000
0.6	0.5258	1.0000
0.7	0.5477	1.0000
0.8	0.5703	0.9858
0.9	0.5938	0.9591
1.0	0.6181	0.9329

Табл. 8. Эффективность θ четвертой организации в расчете №2; $\theta_1(\eta)$ и $\theta_2(\eta)$ – индексы нечеткого числа θ.

Расчет №2 также проведен с данными из таблицы 1, но из анализа исключена первая организация. Результаты расчета представлены в таблицах 6 – 8. Естественно, эффективности второй, третьей и четвертой организаций увеличились по сравнению с расчетом №1. Интересно отметить эффект, обнаруженный в работе [11], что эффективности не являются трапецоидальными нечеткими числами, хотя для всех организаций и затраты, и выпуск – трапецоидальные нечеткие числа.

Для сравнения сделаны еще два расчета, данные для которых представлены в таблице 9. Они близки к данным из таблицы 1, но "убрана" нечеткость. Левые и правые индексы у всех нечетких чисел совпадают, операции над такими нечеткими числами должны сводиться к операциям над обычными действительными числами.

i	x_{1i}	y_{1i}
1	(12 , 12 , 12 , 12)	(10 , 10 , 10 , 10)
2	(30 , 30 , 30 , 30)	(14 , 14 , 14 , 14)
3	(40 , 40 , 40 , 40)	(11 , 11 , 11 , 11)
4	(50 , 50 , 50 , 50)	(17 , 17 , 17 , 17)

Табл. 9. Затраты и выпуск для 4 организаций для расчетов №3 и №4.

Расчет №3 проведен с данными из таблицы 9. Эффективности первой, второй, третьей и четвертой организации оказались равными, соответственно, 1, 0.56, 0.33, 0.408.

Расчет №4 также проведен с данными из таблицы 9, но из анализа исключена первая организация. Эффективности второй, третьей и четвертой организации оказались равными, соответственно, 1, 0.5893, 0.7286.

Отметим, что результаты расчета №3 достаточно хорошо согласуются с результатами расчета №1, а результаты расчета №4 – с результатами расчета №2.

3. Замена вероятностного ограничения в задаче стохастического линейного программирования на детерминированное ограничение

Рассматриваемый метод замены вероятностного ограничения в задаче стохастического линейного программирования на детерминированное ограничение предложен в работе [12]. Для несколько более общего случая этот метод изложен в работе [2].

В линейном программировании через x_i обычно обозначаются переменные, подлежащие определению. В оболочечном анализе данных $x_{j\,i}$ – это затраты, которые в задаче линейного программирования являются коэффициентами. В обоих случаях нами сохраняются эти обычные обозначения. Поэтому x_i и $x_{j\,i}$ никак между собой не связаны.

Вероятностное ограничение, о котором идет речь, имеет вид

$$P\left(\sum_{i=1}^{m}(a_i+\zeta_i)x_i \geq b\right) \geq 1-\varepsilon,$$

где a_i, b, x_i – действительные числа, ζ_i – случайные величины, ε – фиксированное действительное число, $0<\varepsilon<\dfrac{1}{2}$.

И в работе [12], и в работе [2] рассматриваются лишь случайные величины $\zeta_1,\ldots,\zeta_m$, имеющие совместное непрерывное распределение, и функция плотности φ этого совместного распределения предполагается симметричной, то есть

$$\varphi(\zeta)=\varphi(-\zeta) \quad \text{при всех } \zeta\in\mathbb{R}^m.$$

Больше того, в работе [12] функция φ считается логарифмически вогнутой. В [2] показано, что требовать логарифмическую вогнутость функции φ не нужно. Важно лишь некоторое свойство плавающего множества, связанного с функцией плотности. И это свойство имеет место, когда функция плотности логарифмически вогнутая. Но имеет место и для более широкого класса функций плотности, подробнее об этом см. в [2].

В этом параграфе метод замены вероятностного ограничения на детерминированное ограничение приводится для распределений случайного вектора $\zeta = (\zeta_1, \ldots, \zeta_m)$ более общего вида, существование функции плотности не требуется. Схема рассуждений близка к схеме рассуждений из работы [2]. Для ясности изложение включает в себя небольшие повторы.

Во-первых, потребуем, чтобы для любого замкнутого полупространства Q пространства $\mathbb{R}^m$, не содержащего 0, выполнялось условие $P(\zeta \in Q) \leq 0.5$. (Можно было бы потребовать несколько меньшего, а именно, $P(\zeta \in Q) < 1 - \varepsilon$. В данном параграфе ε изначально фиксировано. Но обычно при построении математической модели распределение вероятностей случайного вектора ζ не связывается с ε. Поэтому для наглядности накладывается условие $P(\zeta \in Q) \leq 0.5$.)

Пусть S^{m-1} – единичная сфера пространства $\mathbb{R}^m$. Скалярное произведение в пространстве $\mathbb{R}^m$ обозначим $\langle \cdot, \cdot \rangle$.

При каждом $u \in S^{m-1}$ рассмотрим замкнутые полупространства Q пространства $\mathbb{R}^m$, обладающие следующими свойствами.

1. $P(\zeta \in Q) \geq 1 - \varepsilon$. (Отсюда вытекает, что $0 \in Q$.)

2. Граница полупространства Q ортогональна вектору u, и $\langle v, u \rangle > 0$ для любого $v \in \mathbb{R}^m \setminus Q$.

Наименьшее из полупространств Q, обладающих свойствами 1 и 2, обозначим L_u. (Существование наименьшего полупространства следует из теоремы о непрерывности вероятности.) Границу полупространства L_u обозначим через ∂L_u.

Плавающее множество F_ε случайного вектора ζ определяется следующим образом:

$$F_\varepsilon = \bigcap_{u \in S^{m-1}} L_u.$$

Очевидно, что F_ε – замкнутое выпуклое множество, $0 \in F_\varepsilon$.

Условие, которому должно удовлетворять распределение случайного вектора ζ, и которое предполагается выполненным, такое. При любом $u \in S^{m-1}$ гиперплоскость ∂L_u является опорной для множества F_ε, то есть имеет общую точку с этим множеством. В работе [2] дается пример распределения вероятностей, для которого это условие не выполняется.

Рассматриваемое вероятностное ограничение на вектора $x = (x_1, \ldots, x_m)$ может быть записано в следующем виде

$$P(\zeta \in Q(x)) \geq 1 - \varepsilon,$$

где

$$Q(x) = \left\{ \zeta \in \mathbb{R}^m : \sum_{i=1}^{m} (a_i + \zeta_i) x_i \geq b \right\}.$$

При $x = 0$ либо $Q(x) = \mathbb{R}^m$, либо $Q(x) = \emptyset$ в зависимости от знака b. При $x \neq 0$ множество $Q(x)$ – это полупространство пространства $\mathbb{R}^m$.

Лемма. Пусть $x \in \mathbb{R}^m$. Тогда

$$P(\zeta \in Q(x)) \geq 1 - \varepsilon$$

в том и только том случае, когда

$$F_\varepsilon \subseteq Q(x).$$

Доказательство. Утверждение леммы очевидно при $x = 0$. В этом случае либо $Q(x) = \mathbb{R}^m$ и $P(\zeta \in Q(x)) = 1$, либо $Q(x) = \emptyset$ и $P(\zeta \in Q(x)) = 0$.

Пусть $x \neq 0$. Если $0 \notin Q(x)$, то F_ε не является подмножеством полупространства $Q(x)$, и $P(\zeta \in Q(x)) \leq 0.5$, и утверждение леммы выполняется. Пусть $0 \in Q(x)$. Рассмотрим вектор $u \in S^{m-1}$, обладающий тем свойством, что он ортогонален границе полупространства $Q(x)$ и $\langle v, u \rangle > 0$ для любого $v \in \mathbb{R}^m \setminus Q(x)$.

Если $P(\zeta \in Q(x)) \geq 1 - \varepsilon$, то $L_u \subseteq Q(x)$, и, следовательно, $F_\varepsilon \subseteq Q(x)$.

Если $F_\varepsilon \subseteq Q(x)$, то можно воспользоваться тем, что гиперплоскость ∂L_u является опорной для множества F_ε. Это означает, что $L_u \subseteq Q(x)$. Но тогда

$$P(\zeta \in Q(x)) \geq 1 - \varepsilon.$$

Лемма доказана.

Из леммы (см. [12], [2]) легко следует, что множество точек $x \in \mathbb{R}^m$, удовлетворяющих рассматриваемому вероятностному ограничению, это выпуклое множество. И вероятностное ограничение эквивалентно детерминированному ограничению

$$\langle a, x \rangle + \inf_{z \in F_\varepsilon} \langle z, x \rangle \geq b,$$

где $a = (a_1, \ldots, a_m)$.

Изложенный метод замены вероятностного ограничения на детерминированное ограничение переносится и на вероятностные ограничения вида

$$P\left(\sum_{i=1}^{m} (a_i + \zeta_i) x_i \geq b + \zeta_{m+1}\right) \geq 1 - \varepsilon.$$

Достаточно заметить, что последнее ограничение может быть заменено на два ограничения

$$P\left(\sum_{i=1}^{m+1}(a_i+\zeta_i)x_i \ge b\right) \ge 1-\varepsilon,$$

$$x_{m+1}=-1$$

при $a_{m+1}=0$. Если плавающее множество случайного вектора $\zeta=(\zeta_1,\ldots,\zeta_{m+1})$ обладает тем свойством, что любая из определяющих гиперплоскостей является для этого множества опорной, то первое из приведенных ограничений (вероятностное) может быть заменено на эквивалентное детерминированное ограничение.

4. Оболочечный анализ данных при нечетко-случайных затратах и выпуске

В параграфе 2 при нечетких затратах и выпуске эффективность организации определяется как нечеткое число. Для нахождения этого нечеткого числа решаются задачи линейного программирования (4) и (5).

В этом параграфе эффективность организации также считается нечетким числом, и для нахождения этого нечеткого числа решаются задачи сходные с задачами (4) и (5). Но в данном параграфе $X_{j\,i}^{L}(\eta)$, $X_{j\,i}^{U}(\eta)$, $Y_{k\,i}^{L}(\eta)$, $Y_{k\,i}^{U}(\eta)$ являются не числами, а случайными величинами. Соответственно, ограничения в задачах (4) и (5) заменяются на вероятностные ограничения. Ниже соответствующие изменения рассматриваются для задачи (4). Для задачи (5) изменения проводятся аналогично.

Как и в параграфе 2, рассчитывается эффективность организации с номером r.

Из p ограничений, которые входят в первую группу в задаче (4), рассмотрим j-е ограничение. Будем считать, что случайные величины $\zeta_{j\,1}, \ldots, \zeta_{j\,n}$ независимы, случайная величина $\zeta_{j\,i}$ равномерно распределена на отрезке $[-\sigma_{j\,i}^{input}, \sigma_{j\,i}^{input}]$. Пусть

$$X_{j\,i}^{L}(\eta) = -a_{j\,i}(\eta) - \zeta_{j\,i} \qquad \text{при} \quad i \neq r,$$

$$X_{j\,r}^{U}(\eta) = -a_{j\,r}(\eta) - \zeta_{j\,r},$$

$a_{j\,i}(\eta)$ – действительные числа, $i = 1, \ldots, n$. Тем самым делается предположение о равномерном распределении случайных величин $X_{j\,i}^{L}(\eta)$ при $i \neq r$ и $X_{j\,r}^{U}(\eta)$.

Тогда j-е ограничение из первой группы записывается в виде

$$\sum_{i=1}^{n}(a_{j\,i}(\eta) + \zeta_{j\,i})\lambda_i - (a_{j\,r}(\eta) + \zeta_{j\,r})\theta \geq 0.$$

Это ограничение заменяется на вероятностное ограничение

$$P\left(\sum_{i=1}^{n}(a_{j\,i}(\eta) + \zeta_{j\,i})\lambda_i - (a_{j\,r}(\eta) + \zeta_{j\,r})\theta \geq 0\right) \geq 1 - \varepsilon,$$

где ε, как и в предыдущем параграфе, действительное число, удовлетворяющее условию $0 < \varepsilon < \dfrac{1}{2}$. Если ввести обозначения

$$x_i = \lambda_i \quad \text{при} \quad i \neq r, \quad x_r = \lambda_r - \theta,$$

то вероятностное ограничение записывается в виде

$$P\left(\sum_{i=1}^{n}(a_{j\,i}(\eta) + \zeta_{j\,i})x_i \geq 0\right) \geq 1 - \varepsilon.$$

Совместное распределение случайных величин $\zeta_{j\,1}, \ldots, \zeta_{j\,n}$ – это равномерное распределение на параллелепипеде

$$\prod_{i=1}^{n}[-\sigma_{j\,i}^{input}, \sigma_{j\,i}^{input}].$$

Функция плотности этого совместного распределения является логарифмически вогнутой. Поэтому вероятностное ограничение эквивалентно детерминированному ограничению

$$\langle a_j(\eta), x\rangle + \inf_{z\in F_{j,\varepsilon}^{input}} \langle z, x\rangle \geq 0,$$

где $F_{j,\varepsilon}^{input}$ – плавающее множество для случайного вектора $(\zeta_{j\,1},\ldots,\zeta_{j\,n})$ (см. параграф 3), через $a_j(\eta)$ обозначен n-мерный вектор с координатами $a_{j\,1}(\eta),\ldots,a_{j\,n}(\eta)$. В развернутом виде это детерминированное ограничение имеет вид

$$\sum_{\substack{i=1\\ i\neq r}}^{n} a_{j\,i}(\eta)\lambda_i + a_{j\,r}(\eta)(\lambda_r-\theta)+ \tag{6}$$

$$+\inf_{z\in F_{j,\varepsilon}^{input}} \left\{ \sum_{\substack{i=1\\ i\neq r}}^{n} z_i\lambda_i + z_r(\lambda_r-\theta) \right\} \geq 0.$$

Из q ограничений, которые входят во вторую группу в задаче (4), рассмотрим k-е ограничение. Будем считать, что случайные величины $\xi_{k\,1},\ldots,\xi_{k\,n}$ независимы, случайная величина $\xi_{k\,i}$ равномерно распределена на отрезке $[-\sigma_{k\,i}^{output}, \sigma_{k\,i}^{output}]$. Пусть

$$Y_{k\,i}^{U}(\eta) = c_{k\,i}(\eta) + \xi_{k\,i} \qquad \text{при} \quad i\neq r,$$

$$Y_{k\,r}^{L}(\eta) = c_{k\,r}(\eta) + \xi_{k\,r},$$

$c_{k\,i}(\eta)$ – действительные числа, $i=1,\ldots,n$. Тем самым делается предположение о равномерном распределении случайных величин $Y_{k\,i}^{U}(\eta)$ при $i\neq r$ и $Y_{k\,r}^{L}(\eta)$.

Тогда k-е ограничение из второй группы записывается в виде

$$\sum_{i=1}^{n}(c_{k\,i}(\eta)+\xi_{k\,i})\lambda_i \geq c_{k\,r}(\eta)+\xi_{k\,r}.$$

Это ограничение заменяется на вероятностное ограничение

$$P\left(\sum_{i=1}^{n}(c_{k\,i}(\eta)+\xi_{k\,i})\lambda_i \geq c_{k\,r}(\eta)+\xi_{k\,r}\right) \geq 1-\varepsilon.$$

Если ввести обозначения

$$y_i=\lambda_i \quad \text{при} \quad i\neq r, \quad y_r=\lambda_r-1,$$

то вероятностное ограничение записывается в виде

$$P\left(\sum_{i=1}^{n}(c_{k\,i}(\eta)+\xi_{k\,i})y_i \geq 0\right) \geq 1-\varepsilon.$$

Совместное распределение случайных величин $\xi_{k\,1},\ldots,\xi_{k\,n}$ – это равномерное распределение на параллелепипеде

$$\prod_{i=1}^{n}[-\sigma_{k\,i}^{output},\sigma_{k\,i}^{output}].$$

Функция плотности этого совместного распределения является логарифмически вогнутой. Поэтому вероятностное ограничение эквивалентно детерминированному ограничению

$$\langle c_k(\eta),y\rangle+\inf_{z\in F_{k,\,\varepsilon}^{output}}\langle z,y\rangle \geq 0,$$

где $F_{k,\varepsilon}^{output}$ – плавающее множество для случайного вектора $(\xi_{k\,1},\ldots,\xi_{k\,n})$, через $c_k(\eta)$ обозначен n-мерный вектор с координатами $c_{k\,1}(\eta),\ldots,c_{k\,n}(\eta)$. В развернутом виде это детерминированное ограничение имеет вид

$$\sum_{\substack{i=1\\ i\neq r}}^{n} c_{k\,i}(\eta)\lambda_i + c_{k\,r}(\eta)(\lambda_r-1)+ \qquad (7)$$

$$+\inf_{z\in F_{k,\varepsilon}^{output}}\left\{\sum_{\substack{i=1\\ i\neq r}}^{n} z_i\lambda_i + z_r(\lambda_r-1)\right\}\geq 0.$$

Тогда аналогом задачи линейного программирования (4) является следующая детерминированная задача математического программирования

$$\theta \to \min$$

при условиях (6) для $j = 1, \ldots, p$; (7) для $k = 1, \ldots, q$; $\lambda_i \geq 0$, $\theta \geq 0$.

Библиографический список

[1] Шведов А.С. О нечетко-случайных величинах. Препринт WP2/2013/02. М.: НИУ ВШЭ, 2013.

[2] Шведов А.С. О выпуклости допустимого множества в задаче стохастического линейного программирования. Препринт WP2/2014/01. М.: НИУ ВШЭ, 2014.

[3] Banker R.D., Charnes A., Cooper W.W. Some models for estimating technical and scale inefficiencies in data envelopment analysis // Management Science, 30(9), 1984, 1078 – 1092.

[4] Charnes A., Cooper W.W., Rhodes E. Measuring the efficiency of decision making units // European J. of Operational Research, 2, 1978, 429 – 444.

[5] Cooper W.W., Deng H., Huang Z., Li S.X. Chance constrained programming approaches to technical efficienies and

inefficiencies in stochastic data envelopment analysis // J. of the Operational Research Society, 53, 2002, 1347 – 1356.

[6] Cooper W.W., Huang Z., Li S.X. Satisficing DEA models under chance constraints // Annals of Operations Research, 66, 1996, 279 – 295.

[7] Cooper W.W., Park K.S., Yu G. IDEA and AR-IDEA: Models for dealing with imprecise data in DEA // Management Science, 45(4), 1999, 597 – 607.

[8] Cooper W.W., Seiford L.M., Tone K. Data envelopment analysis. A comprehensive text with models, applications, references and DEA-solver software. 2nd ed. N.Y.: Springer-Verlag, 2006.

[9] Farrell M.J. The measurement of productive efficiency // J. of the Royal Statistical Society, Ser. A, 120, 1957, 253 – 290.

[10] Guo P., Tanaka H. Fuzzy DEA: a perceptual evaluation method // Fuzzy Sets and Systems, 119, 2001, 149 – 160.

[11] Kao C., Liu S.-T. Fuzzy efficiency measures in data envelopment analysis // Fuzzy Sets and Systems, 113, 2000, 427 – 437.

[12] Lagoa C.M., Li X., Sznaier M. Probabilistically constrained linear programs and risk-adjusted controller design // SIAM J. Optim., 15, 2005, 938 – 951.

[13] Olesen O.B., Petersen N.C. Chance constrained efficiency evaluation // Management Science, 41(3), 1995, 442 – 457.

[14] Qin R., Liu Y.-K. A new data envelopment analysis model with fuzzy random inputs and outputs // J. of Applied Mathematics and Computing, 33, 2010, 327 – 356.

[15] Ramezanzadeh S., Memariani M., Saati S. Data envelopment analysis with fuzzy random inputs and outputs: a chance-constrained programming approach // Iranian J. of Fuzzy Systems, 2(2), 2005, 21 – 29.

[16] Sengupta J.K. A fuzzy systems approach in data envelopment analysis // Computers Math. Applic., 24, 1992, 259 – 266.

Препринт WP2/2014/05
Серия WP2
Количественный анализ в экономике

Шведов Алексей Сергеевич

Нечетко-случайные методы
в оболочечном анализе данных