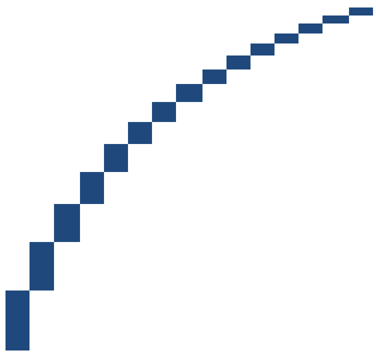
Методы классификации в ABC-анализе



Афанасьев Сергей

11 января 2012

Содержание

Введение

ABC-анализ – один из самых простых методов классификации объектов по заданному параметру. Несмотря на ограничения однопараметрических методов, ABC-анализ, на сегодняшний день, является одним из самых популярных аналитических инструментов. Благодаря этой популярности было разработано множество алгоритмов выделения групп в ABC-анализе, что, с одной стороны, дает аналитику богатый выбор методов в зависимости от целей и технических ресурсов, с другой, может нести в себе скрытые угрозы при неправильном применении того или иного метода

Процедуру проведения ABC-анализа можно разбить на 5 основных этапов схематически представленных ниже.

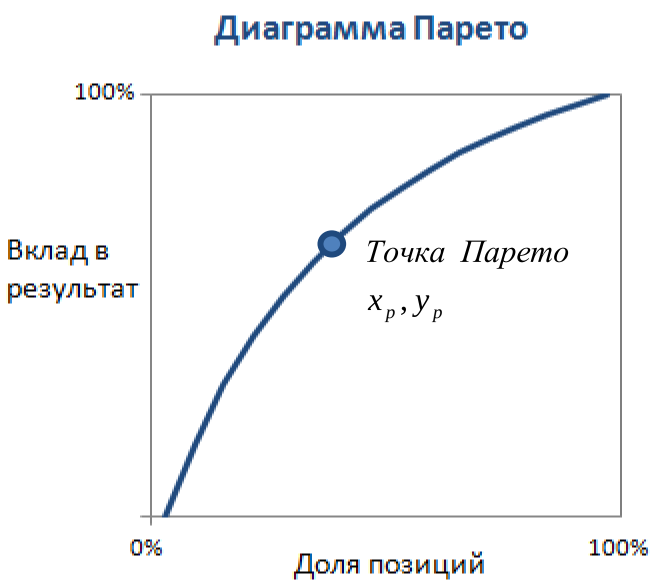
1. На первом этапе определяются цели анализа, от которых зависят дальнейшие действия и проводимые мероприятия.
2. На втором этапе, в зависимости от целей, выбирается параметр классификации объектов.
3. Третий этап включает в себя сбор и подготовку данных для ABC-классификации.
4. На четвертом этапе, с помощью того или иного алгоритма, проводится ABC-классификация объектов по заданному параметру.
5. На последнем, пятом этапе, результаты интерпретируются и принимаются необходимые решения.

В этой статье рассматриваются 7 наиболее часто используемых методов выделения групп в ABC-анализе, демонстрируются их плюсы и минусы, а также проводится сравнительный анализ этих методов.

Методы классификации в ABC-анализе

В основе ABC-анализа лежит правило Парето 20:80, которое формулируется как принцип дисбаланса – небольшая доля объектов дает основной вклад в результат. Таким образом целью ABC-классификации является разбиение множества объектов на три группы по заданному параметру так, что в группу A попадают объекты высоких значений результата, в группу B – объекты средних значений и в группу C – объекты низких значений. При этом учитывается характер дисбаланса.

Для выделения групп A, B и C объекты сортируют по параметру в порядке убывания и строят диаграмму накопленного итога. Оси диаграммы накопленного итога нормируют и приводят к процентным значениям. Полученный график носит название диаграммы Парето и является аналогом кривой Лоренса (кривой кумулятивного распределения).



Для диаграммы Парето вводится понятие *точки Парето*, как точки на диаграмме, сумма координат которой равна 100%:

Определение. Точка Парето – точка на диаграмме Парето с координатами (x_p, y_p), для которых выполняется равенство $x_p + y_p = 100\%$

Диаграмма Парето визуализирует структуру объектов по параметру, а выпуклые свойства кривой позволяют применять различные методы для выделения групп A, B и C.

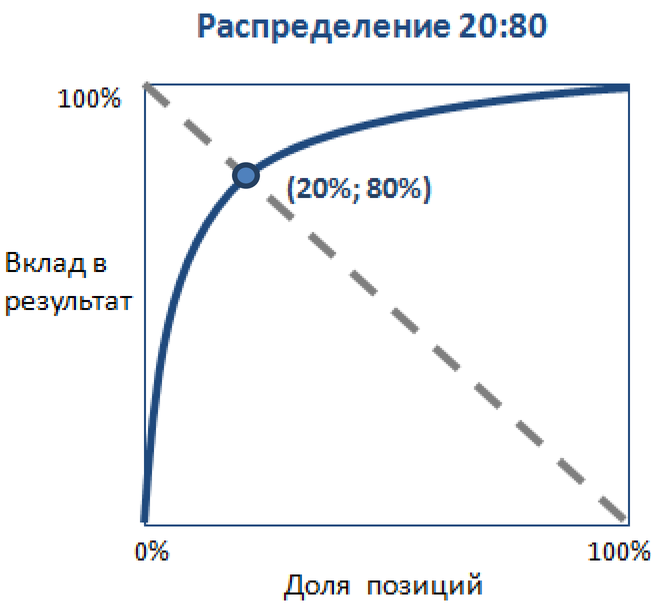
На сегодняшний день в различной литературе наиболее часто встречается описание следующих методов:

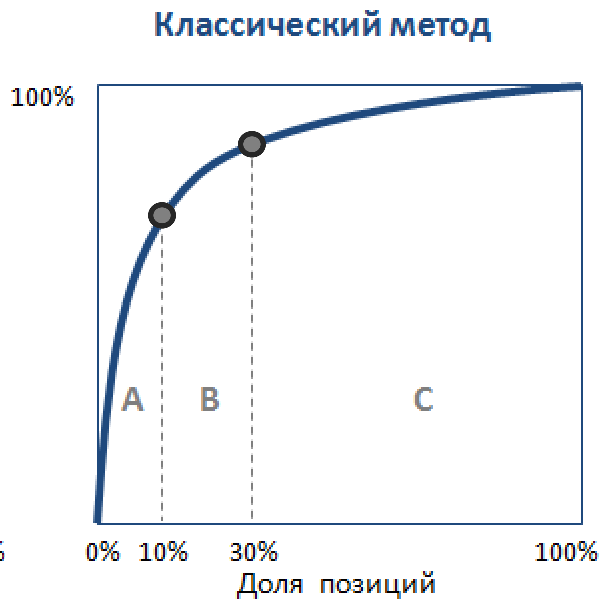
1. Классический метод.
2. Метод суммы.
3. Дифференциальный метод.
4. Метод касательных.
5. Метод многоугольника.
6. Метод петли.
7. Метод треугольника.

Некоторые из этих методов можно назвать продвинутыми, некоторые уже устарели (но до сих пор популярны и используются), а некоторые применимы только в частных случаях и не могут рассматриваться как эффективный инструмент для повседневной практики. Рассмотрим каждый из этих методов более подробно.

1. Классический метод

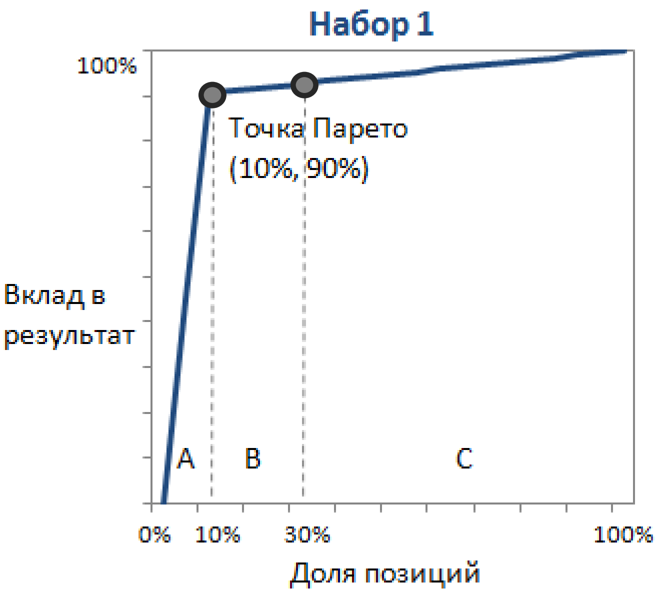
Классический метод (в некоторых источниках – эмпирический) состоит в том, что на построенной диаграмме Парето границы групп A, B и C являются фиксированными вне зависимости от структуры распределения. Такой подход объясняется тем, что согласно исследованиям экономистов и социологов конца XIX начала XX веков (Парето, Джини, Лоренс и др.) большинство процессов близки к распределению 20:80 – 20% объектов дают 80% результата. Поэтому на диаграмме Парето выделяют группу A - обычно это 10% позиций, группу B - 20% позиций, и группу C - 70% позиций.

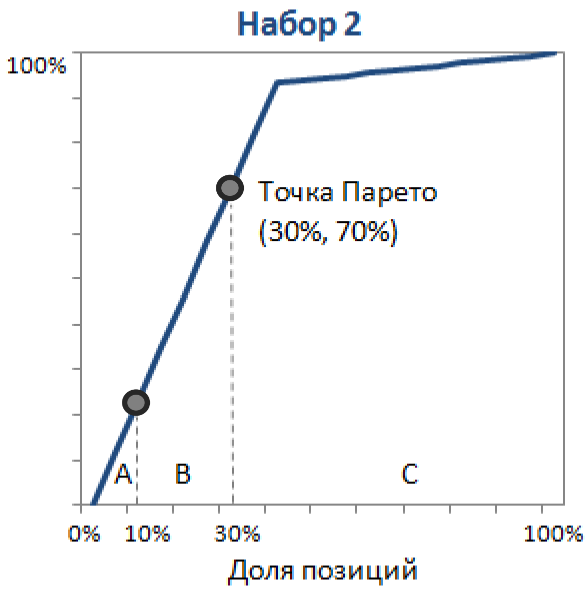




Отметим, что в различной литературе могут быть указаны и другие процентные отношения деления, например, группа A – 15% позиций, B – 20%, C – 65%. Также в некоторых источниках можно увидеть описание методик выделения групп не по доле позиций, а по вкладу в результат, например группа A – 70% вклада в результат, B – 20%, C – 10%. Кардинального различия в этих подходах нет.

Фиксирование границ в классическом методе накладывает существенные ограничения на структуру распределения объектов. Например, для объектов с равномерной структурой распределения (точка Парето = 50%,50%) ранжирование на три группы вообще не имеет практического смысла, поскольку все объекты имеют одинаковые значения результата. Для объектов со структурой распределений сильно отклоняющихся от точки (20%,80%) классический метод также будет давать плохие результаты. На рисунках ниже представлены диаграммы Парето для двух наборов с координатами точки Парето (10%,90%) и (30%,70%), для которых классический метод дает высокую ошибку классификации.

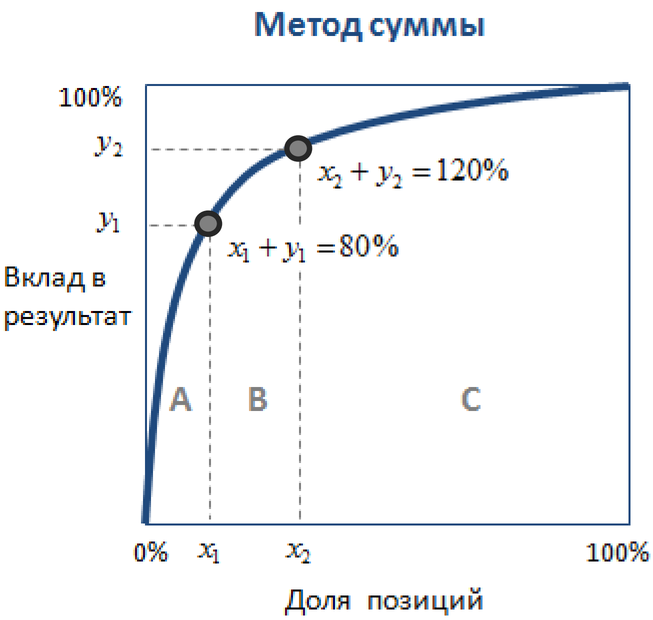




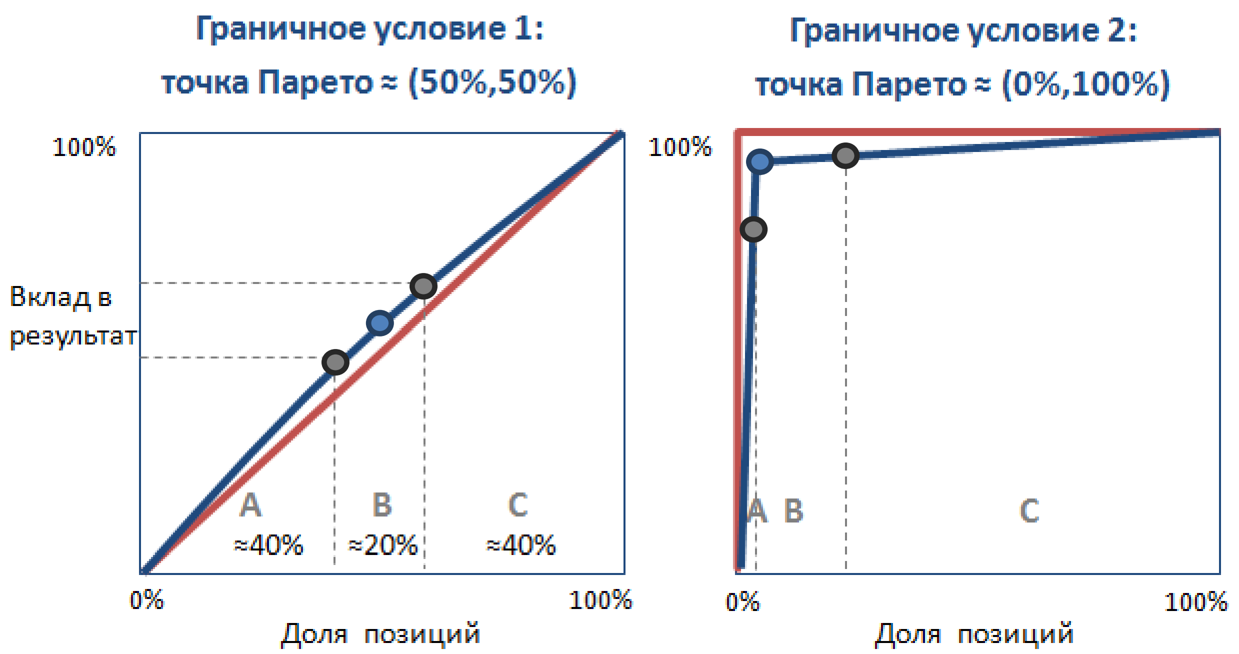
Из плюсов классического метода можно выделить его наглядность и простоту автоматизации. Однако, несмотря на эти плюсы, в общем случае классический метод является сильно упрощенным и может давать довольно высокую погрешность.

2. Метод суммы

Метод суммы является модификаций классического метода, при этом вместо фиксирования границ на одной из осей диаграммы Парето, берутся суммы координат точек диаграммы, а границы разбиения определяют путем фиксации двух значений полученных величин. Например, если отталкиваться от значений классического метода, то в методе суммы граница разделения групп А и В проходит через точку диаграммы, сумма координат которой равна 80% (≈10%+70%), а граница разделения групп В и С через точку с суммой координат 120% (≈30%+90%). Отметим, что в силу выпуклых свойств диаграммы Парето, такие точки существуют и единственны.



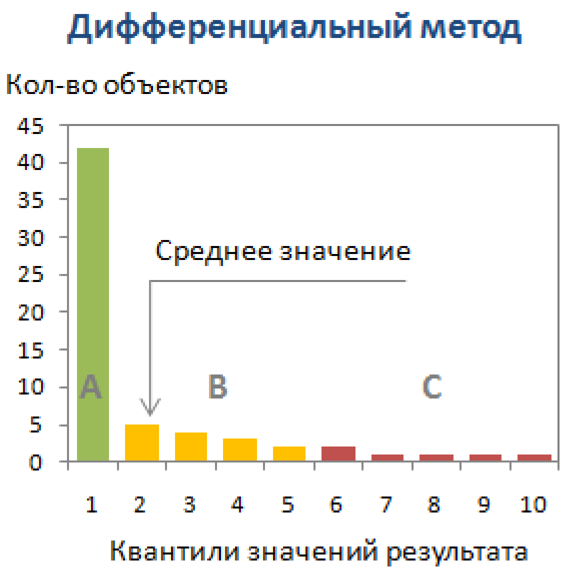
Метод суммы имеет достаточно низкую точность по тем же причинам, что и классический метод – привязка к жестко зафиксированным границам классификации. Так, для равномерных распределений или близких к ним (точка Парето (50%,50%)) метод суммы выделяет в группу А около 40% объектов, в группу В примерно 20% объектов и в группу С порядка 40% объектов. Но как уже было замечено выше, для таких распределений ранжирование объектов на группы вообще нецелесообразно с практической точки зрения. Для распределений, удовлетворяющих второму граничному условию (точка Парето (0%,100%)), результаты также получаются плохими.



На практике часто используют другие значения границ, суммы координат которых равны 100% и 145% соответственно. Это нарушает симметрию ABC анализа относительно точки Парето, однако на фоне такой же низкой точности классификации это различие несущественно. Плюсом метода суммы, как и в классическом методе, можно считать простоту его автоматизации.

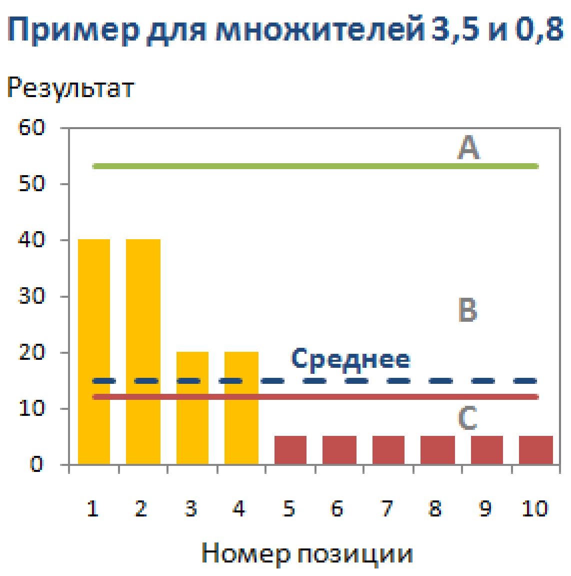
3. Дифференциальный метод

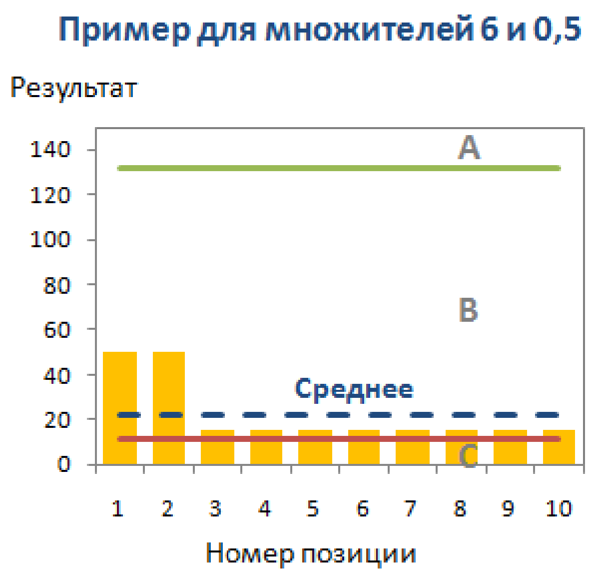
Дифференциальный метод берет свои начала в математической статистике. Суть данного метода заключается в расчете среднего значения результата и определение квантилей для групп A, B и C. Вместо диаграммы Парето для этого метода строят статистическое распределение Парето (частотная гистограмма). Для этого объекты разбивают на категории с одинаковым шагом по значению результата (квантили) и строят частотное распределение, где по оси X откладывают средние значения квантилей, а по оси Y – количество объектов попавших в определенный квантиль (частоту). Количество квантилей зависит от количества объектов и, как правило, варьируется в диапазоне от 10 до 30. Далее к группе A относят объекты попавшие в квантили со значениями в 3,5 раза выше среднего, а в группу C объекты, попавшие в квантили со значениями от 0,8 среднего и ниже (или, что тоже самое, в 1,25 раза ниже среднего). Остальные объекты относят в группу B.



Отметим, что при использовании множителей 3,5 и 0,8 результаты классификации получаются схожими с классическим методом для распределений с точкой Парето (20%;80%). В некоторых источниках приводится описание метода с множителями 6 и 0,5 (группа A – в 6 раз выше среднего, группа C – в 2 раза ниже среднего), что характерно для распределений с точкой Парето (5%,95%).

Минусом дифференциального метода, является то, что фиксированные множители, используемые для выделения групп, накладывают существенные ограничения на структуру распределения (как и в предыдущих двух методах). Так, статистическое распределение Парето может принимать различные формы. Кроме того объекты могут иметь другие распределения, например, нормальное, логнормальное, распределение хи-квадрат и другие. При этом для рационального ранжирования объектов различным формам распределений должны соответствовать различные множители. Поэтому фиксация множителей существенно снижает точность дифференциального метода, что перекрывает плюсы, связанные с простотой автоматизации данного метода. На рисунках ниже приведены примеры наборов, для которых дифференциальный метод со стандартными множителями, используемыми на практике, не работает.





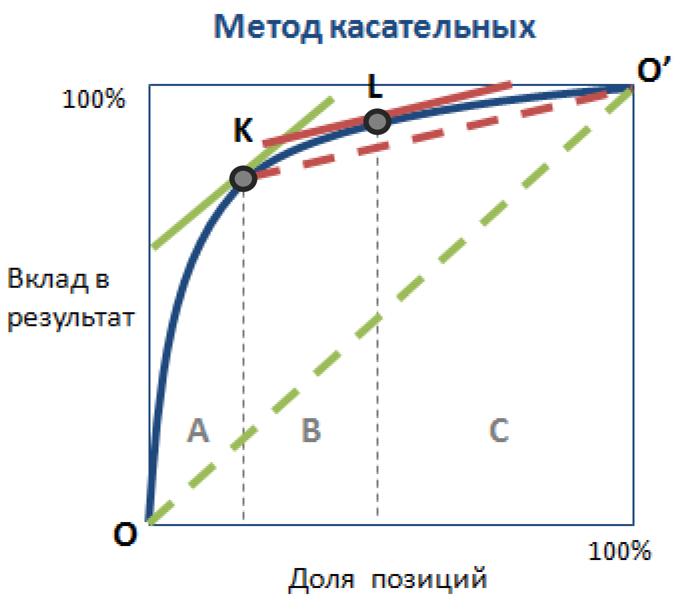
** Раскраска на рисунках соответсвует оптимальному разбиению (красный – группа низких значений, желтый – группа средних значений, зеленый – группа высоких значений).*

4. Метод касательных

Метод касательных, предложенный В.С. Лукинским, относится к классу методов с нефиксированными границами. Суть метода состоит в том, что на диаграмме Парето строят две касательные, которые определяют координаты границ групп А, В и С.

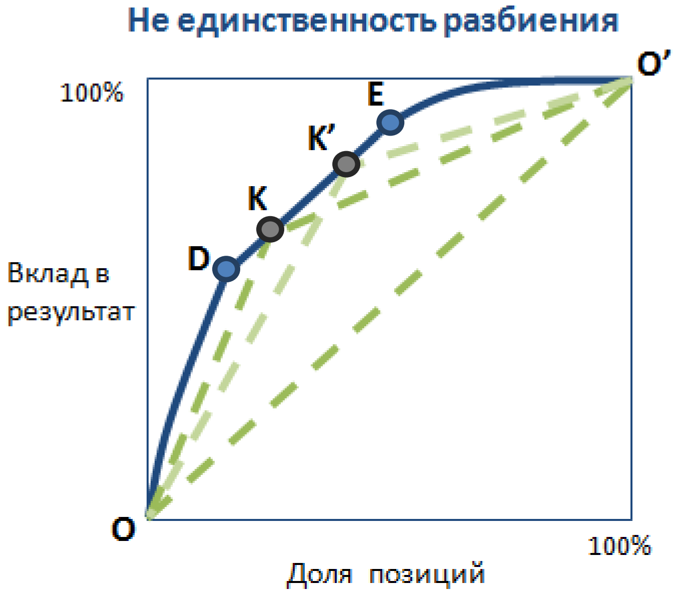
Первую касательную строят как прямую параллельную диагонали ОО′ (см. рисунок ниже). Такая касательная единственна в силу выпуклости диаграммы Парето. Координаты полученной точки касания К определяют границу разбиения групп А и В.

Для нахождения второй границы, точки К и О′ соединяют прямой и строят вторую касательную к диаграмме Парето, как прямую параллельную отрезку КО′. Координаты полученной точки касания L определяют границу разбиения групп В и С.

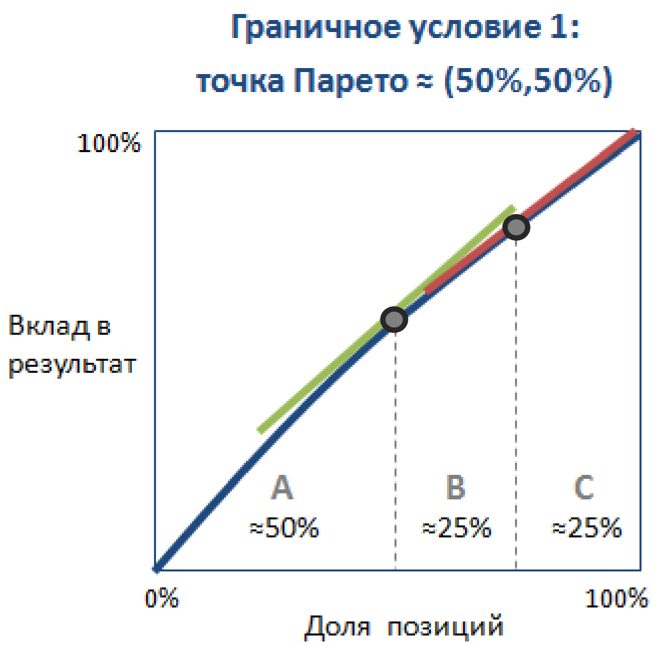


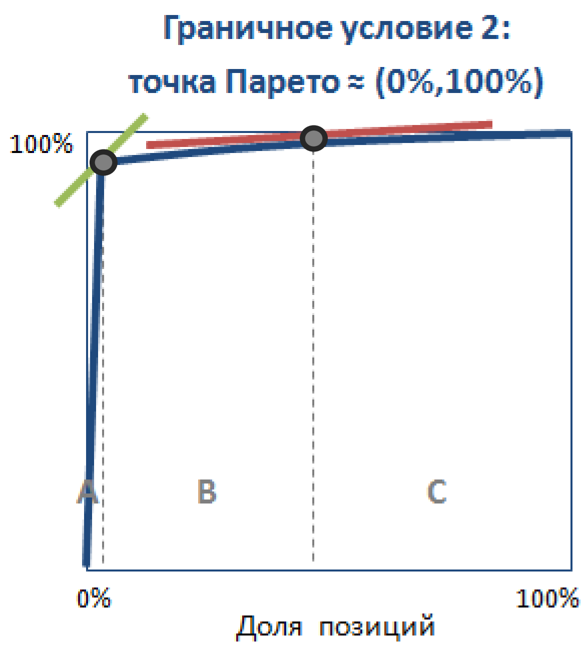
Отметим, что точка K, полученная первой касательной, является точкой минимизации площадей между диаграммой Парето и отрезками OK и KO’. Таким образом, можно считать, что первая касательная оптимально делит множество на две группы, поскольку каждая из двух полученных частей диаграммы наиболее близко лежит к прямолинейному отрезку (площадь минимальна), что говорит о наименьшей изменчивости параметра внутри полученной группы или о минимизации ошибки классификации. Точка L, образованная второй касательной, дает оптимальное разбиение на две группы множества, лежащего между точками K и O’ на диаграмме Парето. Поэтому ошибка классификации для второго разбиения на два множества также минимальна. При этом разбиение всего множества на три группы точками K и L не является оптимальным (минимальная ошибка не равна сумме двух минимальных ошибок).

Одним из основных недостатков метода минимальных площадей является неединственность разбиения для некоторых форм выпуклых кривых (диаграмм Парето). Предположим, что на диаграмме Парето есть прямолинейный отрезок DE параллельный диагонали OO’ (см. рисунок ниже). Тогда первая касательная будет проходить через весь отрезок DE. Таким образом, точки данного отрезка образовывают равные по площади треугольники с основанием OO’. Это значит, что любая точка отрезка DE является оптимальной и разбиение не единственно.



Из минусов данного метода можно также отметить его несимметричность относительно точки Парето, что повышает уровень ошибки. Так же метод плохо работает на распределениях близких к граничным точкам Парето. Для распределений близких к первой граничной точке Парето (50%,50%), то есть близких к равномерному распределению, метод касательных выделяет большую группу А и средних размеров группы В и С. Для таких распределений, как уже отмечалось ранее, АВС-классификация вообще нецелесообразна и все объекты должны быть отнесены в одну группу. Для распределений близких к второй граничной точке Парето (0%,100%), метод касательных достаточно точно выделяет группу А, однако группы В и С выделяются нерационально.

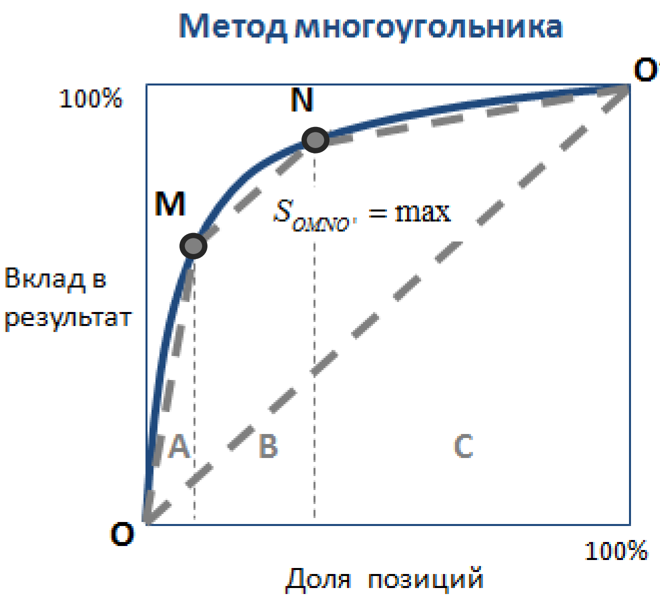


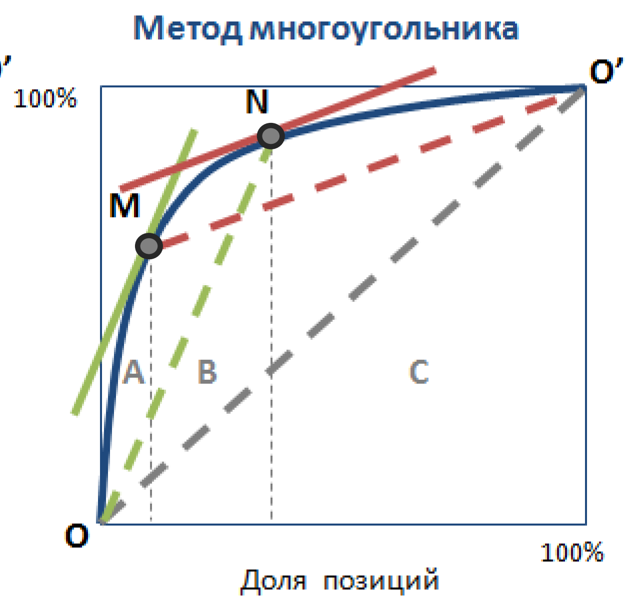


Из плюсов метода касательных можно отметить масштабируемость алгоритма (сложность алгоритма O(cN)) и, как следствие, несложную автоматизацию, а также так же невысокий уровень ошибки в области средних значений точки Парето.

5. Метод многоугольника

Метод многоугольника, называемый также методом двойной касательной, решает задачу оптимального разбиения множества на три группы. Поэтому метод многоугольника можно назвать усовершенствованием метода касательных. Суть метода заключается в нахождении на диаграмме Парето двух точек M и N таких, что площадь многоугольника OMNO’ максимальна, или, что тоже самое, площадь между ломаной OMNO’ и диаграммой Парето минимальна. Минимизация площади позволяет разбить диаграмму Парето на три части таким образом, что каждая из этих частей наиболее близко прилегает к прямолинейному отрезку (площадь минимальна), то есть изменчивость параметра внутри групп минимальна и, как следствие, ошибка классификации тоже минимальна. Графически задача минимизации площади решается через построение двух касательных к диаграмме Парето параллельных отрезкам ON и MO’. Точки M и N определяются как границы разбиения групп A, B и C.

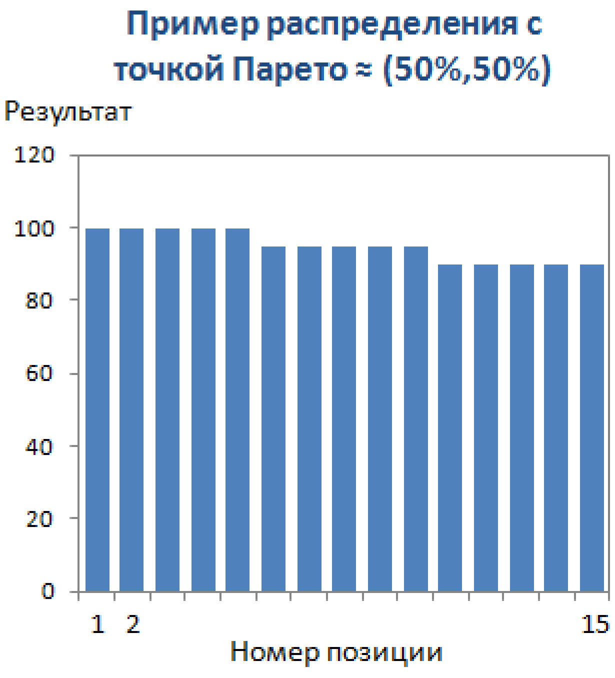


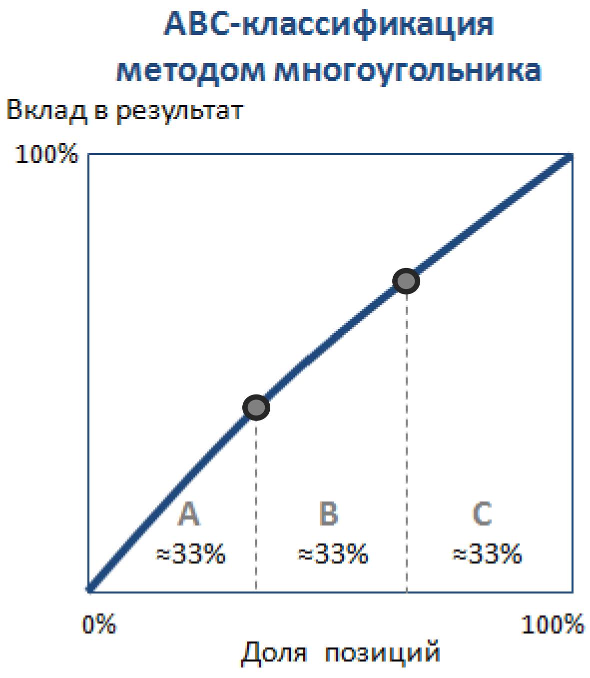


Как и в методе касательных Лукинского, для метода многоугольника существуют формы диаграмм Парето, которые могут содержать прямолинейные участки параллельные отрезкам ON и MO’. Тогда касательные будут проходить через эти отрезки и точки минимизирующие площадь (точки касания) не будут единственны. Не единственность разбиения является одним из главных минусов метода многоугольника.

Существенным недостатком метода можно назвать сложность его автоматизации, поскольку алгоритм имеет сложность $O(N^2)$, то есть не является масштабируемым.

Главным плюсом метода многоугольника является высокая точность классификации. Из рассматриваемых в данной статье методов уровень ошибки метода многоугольника можно считать самым низким. Однако высокая точность классификации не всегда может давать корректные результаты в силу специфики ABC-анализа, что демонстрируется примером ниже, где все объекты незначительно отличаются друг от друга по значению результата и по логике ABC-анализа должны быть отнесены в одну группу.

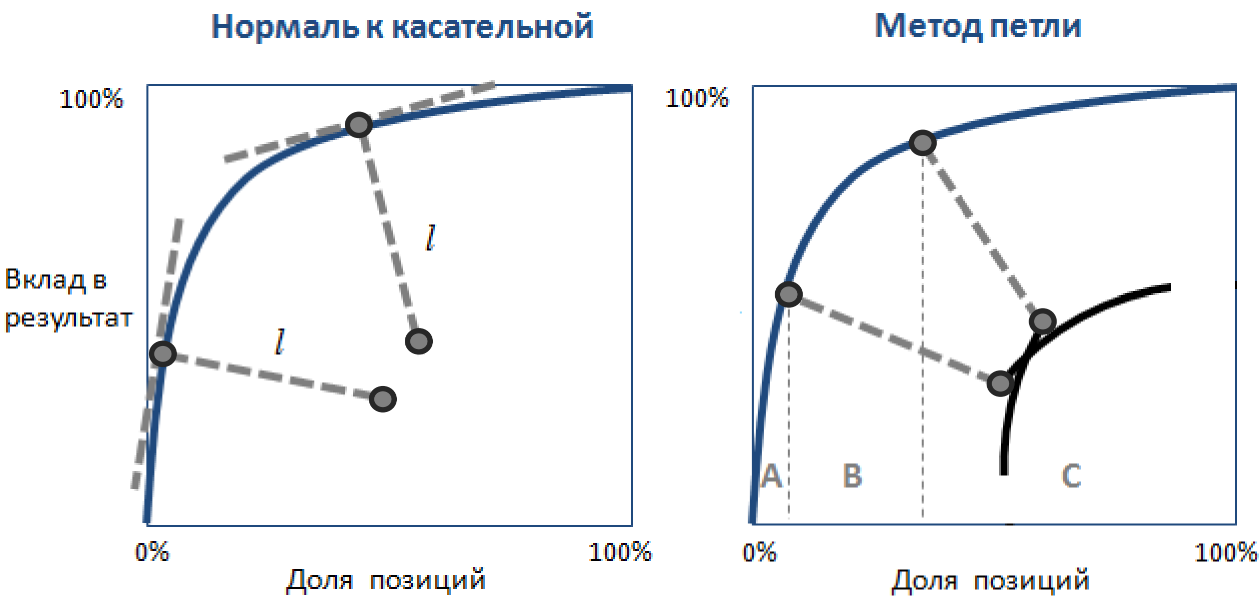




6. Метод петли

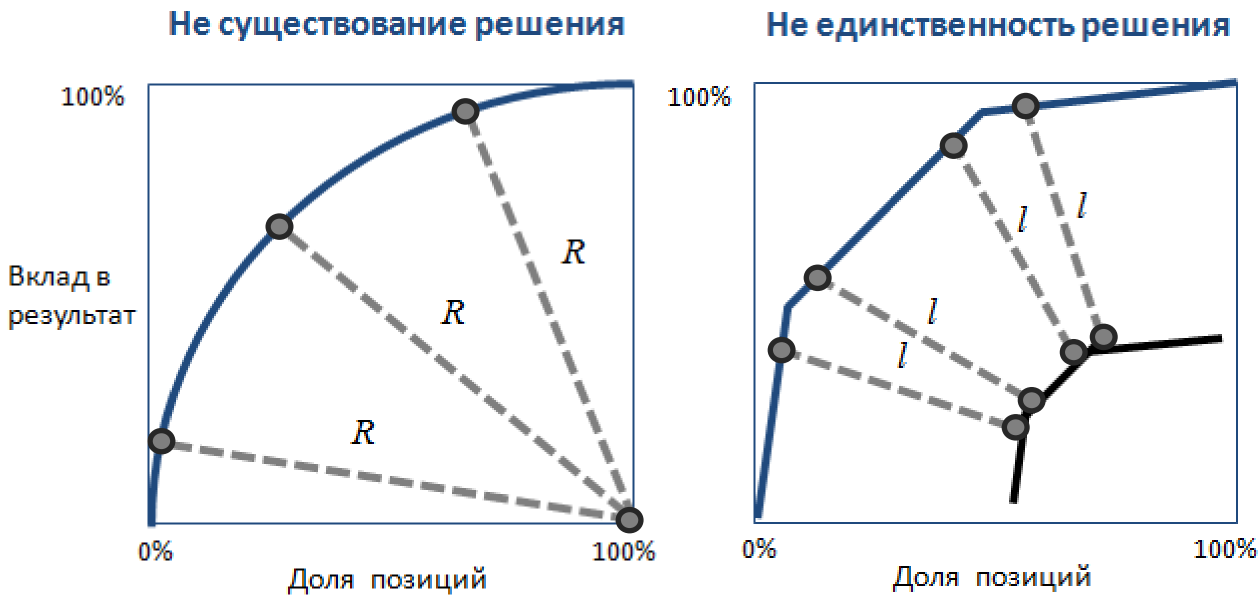
Метод петли, предложенный А.М. Гаджинским, решает задачу оптимального выделения двух групп – высоких и низких значений. В основу метода заложено понятие кривизны кривой, которое определяется ее радиусом. Целью метода является выделение на диаграмме Парето такого участка, который имеет наибольшую кривизну (наименьший радиус кривизны), то есть участка перехода от более вертикальной (крутой) части диаграммы к более горизонтальной (пологой). Такая постановка задачи полностью соответствует целям ABC-классификации, поскольку объекты высоких значений результата располагаются на крутом участке диаграммы, а низких – на пологом.

Алгоритм метода заключается в построении траектории движения конца отрезка, перпендикулярного к касательной диаграммы Парето. При правильном выборе длины отрезка l и существовании единственного участка с наибольшей кривизной для такого отрезка, траектория движения конца нормали должна вырисовывать петлю. Это обусловлено тем, что на крутом и пологом участках диаграммы радиус кривизны больше, чем на участке перегиба. Таким образом, необходимым условием существования петли является условие: длина отрезка l должна быть меньше радиуса кривизны крутого и пологого участков, но больше радиуса кривизны участка перегиба. Точки на диаграмме Парето, соответствующие границам участка перегиба, являются границами разбиения множества на группы А. В и С. Эти точки определяются концами перпендикуляров, выходящих из угловых точек петли



Основными недостатками метода являются несуществование решения и неединственность решения для некоторых форм диаграмм Парето. Так для диаграммы Парето представляющую часть окружности, решения не существует, поскольку все участки диаграммы имеет одинаковую кривизну, т.е. петлю нельзя получить ни при какой длине перпендикуляра. У некоторых форм диаграмм может существовать несколько участков с одинаковой максимальной кривизной. В таких случаях петель, и соответственно решений, будет столько, сколько таких участков содержит диаграмма.

При слишком большом шаге итерации подбора длины нормали, или недостаточной длине нормали (распределения близкие к точке Парето (50%,50%)), петля также может быть не получена.



Из плюсов метода можно отметить высокую точность классификации, а также масштабируемость алгоритма (сложность O(cN)).

7. Метод треугольника

Метод треугольника, как и метод петли, решает задачу оптимального выделения двух групп на диаграмме Парето – высоких и низких значений. В основу метода заложено математическое свойство, при котором множество всех диаграмм Парето, проходящих через одну из точек Парето, лежит в области, которая ограничена двумя предельными диаграммами Парето, имеющих форму двухзвенных ломаных. Это свойство позволяет выделить границы на диаграмме, которые определяют зоны высоких и низких значений. Для этого необходимо построить две прямые, проходящие через точки излома предельных диаграмм, которые при этом параллельны диагонали ((0%;100%),(100%;0%)). Точки пересечения этих прямых и диаграммы Парето определяются за границы разделения групп A, B и C. Уравнения искомых прямых определяются через координаты точки Парето (x_p, y_p).

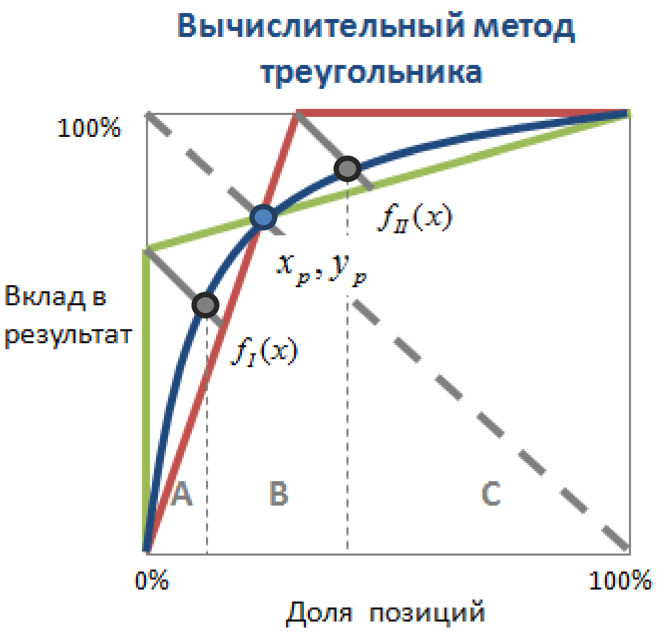
Уравнение прямой:	$f_I(x) = \frac{y_p - x_p}{y_p} - x$	**(1)**

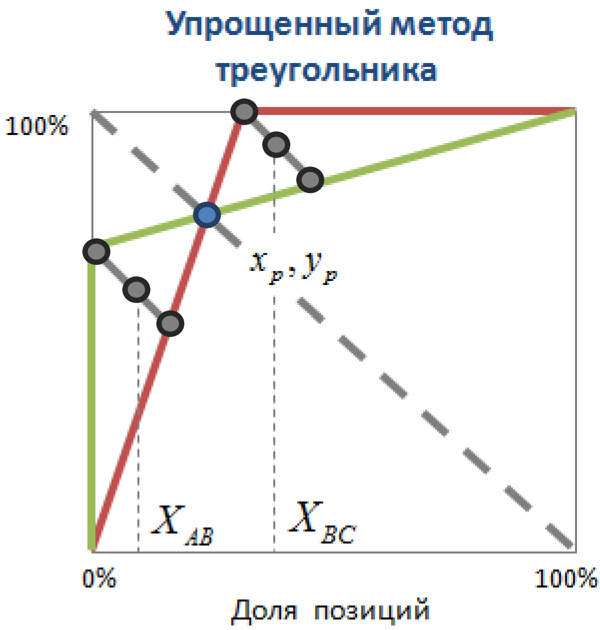
Уравнение прямой:	$f_{II}(x) = \frac{y_p + x_p}{y_p} - x$	**(2)**

Для упрощения вычислений за границы можно брать середины отрезков, образованных построенными прямыми, концы которых лежат на звеньях предельных ломанных. Тогда абсциссы этих точек легко вычисляются по формулам (3) и (4).

Формула:	$X_{AB} = \frac{x_p * y_p - x_p^2}{2 * y_p}$	**(3)**

Формула:	$X_{BC} = \frac{3 * x_p * y_p + x_p^2}{2 * y_p}$	**(4)**





Основным преимуществом метода треугольника является его гибкость – метод показывает высокую точность как в области средних значений точки Парето, так и в области граничных значений. Однако стоит отметить, что в области средних значений метод треугольника проигрывает по точности методу многоугольника.

Необходимо отметить, что метод треугольника удовлетворяет всем целям ABC-анализа, поскольку для любого множества объектов данный метод дает единственное разбиение и удовлетворяет всем граничным условиям, в том числе для равномерных распределений и близких к ним, где все или почти все объекты попадают в одну группу B.

Метод треугольника достаточно просто автоматизируется, а по формулам (3) и (4) можно построить таблицу соответствия координат точек Парето и значений групп A, B и C, что сводит автоматизацию к поиску координат точки Парето.

Сравнительный анализ методов

В сравнительном анализе суммируются все описанные выше достоинства и недостатки методов в рамках 6-ти критериев:

1. Существование разбиения для любого типа распределений;
2. Единственность разбиения для любого типа распределений;
3. Точность классификации в области средних значений;
4. Точность классификации в области точки Парето (50%,50%);
5. Точность классификации в области точки Парето (0%,100%);
6. Сложность технической реализации (автоматизация).

В приведенной ниже таблице знак '–' проставлен по критерию в том случае, если метод не удовлетворяет данному критерию, знак '+' – если удовлетворяет. При этом количество знаков '+' соответствует рейтингу метода по заданному критерию (чем лучше метод удовлетворяет соответствующий критерий, тем больше знаков '+').

	Существование разбиения для любого типа распределений	Единственность разбиения для любого типа распределений	Точность классификации в области средних значений	Точность классификации в области точки Парето (50%,50%)	Точность классификации в области точки Парето (0%,100%)	Сложность технической реализации (автоматизация)
Классический метод	+	+	+	-	-	+ + + +
Метод суммы	+	+	+	-	-	+ + + +
Дифференциальный метод	+	+	+ +	+	+ +	+ + + +
Метод касательных	+	-	+ + +	-	+	+ + +
Метод многоугольника	+	-	+ + + + +	-	+ + + +	+ +
Метод петли	-	-	+ + + +	-	+ + +	+
Метод треугольника	+	+	+ + + +	+ +	+ + +	+ + + +

По данным таблицы видно, что, например, метод ‘петли’ не всегда решает задачу существования разбиения. Метод касательных, метод многоугольник и метод петли не всегда дают единственность разбиения. Наиболее точным методом в области средних значений, а также в области точки Парето (0%,100%) является метод многоугольника, который, однако, дает некорректное разбиение в области точки Парето (50%,50%). Наиболее простыми с точки зрения технической реализации можно считать классический метод, дифференциальный метод, метод суммы и метод треугольника. Наиболее гибким методом, удовлетворяющим всем критериям, является метод треугольника.

Литература

1. В.С. Лукинский «Модели и методы теории логистики».
2. А.М. Гаджинский «Логистика»
3. Т. Василенко «Миф о 80/20», http://www.improvement.ru/
4. С.В. Афанасьев, "Метод треугольника в ABC-анализе" (Маркетинг в России и за рубежом, 2007, N 2), http://www.dis.ru/