УДК 332.012
DOI: 10.33764/2618-981X-2019-3-1-71-77

НОВЫЕ СРЕДСТВА ИЗМЕРЕНИЯ МЕЖРЕГИОНАЛЬНЫХ РАЗЛИЧИЙ

Юрий Петрович Воронов
Институт экономики и организации промышленного производства СО РАН, 630090, Россия, г. Новосибирск, пр. Академика Лаврентьева, 17, кандидат экономических наук, ведущий научный сотрудник, тел. (913)985-19-97, e-mail: corpus-cons@ngs.ru

Матвей Александрович Свиридов
Новосибирский национальный исследовательский государственный университет, 630090, Россия, г. Новосибирск, ул. Пирогова, 2, студент, тел. (923)184-36-99, e-mail: m.sviridov@g.nsu.ru

В статье предложены новые способы измерения различий между показателями экономического развития регионов РФ. В настоящее время для сравнения используются показатели средние, медианные и модальные. При этом выбор из трех вариантов проводится не по формальным критериям, а на основе общих рассуждений. Авторы предлагают строить систему сравнений на базе дивергенций Брегмана. Они уже начинают применяться при сравнении регионов биологами. В экономическом анализе эти меры до сих пор не применялись.

Из всех вариантов дивергенций Брегмана исследуется, преимущественно, дивергенция Кулбака-Лейблера (KL-дивергенция). В качестве критерия качества измерения различий предложены попарные сравнения уровня иммиграции в регионы. Основные расчеты сделаны на примере и уровня заработной платы в регионах юга Западной Сибири.

Сделан вывод о перспективности использования существующих вариантов дивергенции Брегмана при измерении межрегиональных различий, а также о возможности разработки новых вариантов дивергенции, максимально соответствующих сравнительной привлекательности регионов для иммигрантов.

Ключевые слова: межрегиональные различия, асимметрия мер, миграция, дивергенция Брегмана, KL-дивергенция.

NEW INTERREGIONAL DIFFERENCES MEASURE MENTTOOLS

Yuri P. Voronov
Institute for Economics and Industrial Engineering SB RAS, 17, Prospect Akademik Lavrentiev St., Novosibirsk, 630090, Russia, Ph. D., Leading Researcher, phone: (913)985-19-97, e-mail: corpus-cons@ngs.ru

Matvej A. Sviridov
Novosibirsk National Research State University, 2, Pirogova St., Novosibirsk, 630073, Russia, Student, phone: (923)184-36-99, e-mail: m.sviridov@g.nsu.ru

The article proposes new methods of measurement of differences between of RF regions economic development level. Currently, indicators are used for comparison such as mean, median and mode. And the choice of three options is carried out not on formal criteria, but on the basis of general reasoning. The authors propose to build a system of comparisons on the basis of Bregman divergences. They are already beginning to be used when comparing regions by biologists. These measures have not yet been applied in economic analysis. The divergence of Kulback-Leibler (KL-divergence) is investigated, primarily, the variant of the Bregman divergence. Pairwise comparisons of the level of immigration to the regions as a criterion of quality of measurement of differences are proposed. The main calculations are made on the example of the level of wages in the regions of the South of Western Siberia.

The conclusion is made about the prospects of using the existing variants of Bregman divergence, as well as about the possibility of developing new divergence options that best match the comparative attractiveness of the regions for immigrants.

Key words: interregional differences, measures asymmetry, migration, Bregman divergence, KL-divergence.

Введение

Измерение различий в уровне экономического развития регионов имеет определенные недостатки. Сравнение обычно идет через разницу между средними, медианными или модальными показателями.

Когда распределение сильно скошено, средняя существенно отличается от медианы. Чем более симметрично распределение, тем ближе медиана и среднее значение.В исследованиях межрегиональных и межстрановых различий за рубежом чаще используется медиана, в российских исследованиях – средняя.

Рассмотрим на простом примере. Пусть зарплаты сотрудников лаборатории составляют 3, 5,5, 7, 11, 12, 16, 16, 21, 42 и 58 тыс. руб. соответственно. Тогда среднее значение этих зарплат – 17,8, а медиана –12. Различие медианы и средней обусловлено скошенностью распределения [1]. Медиана и среднее совпадают только в случае распределения без асимметрии, однако в современных реалиях очевидно, что распределение зарплат в России далеко от симметричного. Использование средних показателей зарплат завышает фактический уровень зарплат. При измерении различий в уровне заработной платычерез разность средних не учитывается скошенность распределений зарплат.

Если в городе А средняя зарплата 30000 руб., а в городе В – 50000 руб., то при этом методе сравнения различие между средними показателями – 20000. Медианные значения будут соответственно 22000 руб и 35000 руб., разница в этом случае составит 13000 руб.

Целью настоящего исследования является разработка новых средств сравнения показателей, отражающих уровень экономического развития разных регионов. Отсутствие таких средств является недостатком исследований в региональной экономике. В ней пока отсутствуют публикации по новым методам измерения межрегиональных различий, ближайшие по данной теме исследования ведутся биологами и экологами.

Методы и материалы

В данном исследовании используется аппарат информационных мер различий, называемых дивергенциями, в частности, дивергенций Брегмана. Эти меры уже используются в межрегиональных и пространственныхисследованиях в биологии и экологии [2].

Введем содержательное определение дивергенции Брегмана. Пусть задана непрерывно-дифференцируемая выпуклая функция $f:\Omega\to\mathbb{R}$, определенная на замкнутом множестве Ω. Дивергенцией Брегмана, порожденной функцией *f*, называется неотрицательная функция D_f, определенная как

$$D_f(p,q)=f(p)-f(q)-(p-q)\nabla f(q).$$

Дивергенция Брегмана обладает следующими свойствами:

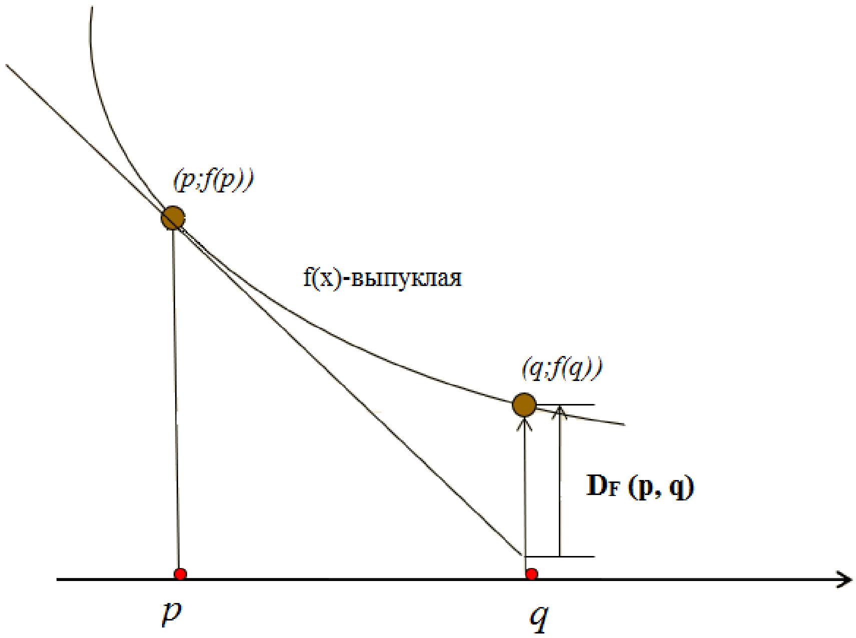
1. Для всех *p, q* выполнено $D_f(p,q)>0$. Это следует из выпуклости функции *f*.

2. $D_f(p,q)$ выпукла по первому аргументу *p*, но не обязательно выпукла по второму аргументу *q*.

3. Для любых $a,b\in\mathbb{R}$ и любых выпуклых функций *f, g* выполнено

$$D_{a\cdot f+b\cdot g}(p,q)=a\cdot D_f(p,q)+b\cdot D_g(p,q).$$

Графически дивергенцию Брегмана можно определить следующим образом: Пусть задана некая выпуклая функция *f(x)* и на оси абсцисс отложены точки p и q. Для определения дивергенции Брегмана от этих двух точек, через точку *(p;f(p))* проводится касательная к функции *f(x)*. Далее, через точку q на оси абсцисс проводится прямая, параллельная оси ординат, и находится точка пересечения этой прямой и касательной к графику, проведенной через точку *(p;f(p))*. Назовем полученную точку А. Тогда $D_f(p,q)$ это расстояние от точки А до точки *(q;f(q))*. Как видно, это значение всегда больше нуля, уменьшается при приближении точек друг к другу, и равно нулю в случае p=q.



Схема, поясняющая дивергенцию Брегмана

Рассмотрим дивергенцию по Брегману подробнее. Пусть на линии заданы две точки p и q. Эвклидово расстояние есть абсолютное значение разницы между позициями этих точек. Это расстояние положительно и симметрично.

Для расчета дивергенции Брегмана нужно сначала определить выпуклую функцию f(x), определенную на плоскости действительных чисел. Для того чтобы измерить различие между p и q с учетом функции f(x), сначала приходится определить линию (q,f(q)), проходящую по касательной к выпуклой функции f(x). Тогда расстояние определяется как различие по вертикали между кривой и линией в точке p.

Поскольку функция f(x) выпукла, можно показать, что если p ближе к q, то различие по вертикали (аналог расстояния) будет меньше. Если q совместится с p, «расстояние» будет равно 0.Прямая линия есть первый элемент ряда Тэйлора, аппроксимирующего функцию f(x) при q. Дивергенция Брегмана есть разница между функцией f(x) и любым расширением ряда в точке p.

Свойства функции:

- единственность. Функция f(x) однозначно определяет Df(p,q)
- положительность. Df(p,q)≥0 для любых p, q
- идентичность. Df(p,p)=0 для любой p

Дивергенция Брегмана заменяет метрику, но она не удовлетворяет ни правилу треугольника, ни правилу симметрии. Ее первоедостоинство в том, что она обобщает квадратичное эвклидово расстояние до класса расстояний, обладающих сходными свойствами. Ее второедостоинство состоит в установлении четкой связи с семействами экспонент распределений и в том, что есть взаимно однозначное соответствие между регулярными семействами экспонент и регулярными дивергенциями Брегмана.

Сам Брегманпредложил дивергенцию в контексте нелинейного (выпуклого) программирования[3]. Но его последователи использовали в качестве основного подхода обобщения эвклидовой метрики не только для усовершенствования алгоритмов оптимизации, но и для определения самой дивергенции Брегмана[4]. Было предложно даже оформить эти исследования как особую науку, информационную геометрию [5]. Однако, новая наука, в большей мере может считаться лишь частью теории вероятностей в исходной постановке А. Н. Колмогорова, которая рассматривалась, в свою очередь, лишь как часть теории меры [6].

Конкретное измерение различия зависит от вида выпуклой функции. Дивергенции широко используются для оценки сходства между двумя объектами. Например, KL-дивергенция (Кульбака-Лейблера) используется в теории информации для сравнения двух распределений вероятностей [7]. Расстояние Итакура-Сайто (IS) используется как мера восприятия разницы между спектрами. Обобщенный класс дивергенций, например дивергенции Брегмана, используются для распознавания образов и кластеризации [8]. В таблице представлены 4 самые распространенные виды дивергенций.

Основные используемые виды дивергенции

Название функции	Функция	Дивергенция
KL-дивергенция	$x\ln x$	$p\ln\frac{p}{q}-p+q$
Расстояние Итакура-Сайто	$-\ln x$	$\frac{p}{q}-\ln\frac{p}{q}-1$
Квадратичное эвклидово расстояние	x^2	$(p-q)^2$
Энтропия фон-Неймана	$tr(X\ln X-X)$	$tr(\mathrm{XlnX}-\mathrm{XlnY}-\mathrm{X}+\mathrm{Y})$

Остановимся поподробнее на KL-дивергенции. Она же – информационная дивергенция, она же относительная энтропия, она же – дискриминационная информация. Это – ассиметричная мера различия между двумя распределениями вероятностей *P* и *Q*. Измеряет математическое ожидание числа дополнительных битов, которые потребуются для кодирования выборки из *P,* когда используется код, основанный на Q.Обычно *P* есть некоторое «истинное» либо теоретическое распределение, а распределение *Q* – аппроксимация или модельное вычисление *P*.

Иногда KL-дивергенцию относят к метрическим мерам, но она ассиметрична. KL-дивергенция от *P* к *Q* не всегда равна KL-дивергенции от *Q* к *P*. KL-дивергенция относится к классу так называемых f-дивергенций. Предложена в 1930-е годы, рассекречена в 1951 году. Изначально KL дивергенция применялась лишь в математике и информатике, однако, с недавних пор началось активное внедрение данной функции в другие сферы деятельности. На сегодняшний день KL дивергенция уже применяется в региональном анализе, но пока только экологами для сравнения составов животного мира и почв. Перед нами встает вопрос об уместности применения подобной функции для сравнения заработных плат в двух различных регионах.

Очевидным является тот факт, что дивергенции Брегмана, и в частности, KL-дивергенция будет давать другие значения, чем обычная разность между средними, модами или медианами. Об этом подробно сказано, например, в исследованиях венгерских и индийских ученых [9].

Результаты

Рассмотрим на примере различия между регионами Южной Сибири по распределениям городов по численности населения. Искажение сравнения при использовании разницы средних значений объясняется следующим. В каждом регионе существует центр, размеры которого существенно превышают численности прочих городов региона.

Если сравнивать через разности средних значений, то максимальное различие выявляется между Омской и Томской областями, а наименьшее между

Кемеровской и Новосибирской. Если сравнивать с помощью KL-дивергенции, то наибольшее различие будет уже между Омской областью и Республикой Алтай, а наименьшее между Новосибирской и Томской областями.

Обсуждение

Таким образом, приведенное выше предложение новых мер межрегиональных отличий представляется полезным во многих отношениях. Прежде всего, это направление имеет множество вариантов развития, которые отчасти уже проработаны в многочисленных исследованиях.

В частности, исследования с применением новых видов дивергенций могут дать принципиально новые результаты в региональном анализе [10]. Наиболее перспективным методом разработки таких функций следует считать порождение их на основе семейств экспонент [11].

И, разумеется, следует учитывать опыт сравнения гистограм (с применение диаграмм Брегмана), накопленный в исследования физических процессов[12], а также в других науках[13].

Заключение

В изложенной выше постановке поиска новой системы оценки межрегиональных отличий снимаются многие недостатки ныне применяемых мер, не учитывающих отклонения фактических распределений от теоретических.

При сравнении полученных результатов с используемыми в настоящее время отмечаются новые аспекты межрегиональных сравнений, ранее ускользавших от внимания исследователей. В частности, довольно хорошо проработана в рамках применения дивергенций Брегмана проблема различной территориальной плотности объектов[15]. С их помощью также может быть по-новому решена проблема кластеризации регионов, построения их иерархических структур (регионы - округа) [16].

Сама по себе зарождающаяся наука «информационная геометрия» способна породить новые виды дивергенций [17]. В частности, неплохие результаты дает применение бета-дивергенции [18], а также энтропии Шеннона [10].

Дальнейшие исследования в данной области могут касаться не только межрегиональных сравнений. Перспективным выглядит применение дивергенций Брегмана для сравнения матриц [19]. Не исключено, что прогресс, который определит все будущие приложения,будет именно в рамках уже упоминавшейся трактовки теории вероятностей [20].

Особое направление будущих исследований связано с использованием триангуляций Делоне и диаграмм Вороного на основе диаграмм Брегмана [21]. продвижение в этом направлении приведет к прогрессу в экономическом районировании.

Работа выполнена по плану НИР ИЭОПП СО РАН, проект XI.170.1. «Формирование основ теории инновационной экономики: операциональные определения, измерения, модели, научно-технологические прогнозы и программы», № АААА-А17-117022250128–5.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Реброва О. Среднее или всё же медиана? Троицкий вариант. – 2011. – № 90. – С. 13.

2. Галенковский П. Количественные методы в экологии и гидробиологии. – [Электронный ресурс] https://zapdoc.site/kolichestvennye-metody-v-ekologii-i-gidrobiologii-bibliograf.html.

3. Bregman L. () The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming // USSR Mathematics and Mathematical Physics. –1967. – Vol. 7. –№. 3, Pp. 200–217.

4. Zhang J. Divergence function, duality, and convex analysis // Neural Computation. – 2004. – № 16. – Pp. 159-195.

5. Amari S., Cichocki A. Information geometry of divergence functions // Bulletin of the polish academy of sciences. Technical sciences. – 2010. – Vol. 58. – No. 1.

6. Amari S. Information geometry on hierarchy of probability distributions // IEEE Transactions on Information Theory. – 2001. – Vol. 47. – Pp. 1701–1711.

7. Kullback S., Leibler R.A. On information and sufficiency // Annals of Mathematical Statistics. –1951. – № 6, Pp. 79–86.

8. Frigyik B. A., Srivastava S., Gupta M. R. Functional Bregman Divergence, Conference Paper // IEEE International Symposium on Information Theory. – № 8. – 2008.

9. Chen P. Bregman Metrics and their Applications, a Dissertation for the Degree of Doctor of Philosophy, University of Florida, Gainesville. – 2007.

10. Wang H., Banerjee A. Bregman alternating direction method of multipliers, The Twenty-eighth Annual Conference on Neural Information Processing, San Diego, CA, USA, 2014.

11. Naudts J. Generalized exponential families and associated entropy functions // Entropy. – 2008. – №. 10. – Pp. 131–149.

12. Битюков С.И., Максимушкина А.В., Смирнова В.В. Сравнение гистограмм в физических исследованиях // Известия вузов. Ядерная энергетика. – 2016. – № 1. – С. 81–90.

13. Cha S.H., Srihari S.N. On measuring the distance between histograms // Pattern Recognition. – 2002. – Vol. 35. – №. 6. – Pp. 1355–1370.

14. Murata N., Fujimoto Yu. Bregman divergence and density integration // Journal of Math-for-industry. – 2009. – № 1. – Pp. 97–104.

15. Banerjee A., Merugu S., Dhillon I. S., Ghosh J., Clustering with Bregman divergences // Journal of Machine Learning Research. – 2005. – № 6. – Pp. 1705–1749.

16. Murata N., Takenouchi T., Kanamori T., Eguchi S. Information geometry of U-Boost and Bregman divergence // Neural Computation. – 2004. – №. 16. – Pp. 1437–1481.

17. Hennequin R., David B., Badeau R. Beta-Divergence as a Subclass of Bregman Divergence// IEEE Signal Processing Letters. – 2011. – Vol. 18. – №. 2. – Pp. 83–86.

18. Lin J. Divergence measures based on the Shannon entropy // IEEE Transactions on Information Theory, 1991. – Vol. IT-37. – Pp. 145–151.

19. Endres D. M., Schindelin J. E. A new metric for probability distributions // IEEE Transactions of Information Theory. – 2003. – Vol. 49. – №. 7, Pp. 1858–1860.

20. Dhillon I. S. Tropp J. A. Matrix Nearness Problems with Bregman Divergences // SIAM Journal on Matrix Analysis and Applications – 2007. – Vol. 29. – №. 4, Pp. 1120–1146.

21. Nielsen F., Boissonnat J.-D., Nock, R. On Bregman Voronoi diagrams // SODA. – 2007. – № 2. – Pp. 746–755.