

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Пензенский государственный университет» (ПГУ)

А. И. Иванов

Нейросетевой многокритериальный
статистический анализ малых выборок.
Проверка гипотезы независимости

Справочник

Пенза
Издательство ПГУ
2023

УДК 519.22; 53; 57.017

ББК 32.818

И20

Р е ц е н з е н т

доктор технических наук, профессор,
ученый секретарь Научно-производственного
предприятия «Рубин» (г. Пенза)

М. М. Бутаев

Иванов, Александр Иванович.

И20

Нейросетевой многокритериальный статистический анализ малых выборок. Проверка гипотезы независимости : справочник / А. И. Иванов. – Пенза : Изд-во ПГУ, 2023. – 218 с.

ISBN 978-5-907666-49-8

Рассматриваются проблемы статистической обработки малых выборок реальных данных. Созданные в прошлом веке критерии проверки гипотезы независимости, к сожалению, дают решения с приемлемой достоверной вероятностью на больших выборках в 200 и более опытов. Как правило, выборки реальных данных оказываются намного меньше. С проблемой нехватки реальных данных сталкиваются медики, экономисты, биологи и представители других научных специализаций. Очень остро эта проблема стоит перед специалистами биометрической идентификации личности человека. Накопленный в России опыт в направлении нейросетевого статистического анализа малых выборок систематизирован и изложен в данном справочнике. Выборки в 16 опытов реальных данных может оказаться вполне достаточно, если вместо одного статистического критерия использовать 160 разных статистических критериев. Объединять разные статистические критерии проще всего, реализовав каждый из критериев соответствующим искусственным нейроном. Тогда однослойная нейросеть из 160 нейронов будет выдавать выходной код с 160-кратной избыточностью. Свертывая столь высокую кодовую избыточность, удается обнаруживать и устранять ошибки. Представлена 21 программная реализация классических статистических критериев прошлого века. Кроме того, приведены 150 реализаций новых статистических критериев, созданных в начале XXI в.

Издание ориентировано на студентов и аспирантов, имеющих начальные знания по математической статистике и программированию.

УДК 519.22; 53; 57.017

ББК 32.818

Индексы цитирования автора:

Scopus – 57189212610;

РИНЦ 744989; SPIN-код 2277-7744.

ISBN 978-5-907666-49-8

© Иванов А. И., 2023

СОДЕРЖАНИЕ

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ.....	19
ВВЕДЕНИЕ.....	33
ФУНКЦИИ ЯЗЫКА MathCAD, ИСПОЛЬЗОВАННЫЕ ПРИ ПРОГРАММИРОВАНИИ В ДАННОМ СПРАВОЧНИКЕ	36
Глава 1. КЛАССИЧЕСКИЕ КРИТЕРИИ, СОЗДАННЫЕ В ПРОШЛОМ ВЕКЕ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ	37
1.1. Корреляционный критерий Эджуорта – Эдлтона – Пирсона (конец XIX в.).....	37
1.2. Критерий Кенуя (1973).....	38
1.3. Критерий Нельсона (1983).....	39
1.4. Критерий Блума – Кифера – Розенбладтта (1961).....	39
1.5. Критерий Кокса – Стюарта (1955)	40
1.6. Критерий Аббе – Линника (1965)	41
1.7. Угловой критерий Олмстеда – Тьюки (1947).....	42
1.8. Критерий Сендвика – Олсона (1982)	43
1.9. Критерий Шахани (1973)	43
1.10. Критерий Кенделла (1971).....	44
1.11. Критерий Спирмена (1971).....	45
1.12. Критерий Шведа – Эйзенхарта (1971)	45
1.13. Сериально-ранговый критерий Рамачандрана – Ранганатана.....	46
1.14. Критерий Ван дер Вандена (1967)	47
1.15. Критерий Шукени – Флори (1973).....	48
1.16. Критерий Бартелса (1982).....	49
1.17. Сериально-ранговый критерий Шахнесси (1981).....	49
1.18. Критерий Эландта (1962).....	50
1.19. Критерий, построенный на статистике Манна – Уитни (1947)	51
1.20. Критерий, построенный Казаковичусом на циклическом сдвиге данных и оценке их автокорреляции (1988)	51
1.21. Дифференциальный вариант критерия, построенного Кноке на циклическом сдвиге данных и оценке коэффициентов автокорреляции (1975).....	52
1.22. Прогнозирование достаточного для практики числа нейрокритериев проверки гипотезы независимости	53
1.23. Оценка корреляционной сцепленности откликов классических статистических критериев.....	56
Список литературы к главе 1	59

Глава 2. ВАРИАЦИИ КРИТЕРИЕВ, БЛИЗКИХ К КЛАССИЧЕСКОМУ КРИТЕРИЮ ЭДЖУОРТА – ЭДЛТОНА – ПИРСОНА.....	61
2.1. Дискретный вариант критерия Эджуорта – Эдлтона – Пирсона ...	61
2.2. Дискретный корреляционный критерий, построенный на использовании двух эллиптических квантователей.....	62
2.3. Первый непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их мгновенными приближениями	63
2.4. Второй непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их приближениями и заменой математических ожиданий минимумами данных в малой выборке.....	63
2.5. Третий непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их приближениями и заменой математических ожиданий максимумами данных в малой выборке	64
2.6. Четвертый непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их приближениями и заменой математических ожиданий $\max(x)$ и $\min(y)$	65
2.7. Пятый непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их приближениями и заменой двух математических ожиданий на $\min(x)$ и на $\max(y)$	66
2.8. Шестой непрерывный аналог критерия Пирсона – $r6R_E_min$, полученный заменой минимальных значений по переменной – x на ее математическое ожидание	66
2.9. Седьмой непрерывный аналог критерия Пирсона – $r7R_min_E$, полученный использованием в исходной формуле математических ожиданий по переменной – y и минимального значения переменной – x	67
2.10. Восьмой непрерывный аналог критерия Пирсона – $r8R_E_max$, полученный заменой экстремумов по переменной – x на математическое ожидание $E(x)$ и использованием максимального значения по второй переменной.....	68
2.11. Девятый непрерывный аналог критерия Пирсона $r9R_max_E$, полученный использованием математического ожидания по переменной – x и максимального значения переменной – y	69

2.12. Десятый непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный нормировкой квадратом суммы дисперсий.....	69
2.13. Корреляционная сцепленность группы аналогов классического критерия Пирсона – Эджуорта – Эдлтона (конец XIX в.).....	70
Список литературы к главе 2	71
Глава 3. НОВЫЕ КРИТЕРИИ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОЛУЧЕННЫЕ ВАРИАЦИЯМИ КЛАССИЧЕСКИХ КРИТЕРИЕВ И ИХ КОМБИНИРОВАНИЕМ	
3.1. Гибрид критерия Олмстеда – Тьюки с критерием Колмогорова – Смирнова	72
3.2. Второй вариант критерия Сендвиги – Олстеда.....	73
3.3. Второй вариант критерия Спирмена.....	73
3.4. Третий вариант критерия Спирмена	74
3.5. Мультипликативный гибрид критериев Ван дер Вандена и Бартлеса	74
3.6. Аддитивная сборка двух вариантов критерия Сендвиги – Олстеда.....	76
3.7. Мультипликативная сборка двух вариантов критерия Сендвиги – Олстеда.....	76
3.8. Модификация критерия Кенделла за счет накапливания квадратов его рангов	77
3.9. Аддитивная сборка критерия Сендвиги – Олстеда и модифицированного критерия Кенделла	78
3.10. Аддитивная сборка критерия Ван дер Вандена и критерия Эландта.....	79
3.11. Аддитивная сборка критерия Шахани и модифицированного критерия Кенделла	80
3.12. Аддитивная сборка критерия Шахани и автокорреляционного критерия Кноке	80
3.13. Аддитивная сборка двух вариантов критерия Кенделла.....	81
3.14. Корреляционная сцепленность в группе критериев, полученных комбинированием классических критериев XX в.	82
Глава 4. НОВЫЕ КРИТЕРИИ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОЛУЧЕННЫЕ УПОРЯДОЧИВАНИЕМ ДАННЫХ ПО ОДНОЙ И ДВУМ ПЕРЕМЕННЫМ ДАННЫХ МАЛОЙ ВЫБОРКИ	
4.1. Континуальный критерий, построенный на упорядочивании одной переменной и сглаживанием второй случайной составляющей	85

4.2. Аддитивная сборка двух континуальных критериев, построенных на упорядочивании одной переменной со сглаживанием случайной составляющей и вычислением ее остатков	86
4.3. Континуальный критерий, построенный на упорядочивании одной переменной с дифференцированием составляющей случайной переменной	87
4.4. Аддитивная сборка двух критериев, построенных на упорядочивании одной переменной с дифференцированием составляющей случайной переменной	88
4.5. Континуальный критерий, построенный на упорядочивании одной переменной и вычислении суммы расстояний Евклида по другой переменной	89
4.6. Континуальный критерий, построенный на упорядочивании одной переменной и накапливании квадратов разностей расстояний между соседними точками по второй переменной.....	90
4.7. Аддитивная сборка двух континуальных критериев, построенных на упорядочивании одной переменной и накапливании квадратов разностей расстояний между соседними точками по второй переменной.....	91
4.8. Континуальный критерий, построенный на упорядочивании одной переменной и накапливании произведения расстояний между соседними точками по второй переменной.....	92
4.9. Континуальный критерий, построенный на упорядочивании одной переменной и вычислении стандартного отклонения произведения расстояний между соседними точками по второй переменной.....	93
4.10. Континуальный критерий, построенный на упорядочивании одной переменной и выявлении максимального значения произведения расстояний между соседними точками по второй переменной.....	94
4.11. Аддитивная сборка континуальных критериев, построенных на упорядочивании двух переменных и оценке стандартного отклонения произведений расстояний между соседними точками по оставшимся случайным переменным	95
4.12. Аддитивная сборка двух континуальных критериев, построенных на упорядочивании одной переменной и оценке стандартного отклонения разности расстояний между соседними точками по второй переменной	95

4.13. Аддитивная сборка двух континуальных критериев, построенных на упорядочивании одной переменной и оценке стандартного отклонения второй производной расстояний между соседними точками по второй переменной	96
4.14. Корреляционная сцепленность группы критериев, построенных на упорядочивании одной или двух переменных малой выборки	97
Список литературы к главе 4	99
Глава 5. НОВЫЕ КРИТЕРИИ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОЛУЧЕННЫЕ ИСПОЛЬЗОВАНИЕМ ЛИНЕЙНЫХ КВАНТОВАТЕЛЕЙ ДВУХМЕРНОГО ПРОСТРАНСТВА	100
5.1. Критерий, построенный на использовании трехуровневых квантователей (две параллельные линии, наклон вверх слева направо)	100
5.2. Критерий, построенный на использовании трехуровневых квантователей (две параллельные линии, наклон вниз слева направо)	101
5.3. Критерий, построенный на использовании двух трехуровневых квантователей (две параллельные линии, перпендикулярные двум другим линиям)	102
5.4. Критерий, построенный на использовании двух квантователей (две параллельные линии, перпендикулярные двум другим линиям) с использованием операции деления	102
5.5. Непрерывный вариант критерия, построенный на использовании двух квантователей (две параллельные линии, перпендикулярные двум другим линиям) с использованием операции деления	103
5.6. Непрерывный вариант критерия, построенный на использовании двух двухуровневых квантователей с использованием операции деления и накоплении произведения модулей координат обнаруженных точек	104
5.7. Дискретный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой	105
5.8. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой числа обнаруженных точек	106

5.9. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой накопленных произведений модулей координат точек	107
5.10. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой накопленных квадратов разностей координат точек	108
5.11. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой накопленных квадратов суммы координат точек	109
5.12. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой четвертой степени разности координат обнаруженных точек.....	110
5.13. Корреляционная сцепленность критериев, построенных из двух перпендикулярных квантователей, собранных из двух параллельных линий.....	111
Список литературы к главе 5	113
Глава 6. НОВЫЕ НЕЙРОКРИТЕРИИ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОСТРОЕННЫЕ НА СТАТИСТИЧЕСКОМ АНАЛИЗЕ ТОЧЕК, ПОПАДАЮЩИХ В СЕКТОРА ДЕКАРТОВОЙ СИСТЕМЫ КООРДИНАТ.....	114
6.1. Дискретный нейрокритерий, построенный на использовании статистического анализа числа точек малой выборки, попадающих в сектора Декартовой системы координат (2022).....	114
6.2. Непрерывный нейрокритерий, построенный на использовании накопления квадратов координат точек малой выборки, попадающих в сектора Декартовой системы координат	115
6.3. Непрерывный нейрокритерий, построенный на использовании накопления модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, и оценке их размаха	115

6.4. Непрерывный нейрокритерий, построенный на использовании накопления квадратов разности координат точек малой выборки, попадающих в сектора Декартовой системы координат, и оценке размаха их вариаций	116
6.5. Непрерывный нейрокритерий, построенный на использовании накопления квадратов произведений координат точек малой выборки, попадающих в разные сектора Декартовой системы координат, с оценкой размаха их вариаций	117
6.6. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, с оценкой размаха их вариаций	118
6.7. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, с аддитивной сверткой	119
6.8. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, с мультипликативной сверткой	120
6.9. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, с обратной мультипликативной сверткой	121
6.10. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, опирающейся на медианы	122
6.11. Непрерывный нейрокритерий, построенный на учете статистик модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат	123
6.12. Непрерывный нейрокритерий, построенный на учете статистик модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, с регуляризацией оценок медиан.....	124

6.13. Корреляционная сцепленность в группе критериев, построенных на статистическом анализе точек, попадающих в сектора Декартовой системы координат	125
Список литературы к главе 6	127
Глава 7. НОВЫЕ НЕЙРОКРИТЕРИИ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОСТРОЕННЫЕ НА СТАТИСТИЧЕСКОМ АНАЛИЗЕ ТОЧЕК, ВЫДЕЛЕННЫХ ПАРАБОЛИЧЕСКИМИ КВАНТОВАТЕЛЯМИ.....	
7.1. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2	128
7.2. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – y^2	129
7.3. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2 и пары квантователей – y^2	130
7.4. Дискретный нейрокритерий, построенный на использовании первой пары из квадратичных квантователей – x^2 и второй пары квантователей – y^2 через аддитивно-мультипликативное свертывание.....	131
7.5. Дискретный нейрокритерий, построенный на использовании первой пары из квадратичных квантователей – x^2 и второй пары квантователей – y^2 через свертывание вычислением максимума	132
7.6. Дискретный нейрокритерий, построенный на использовании первой пары из квадратичных квантователей – x^2 и второй пары квантователей – y^2 через аддитивно-мультипликативное свертывание с вычислением разности максимума и минимума.....	133
7.7. Непрерывный нейрокритерий, построенный на использовании одного квадратичного квантователя – x^2	133
7.8. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей по первой переменной – x^2	134
7.9. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей по второй переменной – y^2	135
7.10. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей – x^2 и двух квадратичных преобразователей по второй переменной – y^2	136

7.11. Непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей по обоим переменным через их мультипликативную свертку.....	137
7.12. Непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей по обоим переменным с дроблением каждого отклика на две составляющие и вычислении стандартного отклонения	138
7.13. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей при накапливании модулей произведения координат обнаруженных точек и вычислении стандартных отклонений.....	139
7.14. Корреляционная сцепленность группы критериев, построенных на использовании квадратичных квантователей.....	140

Глава 8. НОВЫЕ НЕЙРОКРИТЕРИИ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОСТРОЕННЫЕ НА СТАТИСТИЧЕСКОМ АНАЛИЗЕ ТОЧЕК, ВЫДЕЛЕННЫХ ОКНАМИ ИЗ ОКРУЖНОСТЕЙ

8.1. Дискретный критерий, построенный на анализе числа точек малой выборки, попавших в одну окружность единичного радиуса с центром $\{1, 1\}$	142
8.2. Дискретный критерий, построенный на анализе числа точек малой выборки, попавших в две окружности единичного радиуса	143
8.3. Дискретный критерий, построенный на анализе числа точек малой выборки, попавших в четыре окружности единичного радиуса	145
8.4. Непрерывный критерий, построенный на накоплении радиусов точек по отношению к центру $\{0, 0\}$ для четырех окружностей, расположенных в разных четвертях системы координат	145
8.5. Непрерывный критерий, построенный на исследовании точек малой выборки с накоплением среднего радиуса в каждой окружности и делением накопленных данных.....	147
8.6. Непрерывный критерий, построенный на просмотре точек малой выборки, попавших в четыре окружности единичного радиуса, с накоплением значений модулей произведений координат точек	148
8.7. Непрерывный критерий, построенный на просмотре точек малой выборки, попавших в четыре окружности единичного радиуса, с вычислением среднего значения модуля произведений координат точек и деления данных.....	149

8.8. Непрерывный критерий, построенный на просмотре точек малой выборки, попавших в четыре окружности единичного радиуса, с вычислением среднего значения модуля произведений координат обнаруженных точек и деления данных с комбинированием двух последних критериев	150
8.9. Непрерывный критерий, построенный на просмотре точек малой выборки, попавших в четыре окружности с вычислением среднего значения модуля произведений координат точек для окружностей с центрами $\{1, -1\}, \{-1, 1\}$ и накоплением квадратов радиусов в двух других секторах.....	151
8.10. Непрерывный критерий, построенный на оценках статистик точек малой выборки, попавших в четыре окружности, при свертывании промежуточных данных через умножение и деление.....	152
8.11. Корреляционная сцепленность группы нейрокритериев, построенных на использовании квантователей в виде окружностей, размещенных в разных четвертях Декартовой системы координат	153
Глава 9. КРИТЕРИИ, ПОСТРОЕННЫЕ НА ИСПОЛЬЗОВАНИИ ГИПЕРБОЛИЧЕСКИХ ФУНКЦИЙ РАЗДЕЛЕНИЯ ДАННЫХ МАЛОЙ ВЫБОРКИ.....	
9.1. Дискретный критерий, построенный на гиперболическом квантователе данных, попавших в первую четверть Декартовой системы координат.....	155
9.2. Дискретный критерий, построенный на гиперболическом квантователе данных, попавших во вторую четверть Декартовой системы координат.....	156
9.3. Дискретный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих в первую и во вторую четверти Декартовой системы координат.....	156
9.4. Дискретный критерий, построенный на четырех гиперболических квантователях данных малой выборки, размещенных во всех четвертях Декартовой системы координат, с аддитивным свертыванием.....	157
9.5. Дискретный критерий, построенный на четырех гиперболических квантователях данных малой выборки с мультипликативным свертыванием промежуточных данных.....	158
9.6. Непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки с накоплением радиусов точек и аддитивным свертыванием промежуточных данных.....	159

9.7. Непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки с накоплением произведения модулей координат точек и аддитивным свертыванием промежуточных данных.....	160
9.8. Непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки с накоплением произведения модулей координат точек при мультипликативном свертывании промежуточных данных	160
9.9. Непрерывный критерий, построенный на использовании гиперболической функции разделения малой выборки в четырех четвертях Декартовой системы координат с накоплением смещенных радиусов	161
9.10. Непрерывный критерий, построенный на использовании гиперболической функции разделения малой выборки в четырех четвертях Декартовой системы координат с накоплением смещенных радиусов и мультипликативным свертыванием промежуточных данных	162
9.11. Непрерывный критерий, построенный на использовании гиперболической функции разделения малой выборки во всех четвертях Декартовой системы координат с накоплением смещенных произведений модулей координат.....	164
9.12. Непрерывный критерий, построенный на использовании гиперболической функции разделения малой выборки во всех четвертях Декартовой системы координат с накоплением смещенных произведений модулей координат при аддитивном свертывании данных	164
9.13. Корреляционная сцепленность критериев, использующих гиперболические квантователи.....	165
Глава 10. КРИТЕРИИ, ПОСТРОЕННЫЕ НА ИСПОЛЬЗОВАНИИ АНАЛИЗА ПРАВИЛА БАЙЕСА С УЧЕТОМ ОТСУТСТВИЯ ЧУВСТВИТЕЛЬНОСТИ К ЗНАКУ ВОЗМОЖНОЙ КОРРЕЛЯЦИОННОЙ СЦЕПЛЕННОСТИ ДАННЫХ МАЛОЙ ВЫБОРКИ.....	
10.1. Многомерный вариант корреляционного аналога функционалов Байеса	167
10.2. Корреляционный функционал Байеса с накоплением данных в квадратичном пространстве.....	168

10.3. Корреляционный функционал Байеса, использующий сравнение координат точек с их минимальными значениями	168
10.4. Корреляционный функционал Байеса, использующий сравнение координат точек малой выборки с их максимумами	169
10.5. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их максимумами по переменной x и минимумами по переменной y	170
10.6. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с $\max(y)$ и $\min(x)$	171
10.7. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их крайними значениями при мультипликативном свертывании промежуточных данных	171
10.8. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их предельными значениями при аддитивном свертывании промежуточных данных	172
10.9. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их предельными значениями при делении аддитивных сверток	173
10.10. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их предельными значениями при делении аддитивной свертки на значение мультипликативной свертки	174
10.11. Корреляционная сцепленность группы модуль-функционалов, построенных на использовании разных вариантов многомерного правила Байеса, не способных чувствовать знак корреляционной сцепленности анализируемых данных	174
Список литературы к главе 10	176
Глава 11. КРИТЕРИИ, ПОСТРОЕННЫЕ НА ИСПОЛЬЗОВАНИИ МНОГОМЕРНОГО ПРАВИЛА БАЙЕСА В ВАРИАНТАХ, ЧУВСТВИТЕЛЬНЫХ К ЗНАКУ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ	
11.1. Разностно-корреляционный функционал Байеса, чувствительный к знаку корреляционной сцепленности	177

11.2. Аддитивно-корреляционный функционал Байеса, чувствительный к знаку корреляционной сцепленности.....	178
11.3. Усиление разностно-корреляционных функционалов Байеса, чувствительных к знаку корреляционной сцепленности, за счет перехода к накоплению данных в квадратичном пространстве	179
11.4. Разностно-корреляционный функционал Байеса при замене стандартных отклонений размахом изменения данных.....	179
11.5. Разностно-корреляционный функционал Байеса при замене стандартных отклонений размахом данных и замене математических ожиданий минимумами данных	180
11.6. Аддитивно-корреляционный функционал Байеса при замене стандартных отклонений размахом изменения данных и замене математических ожиданий максимумами данных	181
11.7. Аддитивно-корреляционный функционал Байеса с заменой математических ожиданий экстремальными значениями данных.....	182
11.8. Аддитивно-корреляционный функционал Байеса с заменой математических ожиданий экстремальными значениями данных и заменой стандартных отклонений размахом выборки.....	183
11.9. Разностно-корреляционный функционал Байеса со свертыванием четырех оценок по Колмогорову – Смирнову	184
11.10. Корреляционная сцепленность в группе критериев Байеса, чувствительных к знаку корреляционных связей	185
Глава 12. КРИТЕРИИ, ПОСТРОЕННЫЕ НА ФРАКТАЛЬНОМ ЗАПОЛНЕНИИ ОКРУЖНОСТЯМИ, НАКРЫВАЮЩИМИ ПО ДВЕ И ТРИ ТОЧКИ МАЛОЙ ВЫБОРКИ	
12.1. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через усреднение радиусов окружностей, накрывающих половину точек выборки.....	187
12.2. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку стандартного отклонения радиусов	188
12.3. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку максимального значения радиусов	189

12.4. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку математического ожидания радиусов и деление на стандартное отклонение радиусов.....	189
12.5. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку длины ломаной линии, соединяющей все точки малой выборки	190
12.6. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование скользящей хи-квадрат суммы с четырьмя элементами и поиск максимумов	191
12.7. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование скользящей хи-квадрат суммы с четырьмя элементами и поиск минимумов	193
12.8. Дискретный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки, скользящей хи-квадрат суммы с четырьмя элементами (правый «хвост»).....	194
12.9. Дискретный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки, скользящей хи-квадрат суммы с четырьмя элементами (левый «хвост»).....	195
12.10. Дискретный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки, скользящей хи-квадрат суммы с четырьмя элементами и аддитивной сверткой левого и правого «хвостов».....	196
12.11. Непрерывный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки и скользящей хи-квадрат суммы с четырьмя элементами, а также накопления статистик правого «хвоста»	197
12.12. Непрерывный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки и скользящей хи-квадрат суммы с четырьмя элементами и учетом статистик левого «хвоста».....	198

12.13. Непрерывный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки, скользящей хи-квадрат суммы с четырьмя элементами и аддитивным учетом статистик двух «хвостов»	199
12.14. Корреляционная сцепленность фрактальных функционалов, оценивающих статистики ломаных линий, соединяющих точки малой выборки.....	200
Список литературы к главе 12	201
Глава 13. КРИТЕРИИ, ПОСТРОЕННЫЕ НА ПОПАРНОМ СВЕРТЫВАНИИ ЛИБО СВЕРТЫВАНИИ ТРОЕК ДАННЫХ ПОСЛЕДОВАТЕЛЬНОСТИ ТОЧЕК МАЛОЙ ВЫБОРКИ ПРИ ПРОВЕРКЕ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ.....	203
13.1. Критерий накопления модулей коэффициентов ассоциаций последовательности двух точек малых выборок.....	203
13.2. Критерий накопления модулей коэффициентов ассоциаций последовательности трех точек малых выборок	204
13.3. Критерий накопления модулей коэффициентов ассоциаций соседних точек малых выборок без нормирования данных при промежуточных вычислениях	204
13.4. Критерий объединения модулей коэффициентов коллигации Юла двух соседних точек вдоль малой выборки	205
13.5. Критерий объединения коэффициентов сходства пары точек малых выборок	206
13.6. Критерий объединения коэффициентов ассоциации соседних точек малых выборок (перестановка данных в вычисляемых произведениях).....	207
13.7. Вторая перестановка координат в произведениях при синтезе нового критерия, объединяющего коэффициенты ассоциаций координат пары точек.....	208
13.8. Модификация критерия сборки коэффициентов Юла для проверки гипотезы независимости малых выборок	209
13.9. Вторая модификация критерия сборки коэффициентов Юла для проверки гипотезы независимости малых выборок.....	210
13.10. Модификация критерия, обобщающего коэффициенты сходства для проверки гипотезы независимости малых выборок	211

13.11. Вторая модификация критерия сходства для проверки гипотезы независимости малых выборок через использование еще одной перестановки перемножаемых координат соседних точек.....	212
13.12. Критерий, построенный на использовании вычисления коэффициентов ассоциаций двух соседних точек как процедуры предварительной обработки данных с последующим применением ассоциативного разделения независимых и зависимых данных	213
13.13. Корреляционная сцепленность группы критериев, построенных на свертывании координат двух соседних точек малой выборки	214
13.14. Непрерывная аддитивная сборка трех лучших частных критериев при проверке гипотезы независимости малых выборок в 16 опытов	215
13.15. Дискретная сборка трех лучших частных критериев при проверке гипотезы независимости малых выборок в 16 опытов.....	216
Список литературы к главе 13	217

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

Глава первая

1. r – критерий Эджуорта – Эдлтона – Пирсона (конец XIX в.) для проверки гипотезы независимости [1] (с. 37);
2. K_e – критерий Кенуя (1973) [2] для проверки гипотезы независимости (с. 38);
3. N_e – критерий Нельсона (1983) [3] для проверки гипотезы независимости (с. 39);
4. BKR – критерий Блума – Кифера – Розенбладтта (1961) [4] для проверки гипотезы независимости (с. 39);
5. CS – критерий Кокса – Стюарта (1955) [5] для проверки гипотезы независимости (с. 40);
6. AL – критерий Аббе – Линника (1965) [6] для проверки гипотезы независимости (с. 41);
7. OT – угловой критерий Олмстеда – Тьюки (1947) [8] для проверки гипотезы независимости (с. 42);
8. SO – критерий Сендвика – Олсона (1982) [8] (с. 43);
9. Sh – критерий Шахани (1973) [2] для проверки гипотезы независимости (с. 43);
10. K_n – критерий Кенделла (1971) [9] для проверки гипотезы независимости (с. 44);
11. Sp – критерий Спирмена (1971) [10] для проверки гипотезы независимости (с. 45);
12. SE – критерий Шведа – Эйзенхарта (1971) [2] (с. 45);
13. RR – сериальный критерий Рамачандра – Ранганатана (с. 46);
14. $VdrV$ – критерий Ван дер Вандена (1967) [11] (с. 47);
15. SF – критерий Шукени – Флори (1973) [12] (с. 48);
16. Br – ранговый критерий Бартелса (1982) [13] для проверки гипотезы независимости (с. 49);
17. Sa – сериально-ранговый критерий Шахнесси (1981) [14] (с. 49);
18. EI – критерий Эландта (1962) [15] (с. 50);
19. MW – критерий, построенный на статистике Манна – Уитни (1947) [16] (с. 51);
20. $sg14$ – критерий, построенный Казаковичусом на оценке автокорреляции при циклическом сдвиге (1988) [17] (с. 51);
21. $sg14d$ – критерий, построенный Кноке на дифференциальном варианте оценки автокорреляции при циклическом сдвиге (1975) [18] (с. 52).

Глава вторая

22. rR – дискретный эквивалент корреляционного критерия Пирсона – Эджуорта – Эдлтона (2017) [22] (с. 61);

23. eR – дискретный корреляционный критерий с двумя эллиптическими квантователями (2019) [23] (с. 62);
24. $r1R$ – первый непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями (2021) [24] (с. 63);
25. $r2R_min$ – второй непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями, а математических ожиданий на минимальные значения (2021) [24] (с. 63);
26. $r3R_max$ – третий непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями, а математических ожиданий на максимальные значения (2021) [24] данных малой выборки (с. 64);
27. $r4R_max_min$ – четвертый непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями, а математических ожиданий $E(x)$, $E(y)$ на максимальное значение $max(x)$ и минимальное значение $min(y)$ (2021) [24] (с. 65);
28. $r5R_min_max$ – пятый непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями, а математического ожидания $E(x)$ на минимальное значение $min(x)$, а второго математического ожидания $E(y)$ на максимальное значение $max(y)$ (2021) [24] (с. 66);
29. $r6R_E_min$ – шестой непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями, а математического ожидания $E(y)$ на минимальное значение $min(y)$ при сохранении другого математического ожидания $E(x) \leftarrow mean(x)$ как у исходного критерия (с. 66);
30. $r7R_min_E$ – седьмой непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями, а математического ожидания $E(x)$ на минимальное значение $min(x)$ при сохранении другого математического ожидания $E(y) \leftarrow mean(y)$ как у классического критерия (с. 67);
31. $r8R_E_max$ – восьмой непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями, а математического ожидания $E(y)$ на максимальное значение $max(y)$ при сохранении другого математического ожидания $E(x)$ как у классического критерия (с. 68);
32. $r9R_max_E$ – девятый непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями, а математического ожидания $E(x)$ на максимальное значение $max(x)$ при сохранении другого математического ожидания $E(y) \leftarrow mean(y)$ как у классического критерия (с. 69);

33. $r_{10R_E\sigma}$ – десятый непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный заменой в знаменателе произведения стандартных отклонений на квадрат суммы стандартных отклонений (с. 69).

Глава третья

34. OT_KS – гибридный критерий Олмстеда – Тьюки с классическим критерием Колмогорова – Смирнова (с. 72);
35. SO_{\max} – второй вариант критерия Сендвиги – Олстеда, полученный заменой в оригинале вычислительной формулы минимальных значений на их максимальные значения (с. 73);
36. Sp_2 – второй вариант критерия Спирмена, полученный отменой использования преобразования Кенделла через суммирование модуля разности ранга отсчета и его номера по первичному упорядочиванию (с. 73);
37. Sp_3 – третий вариант критерия Спирмена, полученный накоплением модулей разности ранга отсчета и его номера по первичному упорядочиванию, умноженного на преобразования Кенделла (с. 74);
38. $VdrV_Br$ – мультипликативный гибрид критериев Ван дер Вандена и Бартлеса (с. 74);
39. $SOESO$ – аддитивная сборка двух вариантов критериев Сендвиги – Олстеда с разными смещениями математических ожиданий (с. 76);
40. $mSOESO$ – мультипликативная сборка двух вариантов критериев Сендвиги – Олстеда с разными смещениями математических ожиданий (с. 76);
41. Kn_2 – модификация критерия Кенделла за счет накопления квадратов рангов данных (с. 77);
42. aSO_Kn_2 – аддитивная сборка критерия Сендвиги – Олстеда с модифицированным критерием Кенделла (с. 78);
43. $VdrV_E1$ – аддитивная сборка критерия Ван дер Вандена и критерия Эландта (с. 79);
44. aSh_Kn_2 – аддитивная сборка критерия Шахани и модифицированного критерия Кенделла (с. 80);
45. aSh_sg14d – аддитивная сборка критерия Шахани и корреляционного критерия Кноке (с. 80);
46. aKn_2Kn_2 – аддитивная сборка двух вариантов модифицированного критерия Кенделла (с. 81).

Глава четвертая

47. $osrRx$ – континуальный критерий, построенный на упорядочивании переменной – x малой выборки со сглаживанием второй переменной и вычислением ее стандартного отклонения (с. 85);
48. $osrRy$ – континуальный критерий, построенный на упорядочивании переменной – y малой выборки со сглаживанием второй переменной и вычислением ее стандартного отклонения (с. 86);

49. $ostR_{xy}$ – континуальный критерий, построенный на упорядочивании данных малой выборки со сглаживанием второй оставшейся переменной и вычислением суммы стандартных отклонений (с. 87);
50. $odrR_x$ – континуальный критерий, построенный на упорядочивании переменной – x малой выборки с дифференцированием оставшейся случайной переменной (с. 87);
51. $odrR_y$ – континуальный критерий, построенный на упорядочивании переменной – y малой выборки с дифференцированием оставшейся случайной переменной (с. 87);
52. $odrR_{xy}$ – континуальный критерий, построенный на упорядочивании данных малой выборки с дифференцированием второй оставшейся переменной (с. 88);
53. rRE_x – континуальный критерий, построенный на упорядочивании данных малой выборки с суммированием квадратов отсчетов по второй переменной (с. 89);
54. rRE_{dx} – континуальный критерий, построенный на упорядочивании данных малой выборки с суммированием квадратов разностей отсчетов соседних точек по второй переменной (с. 90);
55. rRE_{dxy} – аддитивная сборка двух континуальных критериев, построенных на упорядочивании данных малой выборки с суммированием квадратов разностей отсчетов соседних точек по второй переменной (с. 91);
56. rR_{xx} – континуальный критерий, построенный на упорядочивании данных малой выборки по координате y с суммированием произведений отсчетов соседних точек по второй переменной x (с. 92);
57. rR_{yy} – континуальный критерий, построенный на упорядочивании данных малой выборки по координате x с суммированием произведений отсчетов соседних точек по второй переменной y (с. 92);
58. $rR_{xx\sigma}$ – континуальный критерий, построенный на упорядочивании данных малой выборки по координате y с вычислением стандартного отклонения произведений отсчетов соседних точек по второй переменной x (с. 93);
59. $rR_{yy\sigma}$ – континуальный критерий, построенный на упорядочивании данных малой выборки по координате x с вычислением стандартного отклонения произведений отсчетов соседних точек по второй переменной y ;
60. rR_{xxM} – континуальный критерий, построенный на упорядочивании данных малой выборки по координате y с суммированием произведений отсчетов соседних точек по второй переменной x (с. 94);
61. rR_{yyM} – континуальный критерий, построенный на упорядочивании данных малой выборки по координате x с суммированием произведений отсчетов соседних точек по второй переменной y ;

62. $rR_{xx\sigma}$ – континуальный критерий, построенный на упорядочивании данных малой выборки по координате y с оцениванием стандартного отклонения произведений отсчетов соседних точек по второй переменной x ;
63. $rR_{yy\sigma}$ – континуальный критерий, построенный на упорядочивании данных малой выборки по координате x с оцениванием стандартного отклонения произведений отсчетов соседних точек по второй переменной y ;
64. $rR_{dx\sigma}$ – континуальный критерий, построенный на упорядочивании данных малой выборки по координате y с оцениванием стандартного отклонения разности отсчетов соседних точек по второй переменной x ;
65. $rR_{dy\sigma}$ – континуальный критерий, построенный на упорядочивании данных малой выборки по координате x с оцениванием стандартного отклонения разности отсчетов соседних точек по второй переменной y ;
66. $rR_{dxdy\sigma}$ – аддитивная сборка двух континуальных критериев, построенных на упорядочивании данных малой выборки по одной координате с оцениванием стандартного отклонения разности отсчетов соседних точек по второй переменной, оставшейся случайной (с. 95);
67. $rR_{d^2x\sigma}$ – континуальный критерий, построенный на упорядочивании данных малой выборки по координате y с оцениванием стандартного отклонения второй производной разностей отсчетов соседних трех точек по второй переменной x ;
68. $rR_{d^2y\sigma}$ – континуальный критерий, построенный на упорядочивании данных малой выборки по координате x с оцениванием стандартного отклонения второй производной разностей отсчетов соседних трех точек по второй переменной y ;
69. $rR_{d^2xd^2y\sigma}$ – аддитивная сборка двух континуальных критериев, построенных на упорядочивании данных малой выборки по одной координате с оцениванием стандартного отклонения второй производной разностей отсчетов соседних трех точек по второй переменной, оставшейся случайной (с. 96).

Глава пятая

70. rRL_p – корреляционный критерий, использующий трехуровневый квантователь (две параллельные линии, наклон слева направо) (2020) [28] (с. 100);
71. $rRpL$ – корреляционный критерий, использующий трехуровневый квантователь (две параллельные линии, наклон справа налево) (2020) [28] (с. 101);
72. rRL_{ppL} – корреляционный критерий, использующий два перпендикулярных квантователя, состоящих из двух параллельных линий (с. 102);

73. $rRLppL2$ – корреляционный критерий, использующий два перпендикулярных квантователя, состоящих из двух параллельных линий с использованием операции деления результатов двух квантователей (с. 102);
74. $rRLppL2n$ – непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий с использованием операции деления результатов двух квантователей (с. 103);
75. $rRLppL3n$ – непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий с использованием операции деления результатов двух квантователей и накоплением произведением модулей координат обнаруженных точек (с. 104);
76. $rRLppL4$ – дискретный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий с аддитивной сборкой (с. 105);
77. $rRLppL4n$ – непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий с аддитивной сборкой квадратов координат обнаруженных точек (с. 106);
78. $rRLppL5n$ – непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий с аддитивной сборкой произведения модулей координат обнаруженных точек (с. 107);
79. $rRLppL6n$ – непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий с аддитивной сборкой квадратов разности координат обнаруженных точек (с. 108);
80. $rRLppL7n$ – непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий с аддитивной сборкой квадратов сумм координат обнаруженных точек (с. 109);
81. $rRLppL8n$ – непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий с аддитивной сборкой четвертой степени разности координат обнаруженных точек (с. 110).

Глава шестая

82. $D4rRMm$ – дискретный нейрокритерий, построенный на использовании статистического анализа числа точек малой выборки, попадающих в сектора Декартовой системы координат, и оценкой размаха их вариаций (с. 114);

83. D4rRMmn – непрерывный нейрокритерий, построенный на использовании накопления квадратов координат точек, попадающих в разные сектора Декартовой системы координат, и оценкой размаха их вариаций (с. 115);
84. D4rR2Mmn – непрерывный нейрокритерий, построенный на использовании накопления модулей координат точек, попадающих в разные сектора Декартовой системы координат, и оценкой размаха их вариаций (с. 115);
85. D4rR2pMmn – непрерывный нейрокритерий, построенный на использовании накопления квадратов разностей координат точек, попадающих в разные сектора Декартовой системы координат, и оценкой размаха их вариаций (с. 116);
86. D4rR2прMmn – непрерывный нейрокритерий, построенный на использовании накопления квадратов произведения координат точек, попадающих в разные сектора Декартовой системы координат, и оценкой размаха их вариаций (с. 117);
87. D4rRмпрMmn – непрерывный нейрокритерий, построенный на использовании накопления модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, и оценкой размаха их вариаций (с. 118);
88. D4rRмпран – непрерывный нейрокритерий, построенный на использовании накопления модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, и аддитивной свертки (с. 119);
89. D4rRмпр_ms – непрерывный нейрокритерий, построенный на использовании накопления модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, и мультипликативной их свертки (с. 120);
90. D4rRмпр_sm – непрерывный нейрокритерий, построенный на использовании накопления модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, и обратной мультипликативной их свертки (с. 121);
91. D4rRмпр_Mm – непрерывный нейрокритерий, построенный на использовании накопления модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, с центром в медианах малой выборки (с. 122);
92. D4rRмпр_max – непрерывный нейрокритерий, построенный на учете статистик модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, с центром в медианах малой выборки (с. 123);
93. D4rRмпрS_max – непрерывный нейрокритерий, построенный на учете статистик модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, с учетом стабилизации оценок медиан малой выборки по двум крайним отсчетам правого и левого «хвостов» малой выборки (с. 124).

Глава седьмая

94. Prx2 – дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2 (с. 128);
95. Pry2 – дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – y^2 (с. 129);
96. Prxy2 – дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – y^2 и пары квадратичных квантователей по другой переменной – x^2 (с. 130);
97. Prxy2ам – дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – y^2 и пары квадратичных квантователей по другой переменной – x^2 через аддитивно-мультипликативное свертывание (с. 131);
98. Prxy2амМ – дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – y^2 и пары квадратичных квантователей по другой переменной – x^2 через аддитивно-мультипликативное свертывание с вычислением максимума промежуточных данных (с. 132);
99. Prxy2амМ – дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – y^2 и пары квадратичных квантователей по другой переменной – x^2 через аддитивно-мультипликативное свертывание с вычислением разности максимума и минимума промежуточных данных (с. 133);
100. Prx21 – непрерывный нейрокритерий, построенный на использовании одного квадратичного квантователя – x^2 (с. 133);
101. Prx22 – непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей – x^2 (с. 134);
102. Pry22 – непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей – y^2 (с. 135);
103. Prxy22 – непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей (два по каждой переменной) (с. 136);
104. Prxy22д – непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей (два по каждой переменной) через мультипликативную свертку данных делением (с. 137);
105. Prxy22σ – непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей (два по каждой переменной) с дроблением на два дополнительных состояния (с. 138);
106. Pry2ум – непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей с накоплением модулей произведений координат обнаруженных точек (с. 139);
107. Prxy2ум_max – непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей с накоплением модулей произведений координат обнаруженных точек и вычислении максимума.

Глава восьмая

108. ΦI – дискретный критерий, построенный на подсчете точек, попавших в одну окружность единичного радиуса с координатами центра $\{1, 1\}$ (с. 142);
109. ΦII – дискретный критерий, построенный на подсчете точек, попавших в одну окружность единичного радиуса с координатами центра $\{1, -1\}$ (с. 143);
110. ΦI_{II} – дискретный критерий, построенный на подсчете точек, попавших в две окружности единичного радиуса с координатами центров $\{1, 1\}, \{-1, 1\}$ (с. 143);
111. ΦI_{III} – дискретный критерий, построенный на подсчете точек, попавших в две окружности единичного радиуса с координатами центров $\{1, 1\}, \{-1, -1\}$ (с. 144);
112. ΦI_{IV} – дискретный критерий, построенный на подсчете точек, попавших в четыре окружности единичного радиуса с координатами центров $\{1, 1\}, \{-1, 1\}, \{-1, -1\}, \{1, -1\}$ (с. 145);
113. ΦRI_{IV} – непрерывный критерий, построенный на подсчете точек, попавших в четыре окружности единичного радиуса с координатами центров $\{1, 1\}, \{-1, 1\}, \{-1, -1\}, \{1, -1\}$, и вычислении средних радиусов по отношению центра выборки $\{0, 0\}$ (с. 145);
114. $\Phi RдI_{IV}$ – непрерывный критерий, построенный на подсчете точек, попавших в четыре окружности единичного радиуса с координатами центров $\{1, 1\}, \{-1, 1\}, \{-1, -1\}, \{1, -1\}$, и вычислении средних радиусов по отношению центра выборки $\{0, 0\}$ с делением (с. 147);
115. ΦSI_{IV} – непрерывный критерий, построенный на подсчете точек, попавших в четыре окружности единичного радиуса с координатами центров $\{1, 1\}, \{-1, 1\}, \{-1, -1\}, \{1, -1\}$, и вычислении средних модулей произведений координат точек (с. 148);
116. $\Phi SдI_{IV}$ – непрерывный критерий, построенный на подсчете точек, попавших в четыре окружности единичного радиуса с координатами центров $\{1, 1\}, \{-1, 1\}, \{-1, -1\}, \{1, -1\}$, и вычислении средних модулей произведений координат точек с делением промежуточных данных (с. 149);
117. $\Phi R SдI_{IV}$ – непрерывный критерий, построенный на подсчете точек, попавших в четыре окружности единичного радиуса с координатами центров $\{1, 1\}, \{-1, -1\}$, для вычисления средних модулей произведений координат точек и для окружностей с центрами $\{-1, 1\}, \{1, -1\}$ выполняется накопление квадратов радиусов (с. 150);
118. $\Phi S RдI_{IV}$ – непрерывный критерий, построенный на подсчете точек, попавших в четыре окружности единичного радиуса с координатами центров $\{-1, 1\}, \{1, -1\}$, для вычисления средних модулей произведений координат точек и для окружностей с центрами $\{1, 1\}, \{-1, -1\}$ выполняется накопление квадратов радиусов (перестановка типов вычислений при накапливании между парами окружностей) (с. 151);

119. $\Phi SR_{удI_IV}$ – непрерывный критерий, построенный на подсчете точек, попавших в четыре окружности единичного радиуса с координатами центров $\{-1, 1\}, \{1, -1\}$, для вычисления средних модулей произведений координат точек и для окружностей с центрами $\{1, 1\}, \{-1, -1\}$ выполняется накопление квадратов с произведением накопленных данных (с. 152).

Глава девятая

120. GrI – дискретный критерий, построенный на гиперболическом квантователе данных, попавших в первую четверть Декартовой системы координат (с. 155);

121. $GrII$ – дискретный критерий, построенный на гиперболическом квантователе данных малой выборки, попавших во вторую четверть Декартовой системы координат (с. 156);

122. GrI_II – дискретный критерий, построенный на двух гиперболических квантователях данных малой выборки, попавших в первую и во вторую четверти Декартовой системы координат (с. 156);

123. $GrI_II_III_IV$ – дискретный критерий, построенный на четырех гиперболических квантователях данных малой выборки, размещенных во всех четвертях Декартовой системы координат, с аддитивной сверткой промежуточных данных (с. 157);

124. $GrI_II_III_IVД$ – дискретный критерий, построенный на четырех гиперболических квантователях данных малой выборки, размещенных во всех четвертях Декартовой системы координат, с мультипликативной сверткой промежуточных данных (с. 158);

125. $GrI_II_III_IVR$ – непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки, размещенных во всех четвертях Декартовой системы координат, с накапливанием радиусов и аддитивной сверткой промежуточных данных (с. 159);

126. $GrI_II_III_IVG$ – непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки, размещенных во всех четвертях Декартовой системы координат, с накапливанием произведений модулей координат точек и аддитивной сверткой промежуточных данных (с. 160);

127. $GrI_II_III_IVGD$ – непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки, размещенных во всех четвертях Декартовой системы координат, с накапливанием произведений модулей координат точек и мультипликативной сверткой промежуточных данных через их деление (с. 160);

128. $GrI_II_III_IVR_{см}$ – непрерывный гиперболический критерий, построенный на выявлении точек во всех секторах Декартовой системы координат с накоплением радиусов со смещенным центром оценок (с. 161);

129. $GpI_II_III_IVR_{смД}$ – непрерывный гиперболический критерий, построенный на обнаружении точек во всех секторах Декартовой системы координат с накоплением радиусов со смещенным центром оценок с последующим делением данных (с. 162);
130. $GpI_II_III_IVG_{см}$ – непрерывный гиперболический критерий, построенный на обнаружении точек во всех секторах Декартовой системы координат с накоплением произведений относительно смещенных центров (с. 163);
131. $GpI_II_III_IVG_{смД}$ – непрерывный гиперболический критерий, построенный на обнаружении точек во всех секторах Декартовой системы координат с накоплением произведений координат при смещенном центре оценок с последующим делением промежуточных данных (с. 164).

Глава десятая

132. LS – корреляционный критерий проверки гипотезы независимости Ложникова – Сулавко (Омск, 2017) [32, 33] (с. 167);
133. LS2 – корреляционный критерий проверки гипотезы независимости Байеса, усиленный переходом к накоплению данных в квадратичном пространстве (с. 168);
134. mLs – корреляционный критерий проверки гипотезы независимости многомерным правилом Байеса, усиленный переходом к сравнению данных выборки с ее минимальными значениями (с. 168);
135. MLs – корреляционный критерий проверки гипотезы независимости многомерным правилом Байеса, усиленный переходом к сравнению данных выборки с ее максимумами (с. 169);
136. $MxBLS_{my}$ – корреляционный критерий проверки гипотезы независимости многомерным правилом Байеса, усиленный сравнением данных выборки с $\max(x)$ и $\min(y)$ (с. 170);
137. $mxBLS_{My}$ – корреляционный критерий проверки гипотезы независимости многомерным правилом Байеса, усиленный сравнением данных выборки с $\max(y)$ и $\min(x)$ (с. 171);
138. $MxmBLS$ – корреляционный критерий проверки гипотезы независимости многомерным правилом Байеса, усиленный сравнением данных выборки с их экстремальными значениями и мультипликативным свертыванием промежуточных данных (с. 171);
139. $M\Sigma m_дBLS4$ – корреляционный критерий проверки гипотезы независимости многомерным правилом Байеса, усиленный сравнением данных выборки с их экстремальными значениями и делением аддитивных сверток (с. 172);
140. $M\Sigma xm_дBLS4$ – корреляционный критерий проверки гипотезы независимости многомерным правилом Байеса, усиленный сравнением данных выборки с их экстремальными значениями и делением аддитивной свертки на мультипликативную свертку (с. 173).

Глава одиннадцатая

141. Σr – разностно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных (с. 177);
142. $\Sigma r\pi$ – аддитивно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных (с. 178);
143. $\Sigma r2$ – разностно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных, с накоплением в квадратичном пространстве (с. 179);
144. $\Sigma rEmM$ – разностно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных, с заменой стандартных отклонений размахами (с. 180);
145. ΣrMm – разностно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных, с заменой математических ожиданий минимумами и стандартных отклонений размахами (с. 181);
146. $\pi\Sigma r\mu Mx$ – аддитивно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных, с заменой математических ожиданий экстремумами $\min(y)$, $\max(x)$ и заменой стандартных отклонений размахами (с. 181);
147. $\pi\Sigma r\mu My$ – аддитивно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных, с заменой математических ожиданий экстремумами $\min(x)$, $\max(y)$ и заменой стандартных отклонений размахами (с. 182);
148. $\Sigma r\mu Mx\mu y2$ – аддитивно-мультипликативно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных, с заменой математических ожиданий экстремумами $\min(x)$, $\max(x)$, $\min(y)$, $\max(y)$ и заменой стандартных отклонений размахами (с. 183);
149. $\Sigma rKS4$ – разностно-корреляционный критерий Байеса, чувствительный к знаку корреляционной сцепленности данных, со свертыванием четырех промежуточных параметров по Колмогорову – Смирнову (с. 184).

Глава двенадцатая

150. ΦrE_R5 – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через усреднение радиусов окружностей, накрывающих пять точек малой выборки (с. 187);
151. $\Phi r\sigma_R$ – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку стандартного отклонения всех радиусов (с. 188);
152. Φr_MR – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку максимального значения радиусов (с. 189);

153. $\text{FrE_R5d}\sigma$ – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку математического ожидания радиусов с его делением на стандартное отклонение радиусов (с. 189);
154. Fr_Dr – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку средней длины колена ломаной линии, соединяющей все точки выборки (с. 190);
155. Fr10_M12x4 – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения псевдовыборки случайными перестановками точек (с. 191);
156. Fr10_m12x4 – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения псевдовыборки случайными перестановками точек (с. 193);
157. dFr10_M12x4 – дискретный фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения псевдовыборки случайными перестановками точек и оценки статистик правого «хвоста» хи-квадрат распределения с восемью степенями свободы (с. 194);
158. dFr10_m12x4 – дискретный фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения псевдовыборки случайными перестановками точек и оценки статистик левого «хвоста» хи-квадрат распределения с восемью степенями свободы (с. 195);
159. Fr10_mM12x4 – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения псевдовыборки случайными перестановками точек и оценки аддитивного объединения статистик левого и правого «хвостов» хи-квадрат распределения с четырьмя степенями свободы (с. 196);
160. Fr10_dM12x4 – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения псевдовыборки случайными перестановками точек и оценки мультипликативного объединения статистик правого «хвоста» с накоплением 157 хи-квадрат отсчетов (с. 197);
161. Fr10_dm12x4 – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения псевдовыборки случайными перестановками точек и оценки статистик левого «хвоста» с накоплением 157 хи-квадрат отсчетов (с. 198);

162. Φr_{10_dM12x4} – фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения псевдовыборки случайными перестановками точек и оценки статистик аддитивного объединения левого и правого «хвостов» с накоплением 157 хи-квадрат отсчетов (с. 199).

Глава тринадцатая

163. ac_2 – критерий, построенный на объединении ассоциаций двух соседних точек малой выборки (с. 203);
164. ac_3 – критерий, построенный на объединении ассоциаций трех соседних точек малой выборки (с. 204);
165. ac_{2a} – критерий, построенный на объединении ассоциаций двух соседних точек малой выборки без нормирования их разностей (с. 204);
166. $co_{\text{Ю}2}$ – критерий, построенный на объединении коллигаций Юла рядом стоящих точек малой выборки (с. 205);
167. cx_2 – критерий, построенный на объединении коэффициентов сходства пары точек малой выборки (с. 206);
168. ac_{2X} – критерий, построенный на объединении парных коэффициентов ассоциаций рядом стоящих точек малой выборки при перестановке их координат (с. 207);
169. ac_{2XX} – критерий, построенный на объединении парных коэффициентов ассоциаций рядом стоящих точек малой выборки при перестановке их координат (с. 208);
170. $co_{\text{Ю}2X}$ – критерий, построенный на объединении коллигаций Юла рядом стоящих точек малой выборки через их перестановку (с. 209);
171. $co_{\text{Ю}2XX}$ – критерий, построенный на объединении коллигаций Юла рядом стоящих точек малой выборки через их еще одну (вторую) перестановку (с. 210);
172. cx_{2X} – критерий, построенный на объединении коэффициентов сходства пары точек малой выборки путем перестановок перемножаемых координат (с. 211);
173. cx_{2XX} – критерий, построенный на объединении коэффициентов сходства пары точек малой выборки путем еще одной (второй) перестановки перемножаемых координат (с. 212);
174. $ac_{2\Pi}$ – критерий, построенный на вычислении коэффициентов ассоциаций двух соседних точек малой выборки как функции предварительной обработки и последующего применения этой же функции при накоплении ее модулей для разделения независимых и зависимых данных (с. 213);
175. a_{XXX} – критерий, построенный на аддитивном свертывании трех наиболее эффективных нейрокритериев $\{ac_2, ax_{2X}, ax_{2XX}\}$ (с. 215);
176. XXX – критерий, построенный на дискретном свертывании трех наиболее эффективных нейрокритериев $\{ac_2, ax_{2X}, ax_{2XX}\}$ путем обнаружения и корректирования ошибок в коде нейросети (с. 216).

ВВЕДЕНИЕ

В 1900 г. Пирсон создал свой классический хи-квадрат критерий. Этот несомненный успех привел к тому, что усилиями мировой физико-математической общественности за XX в. было создано примерно 200 статистических критериев. Естественно, что столь значительное число статистических критериев создано с учетом многообразия реальных условий и многообразия решаемых статистических задач, в частности были созданы примерно 20 классических статистических критериев для проверки гипотезы независимости.

В начале XXI в. возник второй всплеск интереса к проблеме проверке гипотезы независимости малых выборок. Это связано с развитием средств автоматического обучения больших сетей искусственных нейронов на малых выборках и появлением первого в мировой практике национального стандарта ГОСТ Р 52633.5–2011¹. Необходимость стандартизации автоматического обучения обусловлена прежде всего жесткими требованиями регуляторов рынка средств информационной безопасности. В России таковыми являются «ФСТЭК России» в части механизмов защиты и «ФСБ России» в части использования в них криптографии. Обычный потребитель может быть уверен в безопасности того или иного средства защиты, если оно сертифицировано испытательными лабораториями первого и второго регуляторов.

В свою очередь испытаны и сертифицированы могут быть только механизмы с прозрачной математикой. Для сетей искусственных нейронов с обогащением данных в линейном пространстве появились семь национальных стандартов, касающихся условий обеспечения их безопасности. При этом всем, кто разбирается в математике, изначально было понятно, что только нейронов с обогащением данных в линейном пространстве недостаточно. Необходимо стандартизировать применение более сложных нейронов, например, выполняющих обогащение данных в квадратичных пространствах. На момент написания этого справочника первый проект национального стандарта² России по автоматическому обучению квадратичных нейронов прошел публичное обсуждение в техническом комитете № 164 «Искусственный интеллект». Как результат, должны появиться последующие уточняющие документы; методические рекомендации, технические спецификации и стандарты более низкого уровня.

¹ ГОСТ Р 52633.5–2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа».

² ГОСТ Р xxxx–2023 «Искусственный интеллект. Нейросетевые алгоритмы в защищённом исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации».

Данный справочник касается вопроса поиска данных со слабой корреляционной зависимостью при обучении (настройке) классических квадратичных нейронов Махалонобиса на малых выборках. В рамках линейной алгебры задача обучения (настройки) квадратичных нейронов Махалонобиса считается плохо обусловленной, так как связана с обращением корреляционной матрицы. Однако эта задача становится лучше обусловленной, если отказаться от обращения корреляционных матриц.

Пользуясь тем, что большие сети искусственных нейронов работают с длинными векторами входных данных, мы можем формировать данные с заранее заданным значением корреляционной сцепленности. В том числе мы можем выбирать слабо коррелированные данные. В этом случае нужно уметь находить слабо коррелированные входные данные, опираясь на малые обучающие выборки в 16 опытов.

Следует отметить, что статистические критерии прошлого века, начиная с хи-квадрат критерия Пирсона, ориентированы на ручное применение. Обязательно с тем или иным критерием создавались специальные таблицы доверительной вероятности. Математик после выполненных вычислений должен был воспользоваться заранее вычисленными таблицами при оценке итоговой доверительной вероятности. Фактически все таблицы существующих критериев есть не что иное, как таблицы компромиссов между ошибками первого и второго рода. Всегда можно свернуть таблицы, зафиксировав соотношения доверительной вероятности и размер обрабатываемой выборки. При этом утрачивается универсальность ручного применения того или иного статистического критерия. Однако это только поверхностный взгляд. На самом деле нам не нужна универсальность в обмен на необходимость в ручных вычислениях. Ручные вычисления всегда крайне дороги и малоэффективны. В 1900 г. ручным вычислениям не было альтернативы. Однако уже в середине XX в. появилась возможность переложить вычисления на ламповые ЭВМ. Во второй половине прошлого века компьютеры бурно развивались, и к началу этого века дефицит вычислительных ресурсов исчез. Сегодня мы можем выполнять очень сложные вычисления. При этом мы можем поменять простоту вычислений с низким доверием на более сложные вычисления, но имеющие более высокий уровень доверия.

К сожалению, вся история предшествующих многовековых ручных вычислений негласно приветствовала вычислительную простоту. Исследователи, будучи сами вычислителями, интуитивно стремились выбирать вычислительную простоту (экономя свои затраты в будущих вычислениях). Сегодня этот подход утратил свою актуальность. Сегодня можно усложнять вычисления в сотни и тысячи раз, нет смысла экономить вычислительные ресурсы компьютеров, когда они есть в любом количестве. Экономить нужно, по-прежнему, ресурсы людей, теперь ресурсы людей-программистов.

В этом контексте и нужен данный справочник. В нем приводятся короткие программы, написанные на языке MathCAD, воспроизводящие работу искусственных нейронов, эквивалентных классическим статистическим критериям. Формально эту работу может выполнить любой человек, обладающий минимальными знаниями по статистике и по программированию. Более сложная работа касается синтеза новых статистических критериев, построенных уже в этом веке. Данные об этих новых статистических критериях пока разбросаны только по специальным журналам нейросетевой биометрической аутентификации, которые не просматривают специалисты по математической статистике, медицине, экономике, биологии, зоологии и ботанике.

Еще одним очень важным моментом является то, что преобразования, используемые математиками при конструировании того или иного статистического критерия, опираются на знание об асимптотическом поведении различных математических конструкций. В этом отношении читать статистические справочники, написанные математиками для математиков, крайне сложно. В части справочников приводятся конкретные примеры расчетов. Следуя этой традиции в представленном читателю варианте, справочник содержит множество коротких программ в несколько строк кода, позволяющих увидеть распределения выходных состояний того или иного критерия. В этом отношении короткие программы справочника играют двойную роль. С одной стороны, они позволяют упростить задачу программистам, с другой стороны, они позволяют любому человеку лично проверить все утверждения, приведенные в справочнике.

ФУНКЦИИ ЯЗЫКА MathCAD, ИСПОЛЬЗОВАННЫЕ ПРИ ПРОГРАММИРОВАНИИ В ДАННОМ СПРАВОЧНИКЕ

$x \leftarrow \text{mom}(16, 0, 1)$	возвращает вектор из 16 случайных значений с нормальным законом распределения с нулевым математическим ожиданием и единичным стандартным отклонением;
$x \leftarrow \text{runif}(16, -3, 3)$	возвращает вектор из 16 случайных значений с равномерным законом распределения значения в интервале от -3 до 3 ;
$\text{md}(0.01)$	возвращает одно случайное число с равномерным распределением в интервале от 0 до $0,01$;
$x \leftarrow \text{sort}(x)$	возвращает упорядоченные значения вектора x по их возрастанию;
$m \leftarrow \text{mean}(x)$	возвращает значение математического ожидания;
$\sigma \leftarrow \text{stdev}(x)$	возвращает значение стандартного отклонения;
$\text{pnorm}(x_1, m, \sigma)$	возвращает значение вероятности появления значения x_1 для нормального закона распределения с математическим ожиданием m и стандартным отклонением σ ;
$\text{dnorm}(x_1, m, \sigma)$	возвращает значение плотности вероятности появления значения x_1 для нормального закона распределения с математическим ожиданием m и стандартным отклонением σ ;
$\text{corr}(x0, x1)$	возвращает значение коэффициента корреляции;
$\text{hist}(\text{int}, h)$	возвращает гистограмму распределения значений вектора h , по числу попаданий в интервалы между возрастающими значениями вектора int ;
$\text{rsort}(A, n)$	возвращает массив, образованный перестановкой столбцов матрицы A в порядке возрастания значений в строке n .

КЛАССИЧЕСКИЕ КРИТЕРИИ, СОЗДАННЫЕ В ПРОШЛОМ ВЕКЕ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ

1.1. Корреляционный критерий Эджуорта – Эдлтона – Пирсона (конец XIX в.)

Интерес к экономичному вычислению коэффициентов парной корреляции существовал давно, однако только в конце XIX в. английские статистики стали широко применять классическую формулу [<https://ru.wikipedia.org/wiki/Корреляция>]:

$$r(x, y) = \frac{1}{n} \sum_{i=1}^n \frac{(E(x) - x_i) \cdot (E(y) - y_i)}{\sigma(x) \cdot \sigma(y)}. \quad (1.1)$$

Очевидно, что эту формулу можно рассматривать как основу некоторого критерия независимости данных двух выборок. Если модуль значений, полученных по формуле (1.1), мал, то данные двух выборок можно считать независимыми. Как альтернативу будем рассматривать зависимые данные с коэффициентом корреляции $r(x, y) = +0.5$. Когда выборки реальных данных велики, функционал (1.1) дает устойчивые оценки. Положение меняется, если экспериментальных данных мало. В этом легко убедиться, воспроизводя зависимые и независимые данные для малых выборок в 16 опытов [1]. На рис. 1.1 отображены программное обеспечение численного эксперимента и его результаты.

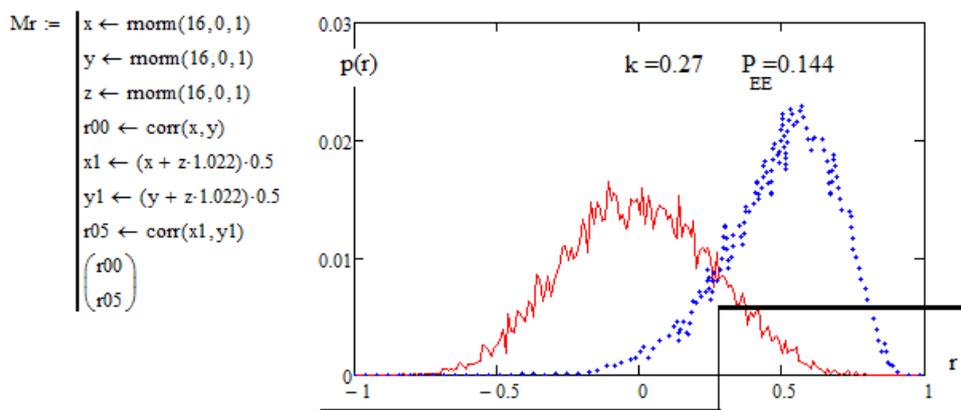


Рис. 1.1. Корреляционный критерий Эджуорта – Эдлтона – Пирсона и соответствующий ему искусственный нейрон

Из рис. 1.1 видно, что для двух независимых выборок в 16 опытов разброс значений оценок коэффициента корреляции лежит в интервале от -0.75 до $+0.75$. Столь значительный разброс препятствует получению достоверных статистических оценок на малых выборках экспериментальных данных. Выходом из создавшегося положения является параллельное использование нескольких статистических критериев и их объединение, например, в форме некоторого нейросетевого обобщения [1].

Классическая формула (1.1) в программном эмуляторе искусственного нейрона выполняет обогащение входных данных малой выборки. Выходной квантователь сравнивает обогащенные данные с порогом $k = 0.27$. При его выходном состоянии «0» нейрон классифицирует положительно коррелированные данные как слабо коррелированные.

1.2. Критерий Кенуя (1973)

Критерий Кенуя является дискретным [2], т.е. он не имеет непрерывного выходного спектра состояний. Спектр имеет конечное число линий (амплитуд вероятности) между которыми отклики критерия появиться не могут. Критерий строится на подсчете пар следующих друг за другом точек в выборке. Подсчитывается число совпадений знаков у разностей соседних точек после их сортировки по одной из координат. Программная реализация критерия Кенуя и результаты численного эксперимента отражены на рис. 1.2.

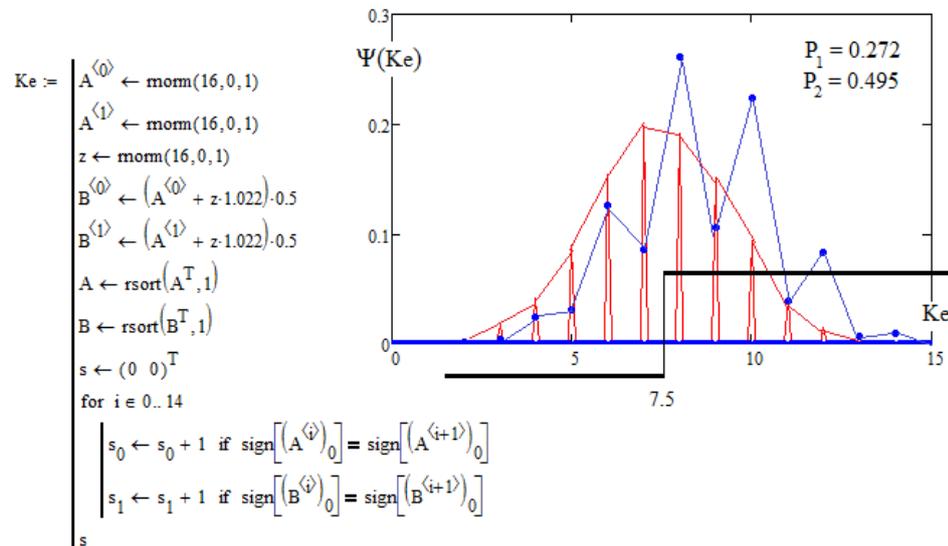


Рис. 1.2. Разделение независимых данных и данных с положительной коррелированностью по критерию Кенуя (1973)

Так как критерий Кенуя дискретен, указать порог для выходного квантователя, обеспечивающий одинаковые вероятности ошибок первого и второго рода, нельзя. В этом контексте приходится указывать вероятности ошибок первого и второго рода для значения порога $k = 7.5$, минимизирующего их разницу. Критерий Кенуя имеет существенно более высокое значение ошибок в сравнении с непрерывным критерием Эджуорта – Эдлтона – Пирсона.

1.3. Критерий Нельсона (1983)

Критерий Нельсона [3] построен на анализе соотношений координат соседних точек. Идет подсчет числа точек, похожих друг на друга по соотношению больше/меньше. Программная реализация критерия Нельсона приведена в левой части рис. 1.3, в правой части этого рисунка приведены результаты численного эксперимента.

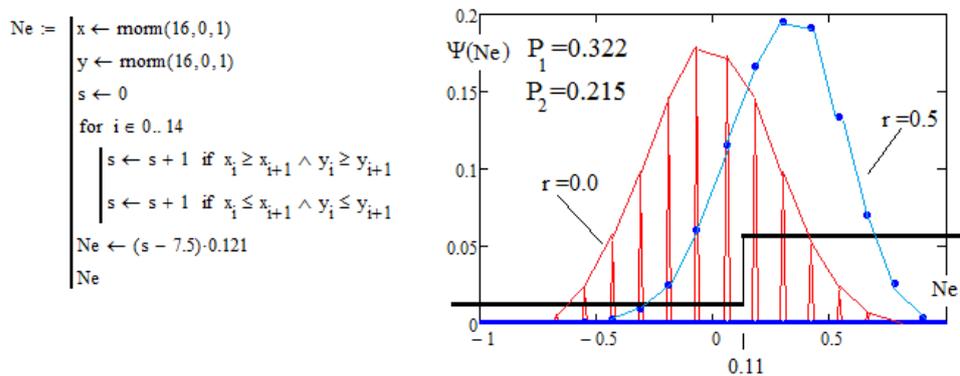


Рис. 1.3. Разделение независимых данных и данных с положительной коррелированностью по критерию Нельсона (1983)

Идея критерия Нельсона используется для обогащения входных данных нейрона, выходной квантователь этого нейрона выполняет сравнение с порогом $k = 0.11$. Нейрон является чувствительным к знаку коэффициента корреляции, т.е. нейроны этого типа не должны использоваться в схемах обработки данных, где нет дополнительных механизмов анализа знака корреляционной сцепленности.

1.4. Критерий Блума – Кифера – Розенбладтга (1961)

Критерий Блума – Кифера – Розенбладтга [4] в вычислительном отношении много сложнее в сравнении с описанными выше критериями. Он построен на сравнении координат каждой точки выборки

со всеми другими точками выборки. Критерий является дискретным, однако в правой части рис. 1.4 отображенные плотности распределения значений выглядят как непрерывные. Это связано с тем, что дискретных спектральных линий в спектрах много, расстояние между ними мало. При наблюдении не всего спектра, а его фрагментов его дискретный характер хорошо наблюдаем.

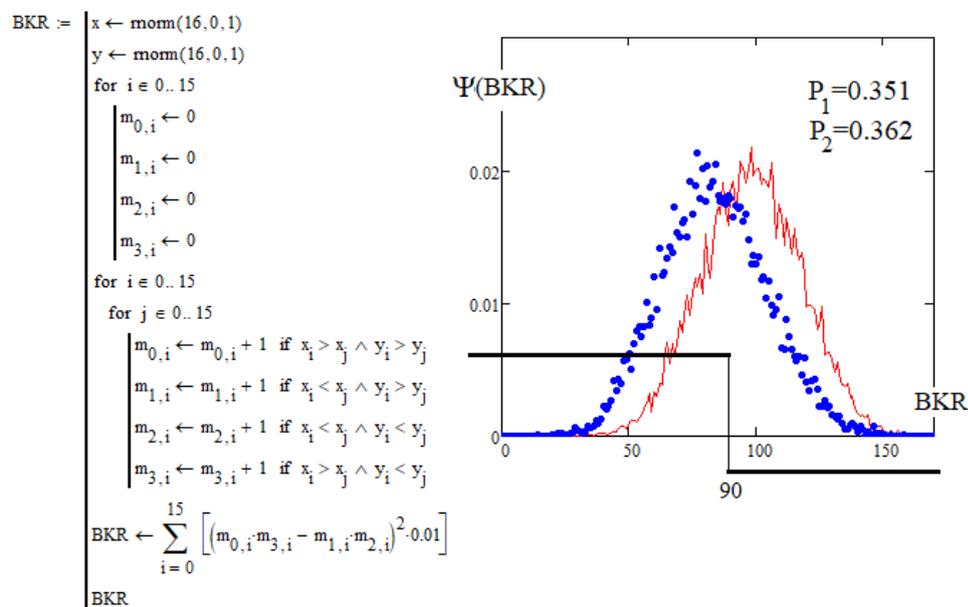


Рис. 1.4. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию Блума – Кифера – Розенблатта

К сожалению, мощность этого критерия по разделению зависимых и независимых данных много ниже мощности классического критерия Эджуорта – Эдлттона – Пирсона.

1.5. Критерий Кокса – Стюарта (1955)

Статистический критерий Кокса – Стюарта создан в 1955 г. [5]. Авторами предложено упорядочивать выборку по одной из переменных. Далее авторы предложили разбивать выборку на три или более подвыборок. Для каждой из подвыборок находят минимум и максимум, которые далее используются как опорные точки для вычисления рангов. В итоговой статистике Кокса – Стюарта все ранги суммируются. Программное описание численного эксперимента приводится в левой части рис. 1.5, результаты численного эксперимента отражены в правой части этого рисунка.

Видно, что значимые амплитуды вероятности спектра критерия Кокса – Стюарта имеют примерно 30 линий. Число линий спектра увеличивается с ростом корреляционной связанности данных малой

выборки. Бинарный нейрон Кокса – Стюарта выполняет квантование через сравнения с порогом $k = 17.5$.

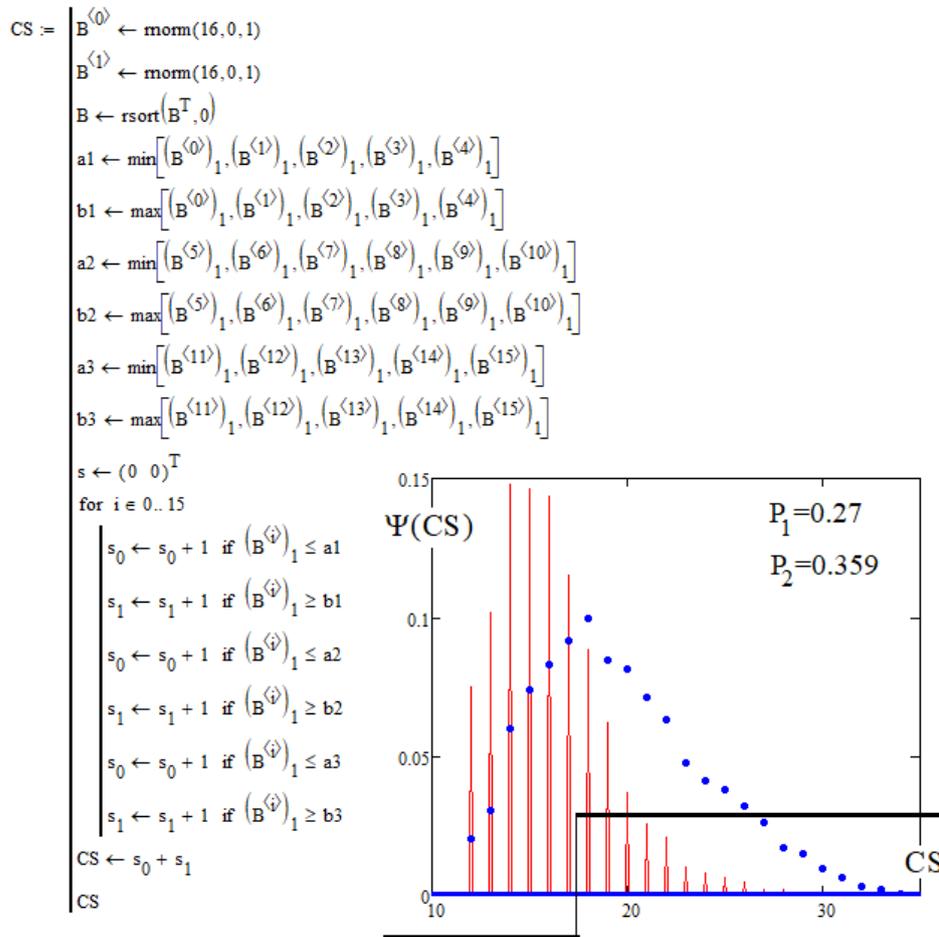


Рис. 1.5. Разделение независимых данных и данных с положительной коррелированностью $\gamma = 0.5$ по критерию Кокса – Стюарта

Критерий и эквивалентный ему искусственный нейрон инвариантны к знаку коэффициентов корреляции.

1.6. Критерий Аббе – Линника (1965)

Критерий Аббе – Линника [6] построен на упорядочивании по возрастанию данных одной из переменных малой выборки. Далее вычисляется отношение суммы квадратов разностей соседних отсчетов по второй, оставшейся случайной переменной, при ее нормировании на ее же дисперсию. В левой части рис. 1.6 дана программная реализация этого критерия, в правой части отражены результаты численного эксперимента.

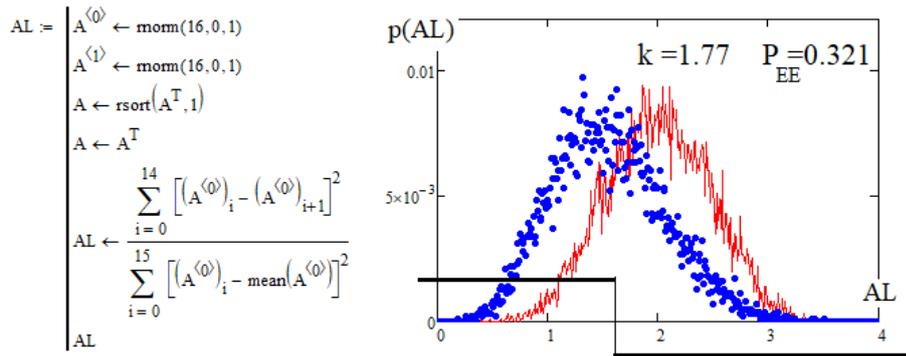


Рис. 1.6. Разделение независимых данных и данных с положительной и отрицательной коррелированностью $|r| = 0.5$ по критерию Аббе – Линника

Очевидным является то, что упорядочивание данных можно выполнять как по первой, так и по второй переменной, т.е. мы имеем возможность использования двух вариантов рассматриваемого статистического критерия. При этом оба критерия дают отклики, слабо связанные между собой. Нейрон нечувствителен к знаку коэффициента корреляции альтернативных данных.

1.7. Угловой критерий Олмстеда – Тьюки (1947)

Критерий Олмстеда – Тьюки [7] построен на вычислении медиан распределения данных по переменной x и по переменной y . Далее ведется подсчет точек выборки, распределенных по углам системы координат Декарта, центр которой находится в точке медиан по двум переменным.

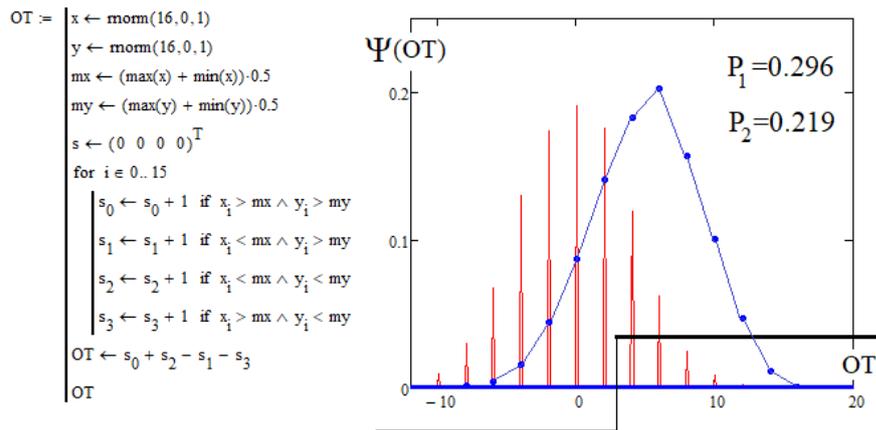


Рис. 1.7. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию Олмстеда – Тьюки

Для выделения зависимых данных выполнено суммирование данных в первой и третьей четвертях системы координат с вычитанием данных, накопленных по второй и четвертой четвертям Декартовой системы координат. Как результат, зависимые данные должны обладать асимметрией. Для положительно коррелированных данных распределение смещается вправо, как это показано на рис. 1.7. Для отрицательно коррелированных данных распределение смещается влево.

1.8. Критерий Сендвика – Олсона (1982)

Критерий Сендвика – Олсона является непрерывным, он создан в 1982 г. [8]. Очевидно, что он получен смещением точки наблюдения от медианы выборки к минимальным значениям данных в малой выборке. Весьма положительным является высокая мощность этого типа критериев, близкая к мощности наиболее часто используемого критерия Эджуорта – Эдлтона – Пирсона (рис. 1.8).

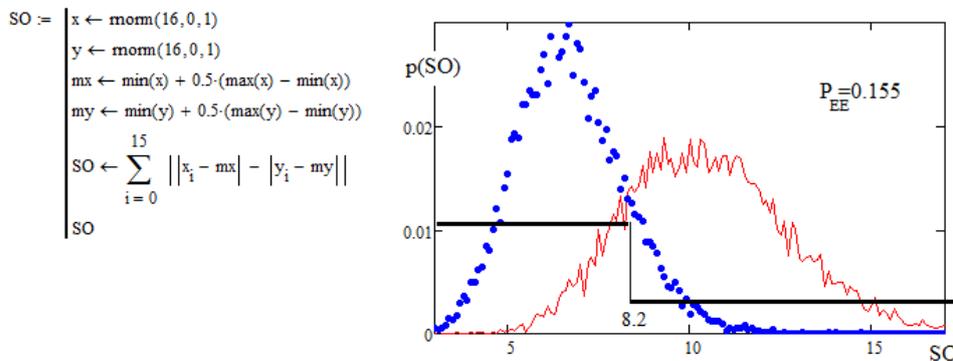


Рис. 1.8. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по модулю критерия Сендвика – Олсона

Так как мощность критерия велика, позднее будут показаны его модификации.

1.9. Критерий Шахани (1973)

Критерий Шахани [2] дискретный, в связи с этим нельзя указать показатель его относительной мощности в виде одинакового значения вероятностей ошибок первого и второго рода. В этом плане для оценки места критерия Шахани целесообразно усреднять вероятности ошибок первого и второго рода $P_{EE} \approx E(0.232, 0.116) = 0.174$. Получается, что критерий Шахани – один из высокоэффективных. Мощность критерия Шахани примерно на 12 % уступает предыдущему критерию.

Оригинальность критерия Шахани обусловлена тем, что его автор в отличие от критерия Сендвика – Олсона использовал две точки, относительно которых выполняется просмотр других точек малой выборки (рис. 1.9).

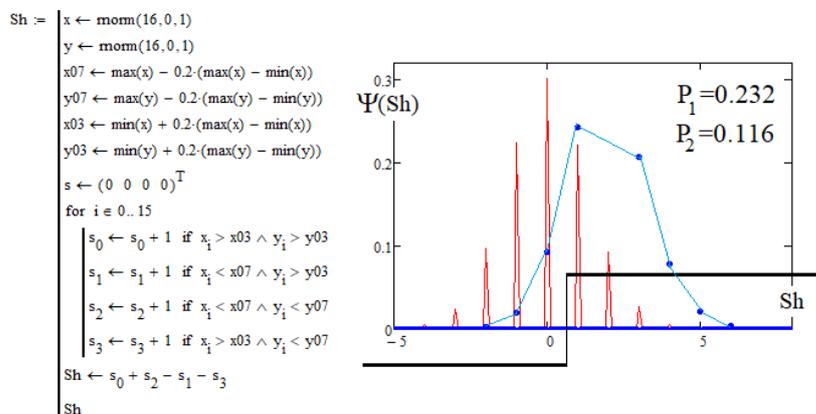


Рис. 1.9. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию Шахани (1973)

Очевидно, что наряду с дискретным вариантом критерия Шахани [2] существуют варианты непрерывного варианта этого критерия.

1.10. Критерий Кенделла (1971)

Ранговый критерий Кенделла [9] является дискретным, однако у него много спектральных линий, они оказываются расположены рядом. Спектральные линии сливаются при их отображении на графике (рис. 1.10). При вычислениях выполняется упорядочивание данных по одной переменной.

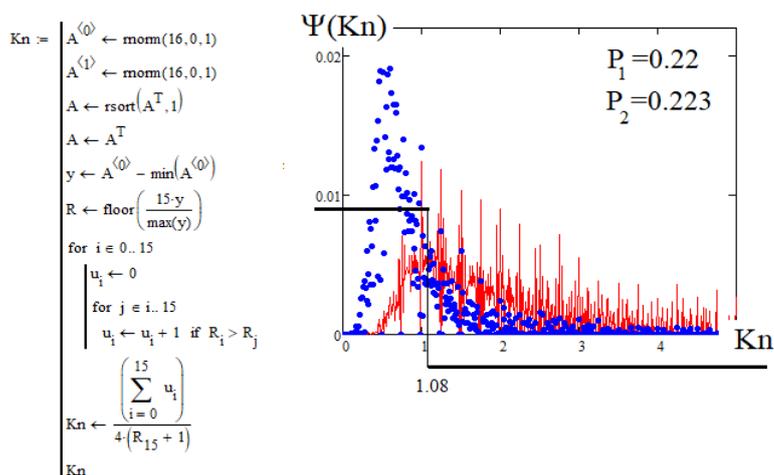


Рис. 1.10. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию Кенделла (1971)

Далее при вычислениях используются нормированные и округленные до целых значений данные. Ведется подсчет инверсий Кенделла для всех последующих отсчетов, которые накапливаются и нормируются.

1.11. Критерий Спирмена (1971)

Ранговый критерий Спирмена [10] является дискретным и может быть построен на предобработке данных по Кенделлу [9]. Отличие состоит в суммировании квадратов разности рангов данных и инверсий Кенделла. На рис. 1.11 представлены программная реализация критерия Спирмена и результаты численного эксперимента.

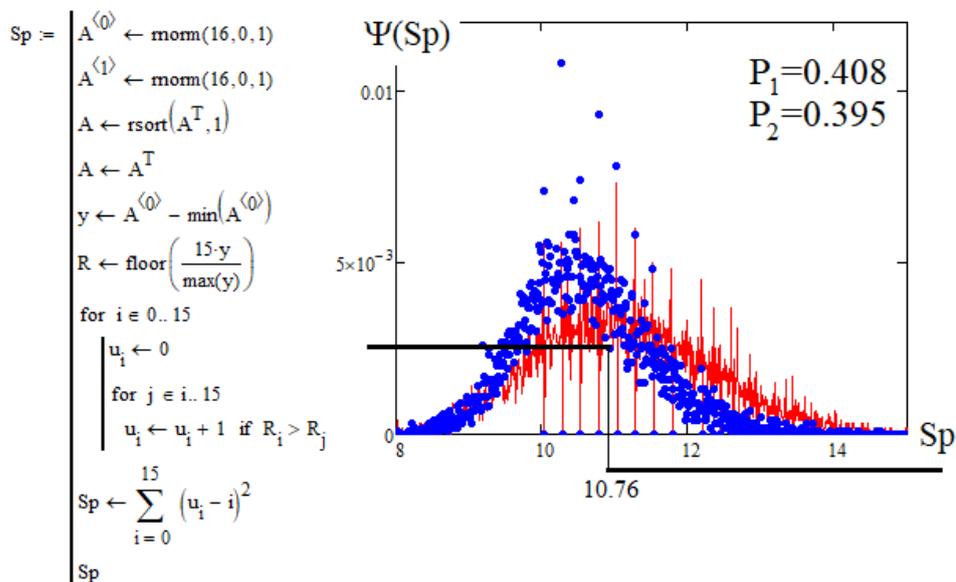


Рис. 1.11. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию Спирмена (1971)

При равных условиях критерий Кенделла имеет мощность больше на 36 % в сравнении с критерием Спирмена. Очевидно также, что предварительная сортировка данных по другой переменной дает еще один вариант критерия Спирмена.

1.12. Критерий Шведа – Эйзенхарта (1971)

Дискретный критерий Шведа – Эйзенхарта [2] построен на подсчете средней длины серий интервалов с последовательностью

одинаковых состояний «0..0» или «1..1». На рис. 1.12 приведены программная реализация критерия и результаты имитационного моделирования.

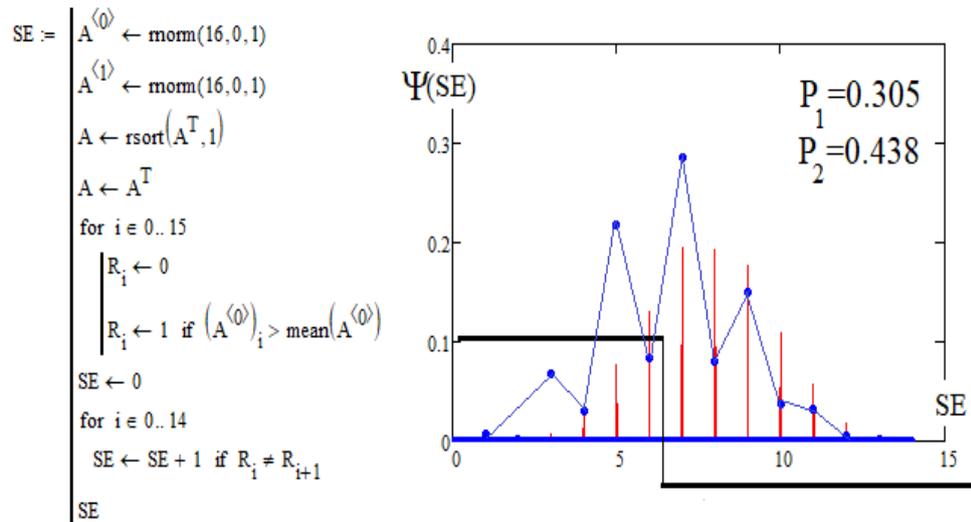


Рис. 1.12. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию Шведа – Эйзенхарта (1971)

Очевидно, могут быть созданы расширения данного критерия, построенные на учете отклонения от наиболее вероятных значений длин интервалов с одинаковыми состояниями. Например, критерий Шведа – Эйзенхарта может быть усилен последующим хи-квадрат статистическим анализом близости к наиболее вероятному статистическому распределению.

1.13. Сериально-ранговый критерий Рамачандрана – Ранганатана

Один из методов повышения эффективности критерия сериальности Шведа – Эйзенхарта предложили Рамачандран и Ранганатан [6]. Они предложили использовать квадрат длин одинаковых серий с учетом умножения их на частоту их появления. Программная реализация этого критерия приведена в левой части рис. 1.13. В правой части этого рисунка приведены результаты численного эксперимента.

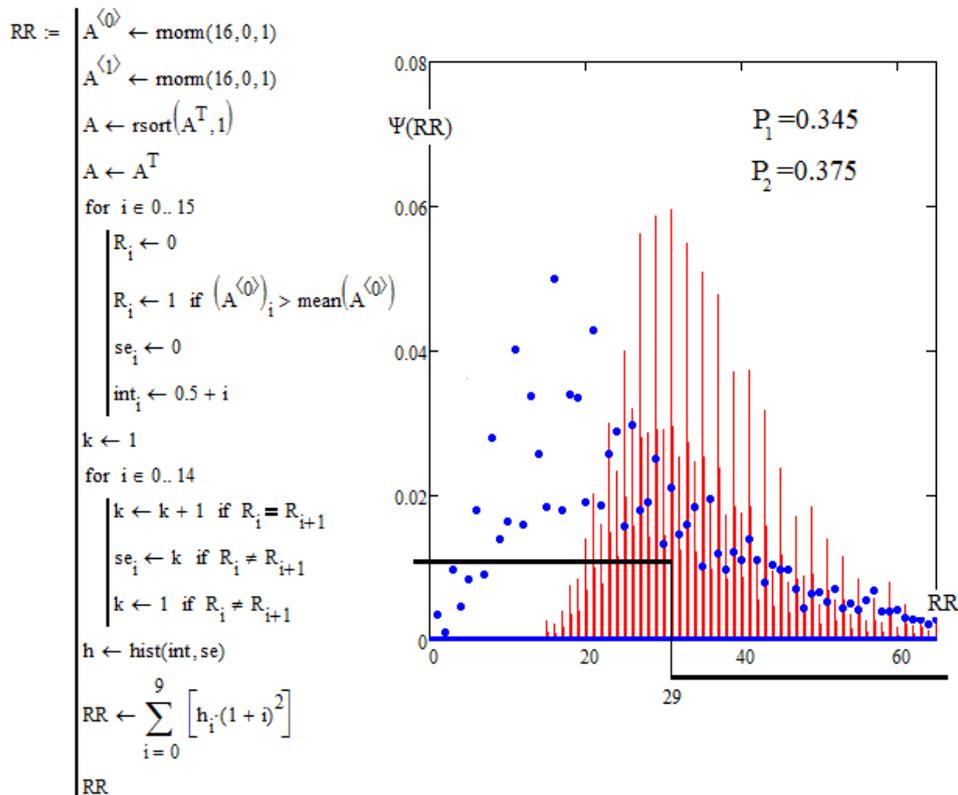


Рис. 1.13. Разделение независимых данных и данных с положительной коррелированностью $\gamma = 0.5$ по критерию Рамачандрана – Ранганатана

Данные этого численного эксперимента примерно на 17 % лучше данных предыдущего критерия.

1.14. Критерий Ван дер Вандена (1967)

Дискретный критерий Ван дер Вандена создан как вариант усиления рангового критерия, построенного как накопление произведений номеров упорядоченных данных по одной переменной и рангов инверсии Кенделла [11]. В левой части рис. 1.14 приведена программная реализация критерия. В правой части данного рисунка приведены результаты численного эксперимента. Рассматриваемый критерий дает вероятности ошибок первого и второго рода примерно на 60 % меньше предыдущего критерия.

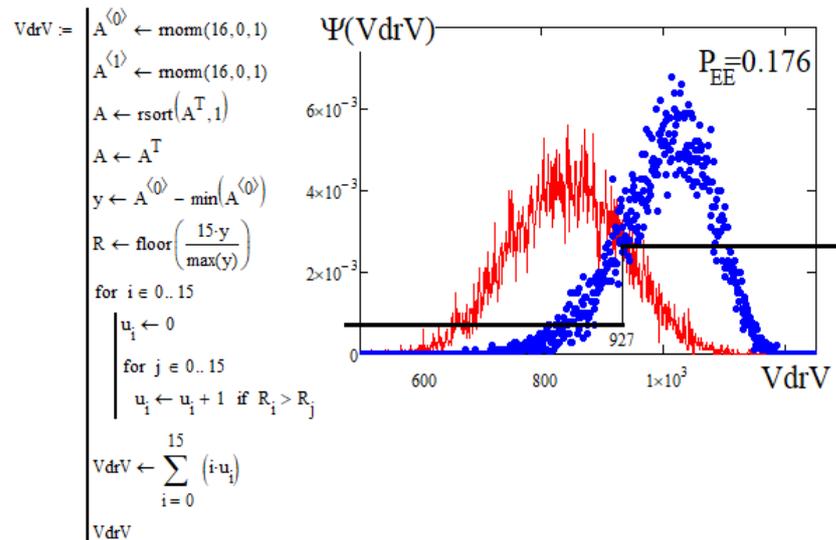


Рис. 1.14. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию Ван дер Вандена

Все это является свидетельством перспективности исследования ветви ранговых критериев, использующих преобразования Кенделла.

1.15. Критерий Шукени – Флори (1973)

Еще один вариант рангового критерия может быть получен по рецепту Шукени – Флори [12], если отказаться от использования преобразований Кенделла и вместо них накапливать произведение номеров упорядоченных данных по одной координате и округленных данных до ближайшего целого значения по оставшейся случайной координате. Программная реализация критерия Шукени – Флори приведена в левой части рис. 1.15.

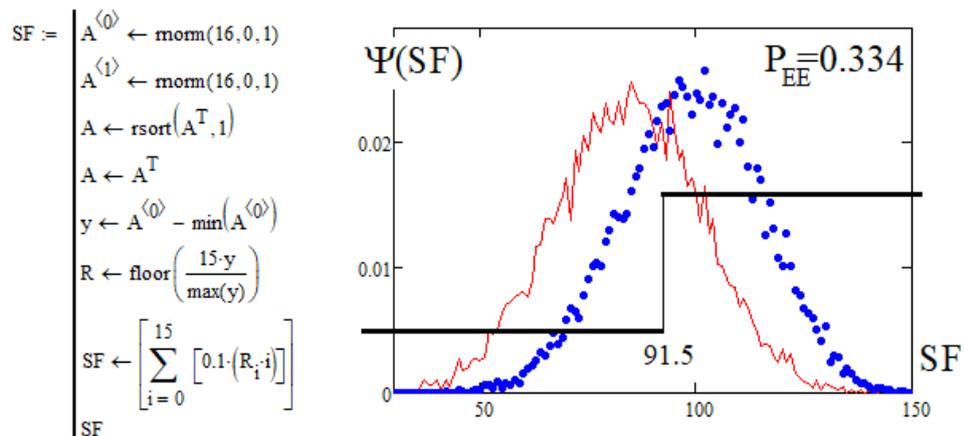


Рис. 1.15. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию Шукени – Флори (1973)

Очевидно, что приведенный выше критерий имеет свой дубликат, построенный упорядочиванием данных по другой переменной. Два критерия с разным упорядочиванием имеют слабую корреляционную сцепленность.

1.16. Критерий Бартелса (1982)

Ранговый критерий Бартелса [13] является дискретным и строится на накоплении квадратов разностей соседних рангов. Его мощность сопоставима с мощностью предыдущего критерия. Смена переменной начальной сортировки приводит к появлению еще одного варианта критерия Br1.

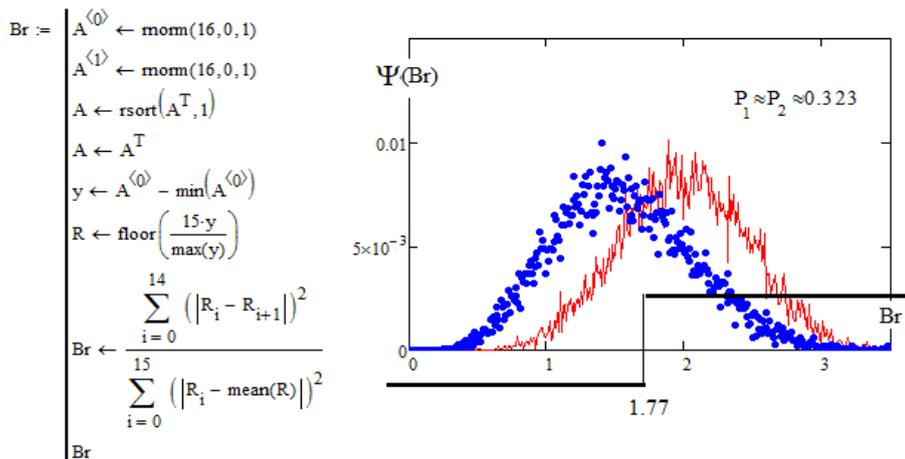


Рис. 1.16. Разделение независимых данных и данных с положительной коррелированностью $\gamma = 0.5$ по критерию Бартелса (1982)

Оба варианта критериев Бартелса имеют отклики с низкой корреляционной сцепленностью $\text{corr}(\text{Br}, \text{Br1}) \approx 0.146$.

1.17. Сериально-ранговый критерий Шахнесси (1981)

Критерий Шахнесси [14], как и предыдущие ранговые критерии, строится на упорядочивании по возрастанию одной из переменных малой выборки. Для оставшейся без упорядочивания переменной выполняется ранжирование путем приведения ее данных к интервалу от 0 до 15 и округления каждого значения до ближайшей целой величины.

Для близких к математическому ожиданию данных присваивается нулевой ранг. Для больших отклонений от математического ожидания присваивается единичный ранг. Для больших отклонений присваивается ранг -2 . Ранги умножаются на обратные номера упорядоченных данных и суммируются. Программная реализация критерия для воздействия на него независимых данных приведена на рис. 1.17.

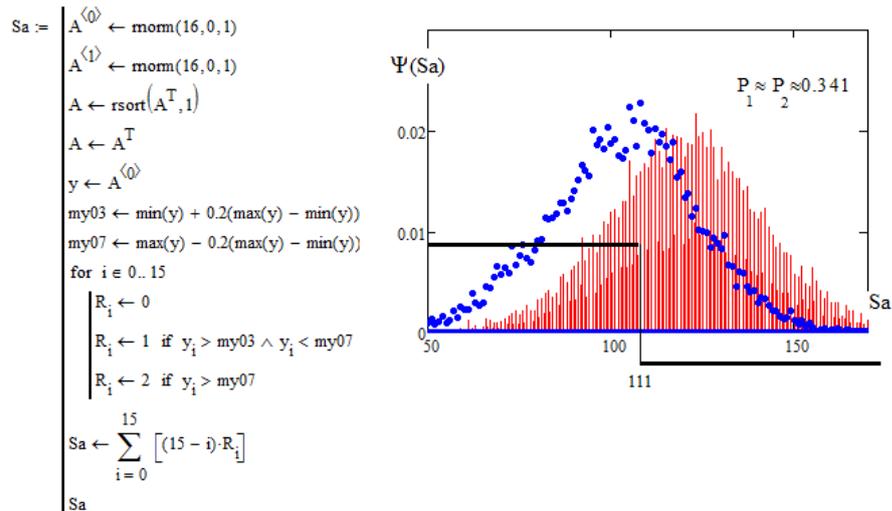


Рис. 1.17. Разделение независимых данных и данных с положительной коррелированностью $\gamma = 0.5$ по сериально-ранговому критерию Шахнесси (1981)

Смена переменной упорядочивания дает еще один вариант корреляционного критерия с низким уровнем корреляционной связанности откликов этих двух функционалов при решении основной их задачи – выявления независимых данных.

1.18. Критерий Эландта (1962)

Дискретный ранговый критерий Эландта [15] построен на определении минимальных значений данных в выборке и их смещении на половину медианы. Единичный ранг точке присваивается, если произведение ее координат оказывается положительным. Сам критерий получен суммированием рангов всех точек. В левой части рис. 1.18 приведена программная реализация критерия.

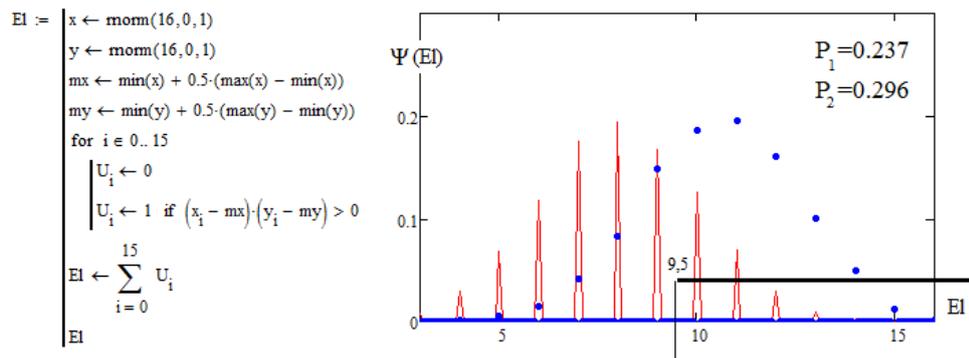


Рис. 1.18. Разделение независимых данных и данных с положительной коррелированностью $\gamma = 0.5$ по критерию Эландта (1962)

1.19. Критерий, построенный на статистике Манна – Уитни (1947)

Дискретный критерий Манна – Уитни [16] построен на формировании ранговой матрицы. Как показано в его программной реализации на рис. 1.19, первоначально формируют нулевую матрицу. Затем присваивают единичный ранг ее элементам, для которых одна из координат точки меньше другой. Сам критерий вычисляют, подсчитывая модуль среднего значения всех элементов ранговой матрицы.

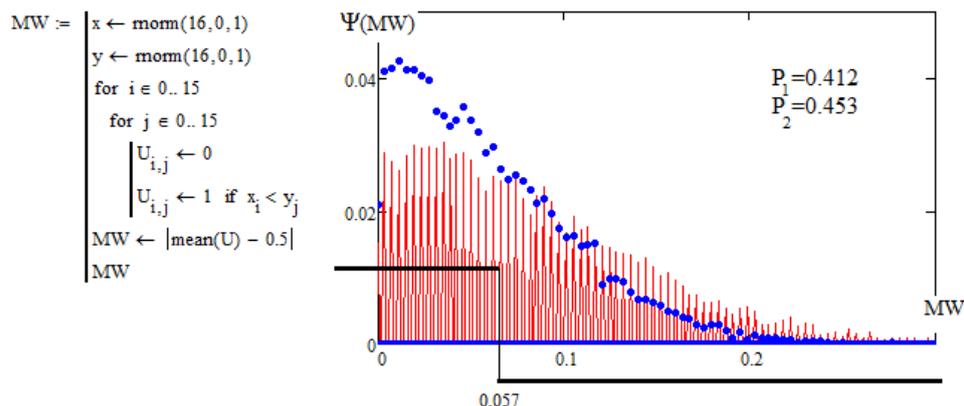


Рис. 1.19. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию, построенному на использовании статистики Манна – Уитни (1947)

Критерий интересен простотой его программной реализации. Очевидно, что он может быть существенно усложнен через заимствование вариантов более сложного ранжирования, использованных в предыдущих ранговых критериях.

1.20. Критерий, построенный Казаковичусом на циклическом сдвиге данных и оценке их автокорреляции (1988)

Критерий построен на упорядочивании данных переменной – x по их возрастанию [17]. В этом случае вторая переменная усиливает свою случайную компоненту. Критерий накапливает произведения соседних отсчетов, обеспечивая достаточно высокий уровень делимости зависимых и независимых данных (рис. 1.20).

Очевидно, что предварительное упорядочивание данных по другой переменной – y даст второй вариант критерия.

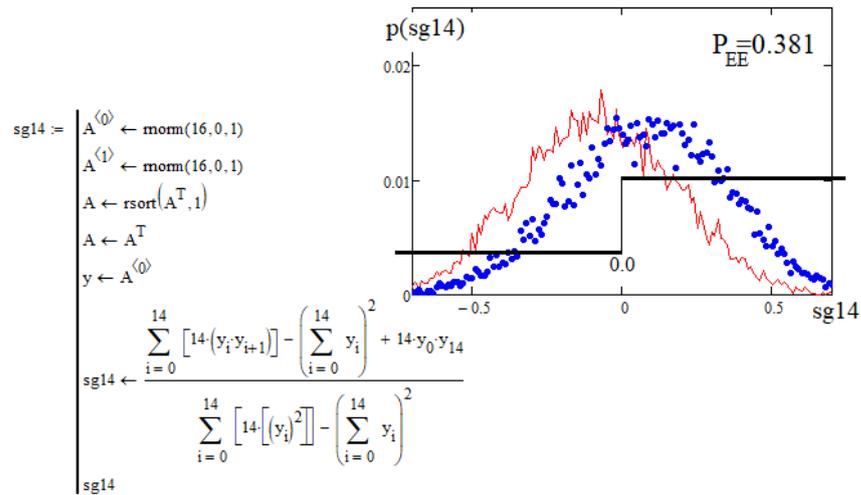


Рис. 1.20. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по критерию, построенному на контроле циклических сдвигов и вычислении коэффициента автокорреляции

Отклики первого и второго вариантов критерия являются слабо зависимыми.

1.21. Дифференциальный вариант критерия, построенного Кноке на циклическом сдвиге данных и оценке коэффициентов автокорреляции (1975)

Существенно усилить мощность автокорреляционных критериев удастся, если воспользоваться их дифференциальными вариантами (рис. 1.21), один из которых описал Кноке [18] в журнале «Биометрика».

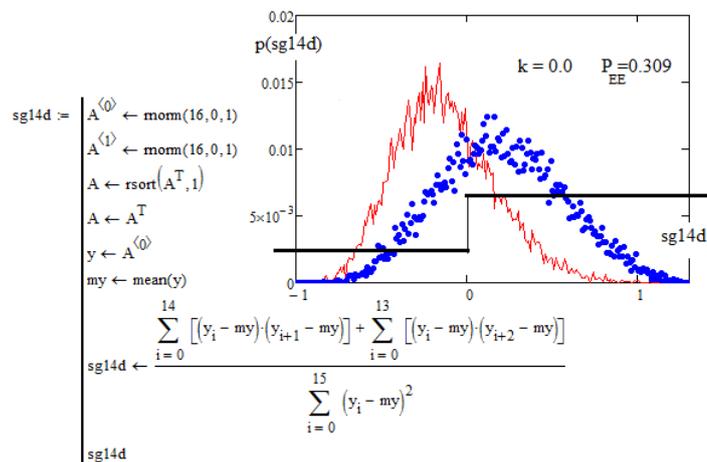


Рис. 1.21. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по дифференциальному варианту критерия, построенному на контроле циклических сдвигов и вычислении коэффициента автокорреляции

Смена переменной, по которой выполняется упорядочивание данных, по-прежнему, дает еще один статистический критерий той же мощности.

1.22. Прогнозирование достаточного для практики числа нейрокритериев проверки гипотезы независимости

Из приведенных выше материалов следует, что в прошлом веке было создано более 20 критериев проверки гипотезы независимости. Большинство из этих критериев пригодно для анализа малых выборок в 16 опытов. Необходимо ответить на вопрос о том, сколько потребуется нейрокритериев для того, чтобы обеспечить приемлемую для практики доверительную вероятность 0.99.

Для решения этой задачи выполним численный эксперимент, результаты которого отображены на рис. 1.22.

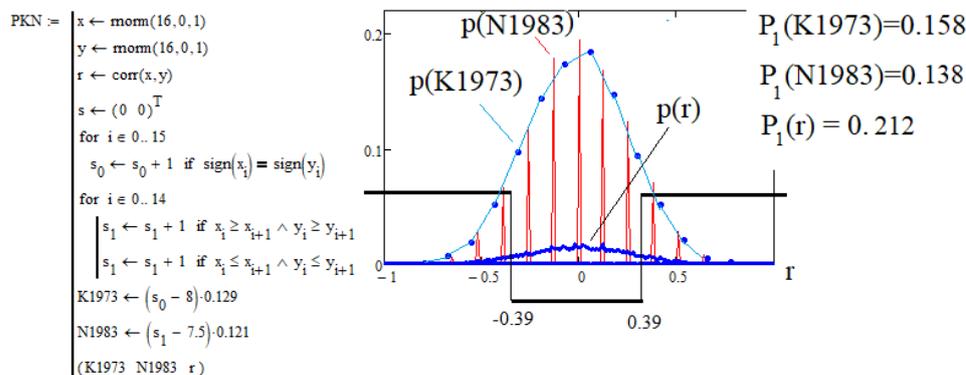


Рис. 1.22. Программное обеспечение и результаты численного моделирования при прогнозировании достаточного для практики числа используемых нейрокритериев

При моделировании использованы классический критерий Эджуорта – Эдлтона – Пирсона, критерий Кенуя 1973 г. и критерий Нельсона 1983 г. Отклики двух последних критериев приведены к масштабу стандартных отклонений, близких к стандартному отклонению независимых данных первого критерия.

Отклики всех критериев квантуются одинаковыми бинарными квантователями с двумя порогами сравнения $\{-0.39, +0.39\}$. При этом каждый из нейрокритериев дает свое значение вероятностей ошибок первого рода, приведенных в правой части рис. 1.22.

При синхронном признании независимости наблюдаемых данных всеми тремя нейрокритериями мы получим код «000». При этом

коды с единицей «1» в одном разряде следует рассматривать как подтверждающие гипотезу независимости. На рис. 1.23 приведен спектр расстояний Хэмминга, используемый при свертывании кодовой избыточности трех нейрокритериев.

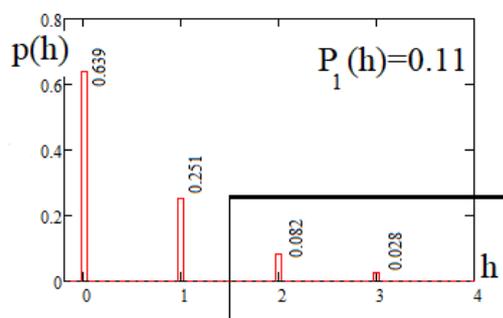


Рис. 1.23. Дискретный спектр амплитуд вероятности четырех линий расстояний Хэмминга между выходными кодами трех нейрокритериев с идеальным кодом «000»

При анализе спектра Хэмминга две первые линии соответствуют подтверждению гипотезы независимости. Две последующие линии отражают ситуацию, когда возникают ошибки первого рода. Одиночно работающий основной критерий Эджуорта – Эдлтона – Пирсона при порогах $\{-0.39, +0.39\}$ дает вероятность ошибок $P_1 \approx 0.212$. Три параллельно работающих нейрокритерия позволяют снизить вероятность ошибок почти в два раза до величины $P_1 \approx 0.11$. Это эквивалентно повышению объема тестовой выборки с 16 до 27 опытов при оценках коэффициентов корреляции.

Из рис. 1.24 видно, что при росте объема малой выборки с 16 до 27 опытов площадь хвостов распределений, находящихся вне порогов квантования, сокращается примерно в два раза.

Две точки вероятностей ошибок для одного нейрокритерия и для трех нейрокритериев позволяют сделать прогноз на ситуацию существенного увеличения числа параллельно используемых нейрокритериев. Прогноз выполнен линейно в логарифмических координатах вероятностей ошибок и числа, объединяемых нейросетью критериев.

Если ориентироваться на линейный прогноз рис. 1.25, то для достижения доверительной вероятности 0.99 потребуется использовать более 250 бинарных нейрокритериев. Снизить число нейрокритериев возможно в случае замены бинарных нейрокритериев на троичные или пятеричные нейрокритерии [19].

Данные прогноза требуют усиления внимания к проблеме синтеза новых статистических критериев в XXI в. При этом для покрытия проблем практики в XXI в. желательно синтезировать примерно в 10 раз больше критериев в сравнении с числом критериев, созданных в прошлом веке.

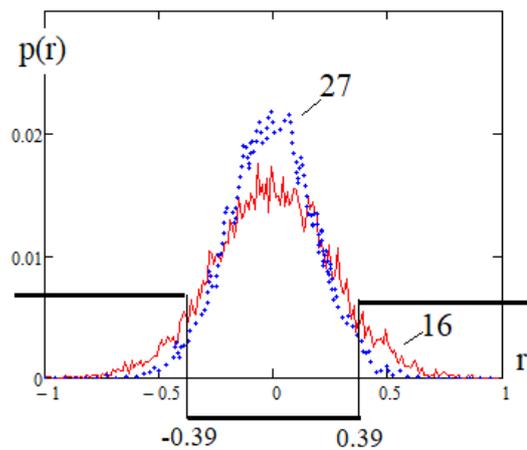


Рис. 1.24. Эффект снижения вероятностей ошибок (сжатия стандартного отклонения данных) при увеличении объема малой выборки с 16 до 27 опытов

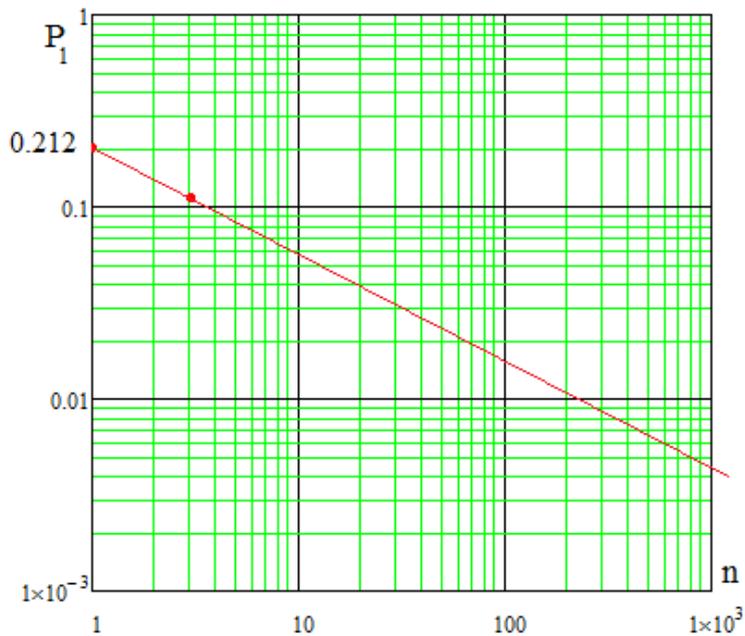


Рис. 1.25. Прогноз снижения вероятности ошибок первого рода при увеличении числа используемых нейрокритериев

Очевидно, что параллельно с ростом числа критериев желательно стараться повысить качество принимаемых ими решений и сократить корреляционные связи между новым критерием и другими известными критериями.

1.23. Оценка корреляционной сцепленности откликов классических статистических критериев

Наклон линии прогноза рис. 1.25 зависит от значения парных коэффициентов корреляции объединяемых нейрокритериев. Если модули коэффициентов парной корреляции снижаются, то наклон линий прогноза увеличивается. Верно и обратное утверждение. Рост модулей коэффициентов парной корреляции приводит к снижению наклона линии прогноза. При предельной корреляционной сцепленности данных $|r| = 1$ наклон линии прогноза рис. 1.25 полностью исчезает. Линия прогноза оказывается параллельна горизонтальным линиям, что эквивалентно полному отсутствию эффекта роста качества принимаемых нейросетевых решений при росте числа объединяемых нейрокритериев.

Предсказать наклон линии прогноза проще всего, выполнив усреднение модулей коэффициентов корреляции, находящихся вне диагонали корреляционной матрицы [20]. В ситуации объединения трех статистических (см. рис. 1.22) среднее значение модулей коэффициентов корреляции мало – $E(0.022, 0.594, 0.003) \approx 0.191$. Именно по этой причине линия прогноза имеет существенный наклон.

Для подобных быстрых оценок целесообразности объединения разных нейрокритериев ниже приведены парные коэффициенты корреляции (рис. 1.26).

Следует отметить, что выбранная в предыдущем параграфе тройка объединяемых нейрокритериев $\{r, Ke, Ne\}$ достаточно удачна, так как дает оценку корреляционной сцепленности – 0.191. Однако если заменить в тройке один нейрокритерий – Ne на другой нейрокритерий – BKR , то мы получим еще одну перспективную тройку нейрокритериев $\{r, Ke, BKR\}$. Для нее показатель корреляционной сцепленности становится практически нулевым. Происходит это из-за того, что все парные коэффициенты корреляции в этой тройке близки нулю $E(0.022, 0.001, 0.008) \approx 0.01$. Мы наблюдаем почти 19-кратное снижение значения коэффициентов равной коррелированности [20] из-за замены одного нейрокритерия на другой.

Еще одним эффективным приемом снижения уровня корреляционной сцепленности данных является повышение размерности решаемой задачи. Вместо тройки нейрокритериев мы можем использовать четверку слабо коррелированных нейрокритериев. В этом случае над диагональю корреляционной матрицы будут находиться шесть модулей коэффициентов корреляции. На рис. 1.26 коэффициенты корреляции перспективных четверок нейрокритериев объединены связывающими линиями, в частности для группы коэффициентов (левого верхнего угла на рис. 1.26) нейропараметров $\{r, Ke, BKR, AL\}$ мы имеем следующие модули коэффициентов корреляции: $\{0.022, 0.001, 0.0256, 0.008, 0.5981, 0.128\}$. Усреднение этих данных дает оценку

коэффициентов равной коррелированности – 0.13. Такой уровень корреляционной сцепленности должен давать существенно лучший прогноз по сравнению с прогнозом рис. 1.25. Это справедливо для всех 10 групп перспективных нейрокритериев на рис. 1.26.

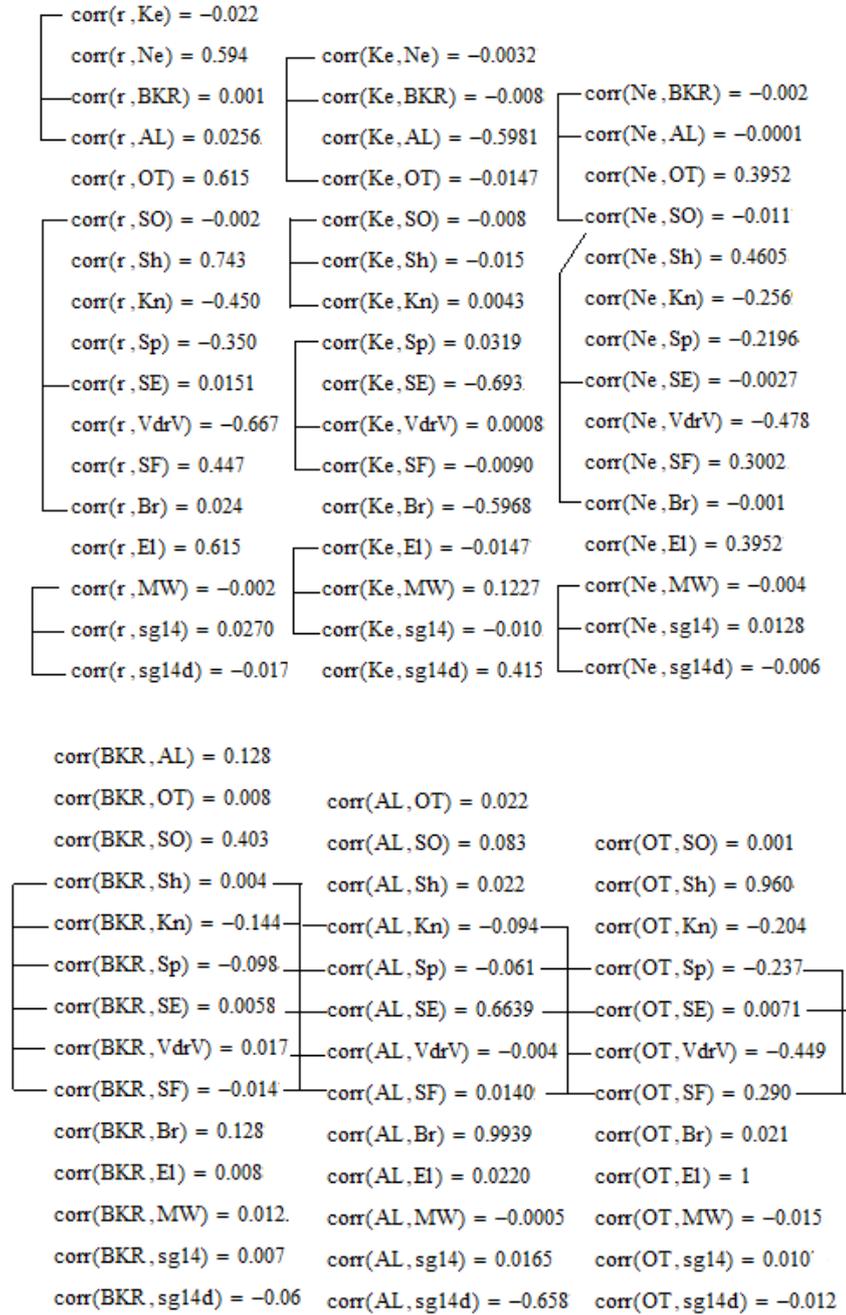


Рис. 1.26. Коэффициенты парной корреляции классических критериев прошлого века (начало)

$\text{corr}(\text{SO}, \text{Sh}) = 0.001$		
$\text{corr}(\text{SO}, \text{Kn}) = -0.116$	$\text{corr}(\text{Sh}, \text{Kn}) = -0.290$	
$\text{corr}(\text{SO}, \text{Sp}) = -0.029$	$\text{corr}(\text{Sh}, \text{Sp}) = -0.275$	$\text{corr}(\text{Kn}, \text{Sp}) = -0.127$
$\left[\begin{array}{l} \text{corr}(\text{SO}, \text{SE}) = 0.001 \\ \text{corr}(\text{SO}, \text{VdrV}) = 0.001 \\ \text{corr}(\text{SO}, \text{SF}) = -0.003 \end{array} \right.$	$\text{corr}(\text{Sh}, \text{SE}) = 0.009$	$\text{corr}(\text{Kn}, \text{SE}) = 0.0002$
	$\text{corr}(\text{Sh}, \text{VdrV}) = -0.525$	$\text{corr}(\text{Kn}, \text{VdrV}) = 0.484$
	$\text{corr}(\text{Sh}, \text{SF}) = 0.3454$	$\text{corr}(\text{Kn}, \text{SF}) = -0.335$
$\text{corr}(\text{SO}, \text{Br}) = 0.082$	$\text{corr}(\text{Sh}, \text{Br}) = 0.0216$	$\text{corr}(\text{Kn}, \text{Br}) = -0.084$
$\text{corr}(\text{SO}, \text{El}) = 0.001$	$\text{corr}(\text{Sh}, \text{El}) = 0.9604$	$\text{corr}(\text{Kn}, \text{El}) = -0.204$
$\text{corr}(\text{SO}, \text{MW}) = -0.069$	$\text{corr}(\text{Sh}, \text{MW}) = -0.013$	$\text{corr}(\text{Kn}, \text{MW}) = -0.004$
$\text{corr}(\text{SO}, \text{sg14}) = 0.0013$	$\text{corr}(\text{Sh}, \text{sg14}) = 0.016$	$\text{corr}(\text{Kn}, \text{sg14}) = -0.042$
$\text{corr}(\text{SO}, \text{sg14d}) = -0.031$	$\text{corr}(\text{Sh}, \text{sg14d}) = -0.014$	$\text{corr}(\text{Kn}, \text{sg14d}) = 0.038$
$\text{corr}(\text{Sp}, \text{SE}) = -0.0354$		
$\text{corr}(\text{Sp}, \text{VdrV}) = -0.319$	$\text{corr}(\text{SE}, \text{VdrV}) = 0.005$	
$\text{corr}(\text{Sp}, \text{SF}) = -0.1964$	$\text{corr}(\text{SE}, \text{SF}) = 0.01221$	$\text{corr}(\text{VdrV}, \text{SF}) = -0.316$
$\text{corr}(\text{Sp}, \text{Br}) = -0.068$	$\text{corr}(\text{SE}, \text{Br}) = 0.66151$	$\text{corr}(\text{VdrV}, \text{Br}) = 0.0011$
$\text{corr}(\text{Sp}, \text{El}) = -0.237$	$\text{corr}(\text{SE}, \text{El}) = 0.00712$	$\text{corr}(\text{VdrV}, \text{El}) = -0.449$
$\text{corr}(\text{Sp}, \text{MW}) = -0.013$	$\text{corr}(\text{SE}, \text{MW}) = -0.002$	$\text{corr}(\text{VdrV}, \text{MW}) = 0.006$
$\text{corr}(\text{Sp}, \text{sg14}) = -0.015$	$\text{corr}(\text{SE}, \text{sg14}) = 0.012$	$\text{corr}(\text{VdrV}, \text{sg14}) = -0.015$
$\text{corr}(\text{Sp}, \text{sg14d}) = 0.055$	$\text{corr}(\text{SE}, \text{sg14d}) = -0.465$	$\text{corr}(\text{VdrV}, \text{sg14d}) = -0.005$
$\text{corr}(\text{SF}, \text{Br}) = 0.0146$		
$\text{corr}(\text{SF}, \text{El}) = 0.2903$	$\text{corr}(\text{Br}, \text{El}) = 0.02115$	
$\text{corr}(\text{SF}, \text{MW}) = 0.011$	$\text{corr}(\text{Br}, \text{MW}) = 0.001$	$\text{corr}(\text{El}, \text{MW}) = -0.015$
$\text{corr}(\text{SF}, \text{sg14}) = -0.004$	$\text{corr}(\text{Br}, \text{sg14}) = 0.016$	$\text{corr}(\text{El}, \text{sg14}) = 0.0107$
$\text{corr}(\text{SF}, \text{sg14d}) = -0.013$	$\text{corr}(\text{Br}, \text{sg14d}) = -0.655$	$\text{corr}(\text{El}, \text{sg14d}) = -0.012$
$\text{corr}(\text{MW}, \text{sg14}) = 0.013$		
$\text{corr}(\text{MW}, \text{sg14d}) = -0.007$	$\text{corr}(\text{sg14}, \text{sg14d}) = -0.008$	

Рис. 1.26. Окончание

При увеличении размера корреляционной матрицы до 7×7 низкий уровень корреляционной сцепленности сохраняется. Примеры значений модулей коэффициентов корреляции такой матрицы выделены на рис. 1.26. Усреднение модулей коэффициентов корреляции дает значение -0.16 . Этот показатель лучше, чем качество данных, на которых построен рис. 1.25.

Список литературы к главе 1

1. Иванов А. И., Золотарева Т. А. Искусственный интеллект в защищенном исполнении: синтез статистико-нейросетевых автоматов многокритериальной проверки гипотезы независимости малых выборок биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 105 с.
2. Кенуй М. Г. Быстрые статистические вычисления. Упрощенные методы оценивания и проверки : пер. с англ. М. : Статистика, 1979.
3. Nelson L. S. A sign test for correlation // *Journal of Quality Technology*. 1983. Vol. 15, № 4. P. 199–200.
4. Blum J. R., Kiefer J., Rosenblatt M. Distribution-free tests of independence based on the sample distribution function // *The Annals of Mathematical Statistics*. 1961. Vol. 32. P. 485–498.
5. Cox D. R., Stuart A. Quick tests for trend in location dispersion // *Biometrika*. 1955. Vol. 42. P. 80–95.
6. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. М. : ФИЗМАТЛИТ, 2006. 816 с.
7. Olmstead P. S., Tukey J. W. A corner test for association // *The Annals Mathematical Statistics*. 1947. Vol. 18. P. 495–513.
8. Sanduik L., Olsson B. A nearly distribution-free test for comparing dispersion in pared samples // *Biometrika*. 1982. Vol. 69, № 32. P. 484–485.
9. Кенделл М. Ранговые корреляции. М. : Статистика, 1975.
10. Идье В., Драйад Д., Джеймс Ф. [и др.]. Статистические методы в экспериментальной физике. М. : Атомиздат, 1976.
11. Гаек Я., Шудак З. Теория ранговых критериев : пер. с англ. М. : Наука, 1971.
12. Schucany W. R., Frawley W. H. A rank test for two group concordance // *Psychometrika*. 1973. Vol. 38. P. 249–258.
13. Bartels R. The rank version of von Neimann's ratio test for randomness // *Journal of the American Statistical Association*. 1982. Vol. 77, № 377. P. 40–46.
14. Shaughessy P. M. Multiple runs distributions: Recurrent and critical values // *Journal of the American Statistical Association*. 1981. Vol. 76, № 375. P. 732–736.
15. Elantd R. C. Exact and approximate power function of the non-parametric test of tendency // *The Annals Mathematical Statistics*. 1962. Vol. 33, № 2. P. 471–481.
16. Mann H. B., Whitney D. R. On a test of whether one of two random variables is stochastically larger than the other // *The Annals Mathematical Statistics*. 1947. Vol. 18. P. 50–60.

17. Казаковичус К. А. Приближенные формулы для статистической обработки результатов механических испытаний // Заводская лаборатория. 1988. Т. 54, № 12. С. 82–85.

18. Knoke J. D. Testing for randomness against autocorrelation: The parametric case // *Biometrika*. 1975. Vol. 62. P. 571–575.

19. Иванов А. И., Савин К. Н., Еременко Р. В. Эффект перехода от применения бинарных искусственных нейронов к троичным нейронам при совместном использовании пяти классических статистических критериев проверки гипотез нормальности или равномерности распределений малых выборок // Вестник Пермского университета. Математика. Механика. Информатика. 2022. Вып. 3 (58). С. 59–67. doi: 10.17072/1993-0550-2022-3-59-67

20. Иванов А. И., Банных А. Г., Серикова Ю. И. Учет влияния корреляционных связей через их усреднение по модулю при нейросетевом обобщении статистических критериев для малых выборок // Надежность. 2020. Вып. 20 (2). С. 28–34. URL: <https://doi.org/10.21683/1729-2646-2020-20-2-28-34>

ВАРИАЦИИ КРИТЕРИЕВ, БЛИЗКИХ К КЛАССИЧЕСКОМУ КРИТЕРИЮ ЭДЖУОРТА – ЭДЛТОНА – ПИРСОНА

2.1. Дискретный вариант критерия Эджуорта – Эдлтона – Пирсона

Непрерывный классический критерий Эджуорта – Эдлтона – Пирсона является одним из самых привычных и одним из самых эффективных. В связи с этим представляет интерес его дискретный аналог [22]. Строится этот критерий размещением центра системы координат Декарта в ожидаемом центре распределения малых выборок и подсчетом числа точек, попавших в каждую из четвертей (рис. 2.1).

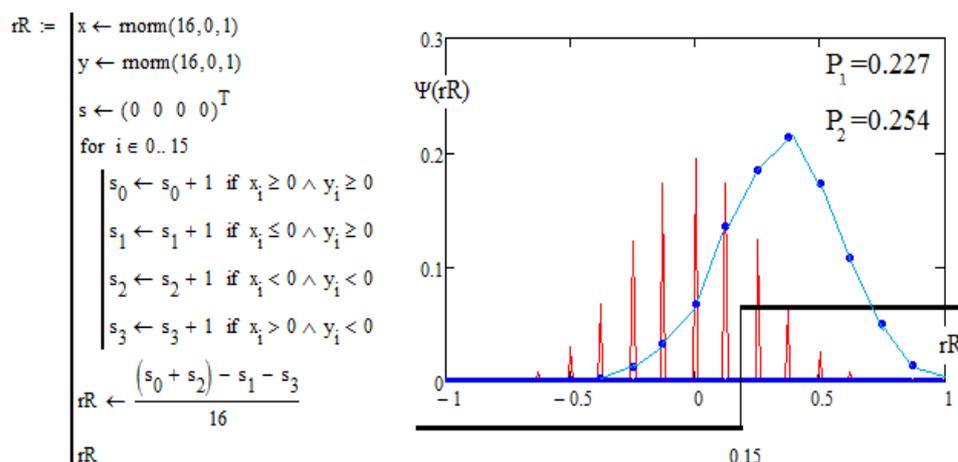


Рис. 2.1. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по дискретному аналогу корреляционного критерия Эджуорта – Эдлтона – Пирсона

Этот критерий похож на критерий Олмстеда – Тьюки (1947) [7]. Отличие состоит только в расположении центра системы координат Декарта. Для критерия Олмстеда – Тьюки центр системы координат размещен в двух медианах распределения. Это дает $\text{corr}(r, OT) \approx 0.615$. Дискретный аналог критерия Эджуорта – Эдлтона – Пирсона имеет сопоставимый уровень коррелированности $\text{corr}(r, Rr) \approx 0.633$.

2.2. Дискретный корреляционный критерий, построенный на использовании двух эллиптических квантователей

Предшествующий дискретный критерий можно рассматривать как два нейрона с двумя бинарными квантователями. При этом две квантующие пространство линии (оси координат) являются ортогональными. Очевидно, что два линейных квантователя можно заменить на два эллиптических квантователя. Программная реализация этой модификации алгоритма приведена в левой части рис. 2.2.

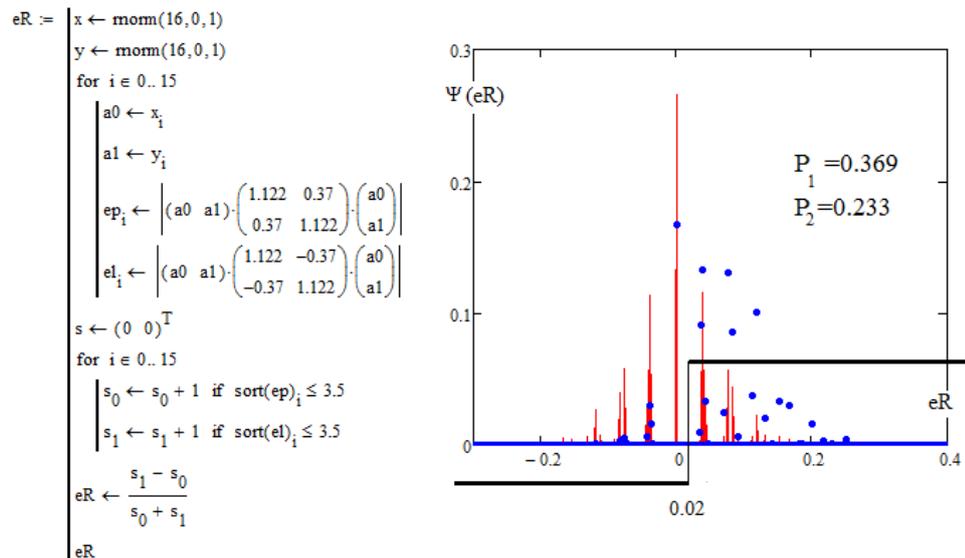


Рис. 2.2. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по дискретному корреляционному критерию с двумя эллиптическими квантователями (2019) [22]

Очевидно, что эллипсы двух квантователей могут иметь разные соотношения малого и большого диаметров, т.е. более сложные в вычислительном отношении квадратичные нейрокритерии имеют большее число возможных вариаций, даже если большие диаметры двух эллипсов перпендикулярны. Мощность предыдущего нейрокритерия с линейными квантователями примерно на 15 % выше в сравнении с нейрокритерием рис. 2.2. Можно ли изменить ситуацию, неизвестно. Для ответа на этот вопрос для нейрокритериев с эллиптическими квантователями должна быть решена задача оптимизации параметров эллипсов.

2.3. Первый непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их мгновенными приближениями

Оба приведенных выше аналога критерия Эджуорта – Эдлтона – Пирсона имеют дискретные спектры выходных состояний и обладают пониженной мощностью по сравнению с их аналогом конца XIX в. Увеличить мощность новых критериев удастся, если вернуться к обработке данных с непрерывным распределением. В левой части рис. 2.3 приведена программная реализация нейрокритерия, являющегося аналогом классического критерия. Его отличие состоит в том, что в знаменателе дробей произведение стандартных отклонений заменено на сумму модулей их мгновенных оценок. Кроме того, в числителе добавлен стабилизатор, исключающий появление маловероятных ситуаций деления на ноль.

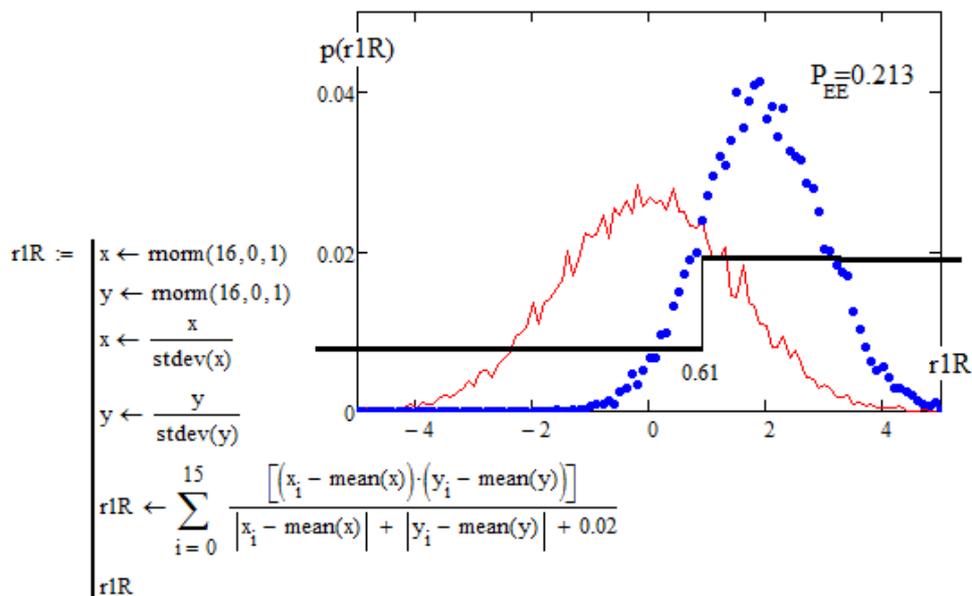


Рис. 2.3. Первый аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой стандартных отклонений их приближениями (2021) [24]

2.4. Второй непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их приближениями и заменой математических ожиданий минимумами данных в малой выборке

Еще один аналог классического критерия Эджуорта – Эдлтона – Пирсона может быть получен, если заменить в его вычислительной

формуле (1.1) математические ожидания на минимальные значения координат точек малой выборки. Соответствующая модификация предыдущего критерия представлена на рис. 2.4.

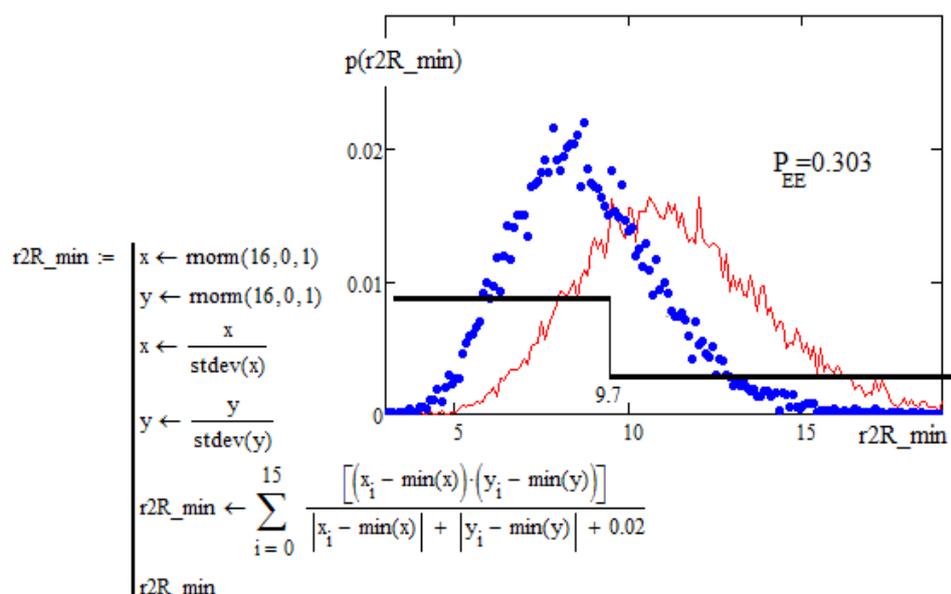


Рис. 2.4. Второй непрерывный аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой математических ожиданий на минимальные значения параметров (2021) [24]

Как следует из данного рисунка, использованная модификация приводит примерно к 30 % потери мощности нового критерия. Однако его использование совместно с другими критериями технически выгодно, если внутри группы корреляционная сцепленность нейрокритериев низкая.

2.5. Третий непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их приближениями и заменой математических ожиданий максимумами данных в малой выборке

Замена минимальных значений данных малой выборки на их максимальные значения дает еще один вариант нового статистического критерия [24]. Результаты имитационного моделирования представлены на рис. 2.5. Мощность нового статистического критерия и предыдущего одинакова, однако их корреляционная сцепленность низкая.

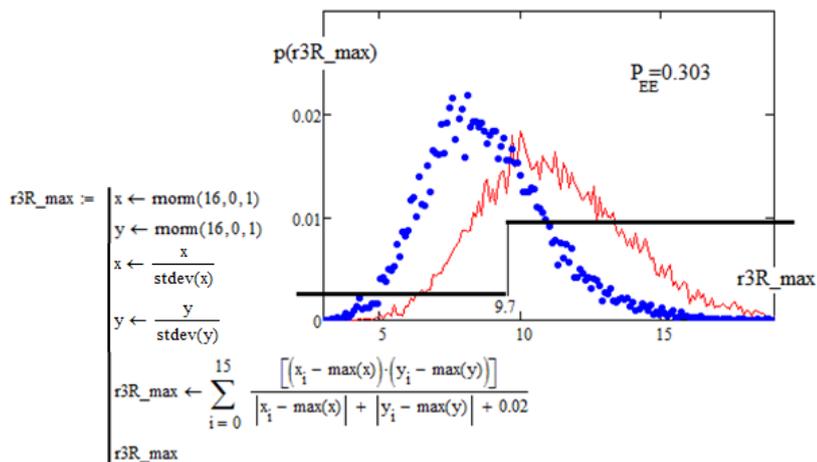


Рис. 2.5. Третий аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой математических ожиданий на максимальные значения параметров (2021) [24] малой выборки

2.6. Четвертый непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их приближениями и заменой математических ожиданий $\max(x)$ и $\min(y)$

Привычные всем математические ожидания в классической формуле (1.1) могут быть заменены парой максимального и минимального значений, как это отражено в программной реализации нового критерия в левой части рис. 2.6. Важным является то, что эта замена позволяет примерно на 30 % увеличить мощность нового критерия.

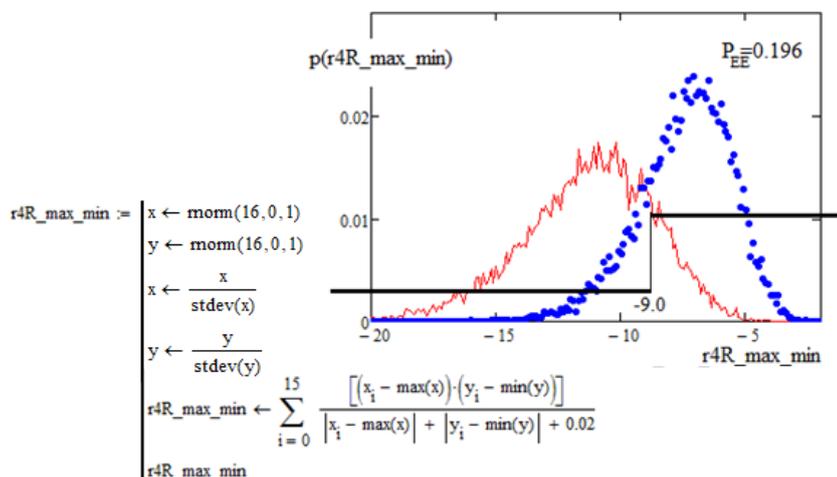


Рис. 2.6. Четвертый непрерывный аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой математических ожиданий $E(x)$ на максимальное значение $\max(x)$ и второго математического ожидания $E(y)$ на минимальное значение $\min(y)$ (2021) [24] в малой выборке

2.7. Пятый непрерывный аналог критерия Пирсона, полученный заменой стандартных отклонений их приближениями и заменой двух математических ожиданий на $\min(x)$ и на $\max(y)$

Смена мест минимума и максимума в вычислительной формуле предыдущего критерия позволяет получить еще один новый статистический критерий (рис. 2.7). Мощности предыдущего и нового критерия совпадают при слабой корреляционной сцепленности их откликов.

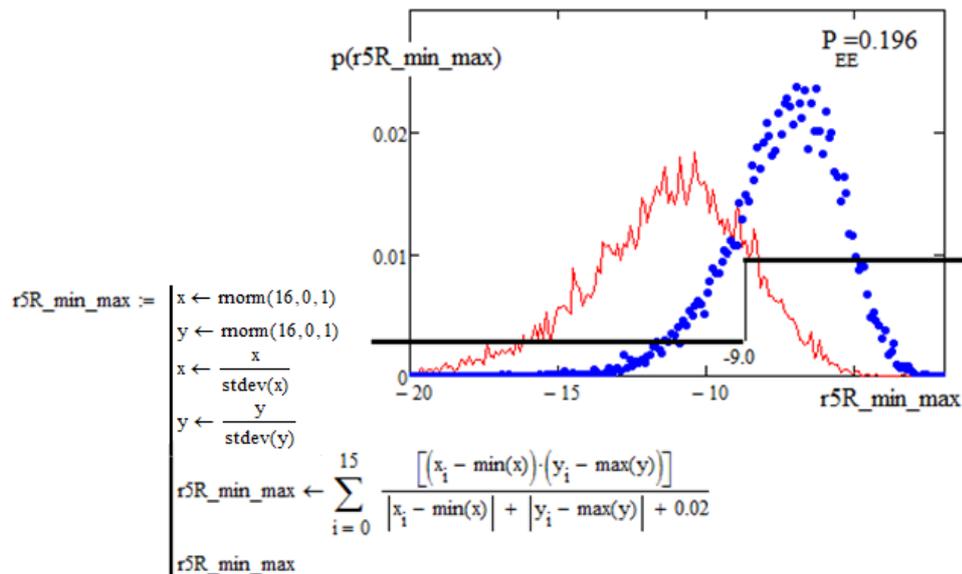


Рис. 2.7. Пятый непрерывный аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой математических ожиданий $E(x)$ на минимальное значение $\min(x)$ и второго математического ожидания $E(y)$ на максимальное значение $\max(y)$ (2021) [24] в малой выборке

2.8. Шестой непрерывный аналог критерия Пирсона – $r6R_E_min$, полученный заменой минимальных значений по переменной – x на ее математическое ожидание

Еще один вариант статистического критерия проверки гипотезы независимости получается, если заменить значения $\min(x)$ на математическое ожидание данных по этой переменной в предыдущем критерии. В нижнем левом углу рис. 2.8 приведена программная реализация этой модификации статистического критерия.

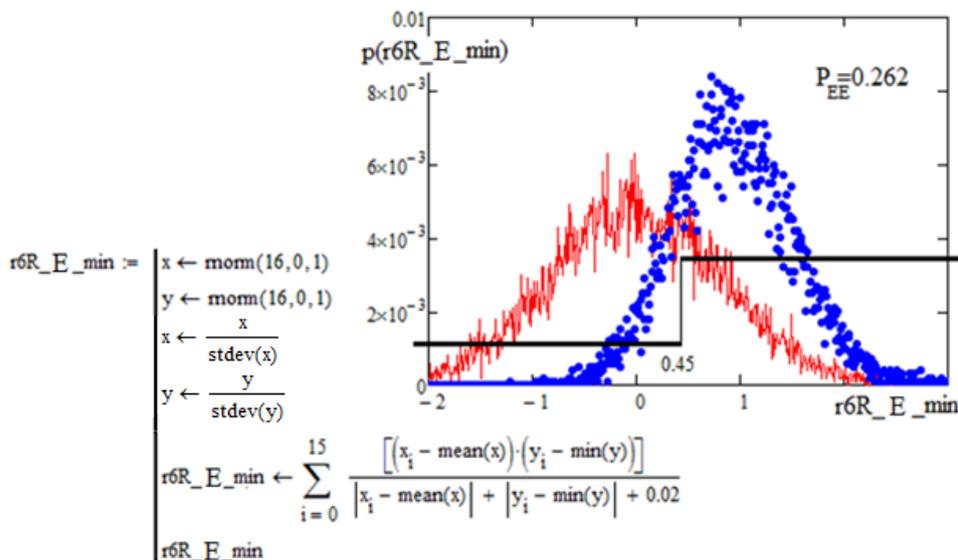


Рис. 2.8. Шестой непрерывный аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой математических ожиданий $E(y)$ на минимальное значение $-\min(y)$ при сохранении второго математического ожидания $E(x)$, как у классического критерия

Результаты численного эксперимента показывают приемлемое качество разделимости этим критерием независимых и зависимых данных, если его сравнивать с большинством классических критериев прошлого века.

2.9. Седьмой непрерывный аналог критерия Пирсона – $r7R_min_E$, полученный использованием в исходной формуле математических ожиданий по переменной – y и минимального значения переменной – x

Следующий вариант статистического критерия с той же мощностью разделения зависимых и независимых данных получается использованием в вычислительной формуле математических ожиданий по переменной – y . Соответственно по переменной – x в вычислительной формуле должно использоваться значение – $\min(x)$.

Программная реализация этого варианта критерия и результаты численных экспериментов приведены на рис. 2.9.

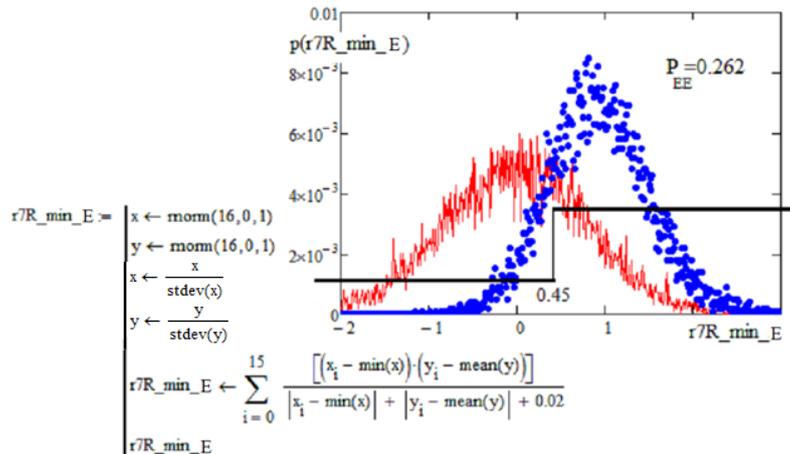


Рис. 2.9. Седьмой непрерывный аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой математических ожиданий $E(x)$ на минимальное значение – $\min(x)$ при сохранении второго математического ожидания $E(y)$, как у исходного критерия

2.10. Восьмой непрерывный аналог критерия Пирсона – $r8R_E_max$, полученный заменой экстремумов по переменной – x на математическое ожидание $E(x)$ и использованием максимального значения по второй переменной

Следующий вариант критерия получается использованием математического ожидания $E(x)$ и максимального значения по другой переменной $\max(y)$. На рис. 2.10 представлены программная реализация этого критерия и результаты численного эксперимента. Мощность этого критерия совпадает с мощностью предыдущего критерия.

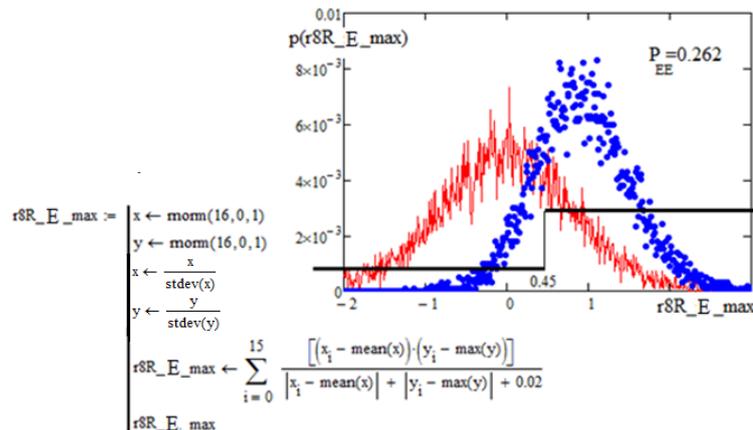


Рис. 2.10. Восьмой аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой математических ожиданий $E(y)$ на максимальное значение $\max(y)$ при сохранении второго математического ожидания $E(x)$, как у исходного критерия

2.11. Девятый непрерывный аналог критерия Пирсона $r9R_max_E$, полученный использованием математического ожидания по переменной – x и максимального значения переменной – y

Девятый вариант критерия получается использованием математического ожидания – $E(y)$ и максимального значения по другой переменной – $\max(x)$. На рис. 2.11 представлены программная реализация этого критерия и результаты численного эксперимента. Мощность этого критерия совпадает с мощностью предыдущего критерия.

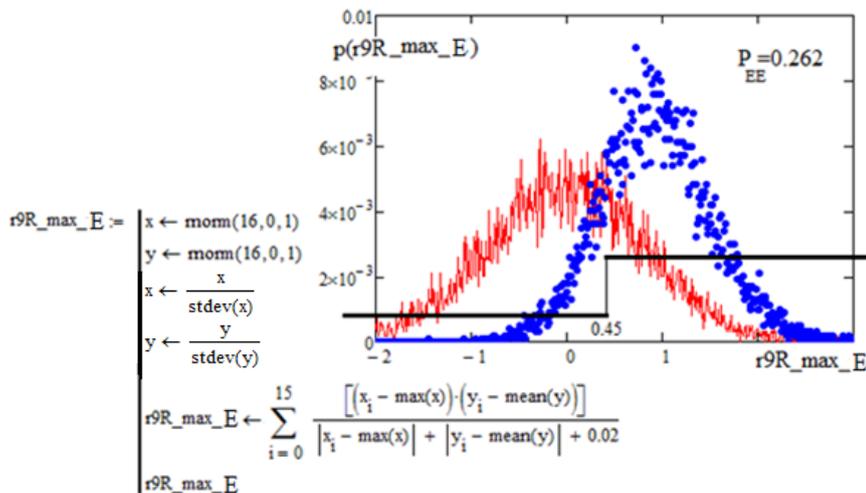


Рис. 2.11. Девятый непрерывный аналог критерия Пирсона – Эджуорта – Эдлтона, полученный заменой математических ожиданий $E(x)$ на максимальное значение – $\max(x)$ при сохранении второго математического ожидания $E(y)$, как у исходного критерия

2.12. Десятый непрерывный аналог классического критерия Пирсона – Эджуорта – Эдлтона, полученный нормировкой квадратом суммы дисперсий

Одним из путей модификации вычислительной формулы классического критерия Пирсона – Эджуорта – Эдлтона является замена в ее числителе произведения стандартных отклонений некоторым их приближением. Например, вместо произведения стандартных отклонений может быть использован квадрат их суммы. В результате мы получаем еще один, десятый, вариант классического статистического критерия. Программная реализация и результаты численного моделирования нового нейрокритерия приведены на рис. 2.12.

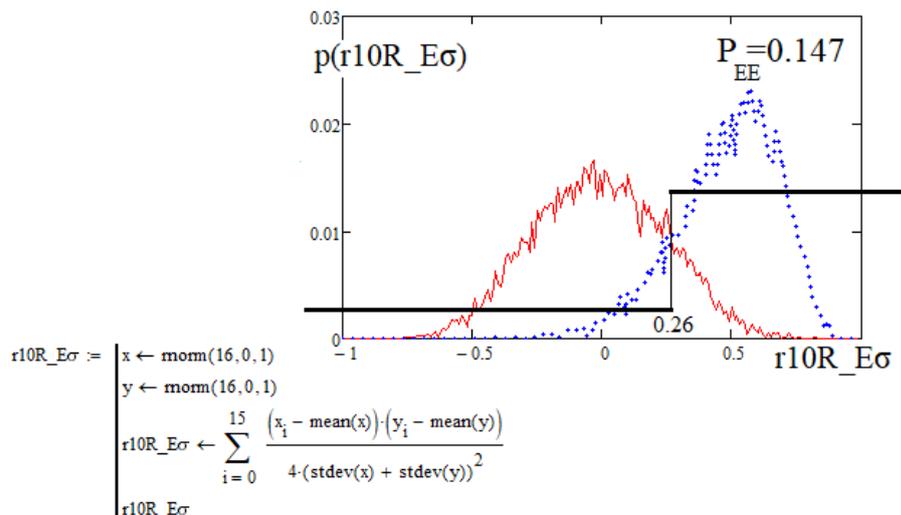


Рис. 2.12. Десятый непрерывный аналог критерия Пирсона – Эджуорта – Эдлтона, построенный через нормирование квадратом суммы стандартных отклонений

2.13. Корреляционная сцепленность группы аналогов классического критерия Пирсона – Эджуорта – Эдлтона (конец XIX в.)

Объединение рассматриваемых новых критериев между собой в тройки, пятерки, семерки следует выполнять с учетом их корреляционных связей, представленных на рис. 2.13. Так как все рассматриваемые критерии принадлежат к одной группе, между ними мы видим достаточно высокий уровень корреляционных связей.

$\text{corr}(r, rR) = 0.641$		
$\text{corr}(r, r1R) = 0.943$	$\text{corr}(rR, r1R) = 0.744$	
$\text{corr}(r, r2R_min) = 0.196$	$\text{corr}(rR, r2R_min) = 0.128$	$\text{corr}(r1R, r2R_min) = 0.193$
$\text{corr}(r, r3R_max) = 0.179$	$\text{corr}(rR, r3R_max) = 0.129$	$\text{corr}(r1R, r3R_max) = 0.178$
$\text{corr}(r, r4R_max_min) = 0.2$	$\text{corr}(rR, r4R_max_min) = 0.129$	$\text{corr}(r1R, r4R_max_min) = 0.191$
$\text{corr}(r, r5R_min_max) = 0.193$	$\text{corr}(rR, r5R_min_max) = 0.114$	$\text{corr}(r1R, r5R_min_max) = 0.183$
$\text{corr}(r, r6R_E_max) = 0.808$	$\text{corr}(rR, r6R_E_max) = 0.504$	$\text{corr}(r1R, r6R_E_max) = 0.769$
$\text{corr}(r, r7R_min_E) = 0.811$	$\text{corr}(rR, r7R_min_E) = 0.499$	$\text{corr}(r1R, r7R_min_E) = 0.771$
$\text{corr}(r, r8R_E_max) = 0.808$	$\text{corr}(rR, r8R_E_max) = 0.504$	$\text{corr}(r1R, r8R_E_max) = 0.769$
$\text{corr}(r, r9R_max_E) = 0.808$	$\text{corr}(rR, r9R_max_E) = 0.506$	$\text{corr}(r1R, r9R_max_E) = 0.767$
$\text{corr}(r, r10R_E\sigma) = 0.991$	$\text{corr}(rR, r10R_E\sigma) = 0.637$	$\text{corr}(r1R, r10R_E\sigma) = 0.952$

Рис. 2.13. Коэффициенты корреляции для вариантов классического критерия Эджуорта – Эдлтона – Пирсона (начало)

Высокий уровень корреляционных связей, характерный для начала, существенно снижается в продолжениях рис. 2.13.

Продолжение коэффициентов корреляции между группой критериев аналогов классической формулы Эджуорта – Эдлтона – Пирсона

corr(r2R_min, r3R_max) = 0.107
 corr(r2R_min, r4R_max_min) = -0.453 corr(r3R_max, r4R_max_min) = -0.47
 corr(r2R_min, r5R_min_max) = -0.474 corr(r3R_max, r5R_min_max) = -0.469 corr(r4R_max_min, r5R_min_max) = 0.103
 corr(r2R_min, r6R_E_max) = -0.0075 corr(r3R_max, r6R_E_max) = 0.3184 corr(r4R_max_min, r6R_E_max) = -0.0041
 corr(r2R_min, r7R_min_E) = 0.319 corr(r3R_max, r7R_min_E) = -0.016 corr(r4R_max_min, r7R_min_E) = 0.0018
 corr(r2R_min, r8R_E_max) = -0.0075 corr(r3R_max, r8R_E_max) = 0.3184 corr(r4R_max_min, r8R_E_max) = -0.0041
 corr(r2R_min, r9R_max_E) = 0.0038 corr(r3R_max, r9R_max_E) = 0.3135 corr(r4R_max_min, r9R_max_E) = 0.3314
 corr(r2R_min, r10R_Eσ) = 0.201 corr(r3R_max, r10R_Eσ) = 0.18 corr(r4R_max_min, r10R_Eσ) = 0.2

Окончание коэффициентов корреляции между аналогами классического критерия XIX в.

corr(r5R_min_max, r6R_E_max) = 0.3278
 corr(r5R_min_max, r7R_min_E) = 0.3305 corr(r6R_E_max, r7R_min_E) = 0.658
 corr(r5R_min_max, r8R_E_max) = 0.3278 corr(r6R_E_max, r8R_E_max) = 1 corr(r7R_min_E, r8R_E_max) = 0.6588
 corr(r5R_min_max, r9R_max_E) = -0.009 corr(r6R_E_max, r9R_max_E) = 0.654 corr(r7R_min_E, r9R_max_E) = 0.3905
 corr(r5R_min_max, r10R_Eσ) = 0.195 corr(r6R_E_max, r10R_Eσ) = 0.814 corr(r7R_min_E, r10R_Eσ) = 0.819
 corr(r8R_E_max, r9R_max_E) = 0.6546
 corr(r8R_E_max, r10R_Eσ) = 0.814 corr(r9R_max_E, r10R_Eσ) = 0.814

Рис. 2.13. Окончание

Список литературы к главе 2

1. Волчихин В. И., Иванов А. И., Сериков А. В., Серикова Ю. И. Квантовая суперпозиция дискретного спектра состояний математической молекулы корреляции для малых выборок биометрических данных // Вестник Мордовского университета. 2017. Т. 27, № 2. С. 230–243.
2. Сериков А. В., Качалин С. В. Корреляционная молекула с эллиптическими квантователями для вычислений на малых обучающих выборках // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. Пенза, 2019. С. 123–129.
3. Volchikhin V. I., Ivanov A. I., Zolotareva T. A., Skudnev D. M. Synthesis of four new neuro-statistical tests for testing the hypothesis of independence of small samples of biometric data // Journal of Physics: Conference Series. 2094 (2021). 032013. doi:10.1088/1742-6596/2094/3/032013

НОВЫЕ КРИТЕРИИ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОЛУЧЕННЫЕ ВАРИАЦИЯМИ КЛАССИЧЕСКИХ КРИТЕРИЕВ И ИХ КОМБИНИРОВАНИЕМ

3.1. Гибрид критерия Олмстеда – Тьюки с критерием Колмогорова – Смирнова

Созданные в XX в. статистические критерии для проверки гипотезы независимости могут использоваться как отдельно, так и совместно. Например, критерий Олмстеда – Тьюки может быть объединен с критерием Колмогорова – Смирнова. На рис. 3.1 приведена программная реализация гибрида двух критериев. На этом же рисунке представлены результаты численного эксперимента по моделированию гибрида критериев.

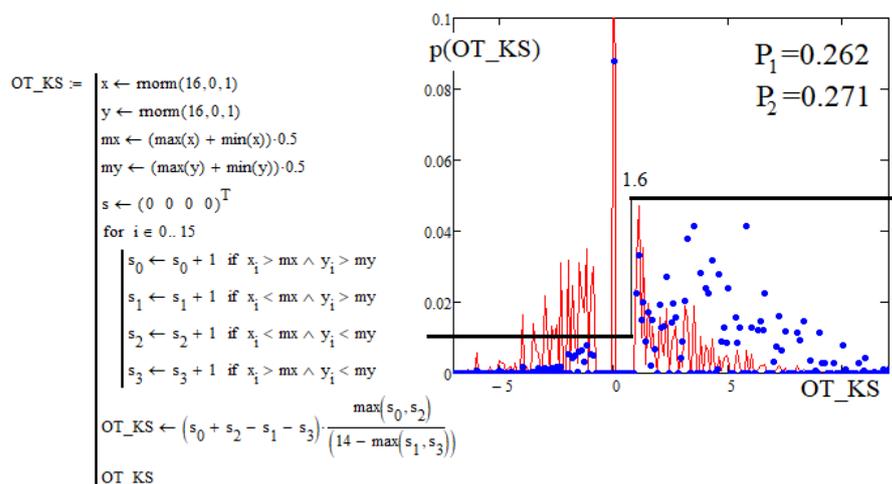


Рис. 3.1. Гибрид критерия Олмстеда – Тьюки
с критерием Колмогорова – Смирнова

Гибрид вполне работоспособен и обладает низкой корреляционной сцепленностью с исходным классическим критерием Олмстеда – Тьюки. Это означает, что многообразие комбинаторных статистических критериев значительно. Возможно использование как комбинаций статистических критериев, созданных в XX в. для проверки гипотезы независимости, так и комбинаций этих критериев с другими критериями, не предназначенными для проверки гипотезы независимости. В нашем случае только критерий Олмстеда – Тьюки решает основную задачу. Критерий Колмогорова – Смирнова создавался для решения других задач.

3.2. Второй вариант критерия Сендвиг – Олстеда

Кроме гибрида с другими критериями, возможна модификация одиночного классического критерия. Например, классический критерий Сендвиг – Олстеда строился на вычислении минимальных значений по двум координатам выборки. Очевидно, что поиск минимальных значений может быть заменен на поиск максимальных значений в малой выборке реальных данных. На рис. 3.2 представлена программная реализация модифицированного критерия Сендвиг – Олстеда.

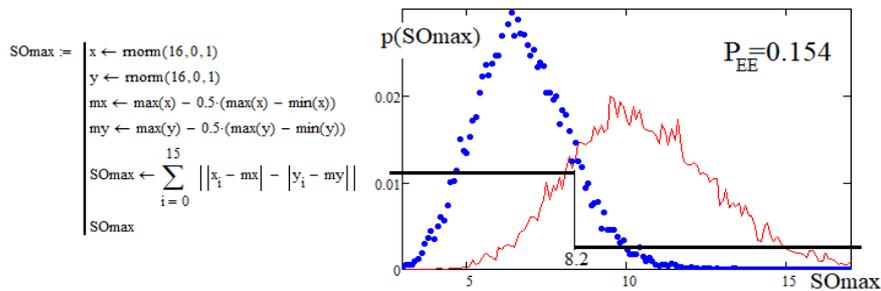


Рис. 3.2. Второй вариант критерия Сендвиг – Олстеда

3.3. Второй вариант критерия Спирмена

Базовый ранговый критерий Спирмена построен на упорядочивании данных по одной переменной, вычислении, округлении второй переменной до целой величины с повторным ранжированием и накоплением рангов. Можно отказаться от повторного ранжирования и накапливать модули разности номера отсчета и его ранга округления. Программная реализация второго варианта критерия Спирмена представлена на рис. 3.3.

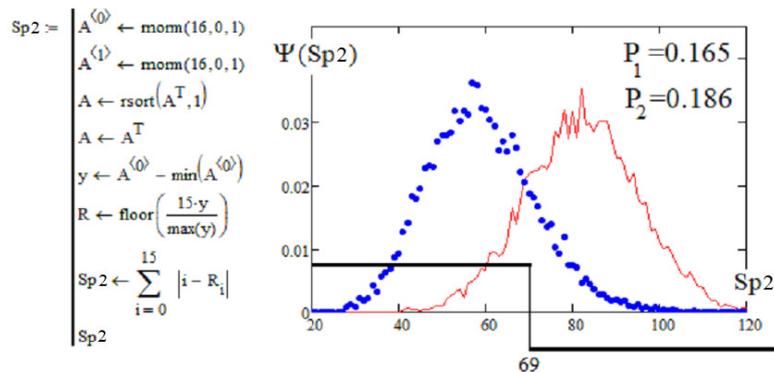


Рис. 3.3. Второй (более мощный) вариант критерия Спирмена, существенно повысивший разделимость независимых и зависимых данных

Результаты численного эксперимента показывают снижение вероятности ошибок первого и второго рода у новой версии критерия более чем в два раза.

3.4. Третий вариант критерия Спирмена

Еще один вариант критерия Спирмена может быть получен, если предыдущий критерий модифицировать, в частности такая модификация может быть получена накоплением модулей разности номеров и рангов округления после их умножения на ранги Кенделла по сравнению с исходным классическим критерием. Третий вариант критерия позволяет снизить вероятности ошибок первого и второго рода чуть меньше двух раз. Программная реализация третьего варианта критерия приведена на рис. 3.4.

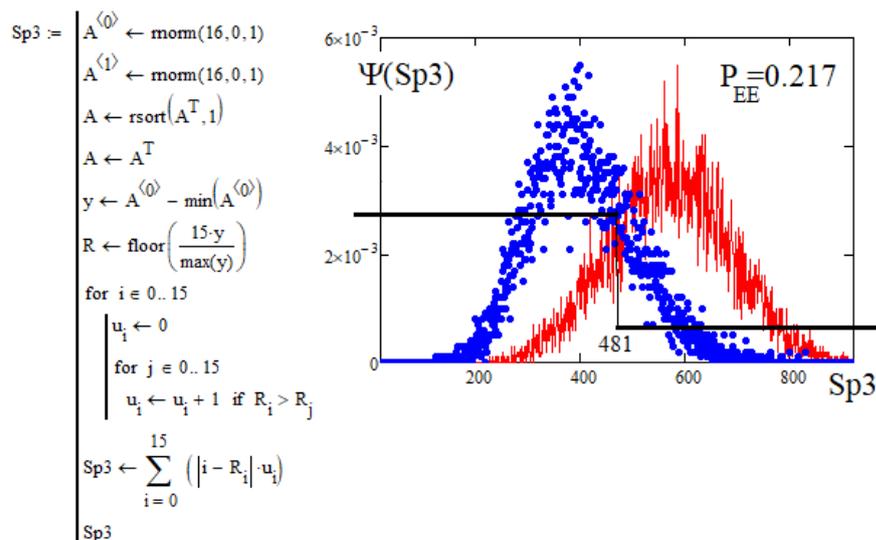


Рис. 3.4. Третий вариант критерия Спирмена

Все три варианта критериев Спирмена дают слабо коррелированные отклики (параграф 3.13).

3.5. Мультипликативный гибрид критериев Ван дер Вандена и Бартлеса

По прогнозам (параграф 1.22) для принятия решений с доверительной вероятности 0.99 при объеме малой выборки в 16 опытов может потребоваться использовать порядка 250 нейрокритериев. Получить столь значительное число статистических критериев можно разными способами. Например, можно использовать комбинирование 21 статистического критерия прошлого века. Потенциально подобное комбинирование двух статистических критериев может дать $(21^2 - 21)/2 = 210$, т.е. комбинирование уже созданных в XX в. критериев должно позволить приблизиться к достижению приемлемой для практики доверительной вероятности 0.99.

Примером подобного синтеза может служить объединение критериев Ван дер Вандена и Бартлеса (рис. 3.5), так как критерии дают разные варианты распределения положительно коррелированных данных. Положительно коррелированные данные имеют распределения откликов, расположенные по разные стороны от распределения независимых данных. В связи с этим используется мультипликативное объединение рассматриваемых критериев. У новой статистической конструкции отклик критерия Ван дер Вандена делится на отклик критерия Бартлеса. При этом вероятность ошибок нового мультипликативного критерия находится в интервале между вероятностями ошибок родительских критериев.

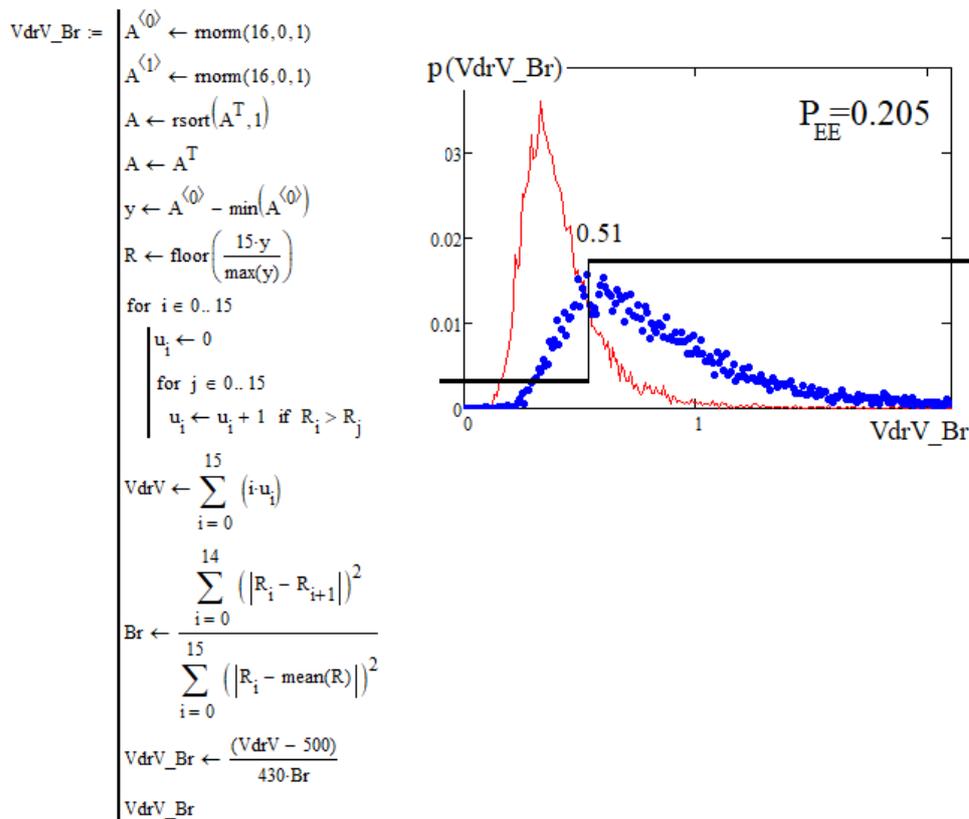


Рис. 3.5. Мультипликативный гибрид критерия Ван дер Вандена и критерия Бартлеса

При синтезе критерия использована сортировка входных данных по одной из переменных. Очевидно, что если использовать сортировку данных по другой переменной, то мы получим еще один вариант нового критерия.

3.6. Аддитивная сборка двух вариантов критерия Сендвиг – Олстеда

Ощутимое снижение вероятностей ошибок первого и второго рода при объединении двух критериев возможно в случае, когда вероятности ошибок у критериев-родителей сопоставимы, а корреляционная сцепленность их откликов низка. Это подтверждается аддитивным объединением двух вариантов критерия Сендвиг – Олстеда. У обоих критериев-родителей вероятности ошибок одинаковы и составляют – 0.154. Программная реализация нового критерия приведена на рис. 3.6.

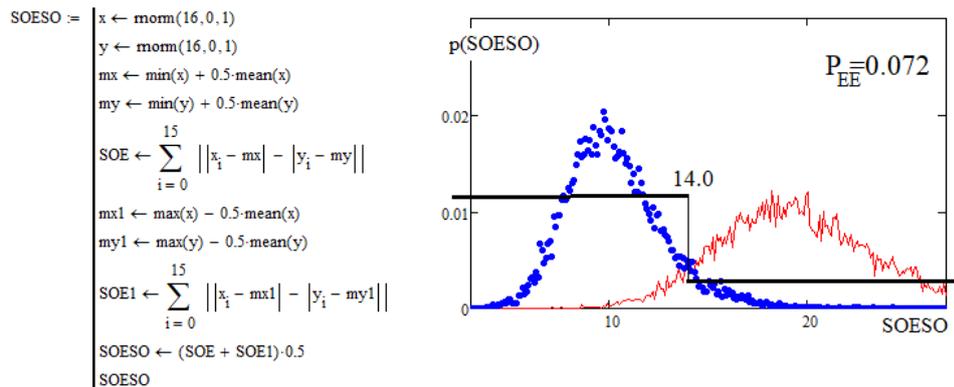


Рис. 3.6. Аддитивная сборка двух вариантов критерия Сендвиг – Олстеда

Численный эксперимент показал почти двухкратное снижение вероятностей ошибок первого и второго рода нового критерия по сравнению с показателями статистических критериев-родителей. Мы получили нейрокритерий с очень высоким уровнем разделимости зависимых и независимых данных малых выборок.

3.7. Мультипликативная сборка двух вариантов критерия Сендвиг – Олстеда

Аналогичных результатов удастся добиться, если заменить аддитивную сборку критериев-родителей на их мультипликативную сборку и вычислять среднее геометрическое. Результаты тестирования нового критерия отражены на рис. 3.7.

Следует подчеркнуть, что в рассмотренной ситуации сборка среднего арифметического хуже на 7.5 % по сравнению с мультипликативной сборкой.

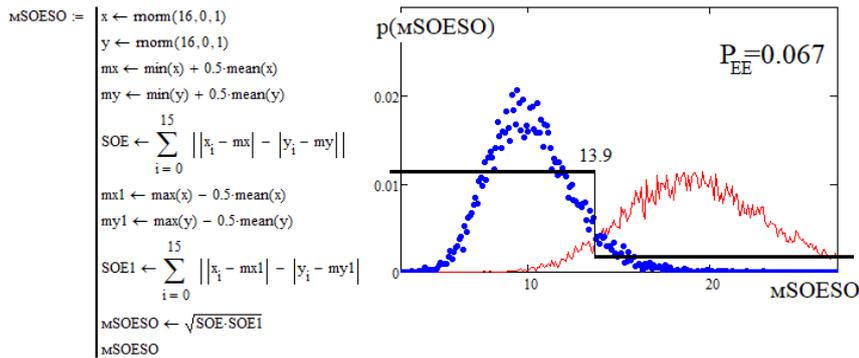


Рис. 3.7. Аддитивная сборка двух вариантов критерия Сендвиг – Олстеда

Высока вероятность того, что именно этот критерий является одним из самых мощных для проверки гипотезы независимости. Мощности нового нейрокритерия много выше мощностей классических нейрокритериев, синтезированных в прошлом веке.

3.8. Модификация критерия Кенделла за счет накапливания квадратов его рангов

Классический критерий Кенделла имеет высокий потенциал. В связи с этим представляют существенный интерес модификации этого критерия. Одной из эффективных модификаций критерия является переход от накапливания повторных сортировок к накапливанию их квадратов. При этом вероятности появления ошибок первого и второго рода у модифицированного критерия снижаются примерно на 21 %.

На рис. 3.8. даны программная реализация модифицированного критерия и результаты численного эксперимента по его моделированию.

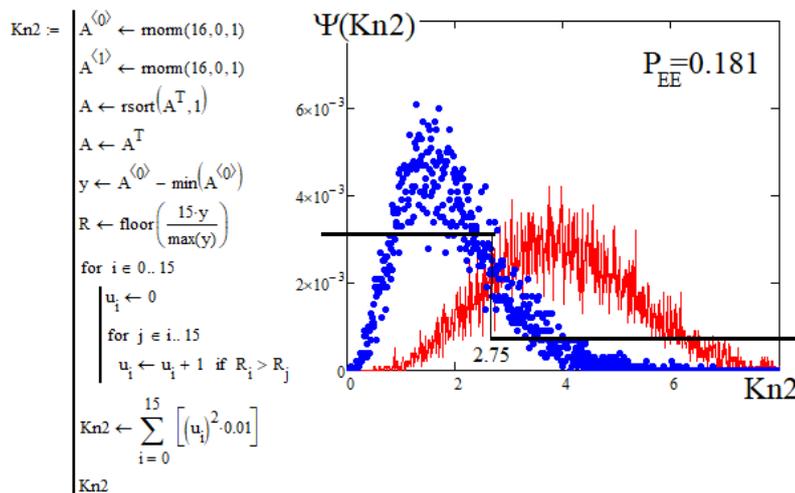


Рис. 3.8. Модифицированный критерий Кенделла за счет накапливания квадратов его рангов данных малой выборки

Очевидно, что, изменив переменную при упорядочивании, мы получим еще один статистический критерий, слабо связанный по его состояниям с рассмотренным выше.

3.9. Аддитивная сборка критерия Сендвиг – Олстеда и модифицированного критерия Кенделла

Критерий Сендвиг – Олстеда и модифицированный критерий Кенделла смещают в левую сторону распределения положительно коррелированных выборок. Это позволяет использовать эти два критерия в паре и взаимно усиливать их. На рис. 3.9 представлена программная реализация аддитивного объединения двух этих критериев. Критерий Сендвиг – Олстеда смещается, а критерий Кенделла масштабируется. В итоге усреднение их данных приводит к некоторому промежуточному значению 0.091 вероятности ошибок первого и второго рода по отношению к аналогичным параметрам критериев-родителей.

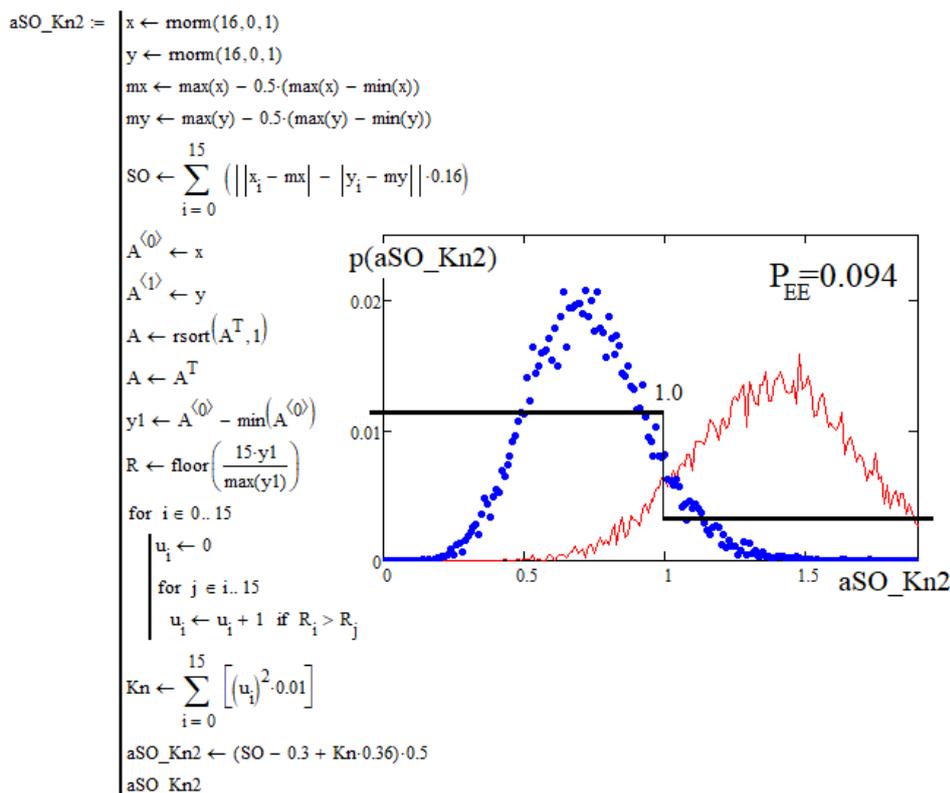


Рис. 3.9. Аддитивное объединение критерия Сендвиг – Олстеда и модифицированного критерия Кенделла

На рис. 3.9 отражены результаты численного моделирования. По аналогии с аддитивным критерием может быть построен и его мультипликативный аналог.

В итоге мы получили еще один новый нейрокритерий с очень высоким уровнем разделимости зависимых и независимых данных малой выборки.

3.10. Аддитивная сборка критерия Ван дер Вандена и критерия Эландта

Аддитивно можно собирать пары любых критериев, сдвигающих распределения положительно коррелированных данных в одну сторону, в частности критерий Ван дер Вандена и критерий Эландта сдвигают положительно коррелированные данные в правую сторону. Программная реализация нового аддитивного критерия приведена на рис. 3.10. На этом же рисунке представлены результаты численного эксперимента.

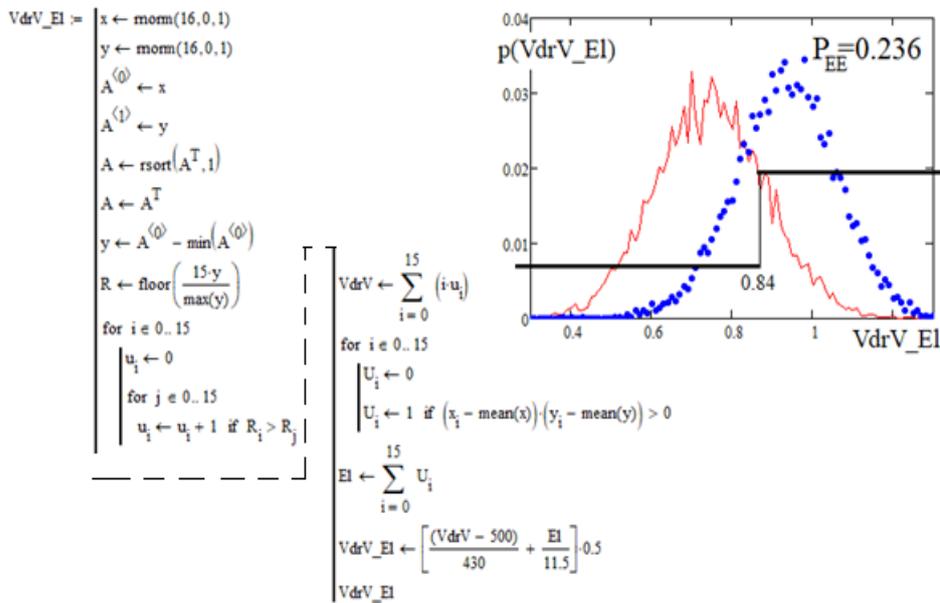


Рис. 3.10. Аддитивная сборка критерия Ван дер Вандена и критерия Эландта

В параграфе 3.14 даны значения коэффициентов корреляции между данным критерием и другими критериями этой группы.

3.11. Аддитивная сборка критерия Шахани и модифицированного критерия Кенделла

Еще одна аддитивная сборка со сдвигом зависимых данных в правую сторону получается, если объединять критерий Шахани и критерий Кенделла. Реализация этого комбинационного критерия приведена на рис. 3.11.

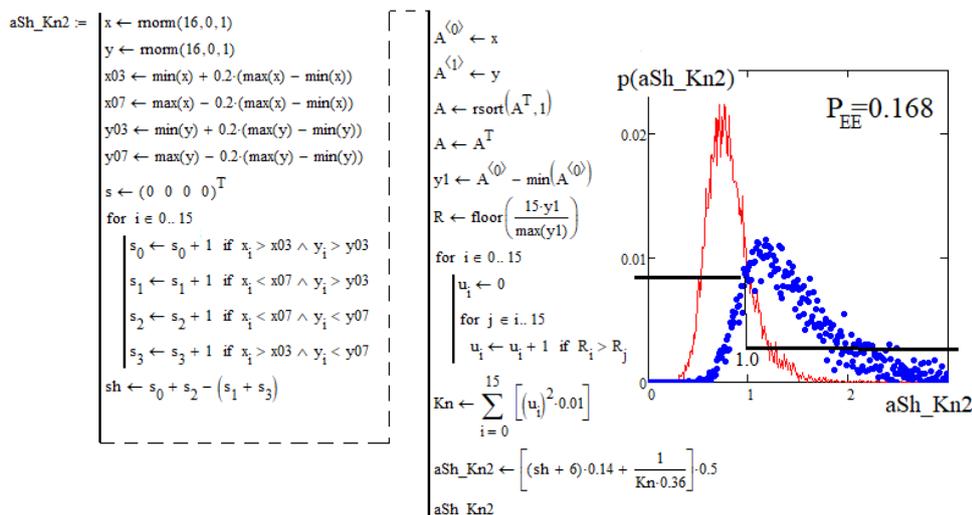


Рис. 3.11. Аддитивная сборка критерия Шахани и модифицированного критерия Кенделла

Следует отметить, что критерий Шахани – Эландта существует в четырех вариантах. Для того, чтобы из первого варианта получить второй вариант критерия, достаточно внести изменения в программную реализацию. Необходимо в реализации критерия Шахани изменить сортировку данных по одной переменной на сортировку данных по другой переменной. Кроме того, параллельно двум аддитивным вариантам критерия могут быть построены два мультипликативных варианта критерия.

3.12. Аддитивная сборка критерия Шахани и автокорреляционного критерия Кноке

В первой главе данного справочника приведены реализации 21 классического статистического критерия, т.е. критерий Шахани может быть скомбинирован как минимум с 20 другими статистическими критериями, созданными в XX в. для проверки гипотезы независимости. Как подтверждение этого тезиса, рассмотрим аддитивную сборку

критерия Шахани и автокорреляционного критерия Кноке. Программная реализация их объединения приведена на рис. 3.12.

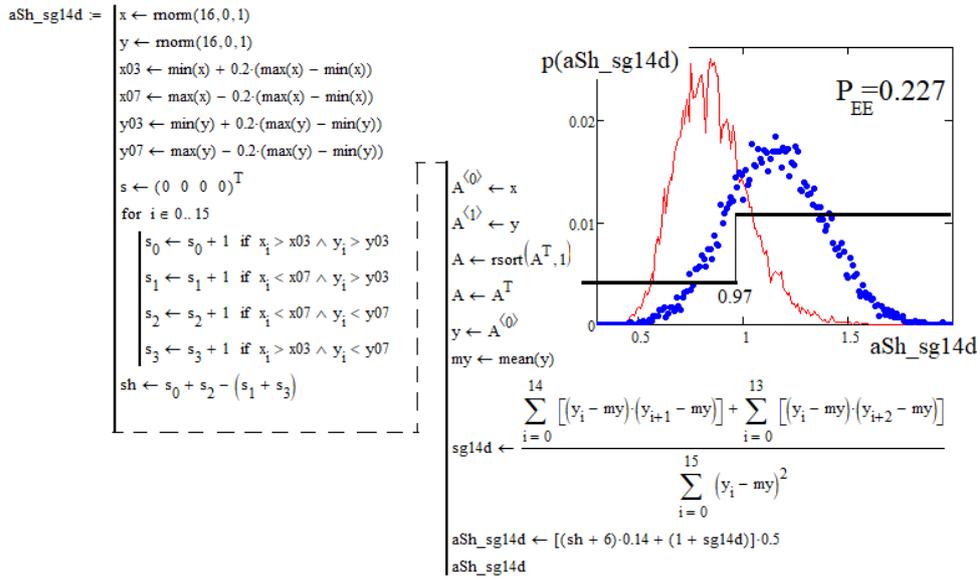


Рис. 3.12. Аддитивная сборка критерия Шахани и корреляционного критерия Кноке

Так как автокорреляционный критерий Кноке построен на упорядочивании данных по одной переменной, замена переменной упорядочивания приводит к появлению другого статистического критерия. Также могут быть получены еще два других критерия, если заменить аддитивную сборку на ее мультипликативный аналог.

3.13. Аддитивная сборка двух вариантов критерия Кенделла

Самой мощной по разделимости зависимых и независимых данных, рассмотренной в данной главе является мультипликативная сборка двух вариантов критериев Сендвиг – Олстеда. Объединяемые критерии похожи друг на друга и отличаются только точками наблюдения одной и той же малой выборки.

Видимо, все сборки пар родственных двухмерных критериев, полученных двумя упорядочиваниями по разным переменным, также должны давать значительный рост разделимости зависимых и независимых данных. Этот тезис хорош и подтверждается на примере аддитивной сборки двух вариантов критерия Кенделла. Программная реализация этого варианта сборки приведена на рис. 3.13.

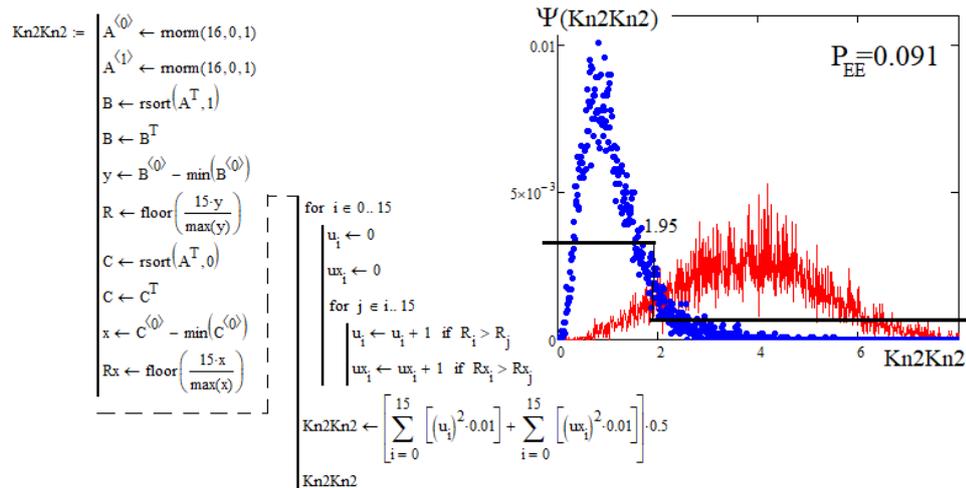


Рис. 3.13. Аддитивная сборка двух вариантов модифицированного критерия Кенделла

Каждый из исходных вариантов модифицированного нейрокритерия Кенделла обеспечивает равные вероятности ошибок первого и второго рода на уровне 0.181. Их аддитивное объединение позволяет получить снижение вероятностей ошибок почти в два раза. Это очень хороший показатель. Мы получили новый статистический нейрокритерий с очень высоким уровнем разделимости зависимых и независимых данных по сравнению с наиболее популярным классическим Эджуорта – Эдлтона – Пирсона.

3.14. Корреляционная сцепленность в группе критериев, полученных комбинированием классических критериев XX в.

Выше была рассмотрена часть критериев (12 критериев), полученных комбинированием внутри группы из 21 классического критерия XX в. Рассмотрим ситуацию, когда эти две группы используются отдельно, преобразуя одни и те же входные данные в выходной избыточный код нейросети. При этом будем сравнивать среднее значение модулей коэффициентов корреляции и среднее значение вероятностей ошибок по той или иной группе нейрокритериев.

При усреднении модулей коэффициентов корреляции воспользуемся 21 первым значением рис. 1.26. Если при вычислениях учитывать только две значащие цифры после точки, то в итоге мы получим следующий результат:

$$\text{mean}(.13, .01, .4, .0, .14, .1, .01, .02, .01, 0.13, .01, .01, .01, .06, .02, .08, .02, .09, .06, .6, .04) = 0.093.$$

По аналогии с рис. 1.26 создадим рис. 3.14 для критериев, полученных комбинированием классических критериев XX в.

$\text{corr}(r, \text{OT_KS}) = 0.573$		
$\text{corr}(r, \text{SOMax}) = -0.377$	$\text{corr}(\text{OT_KS}, \text{SOMax}) = -0.255$	
$\text{corr}(r, \text{Sp2}) = -0.841$	$\text{corr}(\text{OT_KS}, \text{Sp2}) = -0.551$	$\text{corr}(\text{SOMax}, \text{Sp2}) = 0.342$
$\text{corr}(r, \text{Sp3}) = -0.801$	$\text{corr}(\text{OT_KS}, \text{Sp3}) = -0.476$	$\text{corr}(\text{SOMax}, \text{Sp3}) = 0.303$
$\text{corr}(r, \text{VdrV_Br}) = -0.425$	$\text{corr}(\text{OT_KS}, \text{VdrV_Br}) = -0.258$	$\text{corr}(\text{SOMax}, \text{VdrV_Br}) = 0.182$
$\text{corr}(r, \text{SOESO}) = -0.523$	$\text{corr}(\text{OT_KS}, \text{SOESO}) = -0.446$	$\text{corr}(\text{SOMax}, \text{SOESO}) = 0.552$
$\text{corr}(r, \text{mSOESO}) = -0.542$	$\text{corr}(\text{OT_KS}, \text{mSOESO}) = -0.450$	$\text{corr}(\text{SOMax}, \text{mSOESO}) = 0.563$
$\text{corr}(r, \text{Kn2}) = -0.855$	$\text{corr}(\text{OT_KS}, \text{Kn2}) = -0.514$	$\text{corr}(\text{SOMax}, \text{Kn2}) = 0.353$
$\text{corr}(r, \text{aSO_Kn2}) = -0.688$	$\text{corr}(\text{OT_KS}, \text{aSO_Kn2}) = -0.420$	$\text{corr}(\text{SOMax}, \text{aSO_Kn2}) = 0.478$
$\text{corr}(r, \text{VdrV_El}) = 0.250$	$\text{corr}(\text{OT_KS}, \text{VdrV_El}) = 0.360$	$\text{corr}(\text{SOMax}, \text{VdrV_El}) = -0.140$
$\text{corr}(r, \text{aSh_Kn2}) = 0.830$	$\text{corr}(\text{OT_KS}, \text{aSh_Kn2}) = 0.848$	$\text{corr}(\text{SOMax}, \text{aSh_Kn2}) = -0.352$
$\text{corr}(r, \text{aSh_sg14d}) = 0.682$	$\text{corr}(\text{OT_KS}, \text{aSh_sg14d}) = 0.853$	$\text{corr}(\text{SOMax}, \text{aSh_sg14d}) = -0.296$
$\text{corr}(r, \text{Kn2Kn2}) = -0.926$	$\text{corr}(\text{OT_KS}, \text{Kn2Kn2}) = -0.560$	$\text{corr}(\text{SOMax}, \text{Kn2Kn2}) = 0.393$

Рис. 3.14. Коэффициенты корреляции группы нейрокритериев, полученных комбинированием критериев XX в. (начало)

Усреднение модулей первых коэффициентов корреляции рис. 3.14 дает следующий результат:

$\text{mean}(0.57, .37, .84, .8, .43, .52, .54, .86, .69, .25, .83, .68, .92, .25, .55, .48, .26, .45, .51, .42, .36) = 0.551$.

Сравнивая результат усреднения рис. 1.26 и рис. 3.14, мы видим повышение уровня корреляционной сцепленности данных примерно в шесть раз. Такое соотношение коэффициентов корреляционной сцепленности характерно не только для начала рис. 1.26 и рис. 3.14, но и для их окончаний.

Коэффициенты корреляции, полученные комбинированием критериев XX в. (продолжение)

$\text{corr}(\text{Sp2}, \text{Sp3}) = 0.766$		
$\text{corr}(\text{Sp2}, \text{VdrV_Br}) = 0.416$	$\text{corr}(\text{Sp3}, \text{VdrV_Br}) = 0.275$	
$\text{corr}(\text{Sp2}, \text{SOESO}) = 0.498$	$\text{corr}(\text{Sp3}, \text{SOESO}) = 0.417$	$\text{corr}(\text{VdrV_Br}, \text{SOESO}) = 0.223$
$\text{corr}(\text{Sp2}, \text{mSOESO}) = 0.518$	$\text{corr}(\text{Sp3}, \text{mSOESO}) = 0.435$	$\text{corr}(\text{VdrV_Br}, \text{mSOESO}) = 0.237$
$\text{corr}(\text{Sp2}, \text{Kn2}) = 0.806$	$\text{corr}(\text{Sp3}, \text{Kn2}) = 0.902$	$\text{corr}(\text{VdrV_Br}, \text{Kn2}) = 0.296$
$\text{corr}(\text{Sp2}, \text{aSO_Kn2}) = 0.717$	$\text{corr}(\text{Sp3}, \text{aSO_Kn2}) = 0.697$	$\text{corr}(\text{VdrV_Br}, \text{aSO_Kn2}) = 0.276$
$\text{corr}(\text{Sp2}, \text{VdrV_El}) = -0.328$	$\text{corr}(\text{Sp3}, \text{VdrV_El}) = -0.320$	$\text{corr}(\text{VdrV_Br}, \text{VdrV_El}) = 0.132$
$\text{corr}(\text{Sp2}, \text{aSh_Kn2}) = -0.784$	$\text{corr}(\text{Sp3}, \text{aSh_Kn2}) = -0.750$	$\text{corr}(\text{VdrV_Br}, \text{aSh_Kn2}) = -0.386$
$\text{corr}(\text{Sp2}, \text{aSh_sg14d}) = -0.639$	$\text{corr}(\text{Sp3}, \text{aSh_sg14d}) = -0.562$	$\text{corr}(\text{VdrV_Br}, \text{aSh_sg14d}) = -0.495$
$\text{corr}(\text{Sp2}, \text{Kn2Kn2}) = 0.842$	$\text{corr}(\text{Sp3}, \text{Kn2Kn2}) = 0.835$	$\text{corr}(\text{VdrV_Br}, \text{Kn2Kn2}) = 0.435$

Рис. 3.14. Продолжение

Коэффициенты корреляции, полученные комбинированием критериев XX в. (окончание)

```

corr(SOESO, мSOESO) = 0.997
corr(SOESO, Kn2) = 0.448      corr(мSOESO, Kn2) = 0.467
corr(SOESO, аSO_Kn2) = 0.709  corr(мSOESO, аSO_Kn2) = 0.722      corr(Kn2, аSO_Kn2) = 0.784
corr(SOESO, VdrV_EI) = -0.185 corr(мSOESO, VdrV_EI) = -0.191      corr(Kn2, VdrV_EI) = -0.366
corr(SOESO, аSh_Kn2) = -0.525 corr(мSOESO, аSh_Kn2) = -0.539      corr(Kn2, аSh_Kn2) = -0.813
corr(SOESO, аSh_sg14d) = -0.464 corr(мSOESO, аSh_sg14d) = -0.473      corr(Kn2, аSh_sg14d) = -0.600
corr(SOESO, Kn2Kn2) = 0.481  corr(мSOESO, Kn2Kn2) = 0.500      corr(Kn2, Kn2Kn2) = 0.924
corr(аSO_Kn2, VdrV_EI) = -0.307
corr(аSO_Kn2, аSh_Kn2) = -0.659  corr(VdrV_EI, аSh_Kn2) = 0.383
corr(аSO_Kn2, аSh_sg14d) = -0.490 corr(VdrV_EI, аSh_sg14d) = 0.331      corr(аSh_Kn2, аSh_sg14d) = 0.891
corr(аSO_Kn2, Kn2Kn2) = 0.722  corr(VdrV_EI, Kn2Kn2) = -0.256  corr(аSh_Kn2, Kn2Kn2) = -0.827

```

Рис. 3.14. Окончание

Причиной роста корреляционной сцепленности синтезированных критериев является стремление увеличить их мощность по разделности зависимых и независимых данных малой выборки.

Для оценки разделности данных следует воспользоваться усредненными значениями вероятностей ошибок – P_{EE} для непрерывных спектров и минимальным значением ошибок первого и второго рода для дискретных спектров. Данные будем брать из рисунков первой главы для классических критериев XX в. Как результат получим:

$$\text{mean}(0.14, .27, .22, .35, .27, .32, .22, .16, .12, .22, .4, .31, .35, .18, .33, .32, .34, .24, .41, .38, .31) = 0.279.$$

По данным рисунков третьей главы мы также можем получить среднее значение вероятностей ошибок – P_{EE} :

$$\text{mean}(0.26, 0.15, 0.17, 0.22, 0.21, 0.07, 0.07, 0.18, 0.09, 0.24, 0.17, 0.23, 0.09) = 0.165.$$

Получается, что, снизив вероятности ошибок примерно в 1.8 раза за счет комбинации классических критериев, мы наблюдаем существенный рост корреляционной сцепленности данных примерно в шесть раз. Всегда для конечной выборки первоначальных «сырых» критериев проверки гипотезы независимости существуют две отмеченные выше противоположные тенденции. Попытки комбинаторикой снизить вероятности ошибок первого и второго рода всегда приводят к росту корреляционной сцепленности в новой группе критериев с большей разделностью зависимых и независимых данных. В этом отношении принципиально важно изначально иметь достаточно большую группу критериев, построенных в существенно нелинейных по отношению друг к другу двухмерных пространствах накопления данных. Далее будут рассмотрены несколько вариантов решения этой задачи.

НОВЫЕ КРИТЕРИИ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОЛУЧЕННЫЕ УПОРЯДОЧИВАНИЕМ ДАННЫХ ПО ОДНОЙ И ДВУМ ПЕРЕМЕННЫМ ДАННЫХ МАЛОЙ ВЫБОРКИ

4.1. Континуальный критерий, построенный на упорядочивании одной переменной и сглаживанием второй случайной составляющей

Значительная часть классических критериев, построенных в прошлом веке, используют процедуру упорядочивания данных по одной из переменных исходной выборки. Описание части подобных критериев даны в первой главе данного справочника. К ним относятся критерии: Аббе – Линника, Кенделла, Спирмена, Эйзенхарта, Рамачандрана – Ранганатана, Ван дер Вандена, Шукени – Флори, Бартлеса, Шахнесси, Казаковичуса, Кноке.

Очевидно, что вычислительный прием упорядочивания данных малой выборки по одной или двум переменным будет активно использоваться при синтезе новых статистических критериев и в этом веке [1, 25].

В качестве примера рассмотрим критерий, построенный на упорядочивании данных малой выборки по переменной – x , сглаживании данных по трем отсчетам оставшейся случайной переменной – y . Далее производится оценка стандартного отклонения сглаженных данных. На рис. 4.1 приведены программная реализация этого критерия и результаты численного моделирования его статистических параметров.

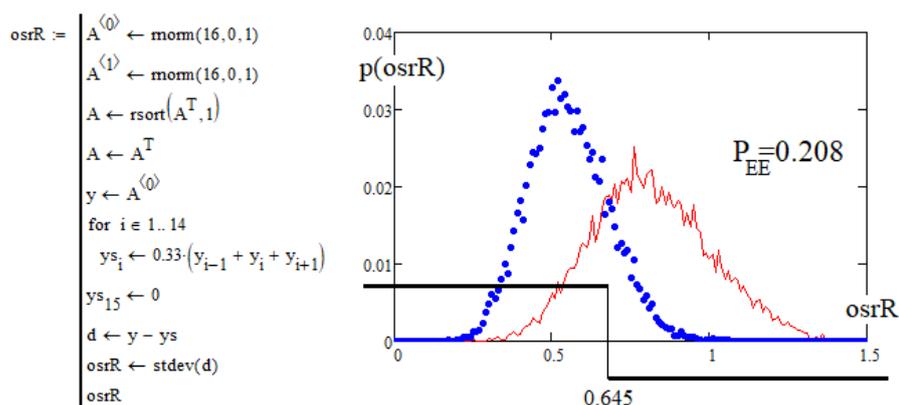


Рис. 4.1. Континуальный критерий, построенный на упорядочивании одной переменной и сглаживании второй оставшейся случайной переменной по трем соседним отсчетам

Очевидно, что в программную реализацию нового критерия могут быть введены изменения, приводящие к первоначальной сортировке данных выборки по другой переменной. В этом случае мы получим другой критерий с той же вероятностью ошибок первого и второго рода $P_{EE} = 0.208$, однако еще один критерий слабо коррелирован с первым критерием $\text{corr}(\text{osrRx}, \text{osrRy}) \approx 0.156$.

4.2. Аддитивная сборка двух континуальных критериев, построенных на упорядочивании одной переменной со сглаживанием случайной составляющей и вычислением ее остатков

В силу того, что два критерия, полученные упорядочиванием данных по первой и второй переменным малой выборки, обладают слабой коррелированностью $\text{corr}(\text{osrRx}, \text{osrRy}) \approx -0.156$, появляется возможность аддитивно объединить два новых критерия. Программное обеспечение аддитивного объединения критериев приводится на рис. 4.2. На этом же рисунке отражены результаты численного моделирования.

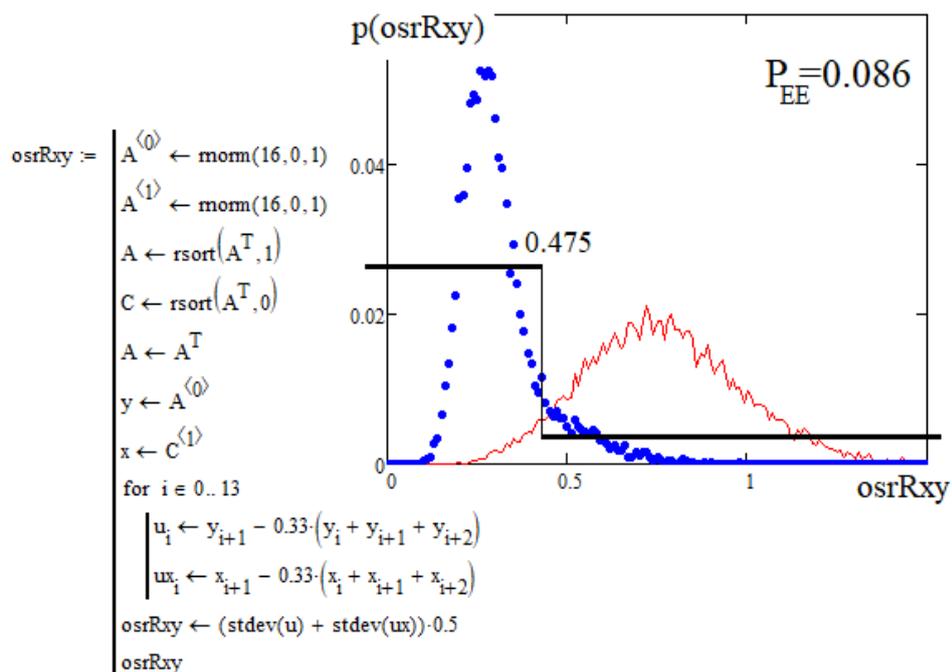


Рис. 4.2. Аддитивная сборка двух критериев, построенных на упорядочивании двух переменных и сглаживании оставшихся случайных данных

Результаты численного моделирования показали возможность снижения вероятностей ошибок двух критериев в 2.4 раза. Это является существенным усилением двух новых статистических критериев. Фундаментальных работ по оптимизации параметров аддитивного свертывания на данный момент пока нет. На данный момент нельзя указать оптимальные параметры аддитивного и мультипликативного свертывания объединяемых нейрокритериев.

Синтезированный выше статистический нейрокритерий более чем в два раза мощнее классического критерия Эджуорта – Эдлтона – Пирсона конца XIX в., что подтверждает эффективность уже выполненных исследований в этом направлении.

4.3. Континуальный критерий, построенный на упорядочивании одной переменной с дифференцированием составляющей случайной переменной

Одним из очевидных способов повышения эффективности свертывания является повышение качества принятия решений двух исходных компонент. Сделать это удастся, если линейное сглаживание случайных данных по трем рядом стоящим соседним точкам заменить преобразованием двойного дифференцирования. Ранее было показано, что использование операций численного дифференцирования достаточно часто приводит к существенному повышению качества принимаемых решений [26, 27].

Этот же технологический прием хорошо работает для рассматриваемых статистических преобразований. Программная реализация нового критерия представлена на рис. 4.3. На этом же рисунке – результаты численного моделирования. В соответствии с ними наблюдается эффект снижения вероятностей ошибок первого и второго рода – P_{EE} на 40.5 %.

Фактически мы получили еще один новый нейрокритерий, эквивалентный самому популярному классическому статистическому критерию Эджуорта – Эдлтона – Пирсона конца XIX в. При этом корреляционная сцепленность классического нейрокритерия Эджуорта – Эдлтона – Пирсона и нового нейрокритерия низка $\text{corr}(r, \text{osr}R_{xy}) \approx -0.011$, т.е. пара двух этих критериев должна давать еще один нейрокритерий с рекордно высокой разделимостью малых выборок с зависимыми и независимыми данными.

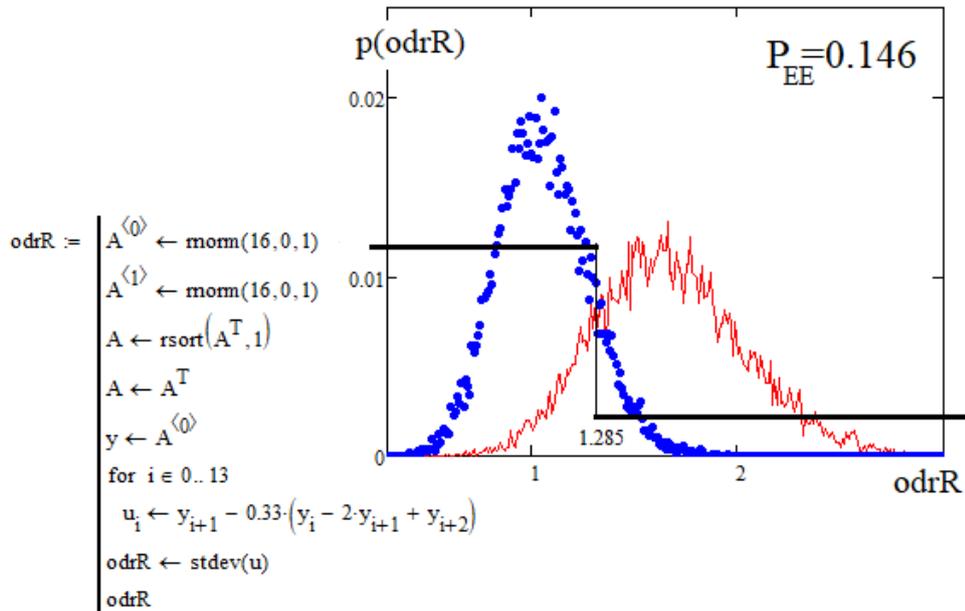


Рис. 4.3. Критерий, построенный на упорядочивании одной переменной и дифференцировании второй переменной

Если изменить переменную упорядочивания в программной реализации рис. 4.3, мы получаем еще один статистический критерий. Исходный критерий и его новый вариант имеют низкое корреляционное сцепление $\text{corr}(\text{odrRx}, \text{odrRy}) = -0.316$.

4.4. Аддитивная сборка двух критериев, построенных на упорядочивании одной переменной с дифференцированием составляющей случайной переменной

В силу низкой корреляционной сцепленности двух исходных статистических критериев их аддитивное объединение дает хороший результат. Программа, воспроизводящая объединение двух статистических критериев для малых выборок независимых данных, приведена на рис. 4.4. По результату численного моделирования получаем оценку $0.146/0.073 = 2$, т.е. мы наблюдаем двукратное снижение вероятности ошибок первого и второго рода при аддитивном объединении двух симметричных статистических критериев с двойным дифференцированием случайных данных.

Мы имеем меньшее снижение вероятностей ошибок первого и второго рода (до двух раз) по сравнению с аналогичным показателем предшествующей аддитивной сборки (снижение вероятности ошибок

P_{EE} в 2.4 раза). Это связано с тем, что коэффициенты корреляции аддитивно сцепляемых критериев имеют разные модули и разные знаки. Видимо, аналогичная ситуация возникает и при мультипликативном объединении рассматриваемых статистических критериев.

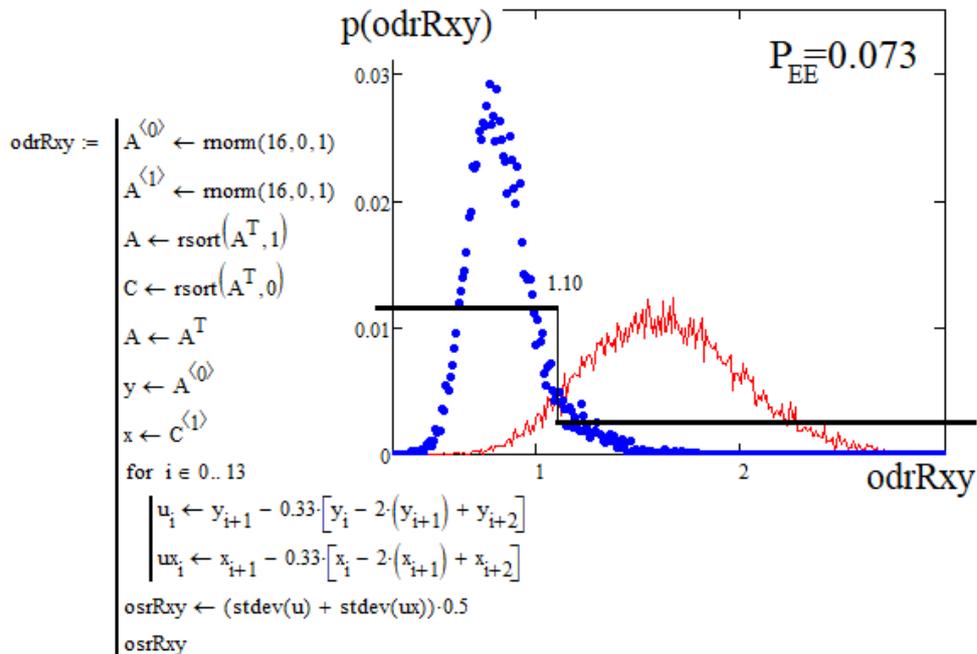


Рис. 4.4. Аддитивная сборка двух критериев, построенных на упорядочивании одной переменной с дифференцированием составляющей случайной переменной

Как результат мы получили еще один вариант нейрокритерия с рекордной для классических нейрокритериев способностью к разделению зависимых и независимых данных.

4.5. Континуальный критерий, построенный на упорядочивании одной переменной и вычислении суммы расстояний Евклида по другой переменной

Еще одним вариантом синтеза новых критериев является подсчет Евклидовых квадратов координат отсчетов случайной составляющей, оставшейся случайной после сортировки данных по другой переменной. Программная реализация нового критерия представлена на рис. 4.5. На этом же рисунке представлены результаты численного эксперимента по моделированию созданного нейрокритерия. К сожалению, вероятность одинаковых ошибок первого и второго рода для созданного критерия оказывается высокой.

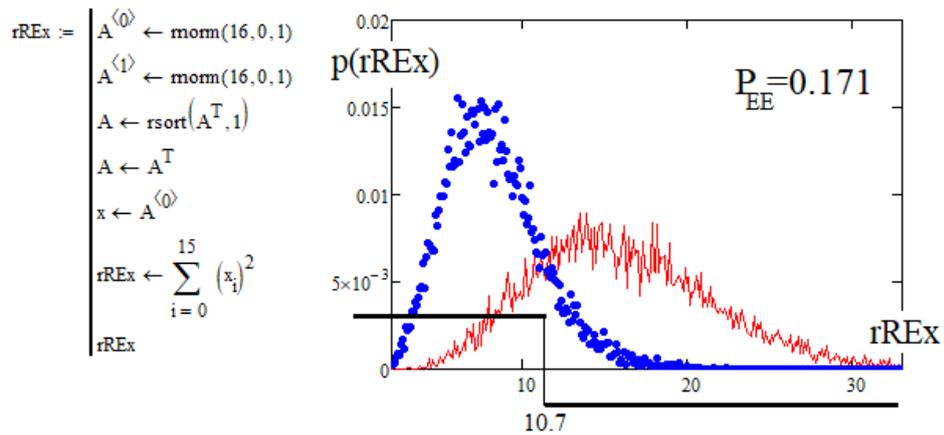


Рис. 4.5. Критерий, построенный на упорядочивании одной переменной и накоплении Евклидовых расстояний для координат второй переменной

4.6. Континуальный критерий, построенный на упорядочивании одной переменной и накоплении квадратов разностей расстояний между соседними точками по второй переменной

Для существенного снижения уровня вероятности ошибок первого и второго рода рассматриваемого класса квадратичных нейрокритериев целесообразно выполнить замену входных данных на их разность (выполнить численное дифференцирование данных). Реализацию этой модификации нейрокритерия поясняет рис. 4.6. Из данных рисунка следует, что снижение вероятности ошибок первого и второго рода в $0.171/0.146 \approx 1.17$ раз.

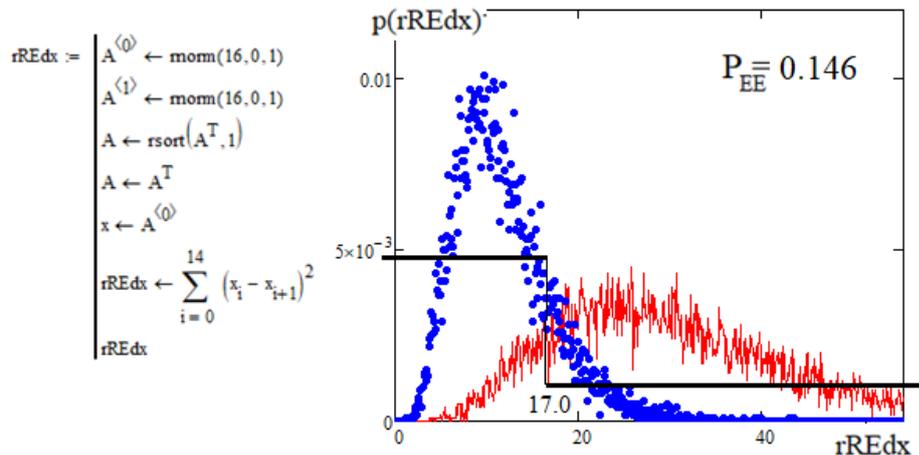


Рис. 4.6. Критерий, построенный на упорядочивании одной переменной и накоплении квадратов расстояний между соседними точками по второй переменной

Если в программной реализации нейрокритерия рис. 4.6 изменить переменную упорядочивания, то мы получим еще один новый нейрокритерий. Исходный критерий и его новый вариант имеют низкое положительное корреляционное сцепление $\text{corr}(rREdx, rREdy) = 0.143$.

4.7. Аддитивная сборка двух континуальных критериев, построенных на упорядочивании одной переменной и накоплении квадратов разностей расстояний между соседними точками по второй переменной

Пользуясь низким уровнем корреляционной сцепленности двух симметричных критериев, мы имеем возможность аддитивно объединить их. Программная реализация этого объединения приведена на рис. 4.7. При этом достигается снижение вероятности ошибок первого и второго рода – P_{EE} примерно на $0.146/0.075 \approx 1.95$.

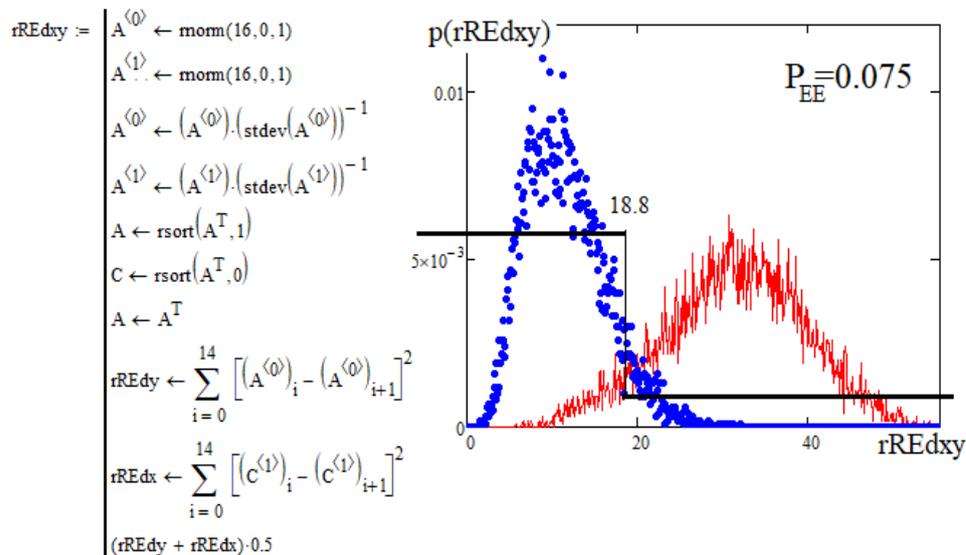


Рис. 4.7. Аддитивная сборка двух критериев, построенных на упорядочивании одной переменной и накоплении квадратов расстояний между соседними точками по второй переменной

Снижение вероятности ошибок первого и второго рода почти в два раза обусловлено, в первую очередь, низкой корреляционной сцепленностью двух объединяемых нейрокритериев. Мы получили нейрокритерий с мощностью много выше, чем мощность классических нейрокритериев.

4.8. Континуальный критерий, построенный на упорядочивании одной переменной и накоплении произведения расстояний между соседними точками по второй переменной

В предыдущей главе (параграфы 3.6 и 3.7) было показано, что аддитивные сборки двух критериев и мультипликативные сборки тех же двух критериев имеют сопоставимые мощности. Кроме того, известно, что дифференцирование случайных данных в ряде случаев дает хороший результат при попытках увеличить мощность статистических критериев [26, 27]. Все это заставляет внимательно относиться не только к разностным схемам предобработки случайных данных [26, 27], но и близким к ним мультипликативным схемам подобной предобработки. В прошлом веке этот тезис в неявной форме был использован в критериях, построенных с перемножением рангов. К таким математическим конструкциям с дискретными распределениями относятся следующие критерии: Эландта 1962 г., Ван дер Вандена 1967 г., Спирмена 1971 г., Шукени – Флори 1973 г., Шахнесси 1981 г.

В связи с этим перспективной является ветвь критериев, построенных на перемножении координат соседних точек оставшихся не упорядоченными данных. Пример реализации такого критерия с непрерывным спектром представлен на рис. 4.8.

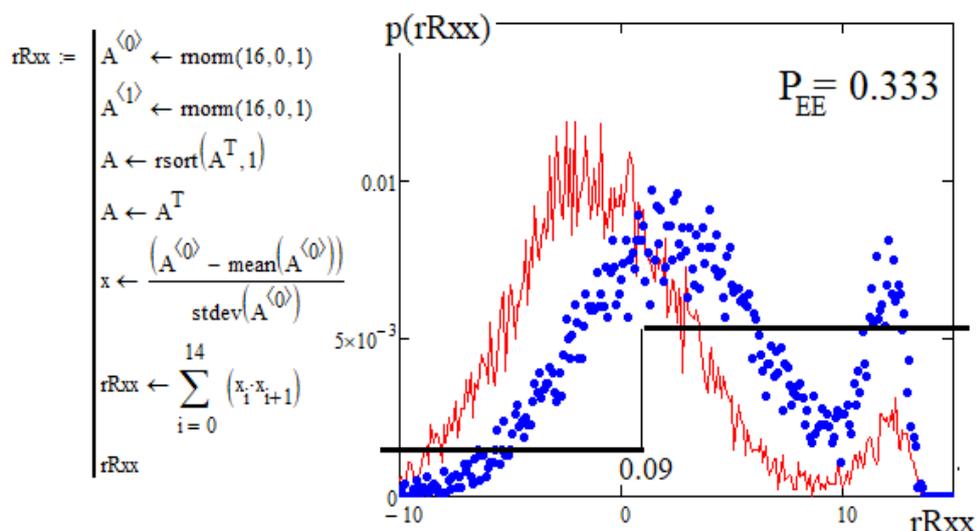


Рис. 4.8. Континуальный критерий, построенный на упорядочивании одной переменной и накоплении произведения расстояний между соседними точками по второй переменной

Численное моделирование нового нейрокритерия показывает его низкую разделяющую способность даже по сравнению с классическими статистическими критериями.

4.9. Континуальный критерий, построенный на упорядочивании одной переменной и вычислении стандартного отклонения произведения расстояний между соседними точками по второй переменной

Существенно улучшить разделяющую способность подобных математических конструкций удастся, если рассматривать каждую мультипликативную пару координат соседних точек как некоторый промежуточный параметр и оценивать их стандартное отклонение.

Пример программной реализации нового нейрокритерия представлен на рис. 4.9. Его тестирование показывает высокий уровень делимости зависимых и независимых данных на малых выборках.

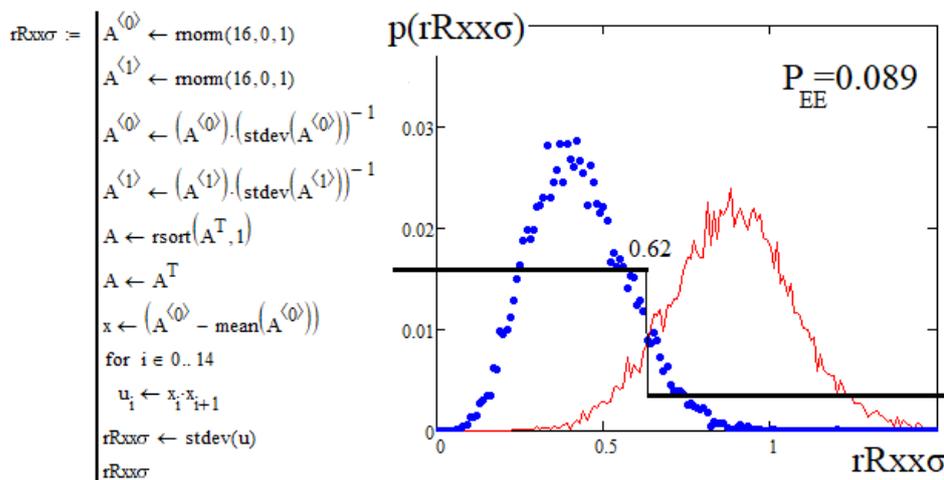


Рис. 4.9. Континуальный критерий, построенный на упорядочивании одной переменной u и вычислении стандартного отклонения произведения расстояний между соседними точками по второй переменной x

Даже используя критерий, построенный на упорядочивании данных по одной переменной, мы получаем рекордный уровень делимости зависимых и независимых данных с вероятностью ошибок первого и второго рода $P_{EE} \approx 0.089$. Для достижения столь высокого уровня делимости данных рассмотренным выше математическим конструкциям нужно было использовать упорядочивание данных по обеим переменным.

Смена переменной упорядочивания позволяет получить новый нейрокритерий, слабо коррелированный с предыдущим критерием $\text{corr}(rR_{xx\sigma}, rR_{yy\sigma}) \approx -0.003$.

4.10. Континуальный критерий, построенный на упорядочивании одной переменной и выявлении максимального значения произведения расстояний между соседними точками по второй переменной

Кроме высокого уровня разделимости зависимых и независимых данных, нейрокритерии, объединяемые в нейросеть, должны иметь как можно меньшую корреляционную сцепленность с соседями. Обычно снижение корреляционной сцепленности дают механизмы выделения максимальных значений (критерий Колмогорова – Смирнова). Программная реализация такого типа модификации приведена на рис. 4.10. Численный эксперимент по моделированию этого критерия показывает относительно низкий уровень его мощности, однако он обладает приемлемым уровнем корреляционной сцепленности с предыдущим нейрокритерием $\text{corr}(rR_{xx\sigma}, rR_{xxM}) \approx 0.465$.

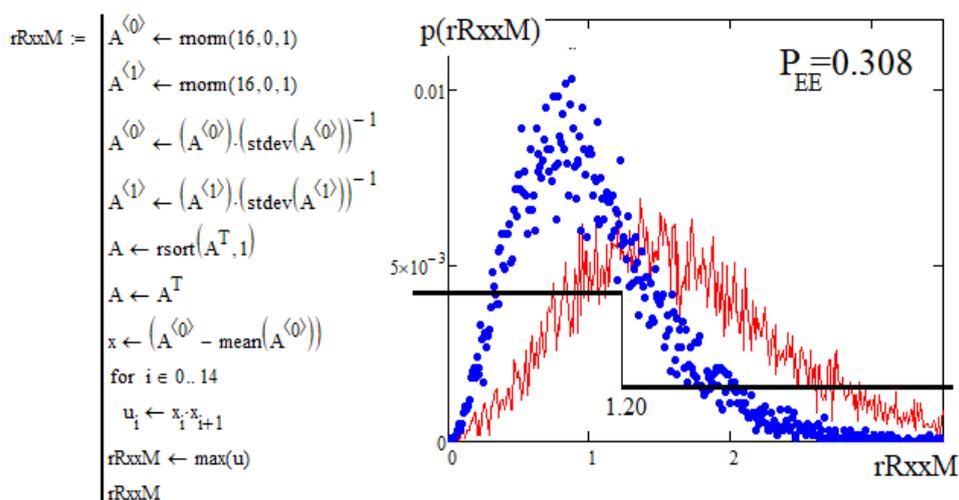


Рис. 4.10. Континуальный критерий, построенный на упорядочивании одной переменной u и выявлении максимального значения произведений расстояний между соседними точками по второй переменной x

Очевидно, что смена переменной упорядочивания данных малой выборки дает новый нейрокритерий, слабо коррелированный со своим аналогом $\text{corr}(rR_{yyM}, rR_{xxM}) \approx -0.0011$.

4.11. Аддитивная сборка континуальных критериев, построенных на упорядочивании двух переменных и оценке стандартного отклонения произведений расстояний между соседними точками по оставшимся случайным переменным

В силу того, что корреляционная сцепленность, рассматриваемых критериев мала, целесообразна их аддитивная или мультипликативная сборка. Пример программной реализации аддитивной сборки приведен на рис. 4.11.

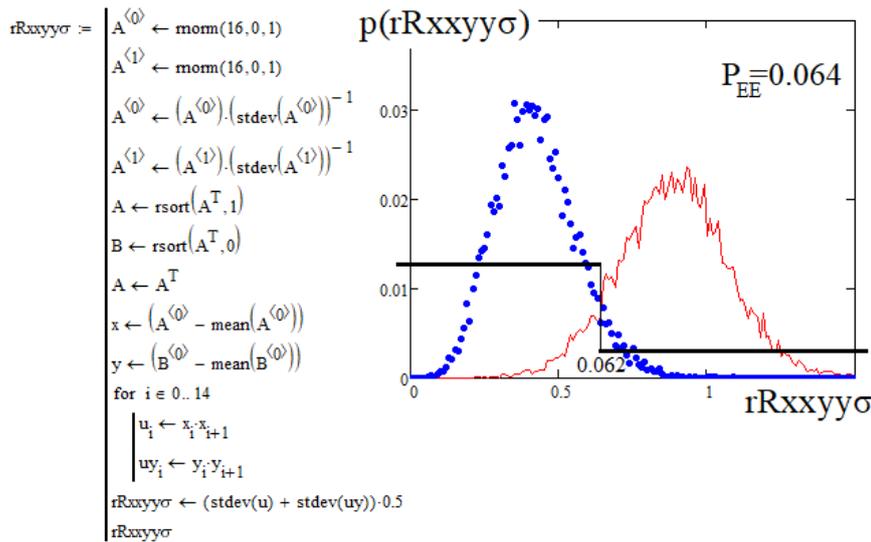


Рис. 4.11. Аддитивная сборка двух статистических критериев, построенных на оценке стандартных отклонений произведений соседних отсчетов малой выборки

Численный эксперимент показал, что объединение критериев позволяет две вероятности ошибок двух критериев $P_{EE} \approx 0.089$ снизить до величины $P_{EE} \approx 0.064$. Наблюдается снижение вероятности ошибок первого и второго рода на 39 %. Видимо, эта аддитивная комбинация двух нейрокритериев является одной из самых эффективных.

4.12. Аддитивная сборка двух континуальных критериев, построенных на упорядочивании одной переменной и оценке стандартного отклонения разности расстояний между соседними точками по второй переменной

Если отказаться от использования произведения координат соседних точек предыдущих критериев и перейти к их разности соседних

отсчетов, то мы получаем еще одну аддитивную сборку. Ее реализация отражена на рис. 4.12. Численный эксперимент показал снижение итоговой вероятности ошибок первого и второго рода.

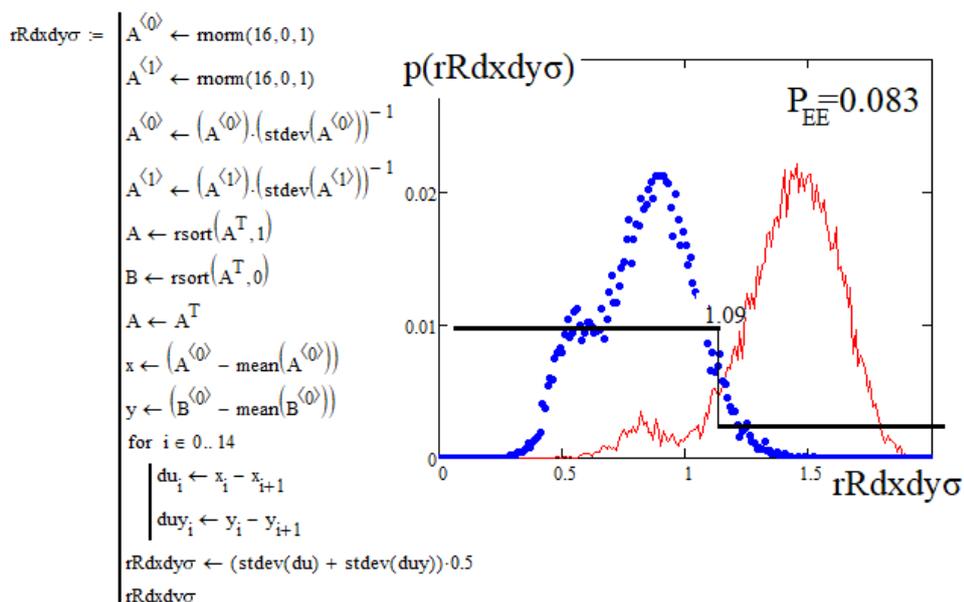


Рис. 4.12. Аддитивная сборка двух статистических критериев, построенных на оценках стандартных отклонений разностей соседних отсчетов малой выборки

Мы видим существенную разницу вероятностей ошибок для двух последних аддитивных сборок. При этом их взаимная корреляционная сцепленность $\text{corr}(rRdxdy\sigma, rRxxuy\sigma) \approx 0.101$ мала. Это делает перспективным еще один уровень объединения этой группы критериев.

4.13. Аддитивная сборка двух континуальных критериев, построенных на упорядочивании одной переменной и оценке стандартного отклонения второй производной расстояний между соседними точками по второй переменной

Если за основу предобработки взять двойное дифференцирование оставшейся случайной компоненты, то получим снижение достижимых вероятностей ошибок. Пример программной реализации этого класса критериев и результаты численного эксперимента представлены на рис. 4.13.

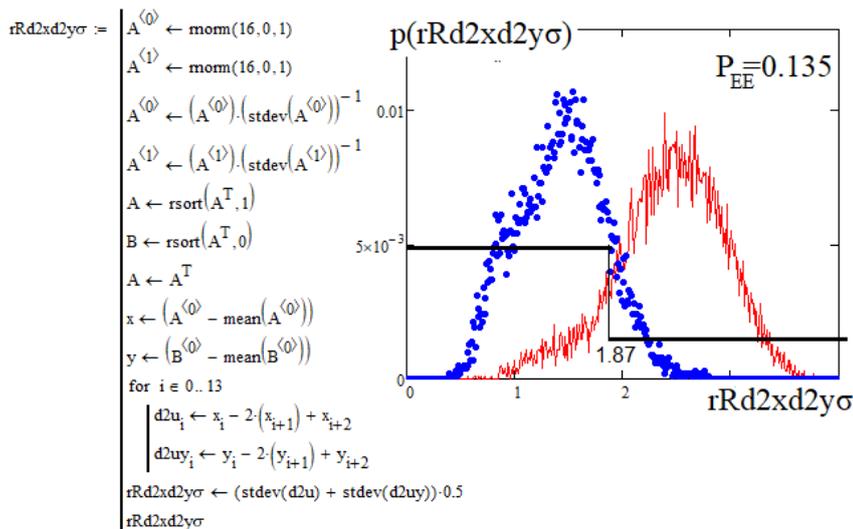


Рис. 4.13. Аддитивная сборка двух статистических критериев, построенных на оценках стандартных отклонений вторых производных расстояний между соседними отсчетами малой выборки

4.14. Корреляционная сцепленность группы критериев, построенных на упорядочивании одной или двух переменных малой выборки

При объединении новых нейрокритериев с уже известными следует учитывать коэффициенты корреляции с классическими критериями и в группе новых близких по содержанию критериев. Эти данные приведены на рис. 414.

$\text{corr}(r, \text{osrR}) = -0.017$		
$\text{corr}(r, \text{osrRxy}) = -0.011$	$\text{corr}(\text{osrR}, \text{osrRxy}) = 0.755$	
$\text{corr}(r, \text{odrR}) = 0.010$	$\text{corr}(\text{osrR}, \text{odrR}) = -0.533$	$\text{corr}(\text{osrRxy}, \text{odrR}) = -0.339$
$\text{corr}(r, \text{odrRxy}) = -0.0006$	$\text{corr}(\text{osrR}, \text{odrRxy}) = -0.315$	$\text{corr}(\text{osrRxy}, \text{odrRxy}) = -0.431$
$\text{corr}(r, \text{rREx}) = 0.0093$	$\text{corr}(\text{osrR}, \text{rREx}) = 0.008$	$\text{corr}(\text{osrRxy}, \text{rREx}) = 0.009$
$\text{corr}(r, \text{rREdx}) = -0.0103$	$\text{corr}(\text{osrR}, \text{rREdx}) = -0.0618$	$\text{corr}(\text{osrRxy}, \text{rREdx}) = -0.5498$
$\text{corr}(r, \text{rREdxy}) = 0.0036$	$\text{corr}(\text{osrR}, \text{rREdxy}) = -0.546$	$\text{corr}(\text{osrRxy}, \text{rREdxy}) = -0.732$
$\text{corr}(r, \text{rRxx}) = 0.0118$	$\text{corr}(\text{osrR}, \text{rRxx}) = 0.0976$	$\text{corr}(\text{osrRxy}, \text{rRxx}) = 0.619$
$\text{corr}(r, \text{rRxxσ}) = -0.0106$	$\text{corr}(\text{osrR}, \text{rRxxσ}) = 0.0749$	$\text{corr}(\text{osrRxy}, \text{rRxxσ}) = 0.0223$
$\text{corr}(r, \text{rRxxM}) = -0.007$	$\text{corr}(\text{osrR}, \text{rRxxM}) = 0.0631$	$\text{corr}(\text{osrRxy}, \text{rRxxM}) = 0.388$
$\text{corr}(r, \text{rRxyyσ}) = 0.0007$	$\text{corr}(\text{osrR}, \text{rRxyyσ}) = 0.0263$	$\text{corr}(\text{osrRxy}, \text{rRxyyσ}) = 0.0413$
$\text{corr}(r, \text{rRdxdyσ}) = 0.0026$	$\text{corr}(\text{osrR}, \text{rRdxdyσ}) = -0.5408$	$\text{corr}(\text{osrRxy}, \text{rRdxdyσ}) = -0.725$
$\text{corr}(r, \text{rRd2xd2yσ}) = -0.005$	$\text{corr}(\text{osrR}, \text{rRd2xd2yσ}) = -0.426$	$\text{corr}(\text{osrRxy}, \text{rRd2xd2yσ}) = -0.575$

Рис. 4.14. Коэффициенты корреляции нейрокритериев, построенных упорядочиванием данных одной и двух переменных (начало)

Первый столбец на рис. 4.14 имеет низкие значения модулей коэффициентов корреляции. Это означает, что группа рассматриваемых критериев хорошо дополняет наиболее часто используемый классический критерий Эджуорта – Эдлтона – Пирсона. Внутри группы новых критериев уровень корреляционной сцепленности низок. Об этом свидетельствует то, что во втором столбце на рис. 4.14 максимальное значение модуля коэффициентов корреляции составляет 0.5408.

Коэффициенты корреляции нейрокритериев,
построенных упорядочиванием данных

$\text{corr}(\text{odrR}, \text{odrRxy}) = 0.783$		
$\text{corr}(\text{odrR}, \text{rREx}) = -0.006$	$\text{corr}(\text{odrRxy}, \text{rREx}) = -0.005$	
$\text{corr}(\text{odrR}, \text{rREdx}) = 0.147$	$\text{corr}(\text{odrRxy}, \text{rREdx}) = 0.677$	$\text{corr}(\text{rREx}, \text{rREdx}) = 0.0017$
$\text{corr}(\text{odrR}, \text{rREdxy}) = 0.699$	$\text{corr}(\text{odrRxy}, \text{rREdxy}) = 0.889$	$\text{corr}(\text{rREx}, \text{rREdxy}) = -0.004$
$\text{corr}(\text{odrR}, \text{rRxx}) = -0.071$	$\text{corr}(\text{odrRxy}, \text{rRxx}) = -0.560$	$\text{corr}(\text{rREx}, \text{rRxx}) = -0.0027$
$\text{corr}(\text{odrR}, \text{rRxx}\sigma) = 0.074$	$\text{corr}(\text{odrRxy}, \text{rRxx}\sigma) = 0.177$	$\text{corr}(\text{rREx}, \text{rRxx}\sigma) = 0.0076$
$\text{corr}(\text{odrR}, \text{rRxxM}) = 0.003$	$\text{corr}(\text{odrRxy}, \text{rRxxM}) = -0.275$	$\text{corr}(\text{rREx}, \text{rRxxM}) = 0.005$
$\text{corr}(\text{odrR}, \text{rRxxyy}\sigma) = 0.174$	$\text{corr}(\text{odrRxy}, \text{rRxxyy}\sigma) = 0.219$	$\text{corr}(\text{rREx}, \text{rRxxyy}\sigma) = 0.010$
$\text{corr}(\text{odrR}, \text{rRdx}\sigma) = 0.701$	$\text{corr}(\text{odrRxy}, \text{rRdx}\sigma) = 0.893$	$\text{corr}(\text{rREx}, \text{rRdx}\sigma) = -0.005$
$\text{corr}(\text{odrR}, \text{rRd2xd2y}\sigma) = 0.734$	$\text{corr}(\text{odrRxy}, \text{rRd2xd2y}\sigma) = 0.934$	$\text{corr}(\text{rREx}, \text{rRd2xd2y}\sigma) = -0.003$

Рис. 4.14. Продолжение

Последнее утверждение подтверждается также тем, что в продолжении и окончании рис. 4.14 редко встречаются модули коэффициентов корреляции, близкие к единице.

Коэффициенты корреляции нейрокритериев,
построенных упорядочиванием данных

$\text{corr}(\text{rREdx}, \text{rREdxy}) = 0.756$		
$\text{corr}(\text{rREdx}, \text{rRxx}) = -0.973$	$\text{corr}(\text{rREdxy}, \text{rRxx}) = -0.719$	
$\text{corr}(\text{rREdx}, \text{rRxx}\sigma) = 0.129$	$\text{corr}(\text{rREdxy}, \text{rRxx}\sigma) = 0.087$	$\text{corr}(\text{rRxx}, \text{rRxx}\sigma) = -0.068$
$\text{corr}(\text{rREdx}, \text{rRxxM}) = -0.586$	$\text{corr}(\text{rREdxy}, \text{rRxxM}) = -0.429$	$\text{corr}(\text{rRxx}, \text{rRxxM}) = 0.632$
$\text{corr}(\text{rREdx}, \text{rRxxyy}\sigma) = 0.072$	$\text{corr}(\text{rREdxy}, \text{rRxxyy}\sigma) = 0.101$	$\text{corr}(\text{rRxx}, \text{rRxxyy}\sigma) = -0.007$
$\text{corr}(\text{rREdx}, \text{rRdx}\sigma) = 0.754$	$\text{corr}(\text{rREdxy}, \text{rRdx}\sigma) = 0.996$	$\text{corr}(\text{rRxx}, \text{rRdx}\sigma) = -0.714$
$\text{corr}(\text{rREdx}, \text{rRd2xd2y}\sigma) = 0.722$	$\text{corr}(\text{rREdxy}, \text{rRd2xd2y}\sigma) = 0.952$	$\text{corr}(\text{rRxx}, \text{rRd2xd2y}\sigma) = -0.667$
$\text{corr}(\text{rRxx}\sigma, \text{rRxxM}) = 0.465$		
$\text{corr}(\text{rRxx}\sigma, \text{rRxxyy}\sigma) = 0.742$	$\text{corr}(\text{rRxxM}, \text{rRxxyy}\sigma) = 0.374$	
$\text{corr}(\text{rRxx}\sigma, \text{rRdx}\sigma) = 0.087$	$\text{corr}(\text{rRxxM}, \text{rRdx}\sigma) = -0.423$	$\text{corr}(\text{rRxxyy}\sigma, \text{rRdx}\sigma) = 0.101$
$\text{corr}(\text{rRxx}\sigma, \text{rRd2xd2y}\sigma) = 0.107$	$\text{corr}(\text{rRxxM}, \text{rRd2xd2y}\sigma) = -0.378$	$\text{corr}(\text{rRxxyy}\sigma, \text{rRd2xd2y}\sigma) = 0.126$
$\text{corr}(\text{rRdx}\sigma, \text{rRd2xd2y}\sigma) = 0.949$		

Рис. 4.14. Окончание

Список литературы к главе 4

1. Волчихин В. И., Иванов А. И., Ахметов Б. Б., Серикова Ю. И. Фрактально-корреляционный функционал, используемый при поиске пар слабо зависимых биометрических данных в малых выборках // Вестник высших учебных заведений. Поволжский регион. Технические науки. 2016. № 4. С. 25–31.
2. Иванов А. И., Малыгин А. Ю., Полковникова С. А. Новый статистический критерий большой мощности, полученный дифференцированием случайных данных малой выборки // Известия высших учебных заведений. Поволжский регион. Технические науки. 2021. № 3. С. 67–74. doi:10.21685/2072-3059-2021-3-7
3. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок : справочник. Пенза : Изд-во ПГУ, 2022. 160 с.

НОВЫЕ КРИТЕРИИ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОЛУЧЕННЫЕ ИСПОЛЬЗОВАНИЕМ ЛИНЕЙНЫХ КВАНТОВАТЕЛЕЙ ДВУХМЕРНОГО ПРОСТРАНСТВА

5.1. Критерий, построенный на использовании трехуровневых квантователей (две параллельные линии, наклон вверх слева направо)

Еще один вариант критерия проверки гипотезы независимости может быть получен использованием двух параллельных линейных квантователей. Разделяющие линии имеют наклон в 45° , расстояние между параллельными линиями двух квантователей составляет – 1.26 по обеим координатам.

Пример программной реализации такого типа критериев приведен на рис. 5.1. На этом же рисунке представлены результаты численного эксперимента.

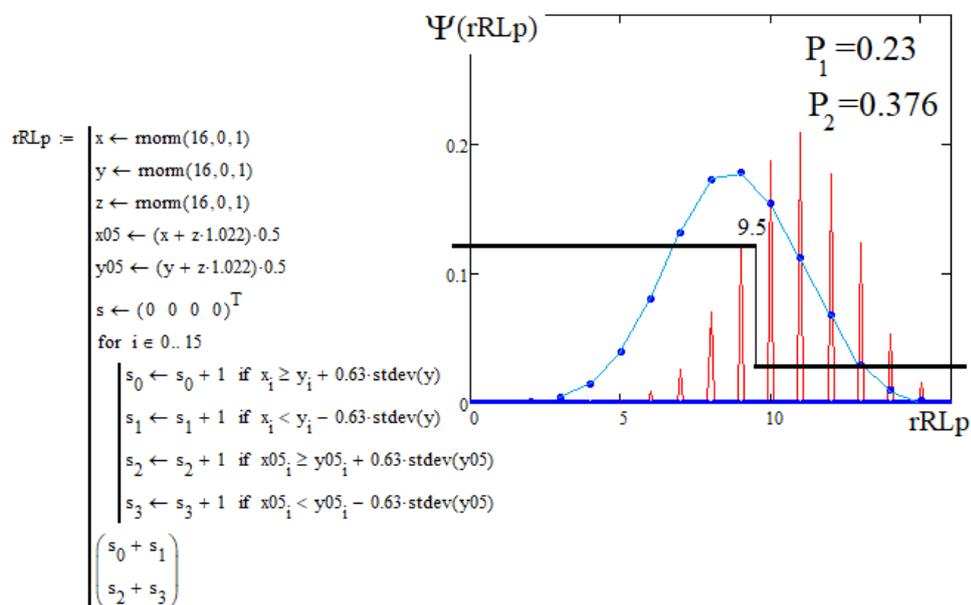


Рис. 5.1. Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по дискретному корреляционному критерию, построенному на троичном квантователе [28] (две параллельные линии, наклон вверх слева направо)

При попадании данных точек малой выборки между двумя линиями квантователей отклик нейрокритерия составляет значение «0». При попадании точек малой выборки вне обеих линий квантователей новый нейрокритерий откликается состоянием «1».

5.2. Критерий, построенный на использовании трехуровневых квантователей (две параллельные линии, наклон вниз слева направо)

Если у двух параллельных линий квантователей изменить наклон, оставив то же расстояние между ними, мы получим еще один новый нейрокритерий. Программная реализация этого критерия приведена на рис. 5.2. На этом же рисунке отражены результаты численного эксперимента.

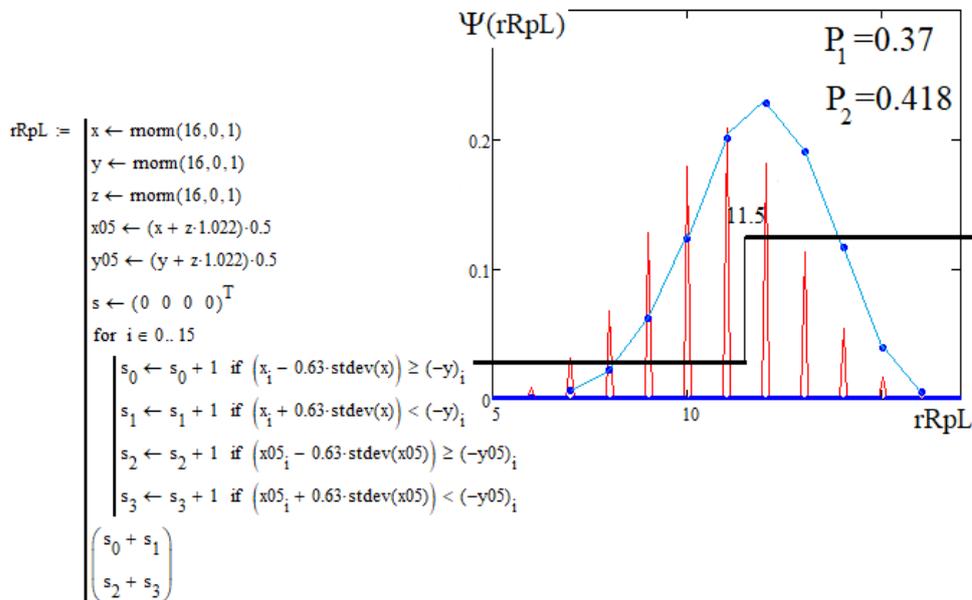


Рис. 5.2 Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по дискретному корреляционному критерию, построенному на троичном квантователе [28] (две параллельные линии, наклон вниз слева направо)

По сравнению с распределением состояний предыдущего критерия отклики, соответствующие коррелированным входным состояниям, переместились на другую сторону от распределений независимых данных.

5.3. Критерий, построенный на использовании двух трехуровневых квантователей (две параллельные линии, перпендикулярные двум другим линиям)

Возможно преобразование последнего нейрокритерия с тем, чтобы распределение зависимых данных переместилось с правой стороны на левую сторону распределения независимых данных. После такого преобразования можно выполнить аддитивную сборку двух последних критериев. Программная реализация нового критерия и результаты численного эксперимента приведены на рис. 5.3.

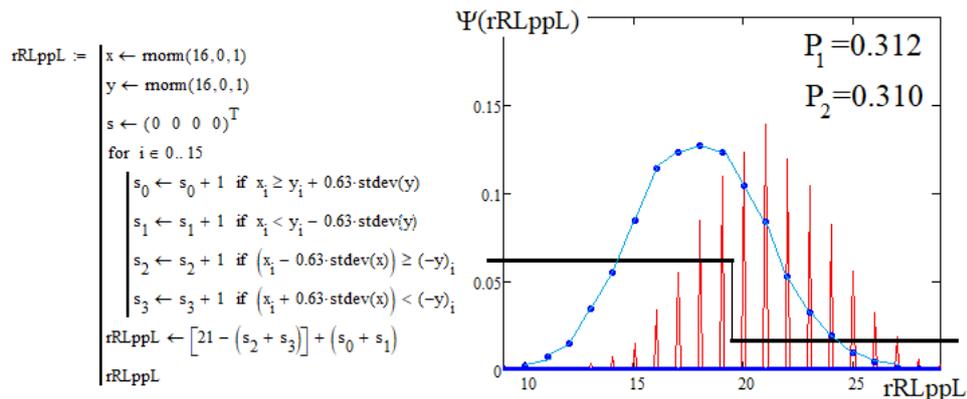


Рис. 5.3 Разделение независимых данных и данных с положительной коррелированностью $r = 0.5$ по дискретному корреляционному критерию, построенному на двух перпендикулярных квантователях, состоящих из двух параллельных линий

Очевидно, что новый критерий существенно уступает по мощности классическому критерию Эджуорта – Эдлттона – Пирсона, однако они имеют низкий уровень корреляционной сцепленности их откликов: $\text{corr}(rRLppL, r) \approx -0.163$.

5.4. Критерий, построенный на использовании двух квантователей (две параллельные линии, перпендикулярные двум другим линиям) с использованием операции деления

Мощность предыдущего критерия увеличена за счет суммирования данных двух типов квантователей. Так как при многокритериальном анализе роль играет не только их мощность, интерес представляет повышение мощности через использование мультипликативного объединения критериев, в частности объединить критерии возможно

их делением. Программная реализация примера подобного критерия представлена на рис. 5.4. На этом же рисунке представлены результаты численного моделирования нейрокритерия.

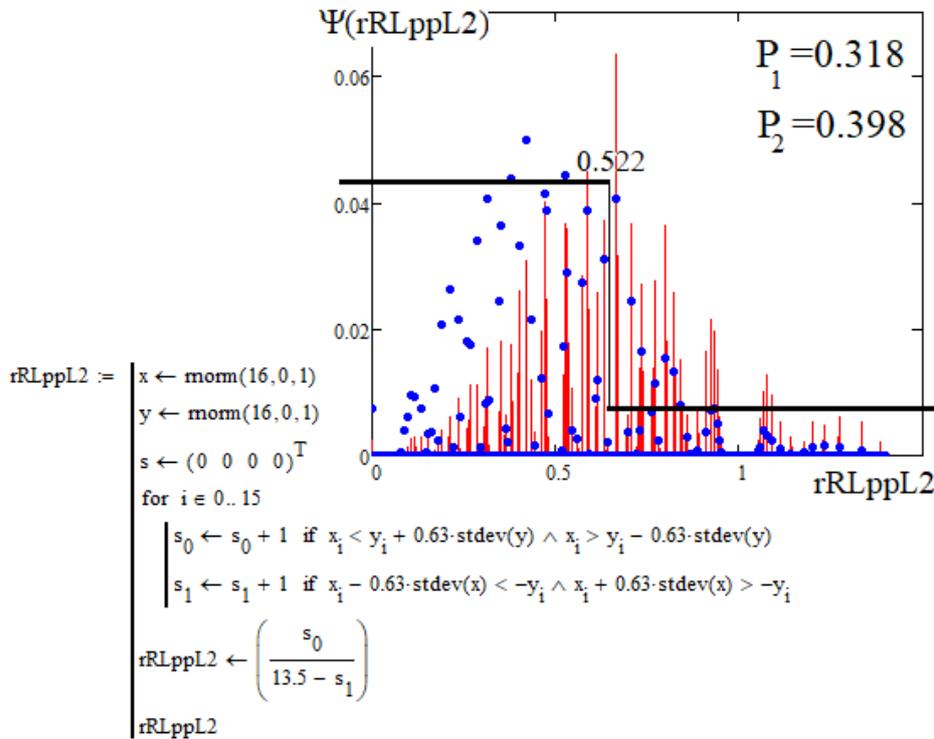


Рис. 5.4 Разделение независимых данных и данных с положительной коррелированностью по дискретному корреляционному критерию, построенному на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием деления

Очевидно, что новый критерий существенно уступает по мощности классическому критерию Эджуорта – Эдлтона – Пирсона, однако они хорошо дополняют друг друга и имеют низкий уровень корреляционной сцепленности: $\text{corr}(rRLppL2, r) \approx 0.058$.

5.5. Непрерывный вариант критерия, построенный на использовании двух квантователей (две параллельные линии, перпендикулярные двум другим линиям) с использованием операции деления

Переход от дискретных критериев к их непрерывным аналогам обычно дает рост качества принимаемых ими решений. Этот тезис подтверждается заменой подсчета точек в предшествующем критерии

накоплением модулей координат, обнаруженных двумя квантователями точек. Реализация этого непрерывного критерия приведена на рис. 5.5.

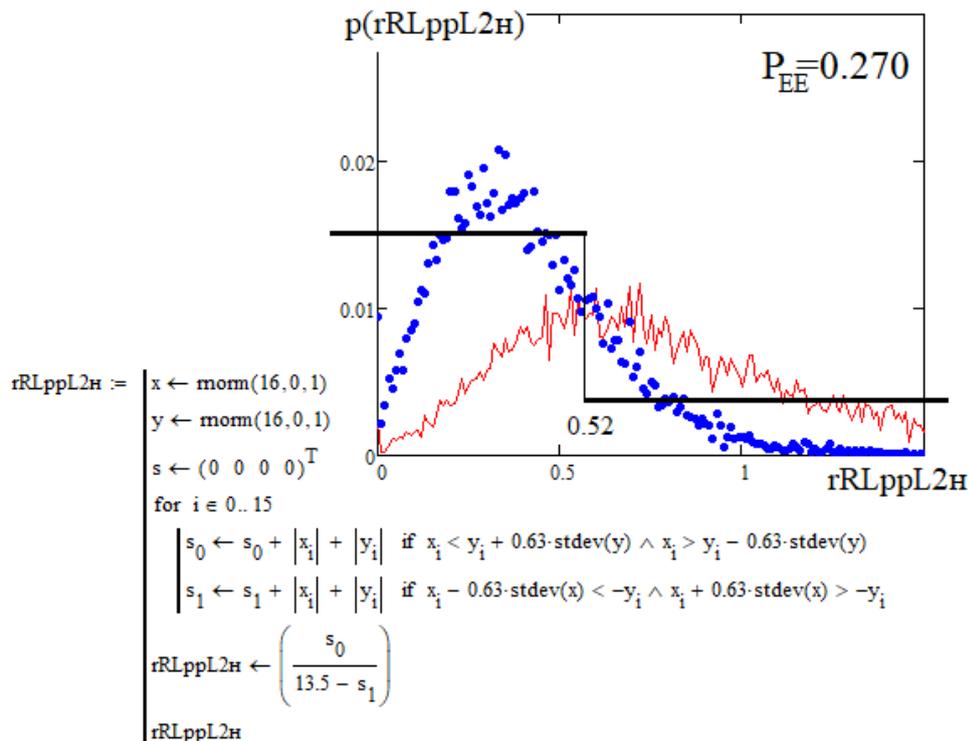


Рис. 5.5 Непрерывный вариант корреляционного критерия, построенного на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием деления

Результаты численного моделирования показали, что получается выигрыш по снижению вероятностей ошибок на 18.5 % по отношению к вероятности ошибок первого рода предшествующего дискретного критерия. Если сравнивать снижение с вероятностью ошибок второго рода, то выигрыш составит 48 %.

5.6. Непрерывный вариант критерия, построенный на использовании двух двухуровневых квантователей с использованием операции деления и накоплении произведения модулей координат обнаруженных точек

Еще одним вариантом непрерывного критерия является замена накапливания суммы модулей координат обнаруженных точек на произведение модулей координат обнаруженных точек. Программная

реализация этой модификации критерия представлена на рис. 5.6. На этом же рисунке отображены результаты моделирования нового нейрокритерия.

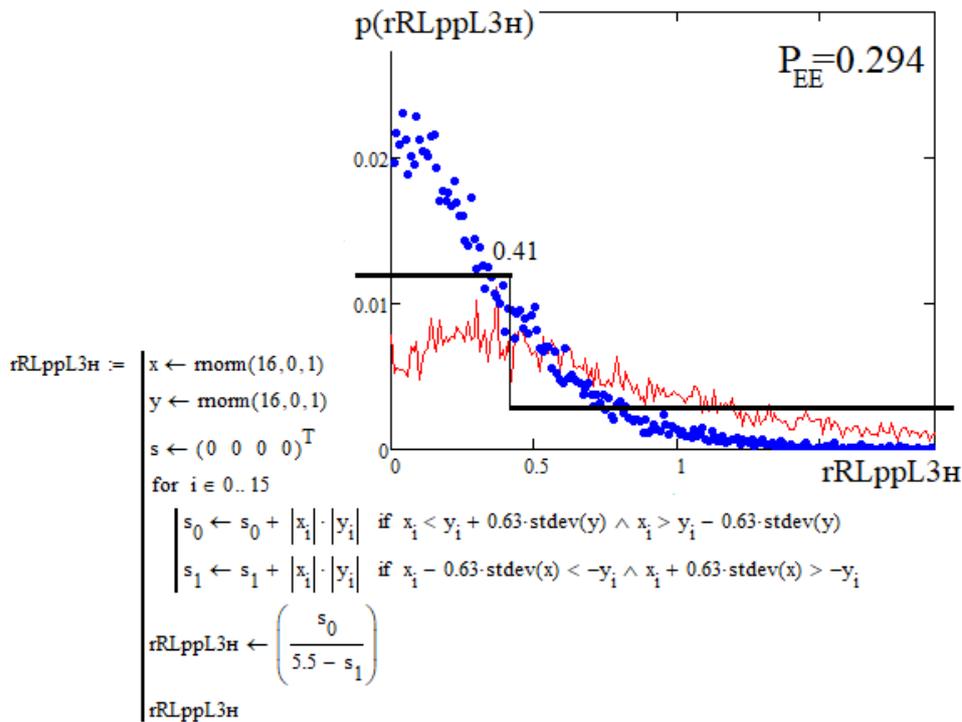


Рис. 5.6. Непрерывный вариант корреляционного критерия, построенного на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием деления и накопления произведений модулей координат обнаруженных точек

5.7. Дискретный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой

В данной главе уже рассмотрены четыре дискретных нейрокритерия, синтезированных для проверки гипотезы независимости. Эти дискретные критерии могут быть дополнены еще одним вариантом нейрокритерия иной сборки. Программная реализация нейрокритерия приведена на рис. 5.7. На этом же рисунке отражены результаты численного моделирования нового нейрокритерия.

По результатам численного моделирования мы видим, что близким к оптимальному является порог квантователя – 9.5. При этом пороге мощность рассматриваемого дискретного критерия оказывается

выше примерно на 7 % в сравнении с предыдущим дискретным критерием параграфа 5.4.

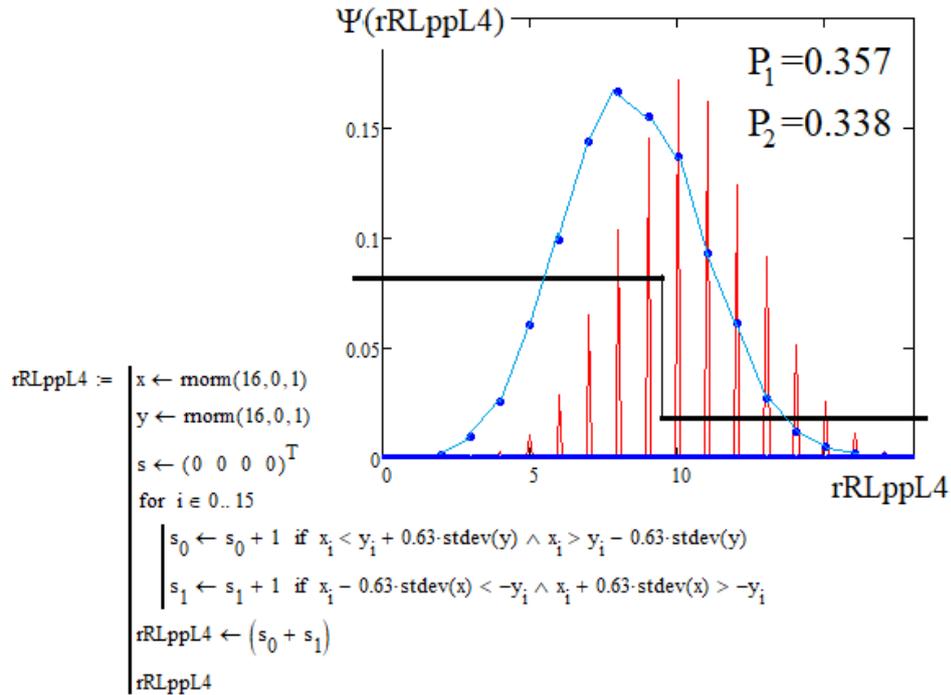


Рис. 5.7. Дискретный корреляционный критерий, построенный на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием аддитивной свертки числа обнаруженных квантователями точек

Положительным свойством двух последних критериев является то, что они имеют низкий уровень корреляционной сцепленности: $\text{corr}(rRLppL4, rRLppL3н) \approx -0.001$.

5.8. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой числа обнаруженных точек

Для перехода от дискретного критерия предыдущего параграфа следует перейти от подсчета числа обнаруженных точек к накапливанию квадратов их координат. Программная реализация этого критерия приведена на рис. 5.8.

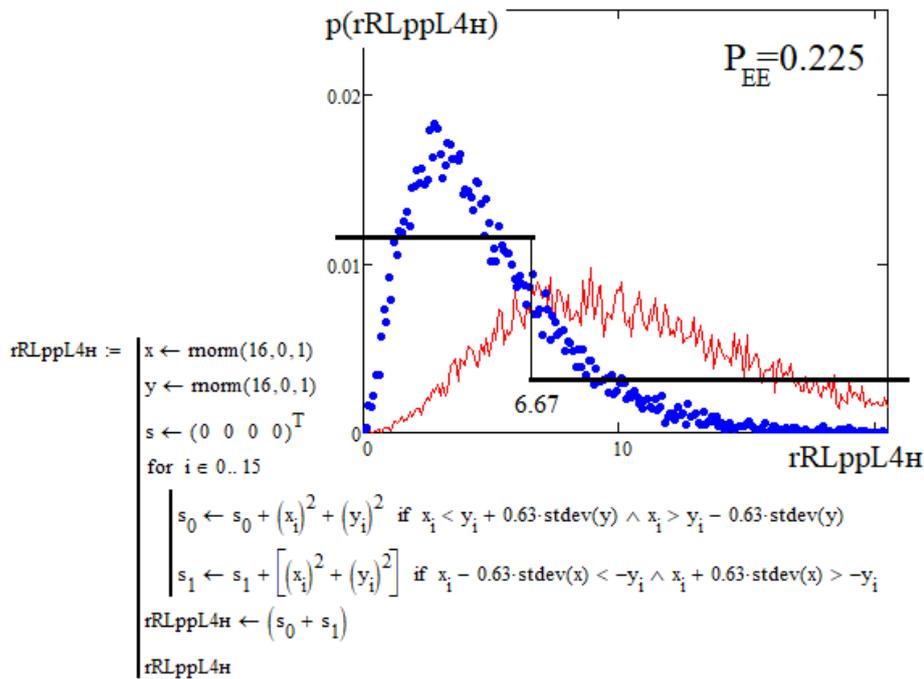


Рис. 5.8. Непрерывный корреляционный критерий, построенный на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием аддитивной свертки суммы квадратов координат точек

Численный эксперимент дает оценку снижения вероятности ошибок первого и второго рода на 51 %. Это подтверждение того, что дискретные (ранговые) критерии проверки гипотезы независимости потенциально слабее непрерывных (континуальных) аналогов этих критериев.

5.9. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой накопленных произведений модулей координат точек

Еще один вариант непрерывного критерия получается, если вместо накапливания квадратов координат применить накапливание модулей произведения координат обнаруженных точек. Пример программной реализации такого нейрокритерия приведен на рис. 5.9. На этом же рисунке – распределения выходных состояний нейрокритерия при воздействии на него зависимыми и независимыми данными малой выборки.

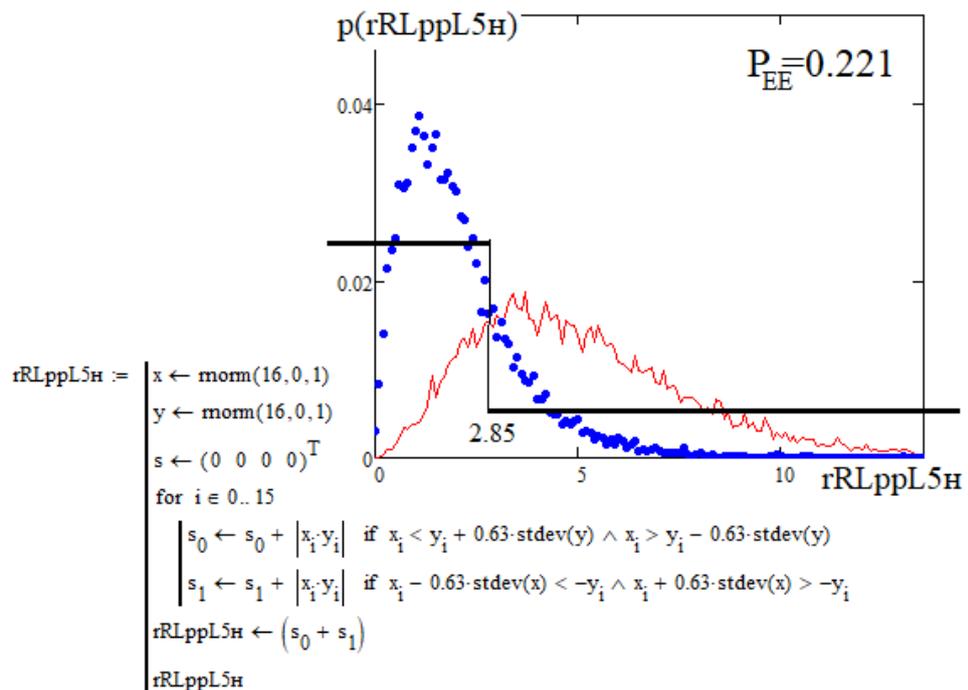


Рис. 5.9. Непрерывный корреляционный критерий, построенный на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием аддитивной свертки суммы произведений модулей координат точек

Следует обратить внимание на то, что два последних квадратичных критерия близки по их вероятностным характеристикам, при этом их корреляционная сцепленность $\text{corr}(rRLppL4н, rRLppL5н) \approx 0.996$ высока. Можно говорить о высоком уровне близости друг другу двух последних критериев.

5.10. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой накопленных квадратов разностей координат точек

Снизить вероятности ошибок первого и второго рода примерно на 23 % при разделении зависимых и независимых данных удастся, если накапливать квадраты разностей координат каждой из обнаруженных точек. Пример программной реализации этого нейрокритерия приведен на рис. 5.10.

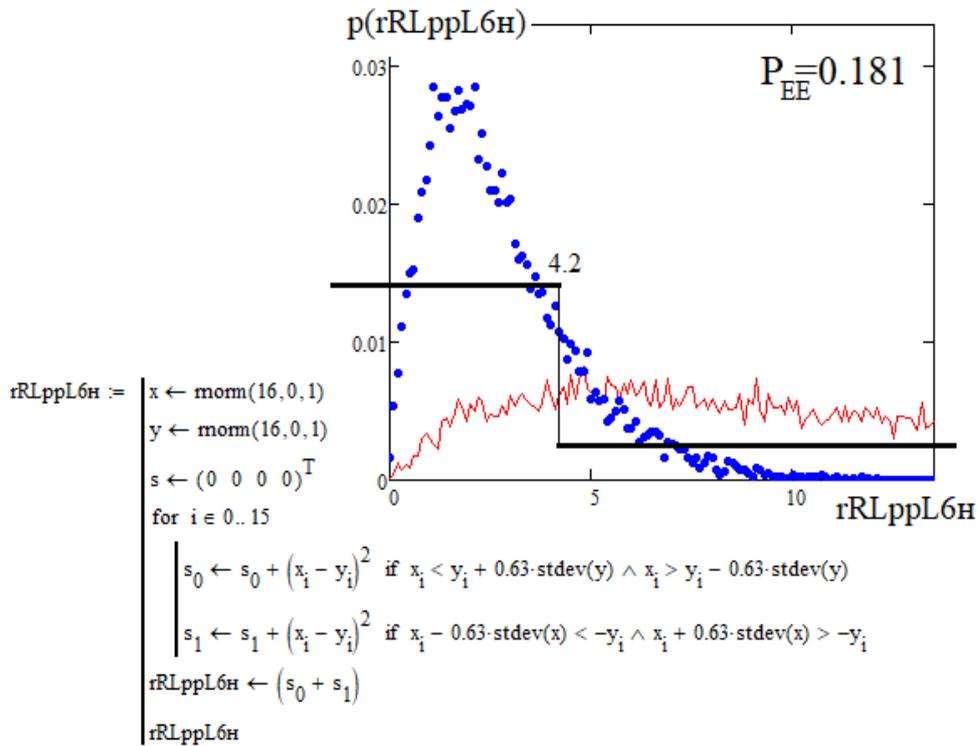


Рис. 5.10. Непрерывный корреляционный критерий, построенный на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием аддитивной свертки суммы квадрата разности координат точек

Важным преимуществом квадратично-разностного критерия является то, что его корреляционная сцепленность с предшествующими критериями существенно снижается $\text{corr}(rRLppL5н, rRLppL6н) \approx 0.7$.

5.11. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой накопленных квадратов суммы координат точек

Вариант критерия, построенного на накапливании квадратов суммы координат точек. К сожалению, этот нейрокритерий имеет низкую разделяемость зависимых и независимых данных. Результаты численного моделирования этого критерия приведены на рис. 5.11.

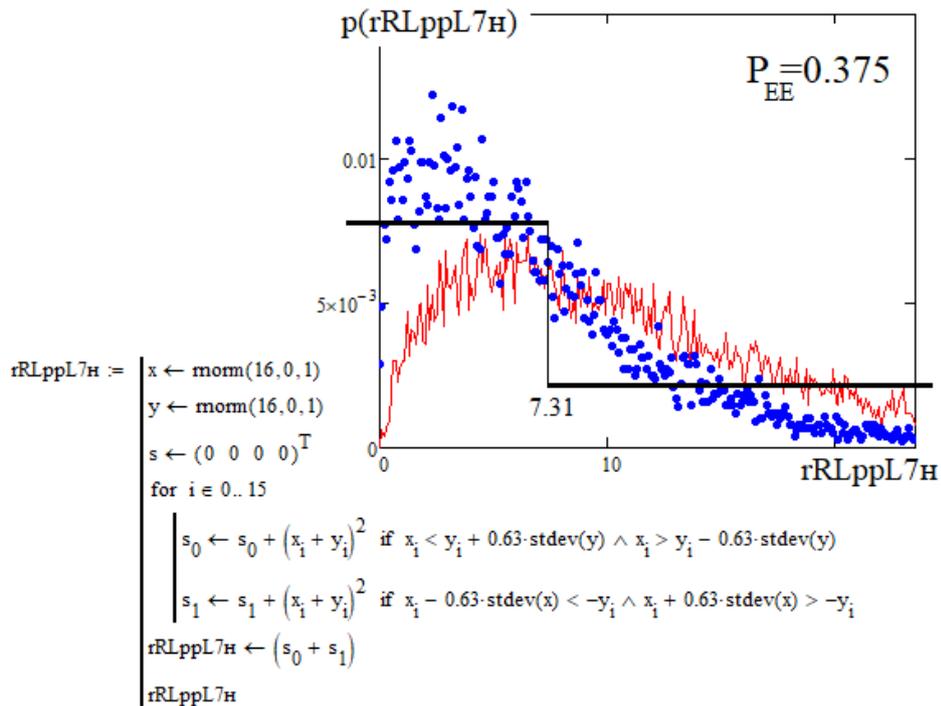


Рис. 5.11. Непрерывный корреляционный критерий, построенный на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием аддитивной свертки накопленных квадратов сумм координат точек

Следует отметить, что два последних критерия хорошо дополняют друг друга, так как имеют очень низкий уровень корреляционной сцепленности их откликов: $\text{corr}(rRLppL6н, rRLppL7н) \approx -0.001$.

5.12. Непрерывный вариант корреляционного критерия, использующего два перпендикулярных квантователя, состоящих из двух параллельных линий, с аддитивной сборкой четвертой степени разности координат обнаруженных точек

Одним из путей синтеза новых критериев является изменение показателя степени накапливаемых данных. Одним из самых эффективных является критерий накапливания квадратов разности координат точек. Этот критерий может быть усилен, если перейти от возведения разности в квадрат к ее возведению в четвертую степень. Программная реализация нейрокритерия и результаты его численного моделирования – на рис. 5.12.

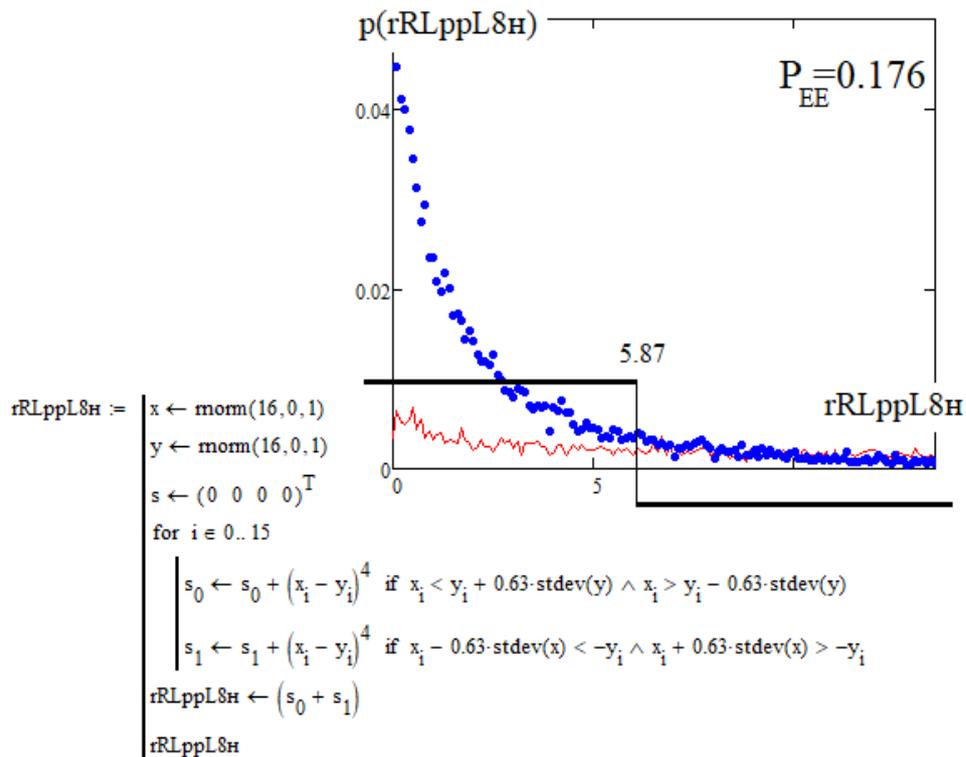


Рис. 5.12. Непрерывный корреляционный критерий, построенный на двух перпендикулярных квантователях, состоящих из двух параллельных линий, с использованием аддитивной свертки накопленных четвертой степени разностей координат точек

Удвоение показателя степени приводит к снижению вероятности ошибок примерно на 3 %.

Последние два критерия хорошо дополняют друг друга, так как имеют очень низкий уровень корреляционной сцепленности откликов: $\text{corr}(rRLppL7н, rRLppL8н) \approx 0.007$.

5.13. Корреляционная сцепленность критериев, построенных из двух перпендикулярных квантователей, собранных из двух параллельных линий

При объединении разных нейрокритериев в группы и свертывании их выходной кодовой избыточности важна близость критериев по уровню делимости зависимых и независимых данных [29]. Желательно объединять нейрокритерии с близкими вероятностями ошибок

первого и второго рода. Однако если модули парной корреляции высоки, то объединять их оказывается бесполезно.

На рис. 5.13 приведены коэффициенты парной корреляции между нейрокритериями рассматриваемой группы. Усреднение модулей данных первого столбца на рис. 5.13 показывает низкий уровень корреляционной сцепленности рассматриваемой группы критериев с классическим критерием Эджуорта – Эдлтона – Пирсона. Усреднение модулей коэффициентов корреляции второго и третьего столбцов на рис. 5.13 показывает низкий уровень корреляционной сцепленности внутри группы рассматриваемых критериев.

$\text{corr}(r, rRp) = -0.337$		
$\text{corr}(r, rRpL) = -0.005$	$\text{corr}(rRpL, rRpL) = 0.020$	
$\text{corr}(r, rRLppL) = -0.163$	$\text{corr}(rRpL, rRLppL) = 0.480$	$\text{corr}(rRpL, rRLppL) = -0.866$
$\text{corr}(r, rRLppL2) = 0.058$	$\text{corr}(rRpL, rRLppL2) = -0.666$	$\text{corr}(rRpL, rRLppL2) = -0.015$
$\text{corr}(r, rRLppL2н) = 0.0006$	$\text{corr}(rRpL, rRLppL2н) = -0.017$	$\text{corr}(rRpL, rRLppL2н) = -0.006$
$\text{corr}(r, rRLppL3н) = -0.009$	$\text{corr}(rRpL, rRLppL3н) = 0.003$	$\text{corr}(rRpL, rRLppL3н) = 0.011$
$\text{corr}(r, rRLppL4) = -0.011$	$\text{corr}(rRpL, rRLppL4) = -0.615$	$\text{corr}(rRpL, rRLppL4) = -0.017$
$\text{corr}(r, rRLppL4н) = -0.011$	$\text{corr}(rRpL, rRLppL4н) = -0.298$	$\text{corr}(rRpL, rRLppL4н) = -0.008$
$\text{corr}(r, rRLppL5н) = -0.010$	$\text{corr}(rRpL, rRLppL5н) = -0.291$	$\text{corr}(rRpL, rRLppL5н) = -0.007$
$\text{corr}(r, rRLppL6н) = -0.474$	$\text{corr}(rRpL, rRLppL6н) = 0.157$	$\text{corr}(rRpL, rRLppL6н) = -0.002$
$\text{corr}(r, rRLppL7н) = 0.469$	$\text{corr}(rRpL, rRLppL7н) = -0.585$	$\text{corr}(rRpL, rRLppL7н) = -0.009$
$\text{corr}(r, rRLppL8н) = -0.388$	$\text{corr}(rRpL, rRLppL8н) = 0.037$	$\text{corr}(rRpL, rRLppL8н) = -0.004$

Рис. 5.13. Коэффициенты корреляции откликов нейрокритериев, собранных из перпендикулярных квантователей из двух параллельных линий (начало)

Усреднении модулей коэффициентов корреляции трех столбцов продолжения рис. 5.13 подтверждает низкий уровень корреляционной сцепленности данных внутри группы исследованных критериев.

Коэффициенты корреляции нейрокритериев, собранных из перпендикулярных квантователей, состоящих из двух линий

$\text{corr}(rRpL, rRLppL2) = -0.367$		
$\text{corr}(rRpL, rRLppL2н) = -0.014$	$\text{corr}(rRpL2, rRLppL2н) = 0.035$	
$\text{corr}(rRpL, rRLppL3н) = 0.002$	$\text{corr}(rRpL2, rRLppL3н) = -0.002$	$\text{corr}(rRpL2н, rRLppL3н) = 0.0006$
$\text{corr}(rRpL, rRLppL4) = -0.314$	$\text{corr}(rRpL2, rRLppL4) = 0.923$	$\text{corr}(rRpL2н, rRLppL4) = 0.041$
$\text{corr}(rRpL, rRLppL4н) = -0.132$	$\text{corr}(rRpL2, rRLppL4н) = 0.422$	$\text{corr}(rRpL2н, rRLppL4н) = 0.072$
$\text{corr}(rRpL, rRLppL5н) = -0.127$	$\text{corr}(rRpL2, rRLppL5н) = 0.411$	$\text{corr}(rRpL2н, rRLppL5н) = 0.072$
$\text{corr}(rRpL, rRLppL6н) = 0.084$	$\text{corr}(rRpL2, rRLppL6н) = 0.215$	$\text{corr}(rRpL2н, rRLppL6н) = 0.054$
$\text{corr}(rRpL, rRLppL7н) = -0.272$	$\text{corr}(rRpL2, rRLppL7н) = 0.382$	$\text{corr}(rRpL2н, rRLppL7н) = 0.048$
$\text{corr}(rRpL, rRLppL8н) = 0.033$	$\text{corr}(rRpL2, rRLppL8н) = 0.189$	$\text{corr}(rRpL2н, rRLppL8н) = 0.035$

Рис. 5.13. Продолжение

Низкий уровень корреляционной сцепленности внутри группы рассматриваемых критериев подтверждается данными рис. 5.13.

Коэффициенты корреляции нейрокритериев, собранных из перпендикулярных квантователей, состоящих из двух линий

$\text{corr}(r_{L3n}, r_{L4}) = -0.001$		
$\text{corr}(r_{L3n}, r_{L4n}) = 0.018$	$\text{corr}(r_{L4}, r_{L4n}) = 0.456$	
$\text{corr}(r_{L3n}, r_{L5n}) = 0.018$	$\text{corr}(r_{L4}, r_{L5n}) = 0.442$	$\text{corr}(r_{L4n}, r_{L5n}) = 0.996$
$\text{corr}(r_{L3n}, r_{L6n}) = 0.012$	$\text{corr}(r_{L4}, r_{L6n}) = 0.322$	$\text{corr}(r_{L4n}, r_{L6n}) = 0.710$
$\text{corr}(r_{L3n}, r_{L7n}) = 0.013$	$\text{corr}(r_{L4}, r_{L7n}) = 0.322$	$\text{corr}(r_{L4n}, r_{L7n}) = 0.703$
$\text{corr}(r_{L3n}, r_{L8n}) = 0.011$	$\text{corr}(r_{L4}, r_{L8n}) = 0.230$	$\text{corr}(r_{L4n}, r_{L8n}) = 0.621$
$\text{corr}(r_{L5n}, r_{L6n}) = 0.708$		
$\text{corr}(r_{L5n}, r_{L7n}) = 0.700$	$\text{corr}(r_{L6n}, r_{L7n}) = -0.001$	
$\text{corr}(r_{L5n}, r_{L8n}) = 0.625$	$\text{corr}(r_{L6n}, r_{L8n}) = 0.867$	$\text{corr}(r_{L7n}, r_{L8n}) = 0.007$

Рис. 5.13. Окончание

Даже для самых сложных критериев рассматриваемой группы продолжает сохраняться низкий уровень корреляционной сцепленности. Так, в окончании рис. 5.13 только два критерия имеют модули коэффициентов корреляции более 0.8.

Список литературы к главе 5

1. Иванов А. И., Сулавко А. Е. Использование сетей корреляционных нейронов с многоуровневым квантованием: защита от извлечения знаний из параметров решающего правила : препринт. Пенза : Изд-во ПГУ, 2020. 48 с.

2. Иванов А. И. Искусственные математические молекулы: повышение точности статистических оценок на малых выборках (программы на языке MathCAD) : препринт. Пенза : Изд-во ПГУ, 2020. 36 с.

НОВЫЕ НЕЙРОКРИТЕРИИ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОСТРОЕННЫЕ НА СТАТИСТИЧЕСКОМ АНАЛИЗЕ ТОЧЕК, ПОПАДАЮЩИХ В СЕКТОРА ДЕКАРТОВОЙ СИСТЕМЫ КООРДИНАТ

6.1. Дискретный нейрокритерий, построенный на использовании статистического анализа числа точек малой выборки, попадающих в сектора Декартовой системы координат (2022)

Следует отметить, что классический угловой критерий Олмстеда – Тьюки (1947) [7] можно рассматривать как критерий, построенный на подсчете числа точек, обнаруженных в четырех четвертях Декартовой системы координат. При этом центр Декартовой системы координат размещен в точке двух медиан распределения данных малой выборки. Если изменить условия и разместить центр системы Декартовых координат в точке математических ожиданий малой выборки, то мы получим еще один критерий, синтезированный уже в начале XXI в. [22].

К сожалению, два упомянутых выше нейрокритерия имеют отклики, сильно коррелированные с классическим критерием Эджуорта – Эдлтона – Пирсона: $\text{corr}(r, OT) \approx 0.615$, $\text{corr}(r, rR) \approx 0.641$. Сильная корреляционная сцепленность является существенным недостатком, который может быть ослаблен (2022) [30] оценкой размаха вариаций обнаруженного ряда чисел.

Реализация нового дискретного критерия представлена на рис. 6.1, на этом же рисунке даны результаты его численного моделирования. Принципиальным является то, что исчезает корреляционная сцепленности нового критерия с классическим критерием Эджуорта – Эдлтона – Пирсона $\text{corr}(r, D4rRMm) \approx -0.0021$.

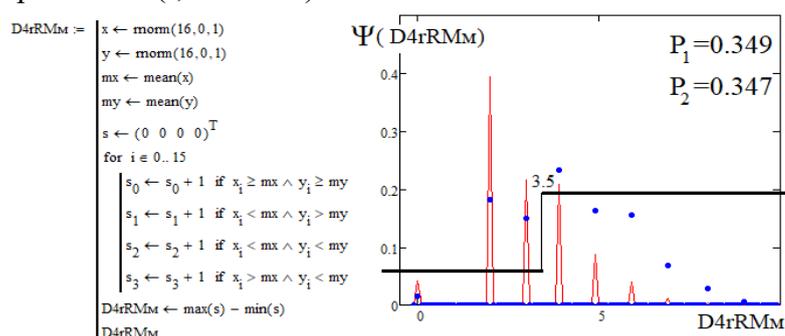


Рис. 6.1. Дискретный критерий размаха вариаций числа точек, попадающих в сектора системы Декартовых координат

6.2. Непрерывный нейрокритерий, построенный на использовании накопления квадратов координат точек малой выборки, попадающих в сектора Декартовой системы координат

К сожалению, дискретный критерий D4rRMm имеет плохую разделимость на независимые и зависимые данные. Повысить уровень разделимости нейрокритерия удастся, если выполнить накапливание квадратов координат точек, попавших в каждую из четвертей Декартовой системы координат. Программная реализация нового нейрокритерия приведена на рис. 6.2.

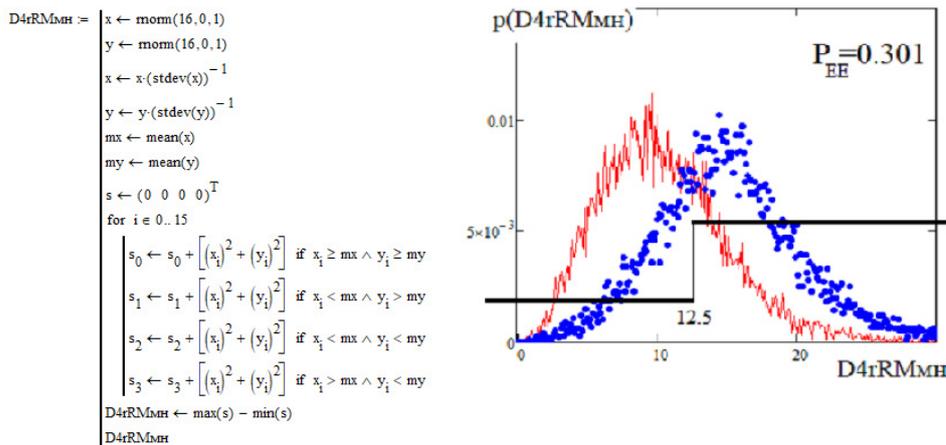


Рис. 6.2. Непрерывный критерий, построенный на накоплении квадратов координат точек, попадающих в разные сектора системы Декартовых координат

Численный эксперимент показал снижение вероятности ошибок первого и второго рода при переходе от дискретного нейрокритерия к непрерывному на 15 %. При этом сохраняется низкая корреляционная сцепленность рассматриваемой группы критериев с классическим критерием Эджуорта – Эдлтона – Пирсона: $\text{corr}(r, D4rRMm) \approx 0.0032$.

6.3. Непрерывный нейрокритерий, построенный на использовании накопления модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, и оценке их размаха

Снизить вероятности ошибок первого и второго рода рассматриваемого класса критериев удастся, если отказаться от накапливания

квадратов координат обнаруженных точек и использовать накапливающие модулей этих координат. В этом случае вероятность появления ошибок снижается примерно на 4 %. При этом сохраняется низкая корреляционная сцепленность рассматриваемой группы критериев с классическим критерием Эджуорта – Эдлтона – Пирсона: $\text{corr}(r, D4rR2M_{\text{Мн}}) \approx 0.002$. Программная реализация нового критерия приведена на рис. 6.3.

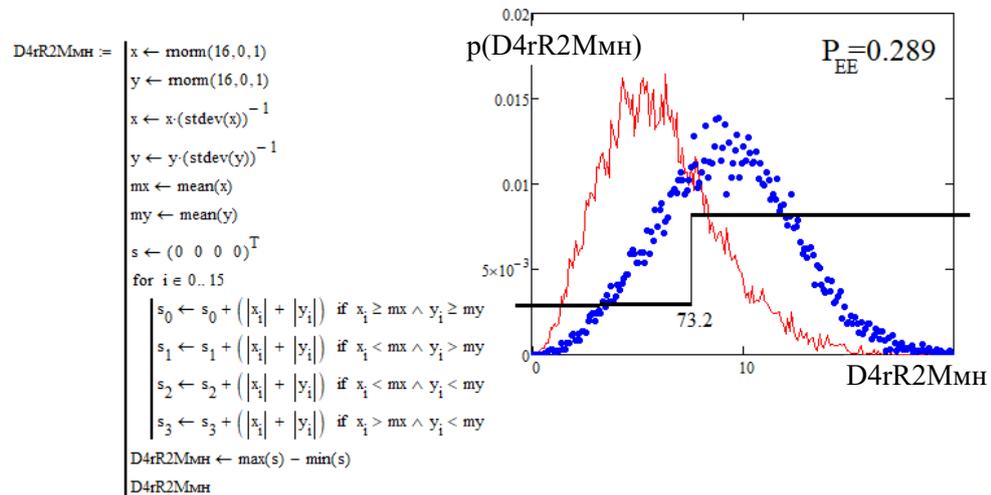


Рис. 6.3. Непрерывный критерий, построенный на накоплении модулей координат точек, попадающих в разные сектора системы Декартовых координат

К сожалению, корреляционная сцепленность двух последних критериев оказывается достаточно высокой:

$$\text{corr}(D4rRM_{\text{Мн}}, D4rR2M_{\text{Мн}}) \approx 0.565.$$

6.4. Непрерывный нейрокритерий, построенный на использовании накопления квадратов разности координат точек малой выборки, попадающих в сектора Декартовой системы координат, и оценке размаха их вариаций

Существенно повысить качество разделения зависимых и независимых данных удастся в том случае, если перейти к накапливанию квадратов разностей зависимых и независимых координат точек, обнаруженных в разных секторах Декартовой системы координат. В этом случае удастся снизить вероятности ошибок нейрокритерия примерно на 38 %. Программная реализация нового критерия приведена на рис. 6.4.

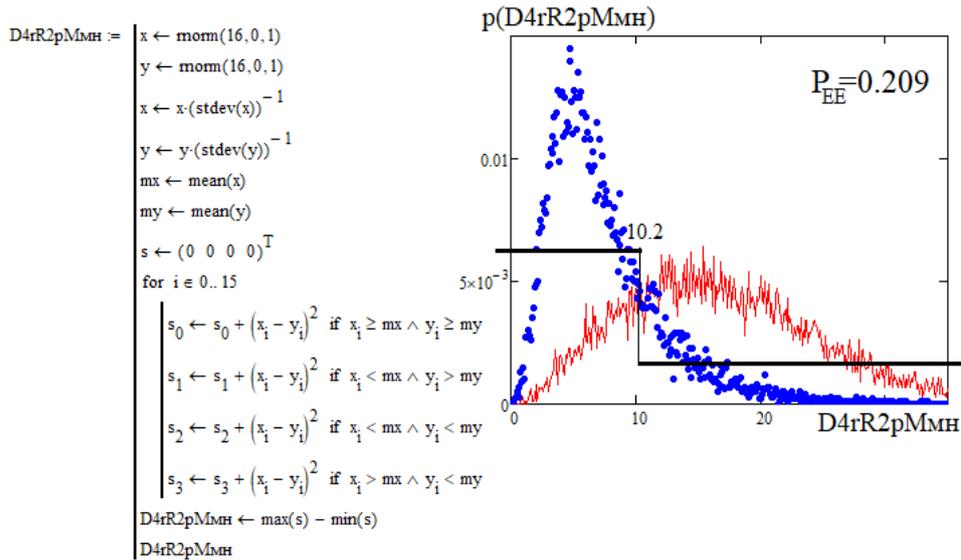


Рис. 6.4. Непрерывный критерий, построенный на накоплении квадратов разности координат точек, попадающих в разные сектора системы Декартовых координат

К сожалению, наряду с ростом качества разделения зависимых и независимых данных растет корреляционная сцепленность нового нейрокритерия с группой классических нейрокритериев $\text{corr}(r, D4rR2pMmn) \approx -0.732$, т.е. по формуле математических вычислений нейрокритерий является оригинальным, а по своей сути он оказался близок к классическому критерию конца XIX в. Эджуорта – Эдлтона – Пирсона.

6.5. Непрерывный нейрокритерий, построенный на использовании накопления квадратов произведений координат точек малой выборки, попадающих в разные сектора Декартовой системы координат, с оценкой размаха их вариаций

Вернуться к выполнению условия низкой корреляционной сцепленности с классикой удастся, если строить очередной нейрокритерий, опираясь на накапливание квадратов произведений координат точек малой выборки, попадающих в разные сектора Декартовой системы координат. Программная реализация нового критерия приведена на рис. 6.5.

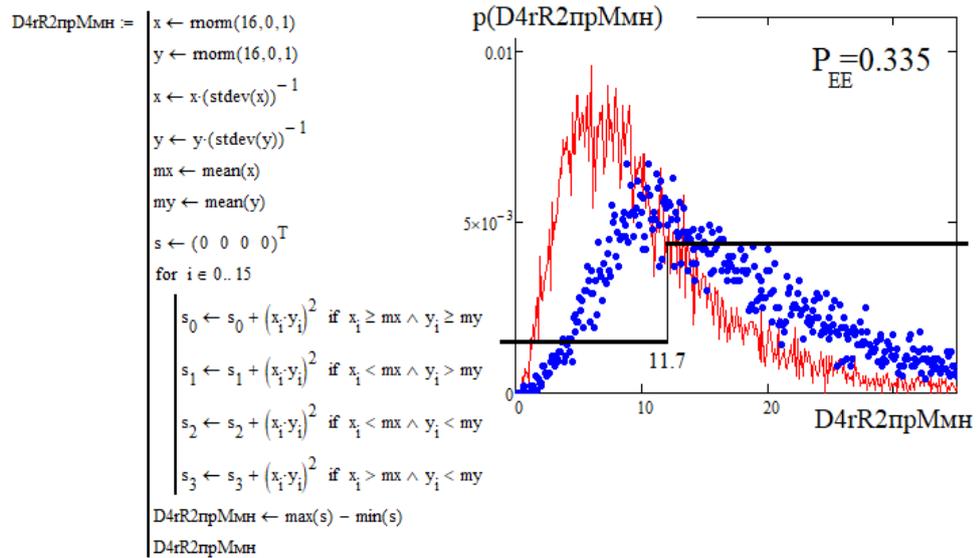


Рис. 6.5. Непрерывный критерий, построенный на накоплении квадратов произведений координат точек, попадающих в разные сектора системы Декартовых координат

К сожалению, разделяющая способность нового нейрокритерия оказывается низкой $P_{EE} \approx 0.335$ в сравнении с классическим критерием $P_{EE} \approx 0.144$. Это затрудняет их совместное использование в одной группе. Низкий уровень корреляционной сцепленности двух критериев $\text{corr}(r, D4rR2прМмн) \approx -0.019$ перестает быть значимым фактором, так как эти критерии трудно использовать в одной группе.

6.6. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, с оценкой размаха их вариаций

Снизить уровень ошибок первого и второго рода на 20 % возможно при одновременном сохранении низкого уровня корреляционной сцепленности. Это достижимо, если перейти к накапливанию модулей произведений точек выборки. Программная реализация нового непрерывного критерия представлена на рис. 6.6, на этом же рисунке даны результаты его численного моделирования.

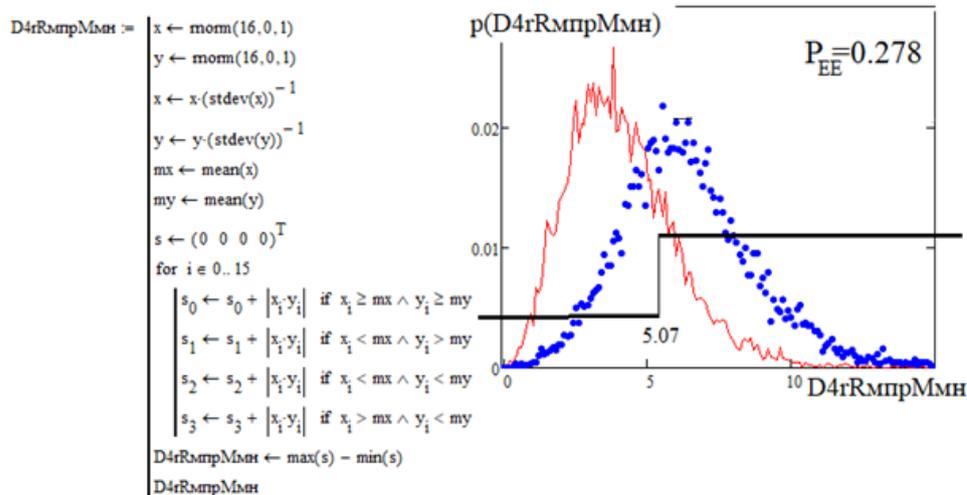


Рис. 6.6. Непрерывный критерий, построенный на накоплении модулей координат точек, попадающих в разные сектора системы Декартовых координат, с оценкой размаха их вариаций

Корреляционная сцепленность нового нейрокритерия и классического критерия конца XIX в. Эджуорта – Эдлттона – Пирсона остается низка: $\text{corr}(r, D4rRmnpMmn) \approx 0.0188$.

6.7. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, с аддитивной сверткой

Следует подчеркнуть, что все нейрокритерии этого параграфа были построены на учете размаха вариаций статистик точек, попавших в разные сектора Декартовой системы координат. Фактически функционал вычисления размаха вариаций статистик точек, попавших в разные сектора системы координат, играет роль свертки четырех промежуточных статистических параметров в один выходной параметр новых критериев.

Заменим функционал вычисления размаха вариаций статистик на аддитивную выходную свертку классического критерия Олмстеда – Тьюки (1947) [7]. В итоге мы получаем статистический критерий с очень высокой разделяющей способностью (почти как у классического критерия конца XIX в. Эджуорта – Эдлттона – Пирсона). Программная реализация этого критерия дана на рис. 6.7.

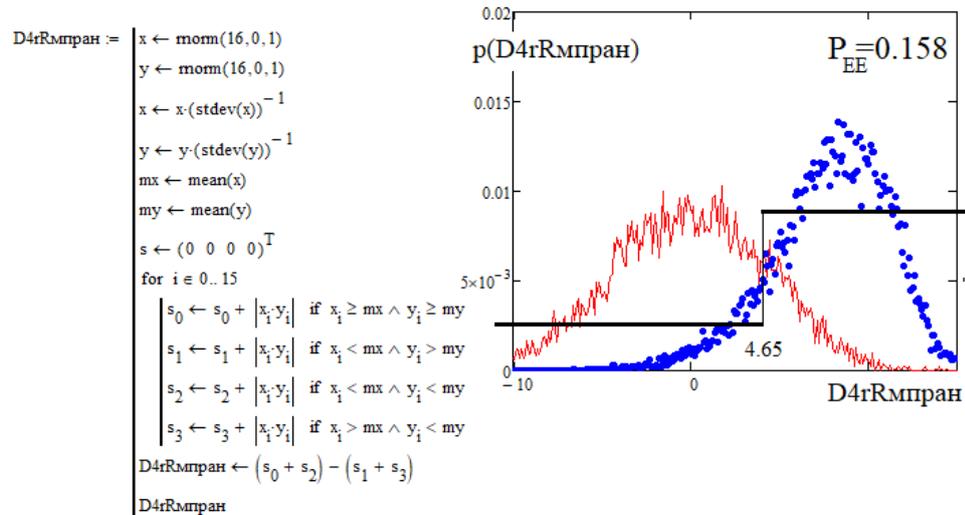


Рис. 6.7. Непрерывный критерий, построенный на накоплении модулей произведений координат точек, попадающих в разные сектора системы Декартовых координат, с последующей аддитивной сборкой

К сожалению, численный эксперимент показал, что для малых выборок этот нейрокритерий является новым только по вычислительной формуле. По своей сути он почти повторяет классический критерий, так как их уровень корреляционной сцепленности составляет $\text{corr}(r, D4rR\text{мплан}) \approx 0.967$.

6.8. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, с мультипликативной сверткой

Тот факт, что аддитивная свертка выходных данных критерия Олмстеда – Тьюки (1947) [7] дает столь высокий уровень корреляционной сцепленности предшествующего критерия с классикой является негативной данностью. Снизить высокий уровень высокой коррелированности критериев вполне возможно, если заменить аддитивную выходную свертку критерия на ее мультипликативный аналог. Программная реализация нового нейрокритерия приведена на рис. 6.8.

Численный эксперимент показал, что новая модификация нейрокритерия сохраняет ту же самую способность к разделению зависимых и независимых данных.

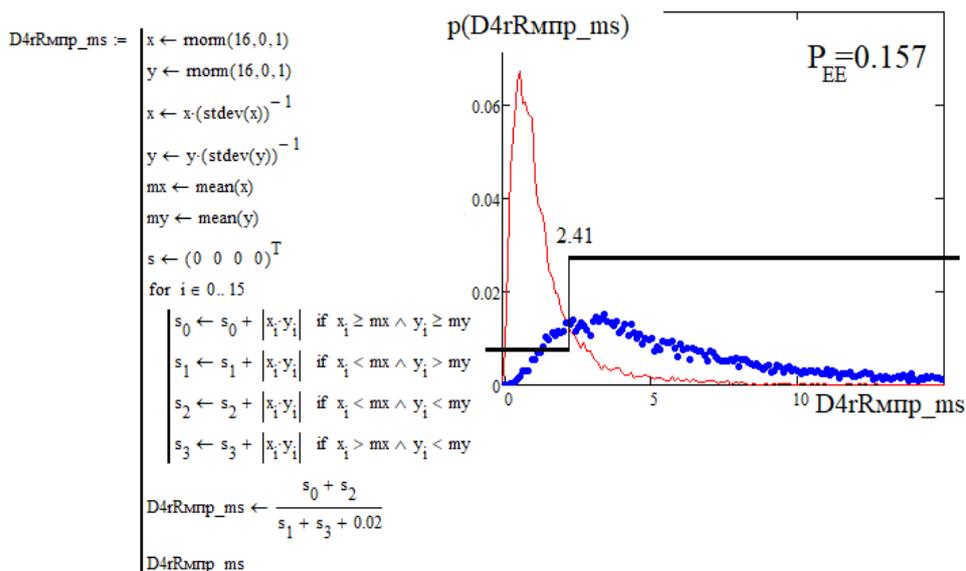


Рис. 6.8. Непрерывный критерий, построенный на накоплении модулей произведений координат точек, попадающих в разные сектора системы Декартовых координат, с последующей мультипликативной сборкой

Однако новый нейрокритерий дает гораздо более низкий уровень корреляционной сцепленности с классикой конца XIX в.: $\text{corr}(r, D4rRmnp_ms) \approx 0.355$.

6.9. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, с обратной мультипликативной сверткой

Возможен еще один вариант этого типа нейрокритерия. Он строится заменой местами числителя и знаменателя в последней мультипликативной свертке предшествующего критерия. Программная реализация нового нейрокритерия приведена на рис. 6.9. Смена местами числителя и знаменателя приводит к изменению знака коэффициента корреляции нового критерия и классического критерия: $\text{corr}(r, D4rRmnp_sm) \approx -0.461$. Уровни качества разделения зависимых и независимых данных для нового и предыдущего нейрокритериев практически совпадают.

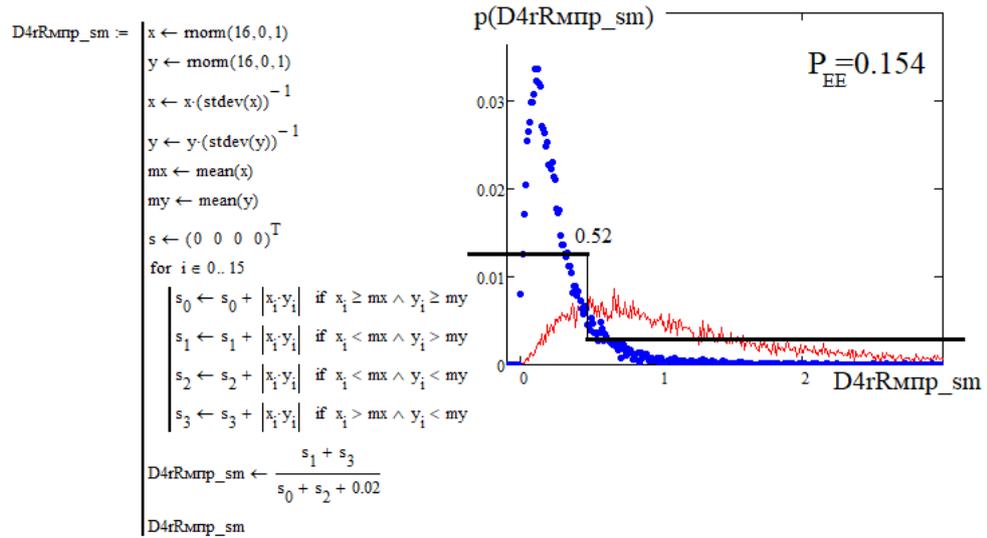


Рис. 6.9. Непрерывный критерий, построенный на накоплении модулей произведений координат точек, попадающих в разные сектора системы Декартовых координат, с обратной мультипликативной сборкой

6.10. Непрерывный нейрокритерий, построенный на использовании накопления произведений модулей координат точек малой выборки, попадающих в сектора Декартовой системы координат, опирающейся на медианы

Все рассмотренные выше модификации статистических критериев были построены, опираясь на использование Декартовой системы координат с центром в точках математического ожидания двух координат малой выборки. Однако классический угловой критерий Олмстеда – Тьюки (1947) [7] предполагает размещение центра Декартовой системы в точках медиан координат точек малой выборки. Формально использование медиан вместо математических ожиданий следует рассматривать как некоторое приближение.

В связи с выше изложенным мы имеем возможность скорректировать все рассмотренные критерии, заменив в них вычисление математических ожиданий на вычисление медиан. Поясним эту ситуацию на примере нейрокритерия параграфа 6.6. Модификация этого критерия остается работоспособной, ее программная реализация приведена на рис. 6.10.

Численный эксперимент показал, что замена оценок математических ожиданий на менее точные оценки медиан приводят

к существенному росту вероятности ошибок новой модификации. Наблюдается повышение вероятностей появления ошибок примерно на 63 % при сохранении низкого уровня корреляционной сцепленности с классикой $\text{corr}(r, D4rRmp_Mm) \approx 0.0573$.

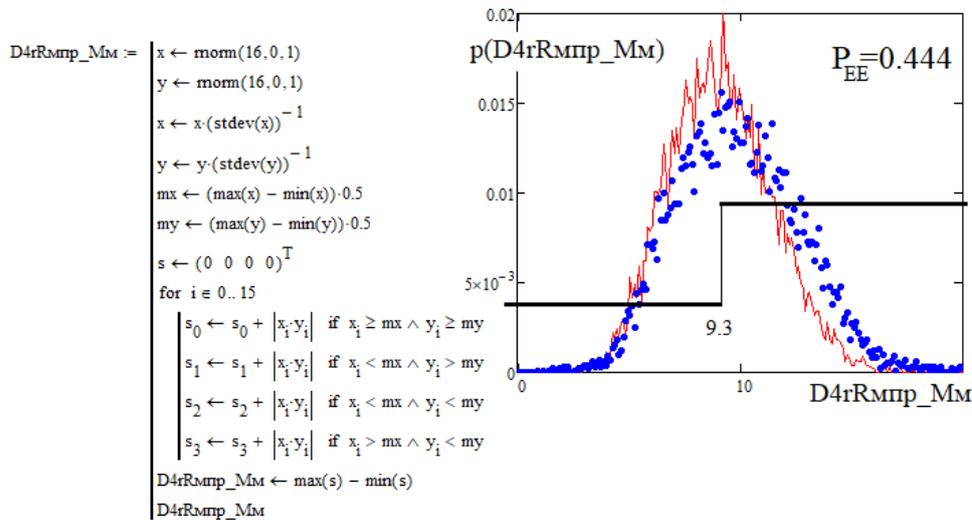


Рис. 6.10. Непрерывный критерий, построенный на накоплении модулей произведений координат точек, попадающих в разные сектора системы Декартовых координат, с центром в медианах малой выборки

Положительным моментом является то, что два последних критерия имеют низкий уровень корреляционной сцепленности:

$$\text{corr}(D4rRmpMmn, D4rRmp_Mm) \approx 0.238.$$

6.11. Непрерывный нейрокритерий, построенный на учете статистик модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат

Оставаясь в классе опоры центра системы Декартовых координат на медианы распределения двухмерных данных малой выборки, возможно увеличить разделяемость зависимых и независимых данных. Для этого приходится объединять все 16 возможных произведений и среди них выявлять максимальное значение модулей. Это позволяет снизить вероятность появления ошибок первого и второго рода примерно на 25 %.

Программная реализация этого варианта критерия приведена на рис. 6.11. К сожалению, за рост качества разделения зависимых

и независимых данных приходится расплачиваться значительным ростом корреляционной сцепленности нового нейрокритерия и его классического аналога: $\text{corr}(r, D4rRmnp_max) \approx 0.46$.

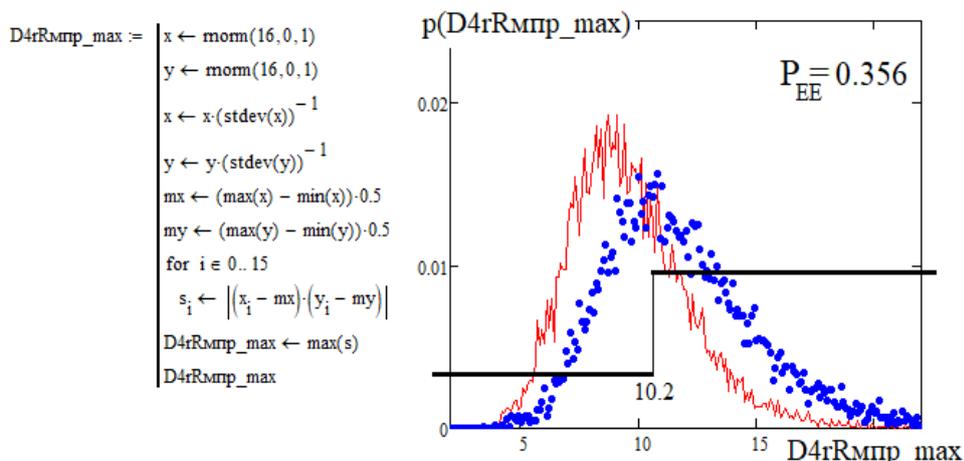


Рис. 6.11. Непрерывный критерий, построенный на накоплении модулей произведений координат точек в системе Декартовых координат, с центром в медианах малой выборки

Два последних критерия имеют низкий уровень корреляционной сцепленности: $\text{corr}(D4rRmnpMm, D4rRmnp_max) \approx -0.031$.

6.12. Непрерывный нейрокритерий, построенный на учете статистик модулей произведения координат точек, попадающих в разные сектора Декартовой системы координат, с регуляризацией оценок медиан

Одним из путей дальнейшего снижения вероятностей ошибок нейрокритерия является регуляризация оценок медиан малых выборок. В простейшем случае регуляризация может быть выполнена усреднением двух крайних точек выборки с правой и с левой стороны. Программная реализация этого варианта критерия приведена на рис. 6.12. Усреднение координат двух крайних точек позволяет снизить вероятность появления ошибок примерно на 3.5 %.

К сожалению, два последних критерия имеют высокий уровень корреляционной сцепленности:

$$\text{corr}(D4rRmnp_max, D4rRmnpS_max) \approx 0.964.$$

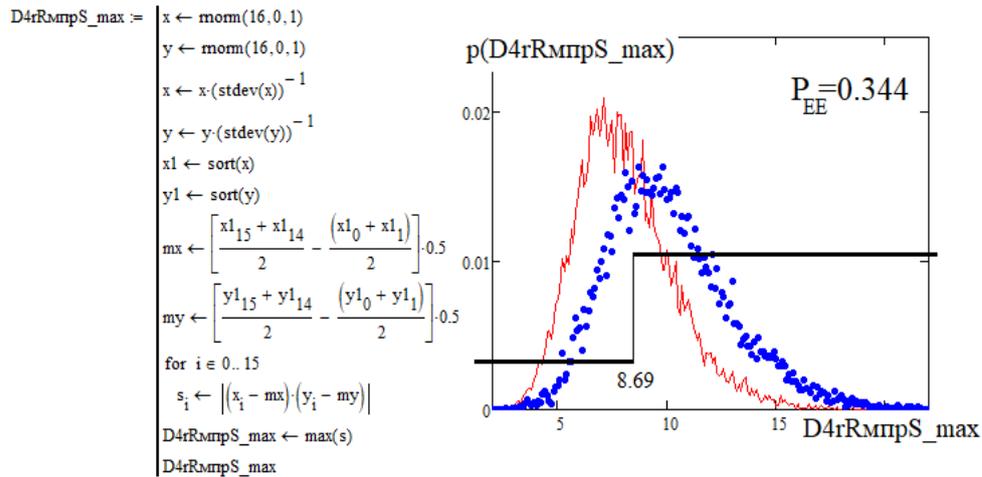


Рис. 6.12. Непрерывный критерий, построенный на накоплении модулей произведений координат точек в системе Декартовых координат, с центром в медианах, которые оцениваются с регуляризацией по четырем крайним отсчетам малой выборки

Однако уровень корреляционной сцепленности последнего критерия с другими близкими критериями низок:

$$\text{corr}(D4rRmpS_max, D4rmp_sm) \approx -0.306,$$

$$\text{corr}(D4rRmpS_max, D4rmp_sm) \approx -0.286.$$

6.13. Корреляционная сцепленность в группе критериев, построенных на статистическом анализе точек, попадающих в сектора Декартовой системы координат

Таким образом, при решении двухмерной задачи проверки гипотезы независимости малых выборок к десяткам уже известных статистических критериев удастся построить группу, содержащую более чем 12 новых статистических критериев. Переход к 12-эквивалентным нейрокритериям позволяет получить нейросетевое решение с 12-кратной избыточностью выходного бинарного кода. Кодовая избыточность может быть свернута любой из известных кодовых конструкций, построенной для обнаружения и устранения в избыточном коде ошибок.

Если бы все 12 новых статистических критериев давали бы независимые (некоррелированные) отклики, то задача проверки гипотезы независимости была бы решена с приемлемой для практики вероятностью – 0.99 и выше для малых выборок в 16 и более опытов.

К сожалению, этого не происходит. Внутри группы откликов из 12 новых статистических критериев существуют значимые корреляционные связи. При объединении нейрокритериев по тройкам или пятеркам следует стремиться к наибольшему уровню независимости их откликов (минимуму модулей их коэффициентов корреляции). На рис. 6.13 приведены оценки парных коэффициентов корреляции откликов рассматриваемой группы новых статистических критериев.

$\text{corr}(r, D4rRMm) = -0.021$		
$\text{corr}(r, D4rRMmn) = -0.021$	$\text{corr}(D4rRMm, D4rRMmn) = 0.318$	
$\text{corr}(r, D4rR2Mmn) = -0.023$	$\text{corr}(D4rRMm, D4rR2Mmn) = 0.562$	$\text{corr}(D4rRMmn, D4rR2Mmn) = 0.827$
$\text{corr}(r, D4rR2pMmn) = -0.732$	$\text{corr}(D4rRMm, D4rR2pMmn) = 0.059$	$\text{corr}(D4rRMmn, D4rR2pMmn) = 0.357$
$\text{corr}(r, D4rR2npMmn) = -0.031$	$\text{corr}(D4rRMm, D4rR2npMmn) = 0.067$	$\text{corr}(D4rRMmn, D4rR2npMmn) = 0.518$
$\text{corr}(r, D4rRmpMmn) = -0.025$	$\text{corr}(D4rRMm, D4rRmpMmn) = 0.209$	$\text{corr}(D4rRMmn, D4rRmpMmn) = 0.774$
$\text{corr}(r, D4rRmpран) = 0.967$	$\text{corr}(D4rRMm, D4rRmpран) = -0.024$	$\text{corr}(D4rRMmn, D4rRmpран) = -0.022$
$\text{corr}(r, D4rRmp_ms) = 0.305$	$\text{corr}(D4rRMm, D4rRmp_ms) = 0.116$	$\text{corr}(D4rRMmn, D4rRmp_ms) = 0.116$
$\text{corr}(r, D4rRmp_sm) = -0.598$	$\text{corr}(D4rRMm, D4rRmp_sm) = 0.220$	$\text{corr}(D4rRMmn, D4rRmp_sm) = 0.233$
$\text{corr}(r, D4rRmp_Mm) = 0.029$	$\text{corr}(D4rRMm, D4rRmp_Mm) = -0.032$	$\text{corr}(D4rRMmn, D4rRmp_Mm) = 0.116$
$\text{corr}(r, D4rRmp_max) = 0.446$	$\text{corr}(D4rRMm, D4rRmp_max) = -0.028$	$\text{corr}(D4rRMmn, D4rRmp_max) = -0.020$
$\text{corr}(r, D4rRmpS_max) = 0.469$	$\text{corr}(D4rRMm, D4rRmpS_max) = -0.031$	$\text{corr}(D4rRMmn, D4rRmpS_max) = -0.046$

Рис. 6.13. Коэффициенты корреляции критериев, построенных на статистиках точек, попавших в сектора Декартовой системы координат (начало)

К сожалению, только простейшие критерии данной главы имеет смысл использовать совместно с наиболее популярным критерием Эджуорта – Эдлтона – Пирсона (верхняя часть первого столбца на рис. 6.13). Внутри группы рассматриваемых критериев корреляционная сцепленность низка (второй и третий столбцы на рис. 6.13).

Коэффициенты корреляции критериев, построенных на статистиках точек, попавших в сектора Декартовой системы координат

$\text{corr}(D4rR2Mmn, D4rR2pMmn) = 0.296$		
$\text{corr}(D4rR2Mmn, D4rR2npMmn) = 0.311$	$\text{corr}(D4rR2pMmn, D4rR2npMmn) = 0.338$	
$\text{corr}(D4rR2Mmn, D4rRmpMmn) = 0.695$	$\text{corr}(D4rR2pMmn, D4rRmpMmn) = 0.397$	$\text{corr}(D4rR2npMmn, D4rRmpMmn) = 0.778$
$\text{corr}(D4rR2Mmn, D4rRmpран) = -0.026$	$\text{corr}(D4rR2pMmn, D4rRmpран) = -0.789$	$\text{corr}(D4rR2npMmn, D4rRmpран) = -0.030$
$\text{corr}(D4rR2Mmn, D4rRmp_ms) = 0.130$	$\text{corr}(D4rR2pMmn, D4rRmp_ms) = -0.221$	$\text{corr}(D4rR2npMmn, D4rRmp_ms) = 0.073$
$\text{corr}(D4rR2Mmn, D4rRmp_sm) = 0.271$	$\text{corr}(D4rR2pMmn, D4rRmp_sm) = 0.492$	$\text{corr}(D4rR2npMmn, D4rRmp_sm) = 0.135$
$\text{corr}(D4rR2Mmn, D4rRmp_Mm) = 0.177$	$\text{corr}(D4rR2pMmn, D4rRmp_Mm) = 0.141$	$\text{corr}(D4rR2npMmn, D4rRmp_Mm) = 0.150$
$\text{corr}(D4rR2Mmn, D4rRmp_max) = -0.062$	$\text{corr}(D4rR2pMmn, D4rRmp_max) = -0.300$	$\text{corr}(D4rR2npMmn, D4rRmp_max) = 0.134$
$\text{corr}(D4rR2Mmn, D4rRmpS_max) = -0.063$	$\text{corr}(D4rR2pMmn, D4rRmpS_max) = -0.307$	$\text{corr}(D4rR2npMmn, D4rRmpS_max) = 0.120$

Рис. 6.13. Продолжение

Третий критерий рассматриваемой группы – D4rR2Mmn в основном имеет коэффициенты корреляции с низким значением модуля.

Как видно из первого столбца продолжения рис. 6.13, только модуль одного коэффициента корреляции имеет значение более 0.5. Эта тенденция сохраняется и для других критериев рассматриваемой группы.

Коэффициенты корреляции критериев, построенных на статистиках точек, попавших в сектора Декартовой системы координат

$\text{corr}(D4rRmpMmn, D4rRmpRan) = -0.024$		
$\text{corr}(D4rRmpMmn, D4rRmp_ms) = 0.107$	$\text{corr}(D4rRmpRan, D4rRmp_ms) = 0.311$	
$\text{corr}(D4rRmpMmn, D4rRmp_sm) = 0.220$	$\text{corr}(D4rRmpRan, D4rRmp_sm) = -0.612$	$\text{corr}(D4rRmp_ms, D4rRmp_sm) = -0.128$
$\text{corr}(D4rRmpMmn, D4rRmp_Mm) = 0.238$	$\text{corr}(D4rRmpRan, D4rRmp_Mm) = 0.034$	$\text{corr}(D4rRmp_ms, D4rRmp_Mm) = 0.002$
$\text{corr}(D4rRmpMmn, D4rRmp_max) = 0.002$	$\text{corr}(D4rRmpRan, D4rRmp_max) = 0.433$	$\text{corr}(D4rRmp_ms, D4rRmp_max) = 0.142$
$\text{corr}(D4rRmpMmn, D4rRmpS_max) = 0.005$	$\text{corr}(D4rRmpRan, D4rRmpS_max) = 0.456$	$\text{corr}(D4rRmp_ms, D4rRmpS_max) = 0.146$
$\text{corr}(D4rRmp_sm, D4rRmp_Mm) = -0.035$		
$\text{corr}(D4rRmp_sm, D4rRmp_max) = -0.286$	$\text{corr}(D4rRmp_Mm, D4rRmp_max) = 0.376$	
$\text{corr}(D4rRmp_sm, D4rRmpS_max) = -0.306$	$\text{corr}(D4rRmp_Mm, D4rRmpS_max) = 0.427$	$\text{corr}(D4rRmp_max, D4rRmpS_max) = 0.964$

Рис. 6.13. Окончание

Тезис о низкой корреляционной сцепленности подтверждается тем, что в продолжении рис. 6.13 только один коэффициент корреляции имеет модуль больше 0.75.

Список литературы к главе 6

1. Серикова Ю. И. Синтез нового статистического критерия для проверки гипотезы независимости, слабо связанного с оценками по классической формуле Эджуорта – Эдлттона – Пирсона // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 59–62.

НОВЫЕ НЕЙРОКРИТЕРИИ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОСТРОЕННЫЕ НА СТАТИСТИЧЕСКОМ АНАЛИЗЕ ТОЧЕК, ВЫДЕЛЕННЫХ ПАРАБОЛИЧЕСКИМИ КВАНТОВАТЕЛЯМИ

7.1. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2

Ранее в данном справочнике были рассмотрены нейрокритерии, использующие линейные квантователи и эллиптические квантователи. Очевидно, что типов квантователей множество. Любая двухмерная функция может быть использована как функция квантования. При этом по разные стороны функции-границы будут разные состояния данных. Как пример, рассмотрим квадратичные квантователи, которые разбивают малую выборку на две группы точек, находящихся ниже и выше параболы.

Реализация такого дискретного критерия может быть построена на подсчете числа точек, лежащих выше обычной параболы («рога» вверх) и ниже перевернутой параболы («рога» вниз). Программная реализация такого критерия приведена на рис. 7.1.

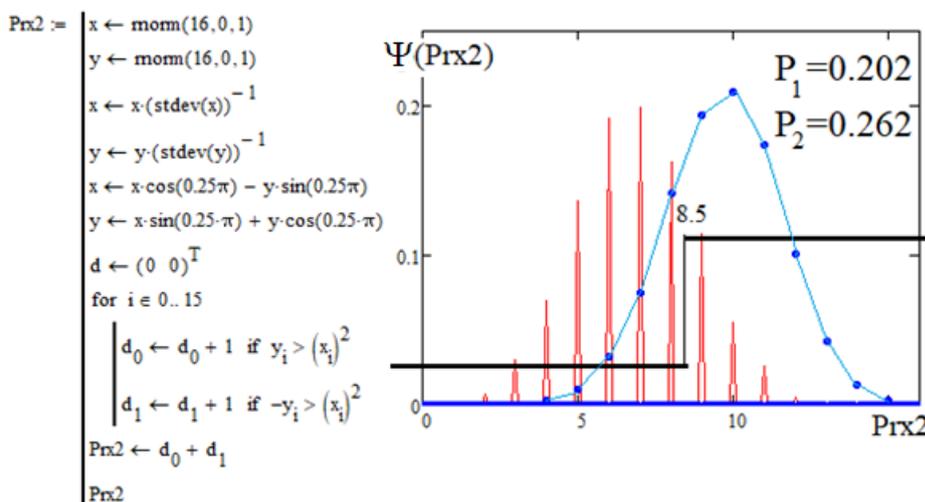


Рис. 7.1. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2

Численный эксперимент показал, что мощность нового критерия достаточно высока. Примерно 30 % классических критериев, синтезированных в XX в., имеют вероятности ошибок первого и второго рода выше в сравнении с новым критерием. Спектр критерия имеет от 11 до 12 линий значимых амплитуд вероятности.

7.2. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – y^2

Дискретный критерий предыдущего параграфа построен на возведении в квадрат координаты – x . Если применить квадратичное преобразование к другой переменной – y , то мы получим еще один вариант дискретного критерия. Программная реализация новой модификации приведена на рис. 7.2.

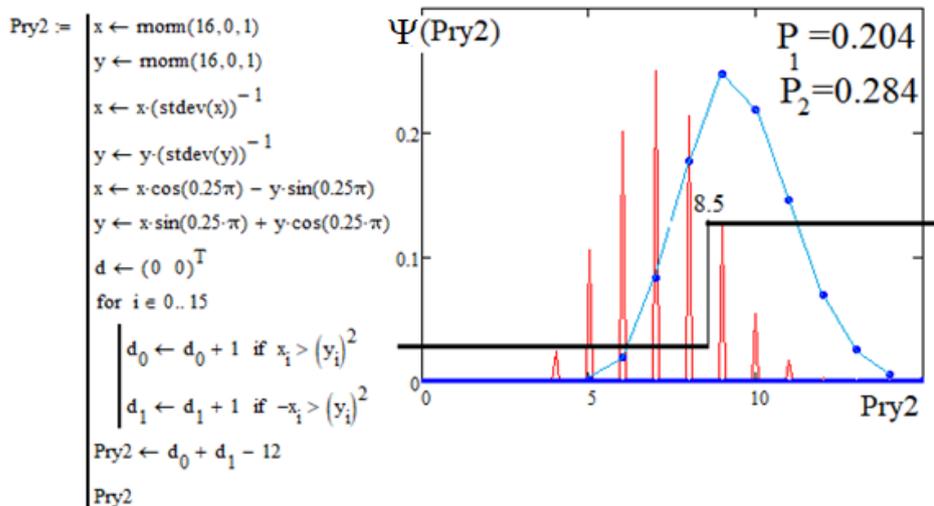


Рис. 7.2. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – y^2

Как и предыдущая модификация, новый критерий имеет достаточно низкий уровень вероятностей ошибок первого и второго рода по сравнению с критериями прошлого века. Дискретный спектр критерия имеет от 8 до 9 линий значимых амплитуд вероятности. Программная реализация за счет специального подбора смещения обеспечивает совпадение порогов $k = 8.5$ выходных квантователей двух последних нейрокритериев.

7.3. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2 и пары квантователей – y^2

Увеличить мощность двух предшествующих критериев удастся, если их объединить. Усиленная модификация строится на двух парах парабол, его программная реализация приведена на рис. 7.3. Численная реализация новой модификации критерия имеет 16 линий значимых амплитуд вероятности. К сожалению, снижение вероятности ошибок у модифицированного критерия незначительно (до 9 %) по сравнению с наиболее мощным первым критерием – P_{rx2} . Видимо, объединение нейрокритериев в группы и свертывание избыточности их кодов будут более выгодными. Свертывание данных простым суммированием в этом случае является мало эффективным.

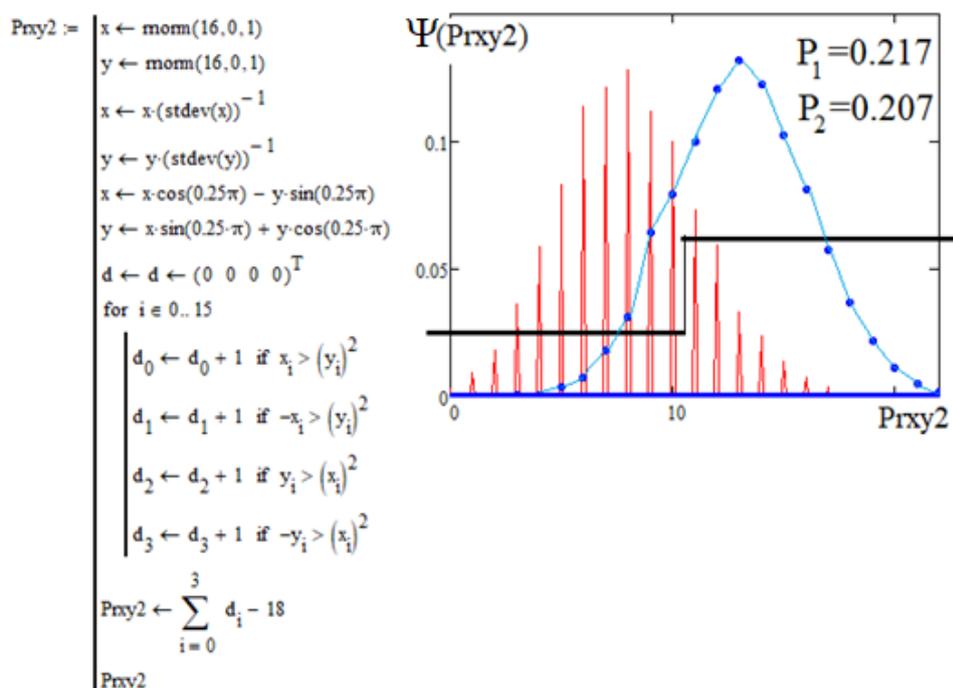


Рис. 7.3. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2 и пары квантователей – y^2

При бинарном квантовании откликов нового критерия с порогом $k = 10.5$ мы получаем новый бинарный нейрокритерий с приемлемым для практики уровнем вероятности ошибок первого и второго рода.

7.4. Дискретный нейрокритерий, построенный на использовании первой пары из квадратичных квантователей – x^2 и второй пары квантователей – y^2 через аддитивно-мультипликативное свертывание

Возможно свертывание промежуточных данных аддитивно-мультипликативной процедурой. Программная реализация такой свертки дана на рис. 7.4. Свертка выполнена попарным перемножением данных четырех компонент с суммированием значений b не повторяющихся парных произведений. Очевидно, что возможна реализация этого типа критериев путем перемножения данных тройками с последующим их суммированием. К сожалению, все варианты этого типа сверток не дают существенного снижения вероятностей ошибок первого и второго рода. Так, по сравнению с предыдущим критерием снижение вероятности ошибок менее 1 %. Кроме того, наблюдается высокий уровень корреляционной сцепленности нового критерия с предыдущим критерием: $\text{corr}(\text{Prxy2}, \text{Prxy2ам}) = 0.987$.

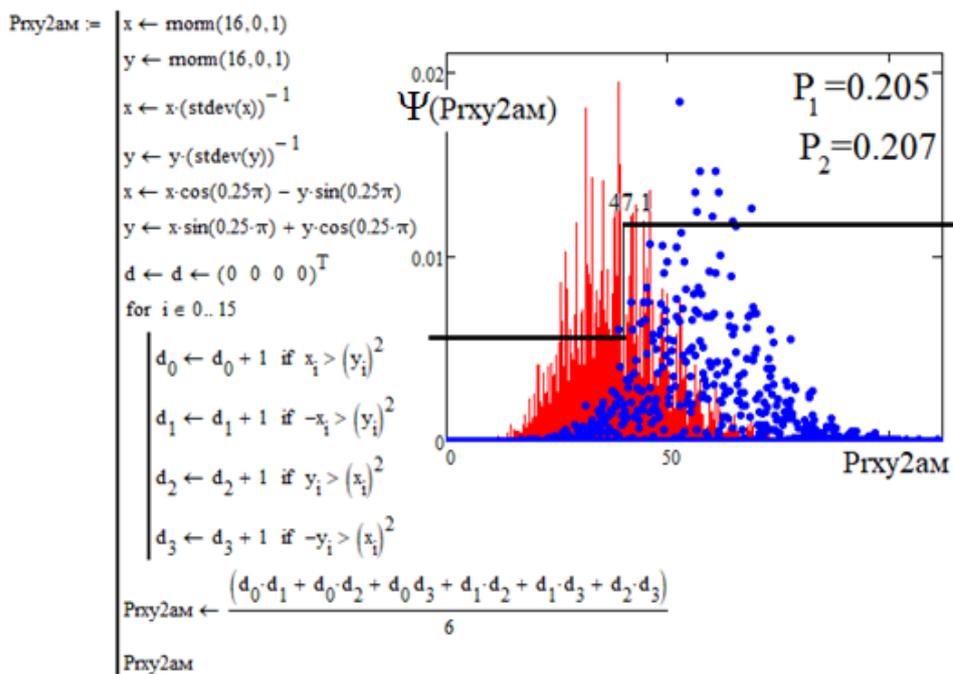


Рис. 7.4. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2 и пары квантователей – y^2 через аддитивно-мультипликативное свертывание

7.5. Дискретный нейрокритерий, построенный на использовании первой пары из квадратичных квантователей – x^2 и второй пары квантователей – y^2 через свертывание вычислением максимума

В случае, если опираться на поиск максимума промежуточных данных (этот прием использован при синтезе критерия Колмогорова – Смирнова), то результат свертывания ухудшается. По сравнению с предыдущим критерием вероятности ошибок первого и второго рода вырастают примерно на 23 %. Программная реализация рассматриваемого критерия приведена на рис. 7.5.

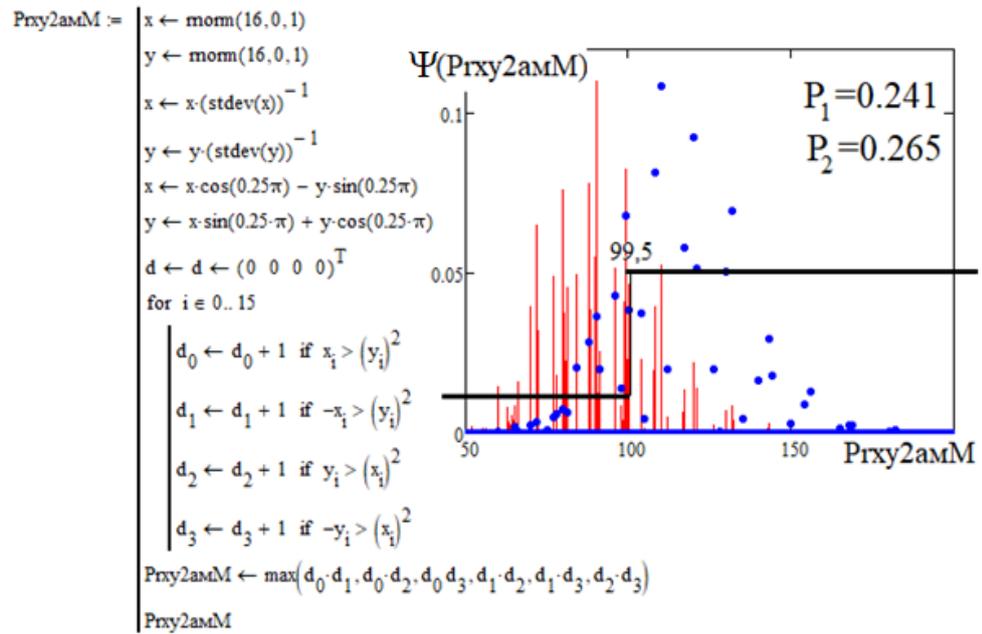


Рис. 7.5. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей по обеим переменным через вычисление максимальных значений парных произведений

Еще одной отрицательной стороной нового критерия является высокий уровень корреляционной сцепленности с близкими критериями: $\text{corr}(\text{Prxy}2, \text{Prxy}2\text{амМ}) = 0.811$, $\text{corr}(\text{Prxy}2\text{ам}, \text{Prxy}2\text{амМ}) = 0.764$. Однако этот критерий имеет и низкий уровень корреляционной сцепленности с рядом иных критериев этой же группы: $\text{corr}(\text{Prxy}2\text{амМ}, \text{Prx}21) = 0.136$, $\text{corr}(\text{Prxy}2\text{амМ}, \text{Prx}22\text{д}) = 0.099$, т.е. при его объединении в одну группу с другими критериями нужно проверить корреляционную сцепленность каждой пары.

7.6. Дискретный нейрокритерий, построенный на использовании первой пары из квадратичных квантователей – x^2 и второй пары квантователей – y^2 через аддитивно-мультипликативное свертывание с вычислением разности максимума и минимума

В случае использования для свертывания промежуточных данных пар произведений максимумов и минимумов промежуточных результатов получается новый статистический критерий. Пример программной реализации нового критерия приведен на рис. 7.6. Этот вариант новой математической конструкции является расширением предшествующей. К сожалению, она дает снижение вероятностных характеристик и в то же время корреляционная сцепленность этих критериев оказывается высокой: $\text{corr}(\text{Prxy2амМм}, \text{Prxy2амМм}) = 0.959$.

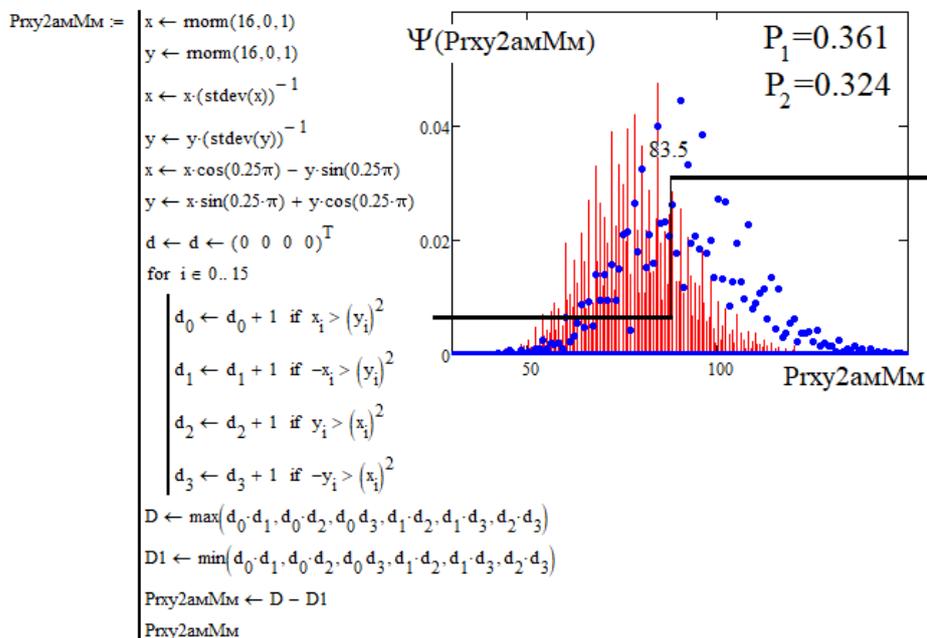


Рис. 7.6. Дискретный нейрокритерий, построенный на использовании пары из квадратичных квантователей – x^2 и пары квантователей – y^2 через аддитивно-мультипликативное свертывание с вычислением разности максимума и минимума

7.7. Непрерывный нейрокритерий, построенный на использовании одного квадратичного квантователя – x^2

Как правило, снизить уровень корреляционной сцепленности группы критериев удастся, если дискретные критерии дополнить их

непрерывными аналогами. Как подтверждение предыдущего тезиса, рассмотрим аналог дискретного критерия Prx2. Этот аналог строится заменой подсчета числа точек, выделенных параболой квадратичного квантователя, на накопление модулей координат этих точек. Программная реализация нового критерия приведена на рис. 7.7.

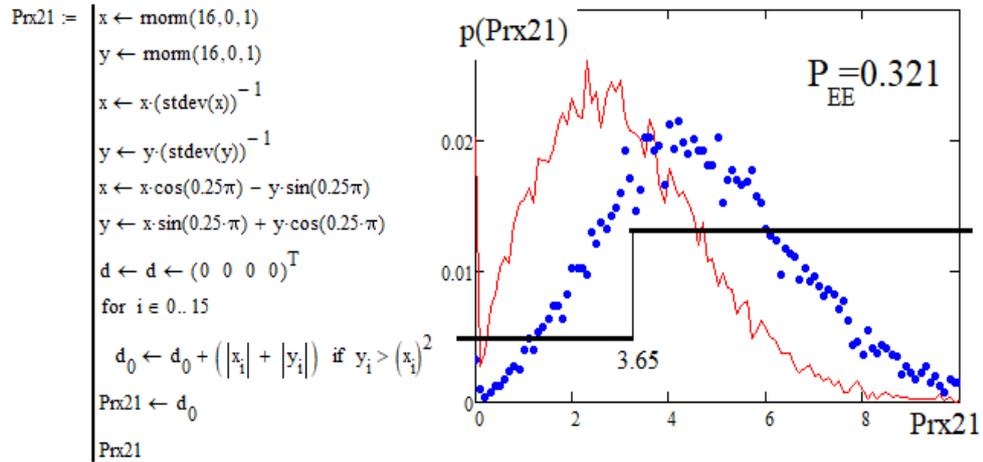


Рис. 7.7. Непрерывный нейрокритерий, построенный на использовании одного квадратичного квантователя $-x^2$

Численный эксперимент по моделированию нового критерия показал, что корреляционная сцепленность критерия-родителя с критерием-потомком значительно уменьшается: $\text{corr}(\text{Prx2}, \text{Prx21}) = 0.318$. Переход от простого подсчета точек к накоплению модулей их координат позволяет устранить на 68.2 % корреляционную сцепленность между двумя близко родственными критериями.

7.8. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей по первой переменной $-x^2$

Очевидно, что повышение числа квадратичных квантователей должно приводить к большему объему получаемой и учитываемой информации. Проиллюстрировать это удастся, если перейти к использованию двух квадратичных квантователей. Программная реализация нового критерия приведена на рис. 7.8.

Численный эксперимент по моделированию нового критерия показал, что удвоение числа квадратичных критериев приводит к снижению уровня вероятности появления ошибок примерно на 30 %.

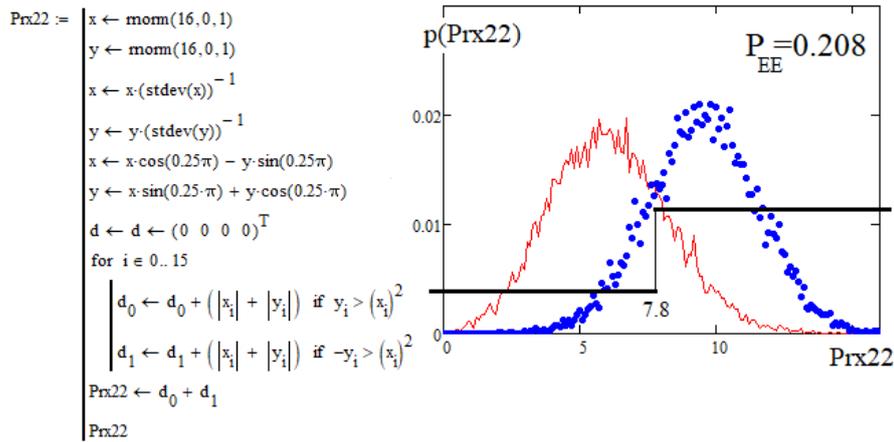


Рис. 7.8. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей по первой переменной – x^2

Одновременно происходит снижение уровня корреляционной сцепленности между близко родственными критериями: $\text{corr}(\text{Prx22}, \text{Prx2}) = 0.756$, $\text{corr}(\text{Prx22}, \text{Prx21}) = 0.431$. Среднее значение модулей этих двух коэффициентов корреляции составит 0.594.

7.9. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей по второй переменной – y^2

Еще один вариант статистического критерия может быть получен, если при анализе данных использовать два квадратичных квантователя по второй переменной. Программная реализация критерия приведена на рис. 7.9.

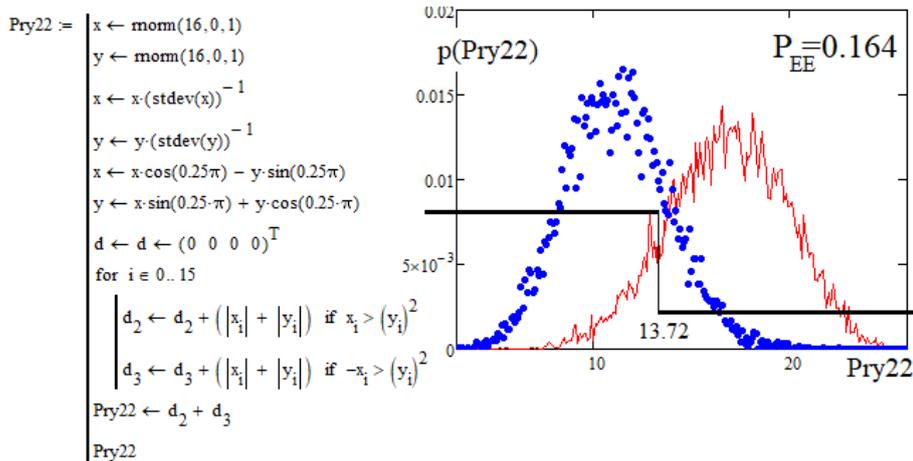


Рис. 7.9. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей по второй переменной – y^2

Численный эксперимент по моделированию нового критерия показал, что переход от смены вариантов применения квадратичных квантователей от одной переменной к другой переменной приводит к смене местами взаимного положения откликов на воздействие нормальными данными и равномерно распределенными данными. Кроме того, снижается вероятность появления ошибок на 26.8 % до величины $P_{EE} \approx 0.164$. Этот эффект снижения вероятностей появления ошибок связан с тем, что выполняется разделение независимых и положительно коррелированных данных. Если на программные реализации двух последних критериев воздействовать отрицательно коррелированными данными, то вероятности появления в двух рисунках поменяются местами.

Сцепленность корреляционных связей родственных критериев повторяет аналогичную сцепленность предыдущего критерия: $\text{corr}(P_{ry22}, P_{ry2}) = 0.851$, $\text{corr}(P_{ry22}, P_{ry21}) = 0.423$. Между собой отклики критериев с квадратичным квантованием по разным переменным слабо связаны: $\text{corr}(P_{ry22}, P_{rx22}) = 0.025$.

7.10. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей – x^2 и двух квадратичных преобразователей по второй переменной – y^2

Дополнительного снижения уровня вероятности ошибок у новых нейрокритериев удастся добиться объединением четырех квадратичных квантователей (по два квантователя на каждую переменную). Пример программной реализации нового критерия приведен на рис. 7.10.

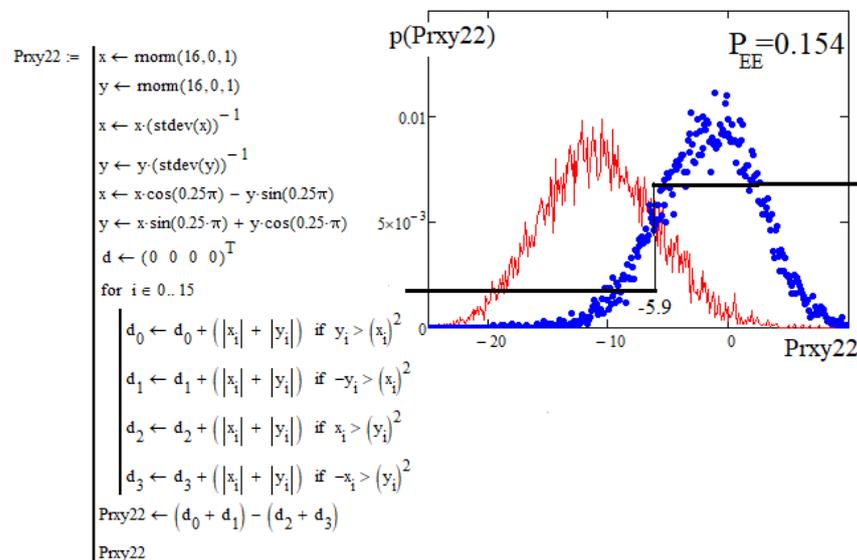


Рис. 7.10. Непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей (два по каждой переменной)

Численный эксперимент по моделированию нового критерия показал, что переход к анализу четырех потоков данных через их аддитивное свертывание позволяет снизить вероятность появления ошибок на 6,5 % до величины $P_{EE} \approx 0.154$. При этом у нового критерия оказываются сильные корреляционные связи с критериями-родителями: $\text{corr}(P_{ry22}, Pr_{xy22}) = -0.909$, $\text{corr}(Pr_{x22}, Pr_{xy22}) = 0.729$.

7.11. Непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей по обеим переменным через их мультипликативную свертку

Кроме аддитивной сборки промежуточных вычислений, возможна их аддитивно-мультипликативная сборка. Пример программной реализации нового критерия приведен на рис. 7.11.

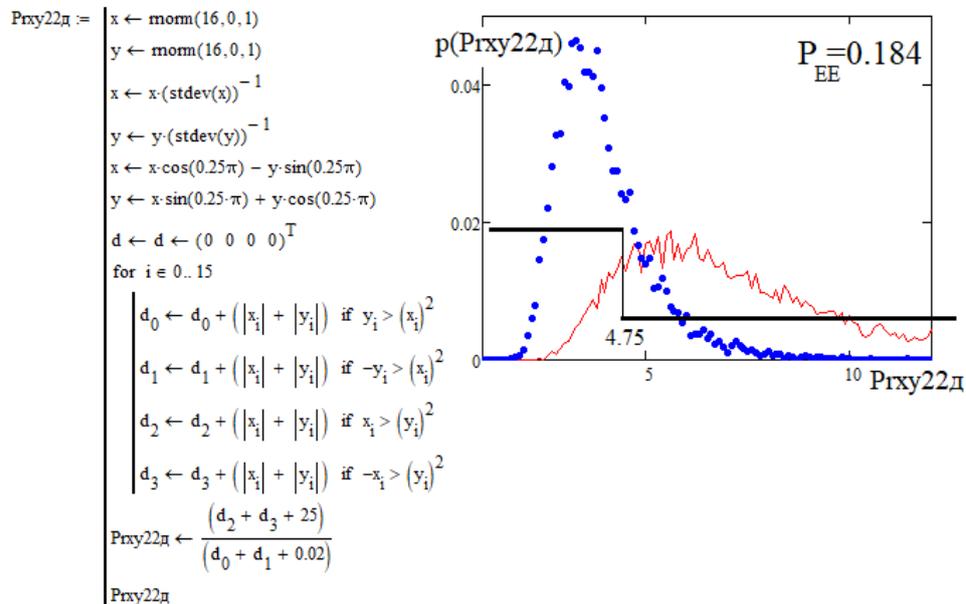


Рис. 7.11. Непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей (два по каждой переменной) с мультипликативным свертыванием

Численный эксперимент по моделированию нового критерия показал, что переход к анализу четырех потоков данных через их аддитивное свертывание, к сожалению, повышает вероятность появления ошибок и высокого уровня корреляционной сцепленности: $\text{corr}(P_{ry22}, Pr_{xy22д}) = 0.448$, $\text{corr}(Pr_{x22}, Pr_{xy22д}) = -0.673$. При этом

для других родственных критериев наблюдается низкий уровень корреляционной сцепленности:

$$\text{corr}(\text{Pr}y_{21}, \text{Pr}x_{y22д}) = 0.169, \text{corr}(\text{Pr}x_{21}, \text{Pr}x_{y22д}) = 0.141.$$

7.12. Непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей по обеим переменным с дроблением каждого отклика на две составляющие и вычислении стандартного отклонения

Одной из возможных ветвей развития критериев проверки гипотезы независимости является дробление числа параметров внутренних вычислений. Например, этого можно добиться делением пополам четырех параметров предыдущего критерия. Программная реализация такого типа критерия приведена на рис. 7.12. Если отклик программной реализации критерия подвергнуть выходному бинарному квантованию, установив порог сравнения $k = 3.01$, то мы получим нейрокритерий с близкими значениями вероятностей ошибок первого и второго рода.

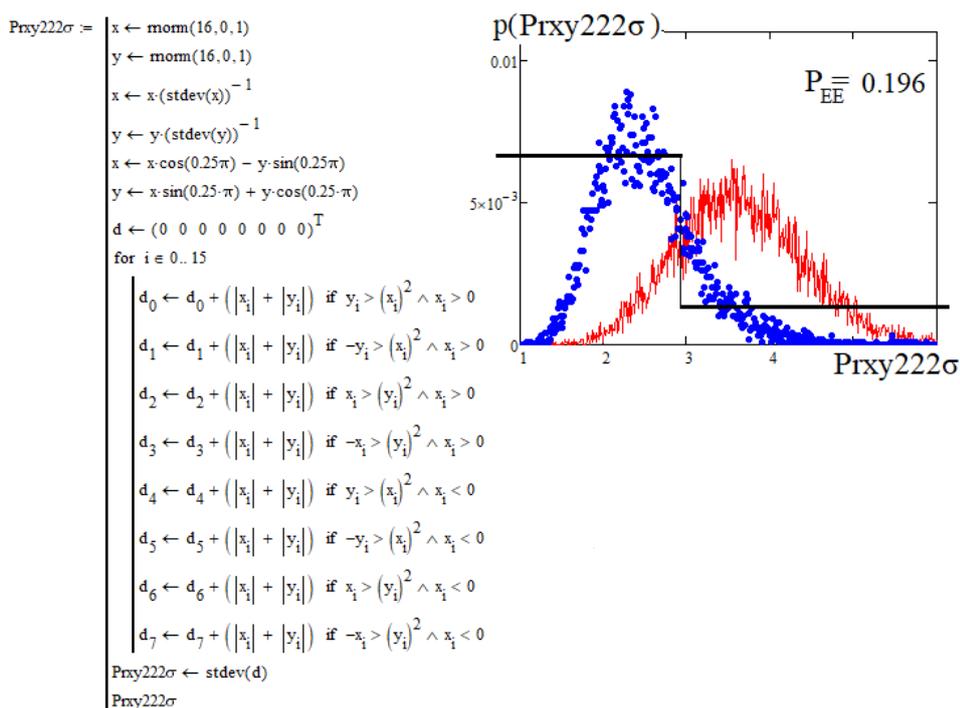


Рис. 7.12. Непрерывный нейрокритерий, построенный на использовании четырех квадратичных квантователей с дроблением каждого отклика на две составляющие

Проведение численного эксперимента по моделированию нового критерия показывает сохранение уровня вероятностей ошибок по отношению к предыдущему критерию, однако они могут обеспечивать низкий уровень корреляционной сцепленности близко родственных критериев:

$$\text{corr}(\text{Prx2}, \text{Prxy222}\sigma) = -0.109, \quad \text{corr}(\text{Prx21}, \text{Prxy222}\sigma) = -0.002, \quad \text{corr}(\text{Prx22}, \text{Prxy222}\sigma) = -0.018.$$

7.13. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей при накоплении модулей произведения координат обнаруженных точек и вычислении стандартных отклонений

Шесть предшествующих критериев были построены, опираясь на использование параболических квантователей двумерного пространства точек малой выборки. Накопление данных велось суммированием модулей координат выделенных групп точек. Возможно использование любого иного алгоритма непрерывного преобразования координат отсеянных точек перед накоплением, в частности может быть использовано перемножение модулей координат отсеянных точек.

Программная реализация такого критерия для двух квадратичных квантователей приведена на рис. 7.13.

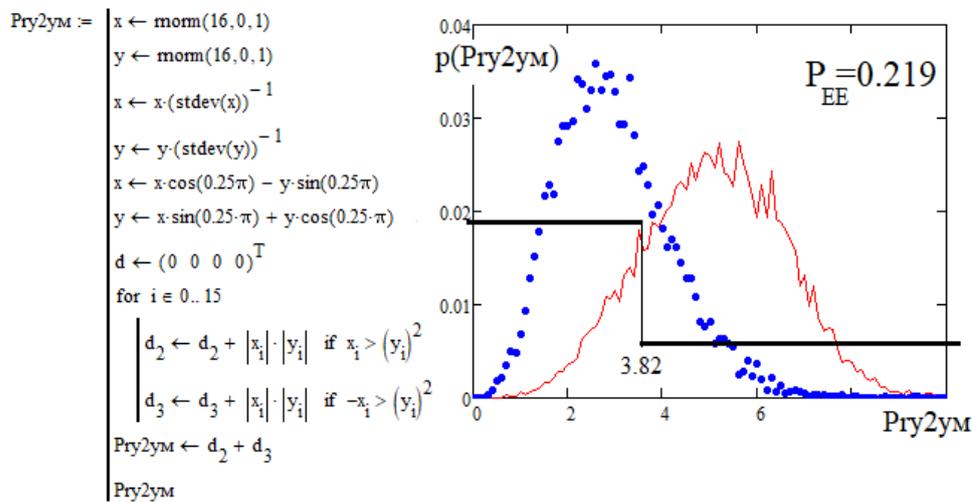


Рис. 7.13. Непрерывный нейрокритерий, построенный на использовании двух квадратичных квантователей с накоплением модулей произведений координат обнаруженных точек

Численный эксперимент по моделированию нового критерия показал, что его отклики слабо коррелированы с критериями, построенными на суммировании модулей координат обнаруженных точек с их входным квантованием по другой координате:

$$\begin{aligned} \text{corr}(\text{Pry2ум}, \text{Prx2}) &\approx -0.091, \text{corr}(\text{Pry2ум}, \text{Prx21}) \approx -0.023, \\ \text{corr}(\text{Pry2ум}, \text{Prx22}) &\approx -0.069. \end{aligned}$$

7.14. Корреляционная сцепленность группы критериев, построенных на использовании квадратичных квантователей

Так как при формировании группы нейрокритериев важен их средний уровень модулей коэффициентов корреляции, они приведены на рис. 7.14.

$\text{corr}(r, \text{Prx2}) = 0.009$		
$\text{corr}(r, \text{Pry2}) = 0.014$	$\text{corr}(\text{Prx2}, \text{Pry2}) = -0.017$	
$\text{corr}(r, \text{Prxy2}) = 0.017$	$\text{corr}(\text{Prx2}, \text{Prxy2}) = 0.443$	$\text{corr}(\text{Pry2}, \text{Prxy2}) = 0.888$
$\text{corr}(r, \text{Prxy2ам}) = 0.015$	$\text{corr}(\text{Prx2}, \text{Prxy2ам}) = 0.474$	$\text{corr}(\text{Pry2}, \text{Prxy2ам}) = 0.858$
$\text{corr}(r, \text{Prxy2амМ}) = 0.020$	$\text{corr}(\text{Prx2}, \text{Prxy2амМ}) = 0.331$	$\text{corr}(\text{Pry2}, \text{Prxy2амМ}) = 0.735$
$\text{corr}(r, \text{Prxy2амМм}) = 0.019$	$\text{corr}(\text{Prx2}, \text{Prxy2амМм}) = 0.204$	$\text{corr}(\text{Pry2}, \text{Prxy2амМм}) = 0.594$
$\text{corr}(r, \text{Prx21}) = 0.003$	$\text{corr}(\text{Prx2}, \text{Prx21}) = 0.318$	$\text{corr}(\text{Pry2}, \text{Prx21}) = -0.005$
$\text{corr}(r, \text{Prx22}) = 0.004$	$\text{corr}(\text{Prx2}, \text{Prx22}) = 0.756$	$\text{corr}(\text{Pry2}, \text{Prx22}) = -0.017$
$\text{corr}(r, \text{Pry22}) = 0.021$	$\text{corr}(\text{Prx2}, \text{Pry22}) = -0.071$	$\text{corr}(\text{Pry2}, \text{Pry22}) = 0.850$
$\text{corr}(r, \text{Prxy22}) = -0.017$	$\text{corr}(\text{Prx2}, \text{Prxy22}) = 0.380$	$\text{corr}(\text{Pry2}, \text{Prxy22}) = -0.790$
$\text{corr}(r, \text{Prxy22д}) = 0.004$	$\text{corr}(\text{Prx2}, \text{Prxy22д}) = -0.674$	$\text{corr}(\text{Pry2}, \text{Prxy22д}) = 0.403$
$\text{corr}(r, \text{Pry2ум}) = 0.011$	$\text{corr}(\text{Prx2}, \text{Pry2ум}) = -0.091$	$\text{corr}(\text{Pry2}, \text{Pry2ум}) = 0.547$

Рис. 7.14. Коэффициенты корреляции группы критериев, построенных на использовании квадратичных квантователей (начало)

Все модули коэффициентов корреляции первого столбца на рис. 7.14 малы. Это означает, что любой из критериев, рассматриваемой группы хорошо совместим с наиболее популярным классическим критерием Эджуорта – Эдлтона – Пирсона.

Первый критерий группы – Prx2 плохо совместим (имеет высокую корреляцию) с двумя критериями (второй столбец на рис. 7.14). Второй критерий группы – Pry2 плохо совместим уже с семью критериями рассматриваемой группы.

НОВЫЕ НЕЙРОКРИТЕРИИ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, ПОСТРОЕННЫЕ НА СТАТИСТИЧЕСКОМ АНАЛИЗЕ ТОЧЕК, ВЫДЕЛЕННЫХ ОКНАМИ ИЗ ОКРУЖНОСТЕЙ

8.1. Дискретный критерий, построенный на анализе числа точек малой выборки, попавших в одну окружность единичного радиуса с центром {1, 1}

Функция, на которой может быть построен квантователь данных малой выборки, может быть любой. При этом использование сложных функций будет приводить к сложным вариантам программной реализации того или иного статистического критерия. Достаточно простой оказывается программная реализация, построенная на использовании окружностей как функций квантования. Пример программной реализации, построенной на использовании одной окружности, приведен на рис. 8.1.

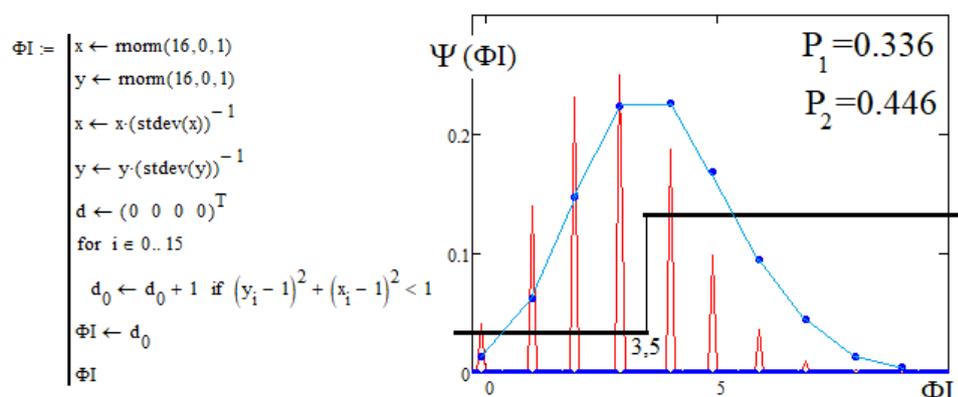


Рис. 8.1. Дискретный спектр амплитуд вероятности появления выходных состояний критерия, построенного на подсчете точек малой выборки, попавших в одну окружность

При использовании второго бинарного квантователя с порогом сравнения $k = 3,5$ мы получаем нейрокритерий с уровнем вероятностей ошибок первого и второго рода, сопоставимым с наиболее слабыми классическими статистическими критериями прошлого века.

Очевидно, что окружности квадратичных квантователей с единичным радиусом могут быть размещены в любой четверти Декартовой системы координат. В этой ситуации мы получаем четыре варианта нейрокритериев. Дискретные спектры у всех нейрокритериев одинаковы и будут повторять распределения рис. 8.1. При этом коэффициенты корреляции между откликами разных вариантов этого критерия будут близки друг другу как по модулю, так и по знаку. Пример оценок коэффициентов корреляции, полученных для 999 малых выборок, приведен на рис. 8.2.

$$\begin{bmatrix} \text{corr}(\Phi I, \Phi II) = -0.243 & \text{corr}(\Phi I, \Phi III) = -0.214 & \text{corr}(\Phi I, \Phi IV) = -0.225 \\ & \text{corr}(\Phi II, \Phi III) = -0.237 & \text{corr}(\Phi II, \Phi IV) = -0.201 \\ & & \text{corr}(\Phi III, \Phi IV) = -0.231 \end{bmatrix}$$

Рис. 8.2. Коэффициенты корреляции откликов нейрокритериев, построенных на использовании одной окружности

Распределение каждого из шести коэффициентов корреляции на рис. 8.2 близко к нормальному закону.

8.2. Дискретный критерий, построенный на анализе числа точек малой выборки, попавших в две окружности единичного радиуса

Высокий уровень вероятностей ошибок первого и второго рода предыдущего критерия может быть снижен, если воспользоваться двумя квантователями, образованными двумя окружностями. Это позволяет снизить уровень ошибок первого и второго рода примерно на 10 %. Программная реализация нового критерия приведена на рис. 8.3.

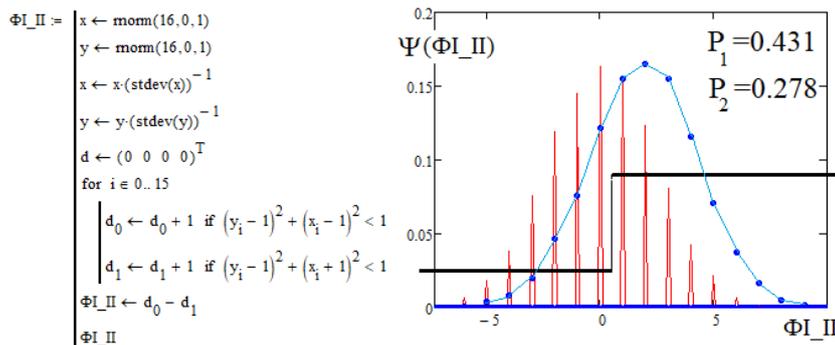


Рис. 8.3. Дискретный спектр амплитуд вероятности появления выходных состояний критерия, построенного на подсчете точек малой выборки, попавших в две окружности

Если над выходными данными нового критерия выполнить бинарное квантование, сравнивая их с порогом $k = 0.5$, то мы получим еще один вариант нейрокритерия для проверки гипотезы независимости.

Если рассматривать единичные окружности первого сектора и третьего сектора Декартовой системы координат, то получается еще один вариант нейрокритерия для проверки гипотезы независимости. Программная реализация нового критерия представлена на рис. 8.4.

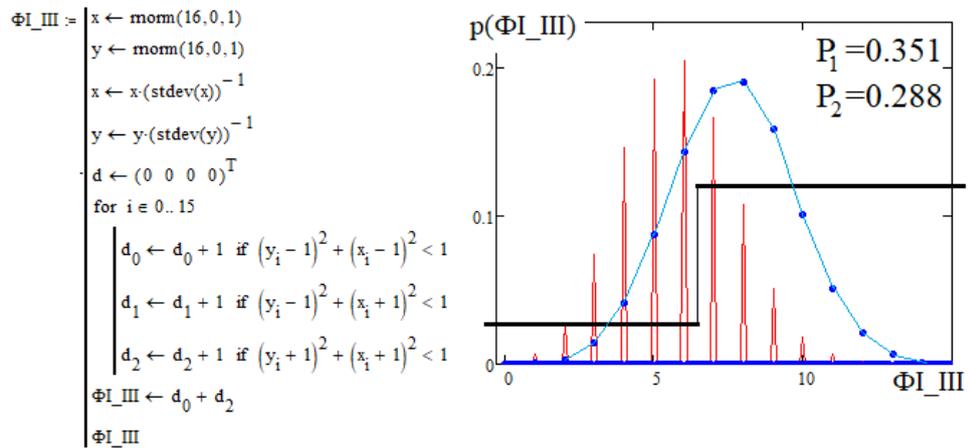


Рис. 8.4. Дискретный спектр амплитуд вероятности появления выходных состояний критерия, построенного на подсчете точек малой выборки, попавших в окружность первой и третьей четвертей

Новый критерий построен на суммировании данных, попавших в две окружности. Это существенное отличие от предыдущего критерия, построенного на вычитании числа точек, попавших в разные окружности.

Критерии, построенные на учете статистик попадания точек малой выборки в две окружности, расположенные в разных секторах Декартовой системы координат, имеют разные коэффициенты парной корреляции. Примеры значений коэффициентов корреляции приведены на рис. 8.5.

$\text{corr}(\Phi_{I_II}, \Phi_{I_III}) = 0.641$	$\text{corr}(\Phi_{I_II}, \Phi_{I_IV}) = 0.512$	$\text{corr}(\Phi_{I_II}, \Phi_{II_III}) = 0.49$	$\text{corr}(\Phi_{I_II}, \Phi_{II_IV}) = -0.631$
	$\text{corr}(\Phi_{I_III}, \Phi_{I_IV}) = 0.635$	$\text{corr}(\Phi_{I_III}, \Phi_{II_IV}) = 0.635$	$\text{corr}(\Phi_{I_III}, \Phi_{II_III}) = -0.595$
		$\text{corr}(\Phi_{I_IV}, \Phi_{II_III}) = -0.499$	$\text{corr}(\Phi_{I_IV}, \Phi_{II_IV}) = -0.642$
			$\text{corr}(\Phi_{II_III}, \Phi_{II_IV}) = 0.015$

Рис. 8.5. Коэффициенты корреляции откликов нейрокритериев, построенных на использовании двух окружностей

8.3. Дискретный критерий, построенный на анализе числа точек малой выборки, попавших в четыре окружности единичного радиуса

Повысить уровень доверия к принимаемым предшествующим критерием решениям на 21 % удастся, если перейти к использованию сразу четырех окружностей единичного радиуса. При этом окружности расположены во всех четвертях системы координат. Пример программной реализации нового статистического критерия приведен на рис. 8.6.

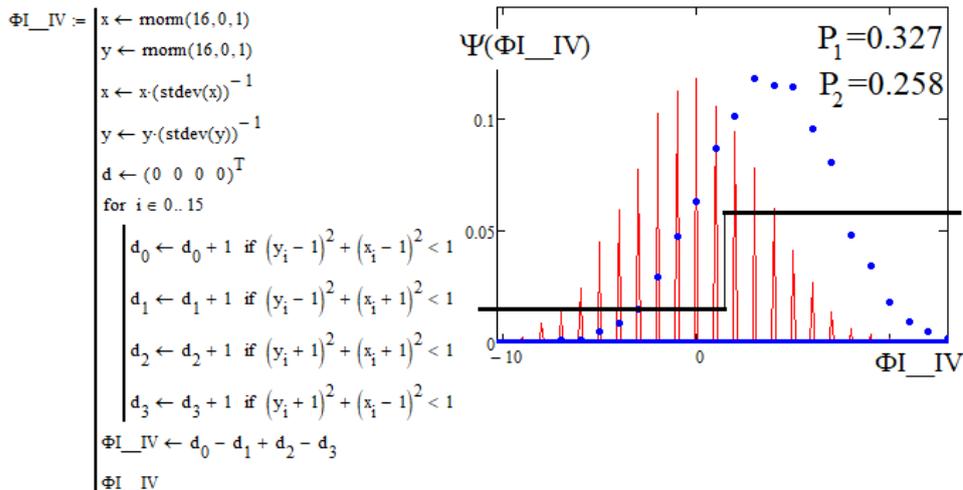


Рис. 8.6. Дискретный спектр амплитуд вероятности появления выходных состояний критерия, построенного на подсчете точек малой выборки, попавших в четыре окружности

При использовании бинарного квантователя с порогом $k = 1.5$ мы получим еще один вариант эффективного нейрокритерия для проверки гипотезы независимости.

К сожалению, рассматриваемый критерий сильно коррелирован с предыдущими критериями: $\text{corr}(\Phi I_IV, \Phi I_III) \approx 0.890$, $\text{corr}(\Phi I_IV, \Phi I_II) \approx 0.718$, $\text{corr}(\Phi I_IV, \Phi II) \approx 0.565$.

8.4. Непрерывный критерий, построенный на накоплении радиусов точек по отношению к центру $\{0, 0\}$ для четырех окружностей, расположенных в разных четвертях системы координат

Существенно повысить качество работы критерия (достигается улучшение на 11 %) получается, если перейти от подсчета числа точек

к накоплению радиусов от центра системы координат $\{0, 0\}$. В каждой четверти системы координат размещается своя окружность с единичным радиусом. Далее накопленные радиусы суммируются с положительным и отрицательным знаками.

Программная реализация нового критерия приведена на рис. 8.7. На этом же рисунке даны распределения откликов нового критерия при воздействии независимыми данными и положительно коррелированными данными. Если знак корреляционной связанности данных обратный, то при их свертывании следует сменить знаки суммируемых членов.

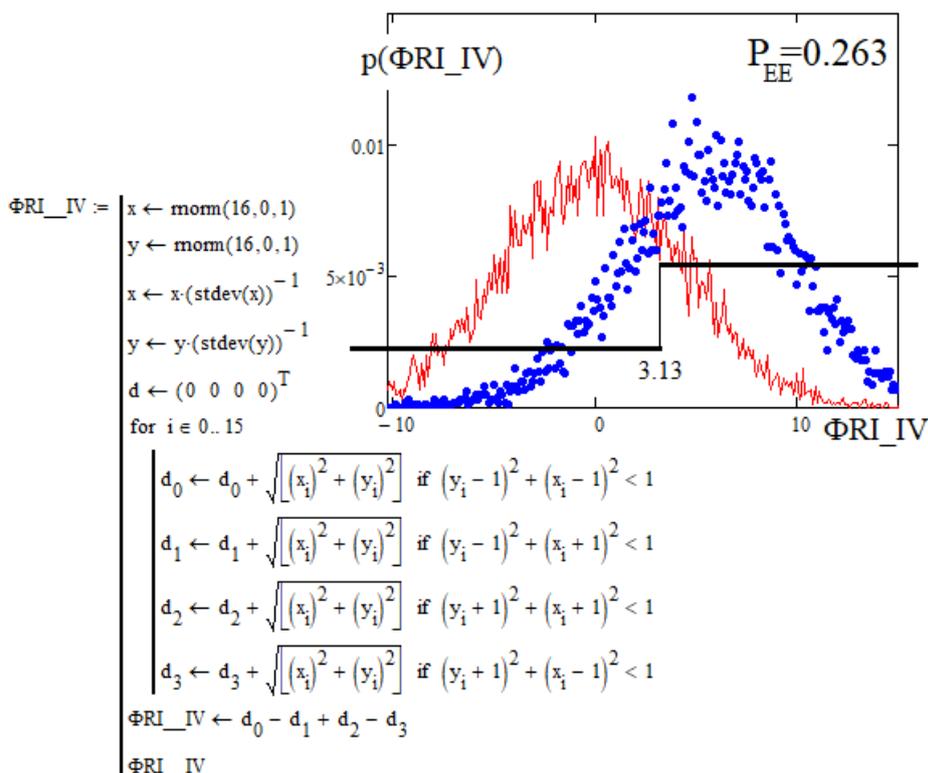


Рис. 8.7. Непрерывный спектр амплитуд вероятности появления выходных состояний критерия, построенного на подсчете точек малой выборки, попавших в четыре окружности, с вычислением среднего радиуса

К сожалению, корреляционная сцепленность двух последних критериев высока:

$$\text{corr}(\Phi\text{RI_IV}, \Phi\text{I_IV}) \approx 0.940.$$

Однако корреляционная сцепленность с другими похожими критериями низка:

$$\text{corr}(\Phi\text{RSдI_IV}, \Phi\text{RI_IV}) \approx 0.122.$$

8.5. Непрерывный критерий, построенный на исследовании точек малой выборки с накоплением среднего радиуса в каждой окружности и делением накопленных данных

Принципиально важным является не только уровень разделимости зависимых и независимых данных. При объединении нейрокритериев в одну нейросеть важным оказывается низкий уровень корреляционной сцепленности разных нейрокритериев. Для того, чтобы существенно снизить уровень корреляционной связанности, аддитивную свертку данных предшествующего критерия заменяют на мультипликативную свертку делением накопленных радиусов.

Программная реализация нового критерия приведена на рис. 8.8. В этом случае корреляционная сцепленность двух последних критериев составляет $\text{corr}(\Phi R_{дI_IV}, \Phi R_{I_IV}) \approx 0.127$, что вполне приемлемо для практики.

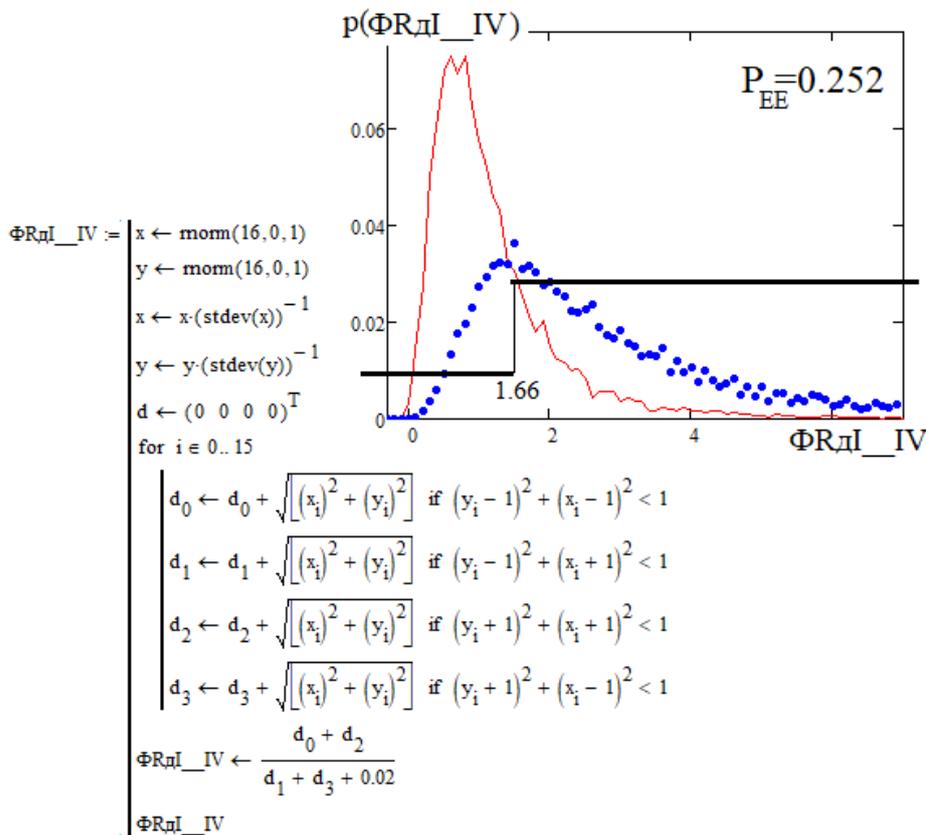


Рис. 8.8. Непрерывный спектр амплитуд вероятности появления выходных состояний критерия, построенного на подсчете точек малой выборки, попавших в четыре окружности, с вычислением среднего радиуса и делением

8.6. Непрерывный критерий, построенный на просмотре точек малой выборки, попавших в четыре окружности единичного радиуса, с накоплением значений модулей произведений координат точек

Еще одним направлением снижения вероятности ошибок первого и второго рода новых критериев является замена процедур накопления радиусов (см. рис. 8.8) на процедуру накопления модулей произведений координат точек, попадающих в квадратичные замкнутые кривые в виде окружностей. Подобная замена увеличивает на 4.1 % уровень разделимости малых выборок зависимых и независимых данных. Программная реализация нового нейрокритерия приведена на рис. 8.9. На этом же рисунке приведены результаты численного тестирования.

Накопленные промежуточные данные четырех квадратичных квантователей аддитивно свертываются и затем сравниваются с порогом $k = 1.21$. Выходное состояние этой математической конструкции «0» соответствует обнаружению ей независимых данных.

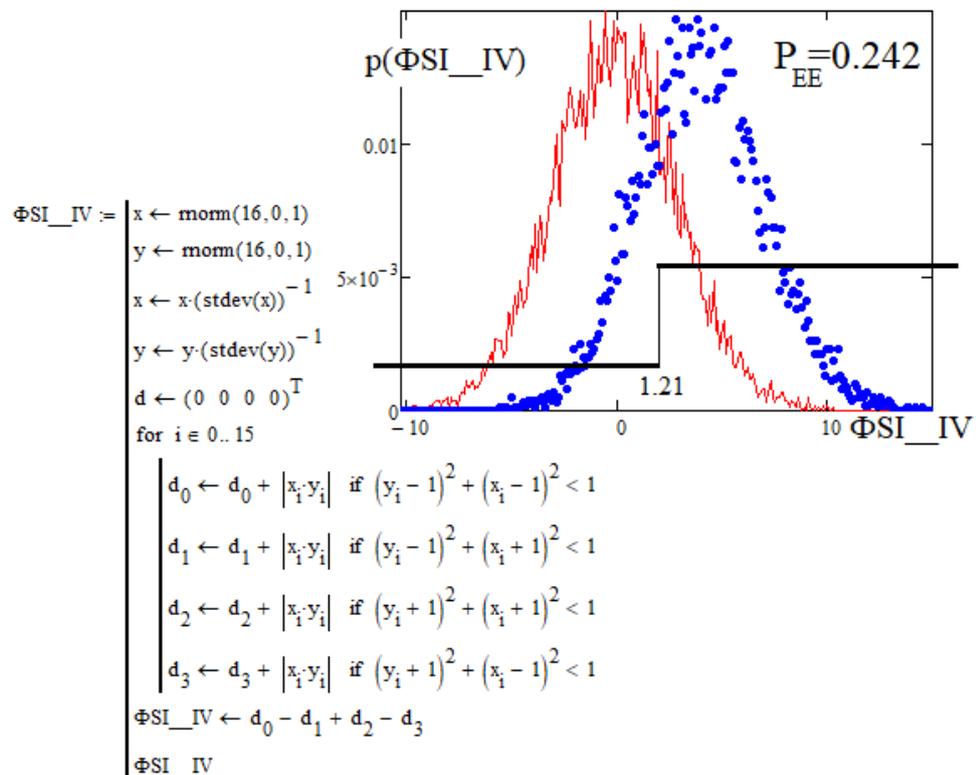


Рис. 8.9. Непрерывный спектр амплитуд вероятности появления выходных состояний критерия, построенного на подсчете точек малой выборки, попавших в четыре окружности, с вычислением модулей произведений координат

8.7. Непрерывный критерий, построенный на просмотре точек малой выборки, попавших в четыре окружности единичного радиуса, с вычислением среднего значения модуля произведений координат точек и деления данных

Аддитивная свертка четырех промежуточных компонент, использованная при синтезе предыдущего нейрокритерия, может быть заменена на мультипликативную свертку данных. Программная реализация этой модификации критерия и результаты численного моделирования статистик его выходных состояний приведены на рис. 8.10.

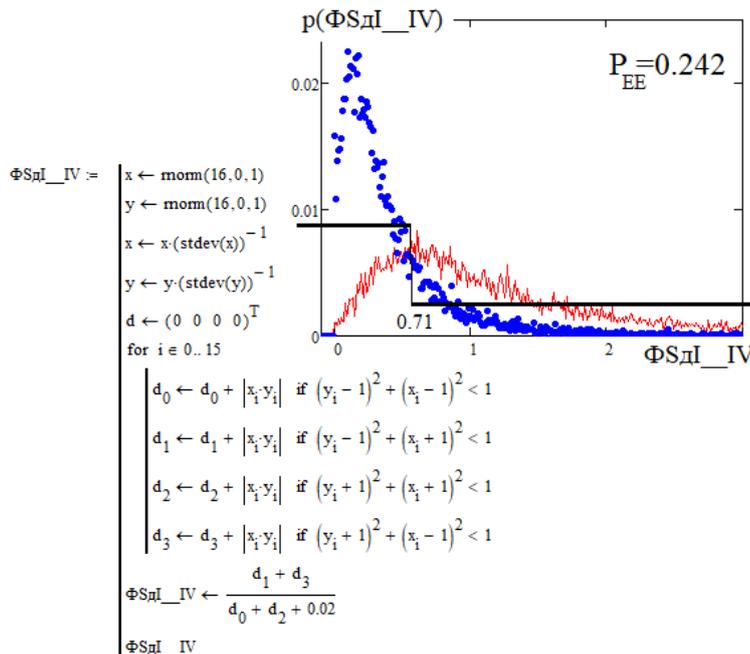


Рис. 8.10. Непрерывный спектр амплитуд вероятности появления выходных состояний критерия, построенного на подсчете точек малой выборки, попавших в четыре окружности, с вычислением модулей произведений координат и их делением

Из данного рисунка видно, что бинарное квантование выходных состояний новой модификации с порогом сравнения $k = 0.71$ приводит к появлению бинарного нейрокритерия, способного эффективно разделять малые выборки зависимых и независимых данных. При этом вероятности ошибок первого и второго рода для двух последних нейрокритериев совпадают и составляют $P_{EE} \approx 0.242$. Важным преимуществом двух последних критериев является их низкий уровень корреляционной сцепленности $\text{corr}(\Phi SI_IV, \Phi SдI_IV) \approx 0.22$. К сожалению, использовать предпоследний нейрокритерий в паре с классическим нейрокритерием Эджуорта – Эдлтона – Пирсона нецелесообразно из-за высокого уровня их взаимной корреляционной сцепленности: $\text{corr}(r, \Phi SI_IV) \approx 0.68$.

8.8. Непрерывный критерий, построенный на просмотре точек малой выборки, попавших в четыре окружности единичного радиуса, с вычислением среднего значения модуля произведений координат обнаруженных точек и деления данных с комбинированием двух последних критериев

Рассмотренные выше преобразования данных позволяют утверждать, что замена процедуры накапливания радиусов $\sqrt{(x_i)^2 + (y_i)^2}$ на процедуру накапливания произведений модулей координат $|x_i| \cdot |y_i|$ обнаруженных точек дает сильно зависимые состояния функционалов: $\text{corr}(\Phi R_{\text{дI_IV}}, \Phi S_{\text{дI_IV}}) \approx 0.961$. Видимо, это происходит из-за смысловых сортировок применения процедур накапливания данных по разным секторам системы координат. При этом способы предобработки данных существенно различаются $\sqrt{(x_i)^2 + (y_i)^2} \neq |x_i| \cdot |y_i|$ по низкому уровню. Используя это, мы можем совместить эти два способа предобработки данных в рамках одного гибридного функционала.

Пример программной реализации такого гибридного функционала приведен на рис. 8.11. Численное моделирование нового нейрокритерия дает приемлемый для практики уровень вероятности появления ошибок первого и второго рода.

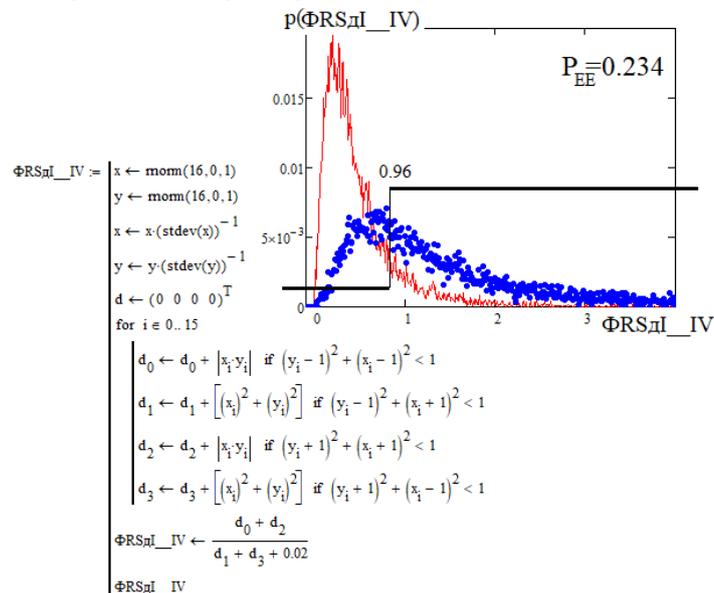


Рис. 8.11. Непрерывный спектр амплитуд вероятности появления выходных состояний критерия, построенного на четырех окружностях, с вычислением модулей произведений координат и накоплением квадратов радиусов

К сожалению, новый нейрокритерий имеет высокий уровень корреляционной сцепленности с функционалами-родителями:

$$\text{corr}(\Phi R_{дI_IV}, \Phi RS_{дI_IV}) \approx 0.968,$$

$$\text{corr}(\Phi S_{дI_IV}, \Phi RS_{дI_IV}) \approx 0.998.$$

8.9. Непрерывный критерий, построенный на просмотре точек малой выборки, попавших в четыре окружности, с вычислением среднего значения модуля произведений координат точек для окружностей с центрами $\{1, -1\}, \{-1, 1\}$ и накоплением квадратов радиусов в двух других секторах

Отмеченная выше положительная ситуация сохраняется, если поменять местами процедуры предобработки данных низкого уровня. Сохраняется правило свертывания промежуточных вычислений, однако процедуры накопления данных меняются между окружностями числителя и знаменателя конечной мультипликативной свертки.

Программная реализация нового варианта критерия представлена на рис. 8.12.

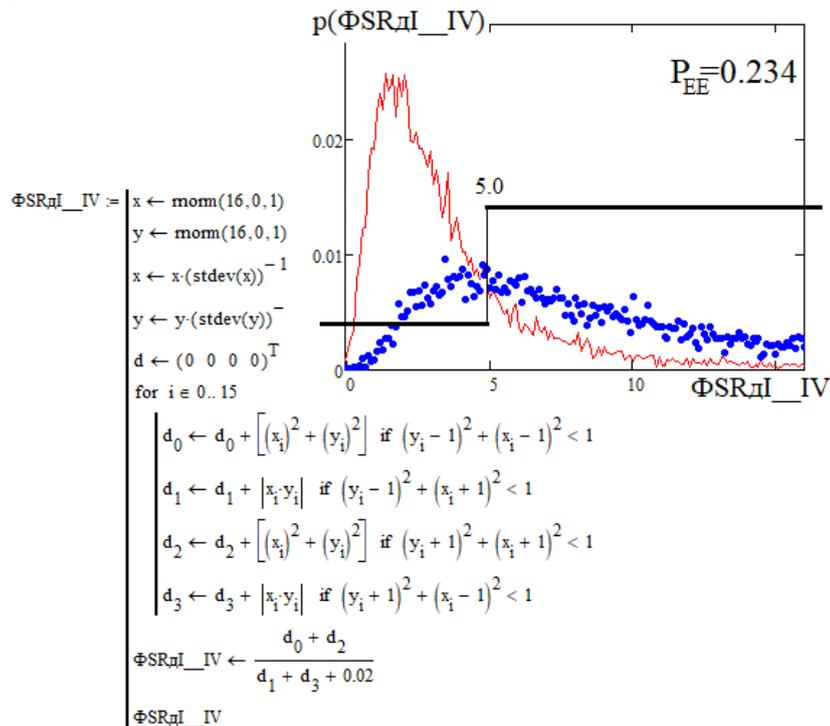


Рис. 8.12. Непрерывный спектр амплитуд вероятности появления выходных состояний критерия, построенного на четыре окружностях, с вычислением модулей произведений координат и накоплением квадратов радиусов

Проведенный численный эксперимент показал полное совпадение вероятностей появления ошибок первого и второго рода двух последних нейрокритериев. К сожалению, уровень корреляционной сцепленности их откликов оказывается высоким:

$$\text{corr}(\Phi\text{SRдI_IV}, \Phi\text{RSдI_IV}) \approx 0.968.$$

8.10. Непрерывный критерий, построенный на оценках статистик точек малой выборки, попавших в четыре окружности, при свертывании промежуточных данных через умножение и деление

Снизить уровень корреляционной сцепленности с функционалами-родителями удастся, если изменить вид выходного свертывания промежуточных данных. Например, суммирование промежуточных данных может быть заменено на их умножение. При этом мы теряем уровень делимости зависимых и независимых данных, но существенно снижаем корреляционную связь с откликами функционалов-родителей:

$$\text{corr}(\Phi\text{РдI_IV}, \Phi\text{RSудI_IV}) \approx 0.205,$$

$$\text{corr}(\Phi\text{СдI_IV}, \Phi\text{RSудI_IV}) \approx 0.306.$$

Пример программной реализации этого варианта нейрокритерия приведен на рис. 8.13.

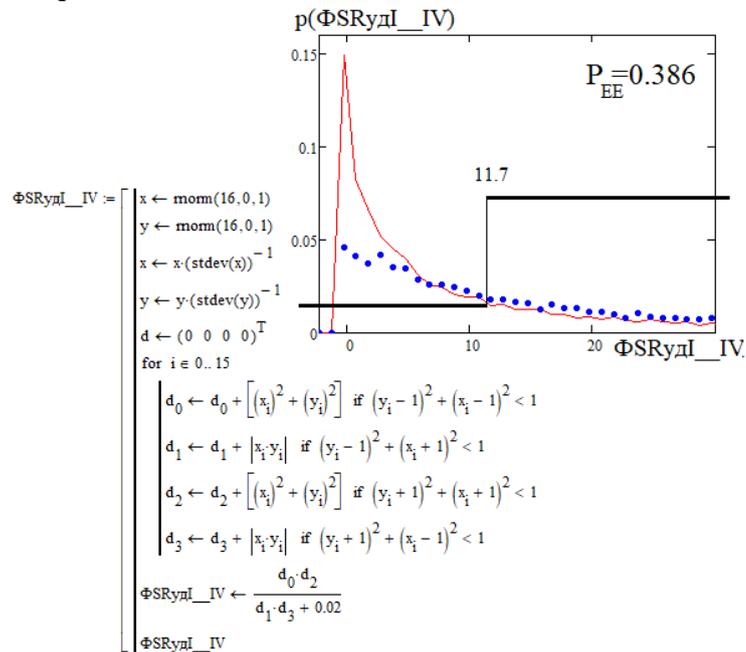


Рис. 8.13. Непрерывный спектр амплитуд вероятности появления выходных состояний критерия, построенного на четырех окружностях, со сверткой через умножение и деление

8.11. Корреляционная сцепленность группы нейрокритериев, построенных на использовании квантователей в виде окружностей, размещенных в разных четвертях Декартовой системы координат

Достижимое качество работы нейросетевого объединения существенно зависит от числа нейрокритериев, собранных в одну нейросеть, и от уровня их взаимной корреляционной сцепленности. При высоком уровне модулей коэффициентов парной корреляции бесполезно наращивать число объединяемых нейрокритериев. Целесообразно объединять нейрокритерии с низким уровнем модулей коэффициентов парной корреляции. Значения коэффициентов парной корреляции приведены на рис. 8.14.

$\text{corr}(r, \Phi I) = 0.288$		
$\text{corr}(r, \Phi II) = -0.292$	$\text{corr}(\Phi I, \Phi II) = -0.229$	
$\text{corr}(r, \Phi I_{II}) = 0.37$	$\text{corr}(\Phi I, \Phi I_{II}) = 0.78$	$\text{corr}(\Phi II, \Phi I_{II}) = -0.788$
$\text{corr}(r, \Phi I_{III}) = 0.446$	$\text{corr}(\Phi I, \Phi I_{III}) = 0.63$	$\text{corr}(\Phi II, \Phi I_{III}) = -0.366$
$\text{corr}(r, \Phi I_{IV}) = 0.504$	$\text{corr}(\Phi I, \Phi I_{IV}) = 0.561$	$\text{corr}(\Phi II, \Phi I_{IV}) = -0.565$
$\text{corr}(r, \Phi RI_{IV}) = 0.613$	$\text{corr}(\Phi I, \Phi RI_{IV}) = 0.53$	$\text{corr}(\Phi II, \Phi RI_{IV}) = -0.534$
$\text{corr}(r, \Phi RII_{IV}) = 0.064$	$\text{corr}(\Phi I, \Phi RII_{IV}) = 0.076$	$\text{corr}(\Phi II, \Phi RII_{IV}) = -0.076$
$\text{corr}(r, \Phi SII_{IV}) = 0.68$	$\text{corr}(\Phi I, \Phi SII_{IV}) = 0.425$	$\text{corr}(\Phi II, \Phi SII_{IV}) = -0.425$
$\text{corr}(r, \Phi SIII_{IV}) = 0.143$	$\text{corr}(\Phi I, \Phi SIII_{IV}) = 0.122$	$\text{corr}(\Phi II, \Phi SIII_{IV}) = -0.127$
$\text{corr}(r, \Phi RSIII_{IV}) = 0.067$	$\text{corr}(\Phi I, \Phi RSIII_{IV}) = 0.072$	$\text{corr}(\Phi II, \Phi RSIII_{IV}) = -0.072$
$\text{corr}(r, \Phi SRIII_{IV}) = 0.137$	$\text{corr}(\Phi I, \Phi SRIII_{IV}) = 0.121$	$\text{corr}(\Phi II, \Phi SRIII_{IV}) = -0.127$
$\text{corr}(r, \Phi SRyIII_{IV}) = 0.274$	$\text{corr}(\Phi I, \Phi SRyIII_{IV}) = 0.211$	$\text{corr}(\Phi II, \Phi SRyIII_{IV}) = -0.228$

Рис. 8.14. Коэффициенты корреляции группы нейрокритериев,
построенных с использованием окружностей
наблюдения (начало)

Показатели первого столбца на рис. 8.14 свидетельствуют о низком уровне корреляционной сцепленности новой группы критериев и самого мощного классического критерия Эджуорта – Эдлттона – Пирсона. Примерно такая же ситуация возникает и по отношению к первым двум критериям рассматриваемой группы.

$\text{corr}(\Phi I_{II}, \Phi I_{III}) = 0.634$		
$\text{corr}(\Phi I_{II}, \Phi I_{IV}) = 0.718$	$\text{corr}(\Phi I_{III}, \Phi I_{IV}) = 0.89$	
$\text{corr}(\Phi I_{II}, \Phi RI_{IV}) = 0.678$	$\text{corr}(\Phi I_{III}, \Phi RI_{IV}) = 0.834$	$\text{corr}(\Phi I_{IV}, \Phi RI_{IV}) = 0.94$
$\text{corr}(\Phi I_{II}, \Phi R_{дI_{IV}}) = 0.097$	$\text{corr}(\Phi I_{III}, \Phi R_{дI_{IV}}) = 0.109$	$\text{corr}(\Phi I_{IV}, \Phi R_{дI_{IV}}) = 0.129$
$\text{corr}(\Phi I_{II}, \Phi SI_{IV}) = 0.542$	$\text{corr}(\Phi I_{III}, \Phi SI_{IV}) = 0.658$	$\text{corr}(\Phi I_{IV}, \Phi SI_{IV}) = 0.746$
$\text{corr}(\Phi I_{II}, \Phi S_{дI_{IV}}) = 0.159$	$\text{corr}(\Phi I_{III}, \Phi S_{дI_{IV}}) = 0.177$	$\text{corr}(\Phi I_{IV}, \Phi S_{дI_{IV}}) = 0.212$
$\text{corr}(\Phi I_{II}, \Phi RS_{дI_{IV}}) = 0.092$	$\text{corr}(\Phi I_{III}, \Phi RS_{дI_{IV}}) = 0.101$	$\text{corr}(\Phi I_{IV}, \Phi RS_{дI_{IV}}) = 0.122$
$\text{corr}(\Phi I_{II}, \Phi SR_{дI_{IV}}) = 0.159$	$\text{corr}(\Phi I_{III}, \Phi SR_{дI_{IV}}) = 0.175$	$\text{corr}(\Phi I_{IV}, \Phi SR_{дI_{IV}}) = 0.212$
$\text{corr}(\Phi I_{II}, \Phi SR_{удI_{IV}}) = 0.28$	$\text{corr}(\Phi I_{III}, \Phi SR_{удI_{IV}}) = 0.332$	$\text{corr}(\Phi I_{IV}, \Phi SR_{удI_{IV}}) = 0.384$

Рис. 8.14. Продолжение

Внутри рассматриваемой группы критериев редко встречаются коэффициенты корреляции с высоким уровнем модулей коэффициентов корреляции. Среднее значение модулей коэффициентов корреляции составляет 0.391.

$\text{corr}(\Phi RI_{IV}, \Phi R_{дI_{IV}}) = 0.127$		
$\text{corr}(\Phi RI_{IV}, \Phi SI_{IV}) = 0.903$	$\text{corr}(\Phi R_{дI_{IV}}, \Phi SI_{IV}) = 0.104$	
$\text{corr}(\Phi RI_{IV}, \Phi S_{дI_{IV}}) = 0.229$	$\text{corr}(\Phi R_{дI_{IV}}, \Phi S_{дI_{IV}}) = 0.961$	$\text{corr}(\Phi SI_{IV}, \Phi S_{дI_{IV}}) = 0.22$
$\text{corr}(\Phi RI_{IV}, \Phi RS_{дI_{IV}}) = 0.124$	$\text{corr}(\Phi R_{дI_{IV}}, \Phi RS_{дI_{IV}}) = 0.997$	$\text{corr}(\Phi SI_{IV}, \Phi RS_{дI_{IV}}) = 0.108$
$\text{corr}(\Phi RI_{IV}, \Phi SR_{дI_{IV}}) = 0.226$	$\text{corr}(\Phi R_{дI_{IV}}, \Phi SR_{дI_{IV}}) = 0.963$	$\text{corr}(\Phi SI_{IV}, \Phi SR_{дI_{IV}}) = 0.21$
$\text{corr}(\Phi RI_{IV}, \Phi SR_{удI_{IV}}) = 0.399$	$\text{corr}(\Phi R_{дI_{IV}}, \Phi SR_{удI_{IV}}) = 0.205$	$\text{corr}(\Phi SI_{IV}, \Phi SR_{удI_{IV}}) = 0.359$
$\text{corr}(\Phi S_{дI_{IV}}, \Phi RS_{дI_{IV}}) = 0.968$		
$\text{corr}(\Phi S_{дI_{IV}}, \Phi SR_{дI_{IV}}) = 0.998$	$\text{corr}(\Phi RS_{дI_{IV}}, \Phi SR_{дI_{IV}}) = 0.968$	
$\text{corr}(\Phi S_{дI_{IV}}, \Phi SR_{удI_{IV}}) = 0.306$	$\text{corr}(\Phi RS_{дI_{IV}}, \Phi SR_{удI_{IV}}) = 0.208$	$\text{corr}(\Phi SR_{дI_{IV}}, \Phi SR_{удI_{IV}}) = 0.304$

Рис. 8.14. Окончание

В целом вариация модулей коэффициентов корреляции значительна: от 0.104 до 0.998. Это делает задачу эмпирического подбора сочетаний нейрокритериев актуальной.

КРИТЕРИИ, ПОСТРОЕННЫЕ НА ИСПОЛЬЗОВАНИИ ГИПЕРБОЛИЧЕСКИХ ФУНКЦИЙ РАЗДЕЛЕНИЯ ДАННЫХ МАЛОЙ ВЫБОРКИ

9.1. Дискретный критерий, построенный на гиперболическом квантователе данных, попавших в первую четверть Декартовой системы координат

Критерий построен на том, что при объеме малой выборки в 16 опытов независимые данные обеспечивают попадания в первую четверть меньшего числа точек в сравнении с положительно коррелированными данными. При этом разделяющая точки гипербола не захватывает точки с малыми значениями координат, лежащими ниже гиперболы.

Пример программной реализации критерия приведен на рис. 9.1. На этом же рисунке приведены результаты имитационного моделирования.

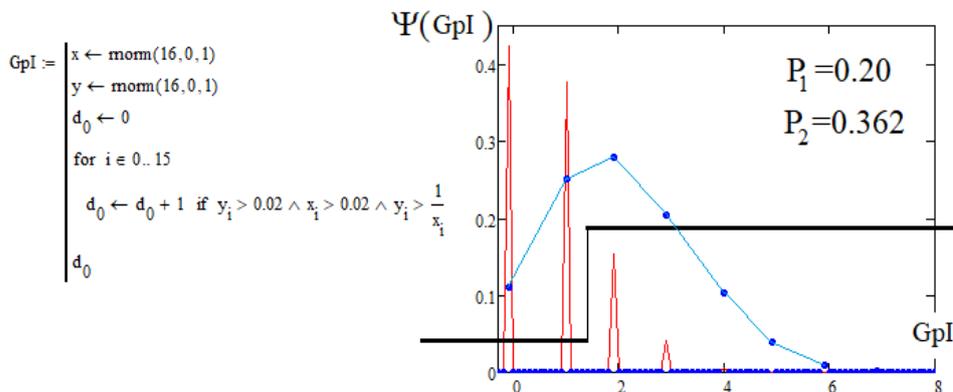


Рис. 9.1. Дискретный критерий, построенный на гиперболическом квантовании данных, попадающих в первую четверть системы координат

Бинарный нейрокритерий для разделения независимых данных и положительно коррелированных данных $r = 0.5$ получается, если установить порог сравнения выходного квантователя $k = 1.5$. В этом случае нейрокритерий дает состояние «0» при подтверждении гипотезы независимости данных тестируемой выборки.

9.2. Дискретный критерий, построенный на гиперболическом квантователе данных, попавших во вторую четверть Декартовой системы координат

По аналогии с предыдущим дискретным критерием может быть построен еще один вариант критерия, ориентированный на анализ точек малой выборки, попадающих во вторую четверть Декартовой системы координат. Пример программной реализации такого критерия приведен на рис. 9.2.

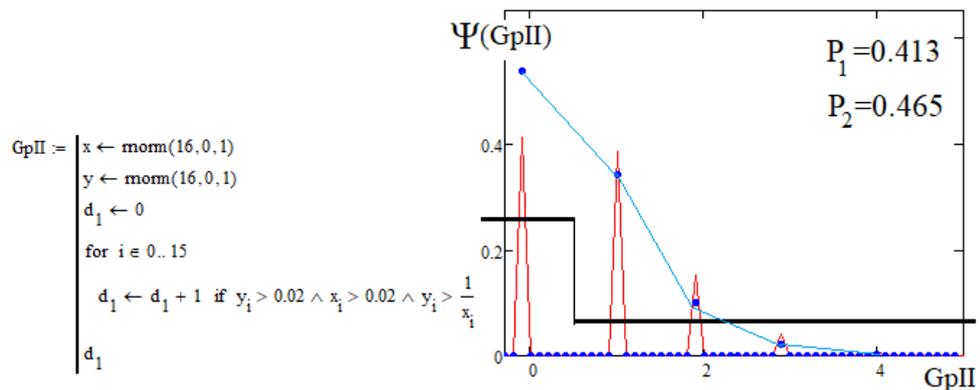


Рис. 9.2. Дискретный критерий, построенный на гиперболическом квантовании данных, попадающих во вторую четверть Декартовой системы координат

Из данного рисунка видно, что разделимость независимых и положительно коррелированных данных существенно ухудшилась. Если коррелированность данных сделать отрицательной, то разделимость растет и становится такой же, как у предыдущего критерия. Два последних критерия чувствительны к знаку корреляции зависимых данных. В связи с этим до их использования необходимо оценить ожидаемый знак коэффициента корреляции и выбрать под него наиболее мощные сочетания нейрокритериев.

9.3. Дискретный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих в первую и во вторую четверти Декартовой системы координат

Аддитивным объединением удастся связать в один критерий два предыдущих критерия, что увеличивает его показатели. Программная реализация еще одного варианта критерия приведена на рис. 9.3.

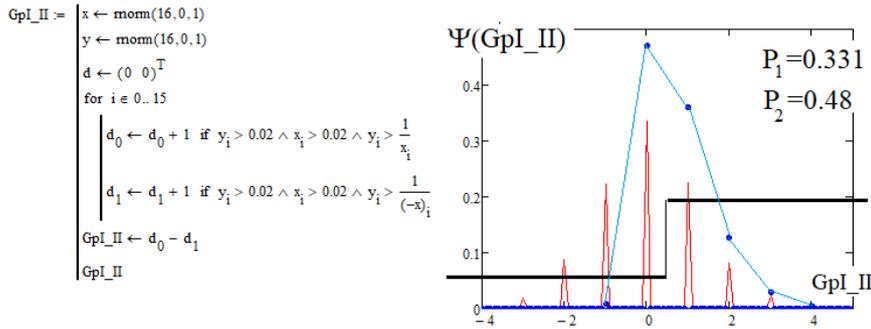


Рис. 9.3. Дискретный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих в первую и во вторую четверти Декартовой системы координат

Из данного рисунка видно, что у нового критерия число спектральных линий независимых данных увеличивается примерно в два раза. Число спектральных линий зависимых данных остается на том же уровне. Уровень разделимости независимых и зависимых данных оказывается между показателями лучшего и худшего из объединяемых критериев.

9.4. Дискретный критерий, построенный на четырех гиперболических квантователях данных малой выборки, размещенных во всех четвертях Декартовой системы координат, с аддитивным свертыванием

Если мы размещаем в каждой четверти Декартовой системы координат по одному гиперболическому квантователю, то получим дискретный критерий, имеющий 16 значимых амплитуд вероятности как для независимых, так и для зависимых данных. Программная реализация нового критерия и результаты проведенных численных экспериментов по его моделированию приведены на рис. 9.4.

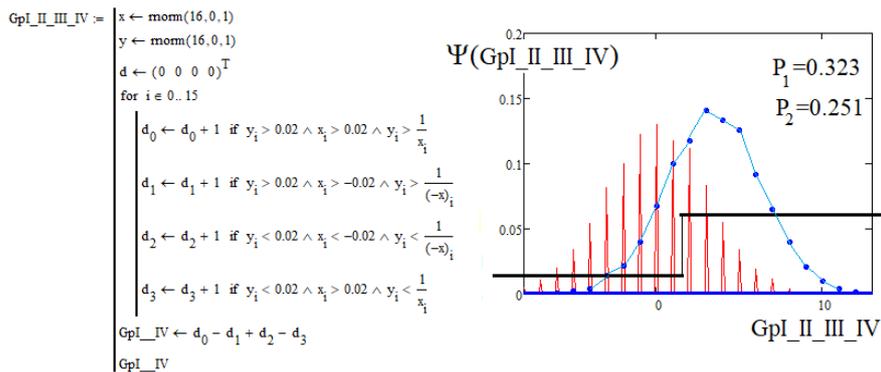


Рис. 9.4. Дискретный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат

При пороге выходного квантователя $k = 1.5$ нейрокритерий достаточно хорошо различает независимые и положительно зависимые данные. Обеспечиваемый им уровень вероятностей ошибок первого и второго рода близок к среднему значению классических статистических критериев, созданных в прошлом веке.

9.5. Дискретный критерий, построенный на четырех гиперболических квантователях данных малой выборки с мультипликативным свертыванием промежуточных данных

Кроме аддитивного свертывания четырех промежуточных параметров, может быть использовано их мультипликативное свертывание, например, в форме деления двух накопленных чисел. Программная реализация этого критерия приведена на рис. 9.5 и дает рост делимости независимых и положительно коррелированных данных примерно на 8 %.

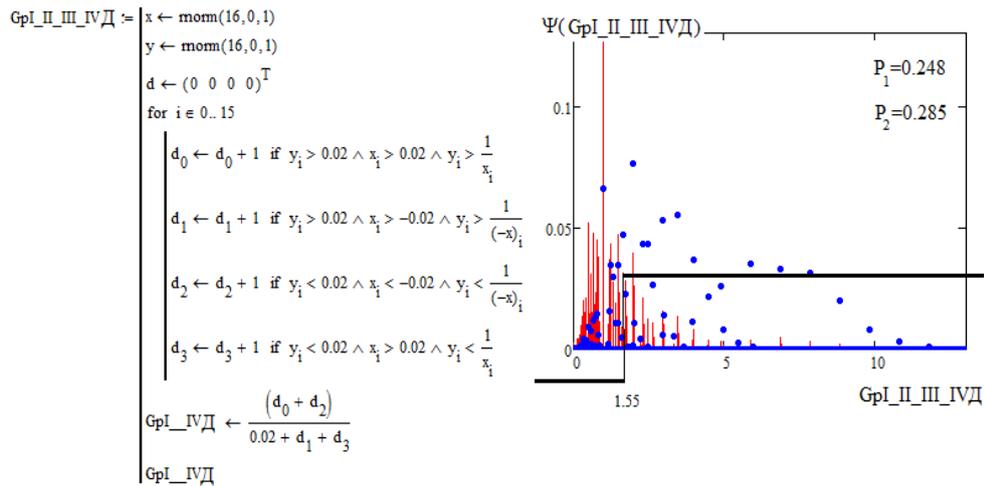


Рис. 9.5. Дискретный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат с мультипликативным свертыванием промежуточных вычислений

Главным достоинством двух последних критериев является то, что их отклики слабо коррелированы: $\text{corr}(GpI_II_III_IV, GpI_II_III_IVD) \approx -0.19$.

9.6. Непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки с накоплением радиусов точек и аддитивным свертыванием промежуточных данных

Подсчет точек, оказавшихся по нужную сторону от четырех гиперболических квантователей, является достаточно грубой оценкой статистик. Округление промежуточных данных до целых значений снижает точность вычислений. Верно и обратное. Переход от дискретных вычислений к более точным вычислениям с непрерывными данными обычно приводит к снижению уровня ошибок первого и второго рода непрерывного критерия.

В нашем случае переход к непрерывным оценкам может быть осуществлен заменой подсчета числа точек накоплением радиусов до обнаруженных точек. Программная реализация критерия приведена на рис. 9.6.

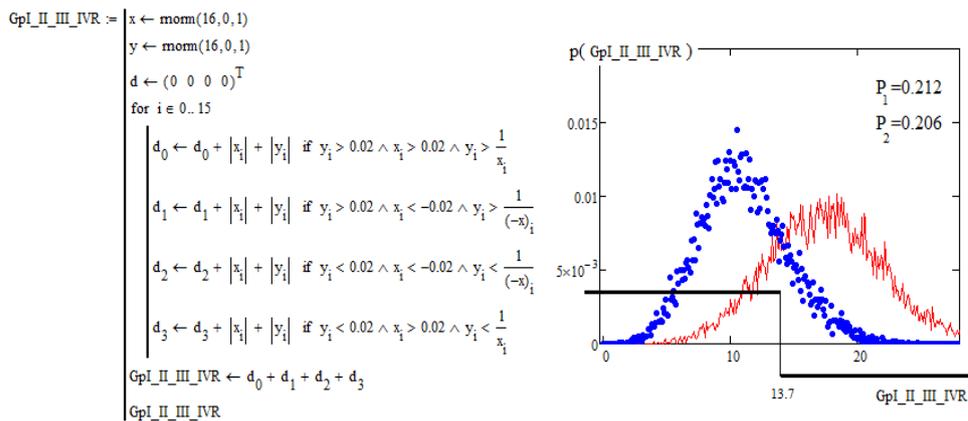


Рис. 9.6. Непрерывный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат, накоплением радиусов и аддитивным свертыванием промежуточных вычислений

Численное моделирование нового критерия показывает снижение вероятности появления ошибок на 36.7 %. Сравнение ведется по отношению к показателям дискретного критерия с аддитивным свертыванием (см. рис. 9.4). Важным является также то, что отклики сравниваемых критериев имеют низкое значение модуля коэффициента корреляции: $\text{corr}(\text{GpI_II_III_IV}, \text{GpI_II_III_IVR}) \approx -0.0005$.

9.7. Непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки с накоплением произведения модулей координат точек и аддитивным свертыванием промежуточных данных

Еще одним направлением синтеза новых критериев является переход от подсчета числа точек к накоплению произведений модулей их координат. Скорректированное программное обеспечение нового критерия приведено на рис. 9.7. Численный эксперимент показал, что для нового критерия уровень разделенности зависимых и независимых данных снижается примерно на 10 %.

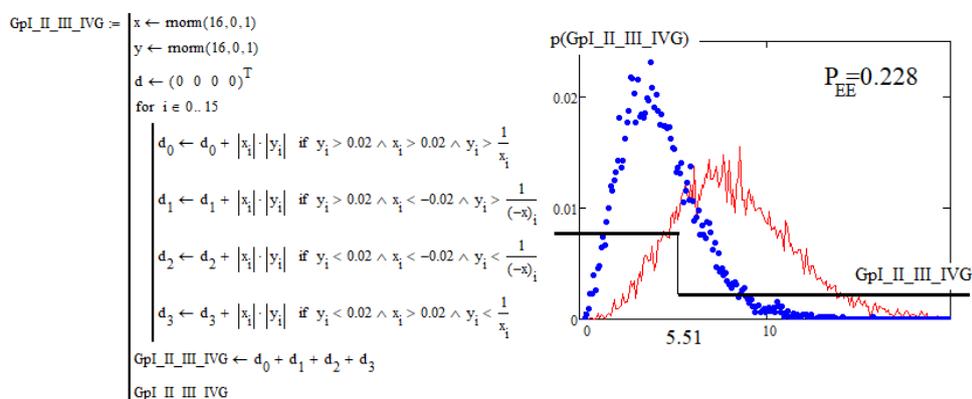


Рис. 9.7. Непрерывный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат, с накоплением произведений координат точек и аддитивным свертыванием промежуточных вычислений

Принципиально важным является то, что переход от накапливания радиусов к накоплению произведений координат, обнаруженных гиперболическими квантователями точек, снижает модуль их корреляционного связывания: $\text{corr}(\text{GpI_II_III_IVR}, \text{GpI_II_III_IVG}) \approx -0.0007$.

9.8. Непрерывный критерий, построенный на четырех гиперболических квантователях данных малой выборки с накоплением произведения модулей координат точек при мультипликативном свертывании промежуточных данных

Существенно снизить вероятности ошибок первого и второго рода для гиперболических критериев удастся, если перейти

от аддитивного свертывания к мультипликативному свертыванию промежуточных вычислений. В рассматриваемой ситуации происходит снижение вероятностей ошибок на 40.7 %, что является значительным результатом.

Программная реализация критерия с мультипликативным свертыванием приведена на рис. 9.8. Реализация численного эксперимента показала очень низкое значение модуля коэффициента корреляции: $\text{corr}(\text{GpI_II_III_IVG}, \text{GpI_II_III_IVGD}) \approx -0.042$.

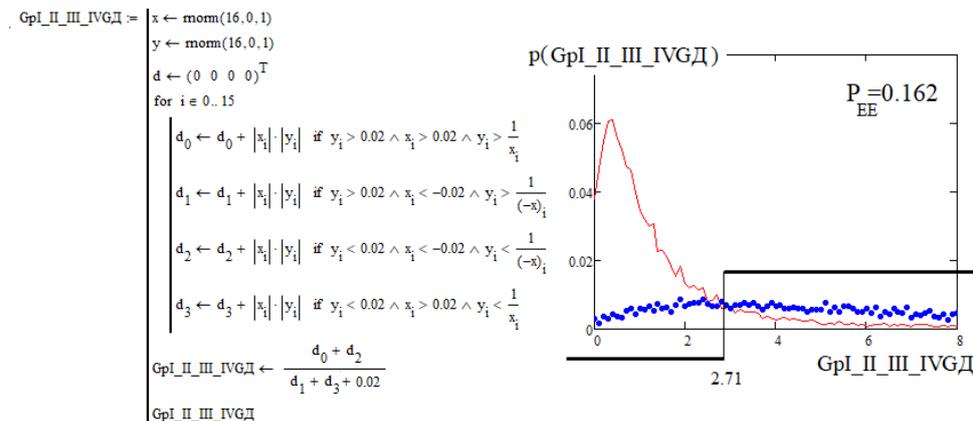


Рис. 9.8. Непрерывный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат, с накоплением произведений координат точек и мультипликативным свертыванием промежуточных вычислений

Низкий уровень вероятностей ошибок двух последних критериев и низкий уровень их корреляционных связей делают эти критерии сильными при их нейросетевом объединении. Кроме того, вероятность появления ошибок первого и второго рода для рассматриваемого критерия близка к наилучшему показателю классического статистического критерия Эджуорта – Эдлтона – Пирсона.

9.9. Непрерывный критерий, построенный на использовании гиперболической функции разделения малой выборки в четырех четвертях Декартовой системы координат с накоплением смещенных радиусов

Важным технологическим моментом является то, что центр, от которого выполняется отсчет радиусов обнаруженных точек, может быть смещен. Программная реализация критерия с единичным смещением точки оценки радиусов приведена на рис. 9.9.

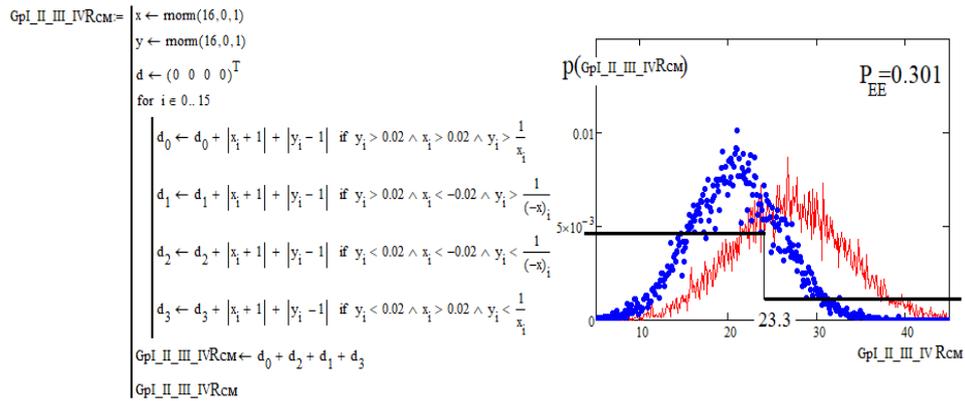


Рис. 9.9. Непрерывный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат, с накоплением смещенных радиусов

Принципиально важным является тот факт, что простая операция смещения точки вычисления радиусов позволяет значительно понижать уровень корреляционной сцепленности. К нашему случаю удается снизить уровень корреляционной сцепленности до значения $\text{corr}(\text{GpI_II_III_IVR}, \text{GpI_II_III_IVRcm}) \approx 0.777$ примерно на 22.3 %. Увеличение смещения точки отсчета радиусов приводит к усилению эффекта снижения уровня корреляционной сцепленности.

9.10. Непрерывный критерий, построенный на использовании гиперболической функции разделения малой выборки в четырех четвертях Декартовой системы координат с накоплением смещенных радиусов и мультипликативным свертыванием промежуточных данных

Еще одним способом снижения уровня корреляционной сцепленности между критериями-родителями и критериями-потомками являются одновременное смещение центров вычисления радиусов и переход от аддитивной к мультипликативной свертке делением. Пример программной реализации такого критерия приведен на рис. 9.10.

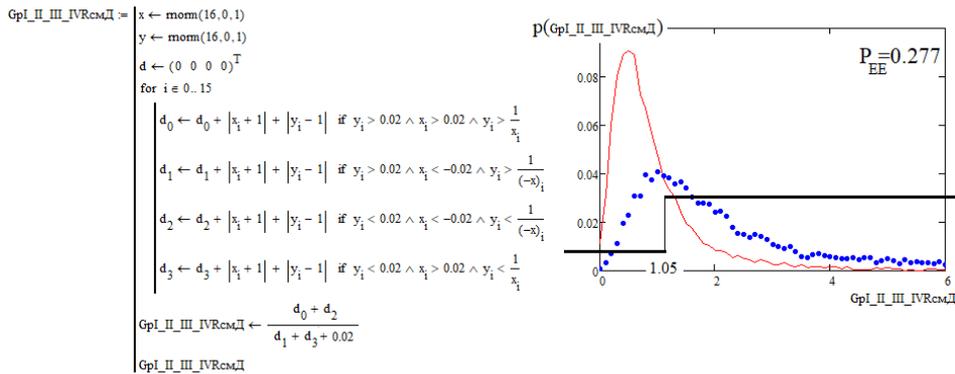


Рис. 9.10. Непрерывный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат, с накоплением смещенных радиусов и мультипликативным объединением данных

Численный эксперимент по моделированию критерия показывает снижение уровня вероятности ошибок на 8.7 %. При этом уровень корреляционной сцепленности критерия-потомка с критерием-родителем существенно снижается: $\text{corr}(\text{GpI_II_III_IVR}, \text{GpI_II_III_IVRcmД}) \approx 0.164$.

9.11. Непрерывный критерий, построенный на использовании гиперболической функции разделения малой выборки во всех четвертях Декартовой системы координат с накоплением смещенных произведений модулей координат

Рассмотренный выше технический прием смещения центра, от которого отсчитываются радиусы, обнаруженных точек, применим для нейрокритериев, построенных на многие типы преобразований, в том числе принцип работает для критериев с гиперболическими квантователями, построенными на умножении модулей координат точек. Пример программной реализации такого критерия приведен на рис. 9.11.

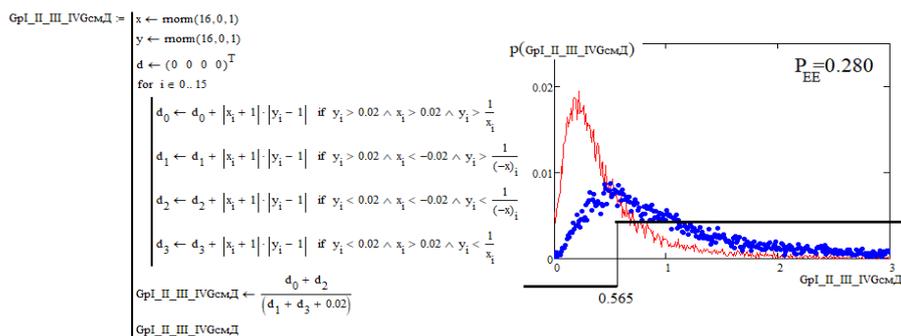


Рис. 9.11. Непрерывный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат, с произведений модулей смещенных координат

Численный эксперимент по моделированию этого критерия показывает низкий уровень корреляционной сцепленности исходного критерия-родителя и нового критерия-потомка:

$$\text{corr}(\text{GpI_II_III_IVG}, \text{GpI_II_III_IVG}_{\text{смД}}) \approx -0.040.$$

9.12. Непрерывный критерий, построенный на использовании гиперболической функции разделения малой выборки во всех четвертях Декартовой системы координат с накоплением смещенных произведений модулей координат при аддитивном свертывании данных

Уровень корреляционной сцепленности снижается, если исходный критерий-родитель отличается от критерия-потомка только смещением точки отсчета расстояния до обнаруженных точек. Пример программной реализации критерия со смещением центров вычисления радиусов с последующим аддитивным свертыванием промежуточных данных приведен на рис. 9.12.

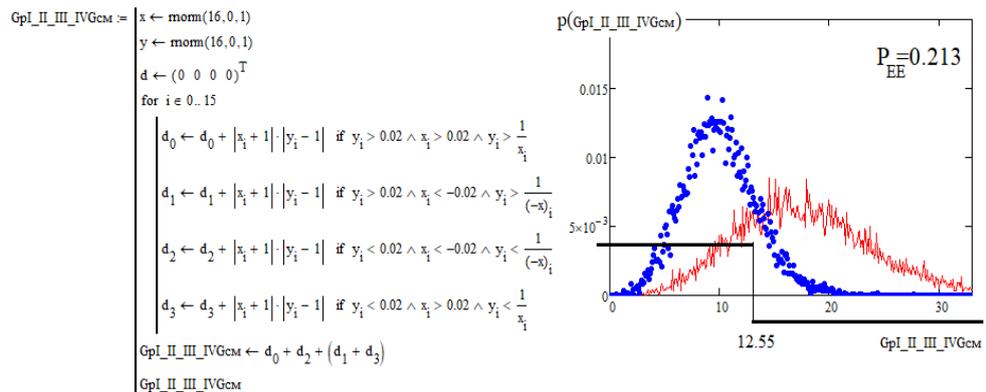


Рис. 9.12. Непрерывный критерий, построенный на гиперболическом квантовании данных, одновременно попадающих во все четверти Декартовой системы координат, с произведений модулей смещенных координат с аддитивным свертыванием данных

Численный эксперимент по моделированию этого критерия показывает «средний» уровень корреляционной сцепленности исходного критерия-родителя и нового критерия-потомка:

$$\text{corr}(\text{GpI_II_III_IVG}, \text{GpI_II_III_IVG}_{\text{см}}) \approx 0.488.$$

9.13. Корреляционная сцепленность критериев, использующих гиперболические квантователи

Сборка группы критериев целесообразна, если модули коэффициентов корреляции их откликов малы. Для рассматриваемой группы критериев коэффициенты парной корреляции приведены на рис. 9.13.

$\text{corr}(r, \text{GpI}) = 0.399$		
$\text{corr}(r, \text{GpII}) = -0.41$	$\text{corr}(\text{GpI}, \text{GpII}) = -0.044$	
$\text{corr}(r, \text{GpI_II}) = 0.56$	$\text{corr}(\text{GpI}, \text{GpI_II}) = 0.41$	$\text{corr}(\text{GpII}, \text{GpI_II}) = -0.729$
$\text{corr}(r, \text{GpI_II_III_IV}) = 0.638$	$\text{corr}(\text{GpI}, \text{GpI_II_III_IV}) = 0.297$	$\text{corr}(\text{GpII}, \text{GpI_II_III_IV}) = -0.312$
$\text{corr}(r, \text{GpI_II_III_IVD}) = 0.144$	$\text{corr}(\text{GpI}, \text{GpI_II_III_IVD}) = 0.054$	$\text{corr}(\text{GpII}, \text{GpI_II_III_IVD}) = -0.071$
$\text{corr}(r, \text{GpI_R}) = 0.419$	$\text{corr}(\text{GpI}, \text{GpI_R}) = 0.978$	$\text{corr}(\text{GpII}, \text{GpI_R}) = -0.042$
$\text{corr}(r, \text{GpI_II_III_IVR}) = -0.004$	$\text{corr}(\text{GpI}, \text{GpI_II_III_IVR}) = 0.367$	$\text{corr}(\text{GpII}, \text{GpI_II_III_IVR}) = 0.347$
$\text{corr}(r, \text{GpI_G}) = 0.426$	$\text{corr}(\text{GpI}, \text{GpI_G}) = 0.901$	$\text{corr}(\text{GpII}, \text{GpI_G}) = -0.038$
$\text{corr}(r, \text{GpI_II_III_IVG}) = -0.002$	$\text{corr}(\text{GpI}, \text{GpI_II_III_IVG}) = 0.396$	$\text{corr}(\text{GpII}, \text{GpI_II_III_IVG}) = 0.372$
$\text{corr}(r, \text{GpI_II_III_IVGD}) = 0.221$	$\text{corr}(\text{GpI}, \text{GpI_II_III_IVGD}) = 0.087$	$\text{corr}(\text{GpII}, \text{GpI_II_III_IVGD}) = -0.108$

Рис. 9.13. Корреляционные коэффициенты группы нейрокритериев, построенных на гиперболических квантователях (начало)

Первый столбец на рис. 9.13 имеет наибольшее значение модуля коэффициентов корреляции – 0.638. Это означает, что все критерии можно использовать к паре с наиболее часто используемым классическим критерием Эджуорта – Эдлтона – Пирсона. Максимальное значение модуля коэффициентов корреляции во втором столбце – 0.396, т.е. нейрокритерий – GpI хорошо совместим с любым критерием рассматриваемой группы.

Третий столбец имеет максимальное значение модуля для пары $\text{corr}(\text{GpII}, \text{GpI_II}) = -0.729$. Использование только этой пары нейрокритериев нежелательно. Все остальные пары нейрокритериев являются перспективными.

$\text{corr}(\text{GpI_II}, \text{GpI_II_III_IV}) = 0.428$		
$\text{corr}(\text{GpI_II}, \text{GpI_II_III_IVD}) = 0.074$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_II_III_IVD}) = 0.19$	
$\text{corr}(\text{GpI_II}, \text{GpI_R}) = 0.717$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_R}) = 0.306$	$\text{corr}(\text{GpI_II_III_IVD}, \text{GpI_R}) = 0.033$
$\text{corr}(\text{GpI_II}, \text{GpI_II_III_IVR}) = 0.003$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_II_III_IVR}) = -0.000$	$\text{corr}(\text{GpI_II_III_IVD}, \text{GpI_II_III_IVR}) = -0.064$
$\text{corr}(\text{GpI_II}, \text{GpI_G}) = 0.658$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_G}) = 0.28$	$\text{corr}(\text{GpI_II_III_IVD}, \text{GpI_G}) = 0.03$
$\text{corr}(\text{GpI_II}, \text{GpI_II_III_IVG}) = 0.000$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_II_III_IVG}) = -0.007$	$\text{corr}(\text{GpI_II_III_IVD}, \text{GpI_II_III_IVG}) = -0.0435$
$\text{corr}(\text{GpI_II}, \text{GpI_II_III_IVGD}) = 0.142$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_II_III_IVGD}) = 0.238$	$\text{corr}(\text{GpI_II_III_IVD}, \text{GpI_II_III_IVGD}) = 0.849$
$\text{corr}(\text{GpI_II}, \text{GpI_Rcm}) = 0.7181$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_Rcm}) = 0.307$	$\text{corr}(\text{GpI_II_III_IVD}, \text{GpI_Rcm}) = 0.032$
$\text{corr}(\text{GpI_II}, \text{GpI_II_III_IVRcm}) = 0.185$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_II_III_IVRcm}) = -0.168$	$\text{corr}(\text{GpI_II_III_IVD}, \text{GpI_II_III_IVRcm}) = -0.096$
$\text{corr}(\text{GpI_II}, \text{GpI_II_III_IVRcmD}) = 0.063$	$\text{corr}(\text{GpI_II_III_IV}, \text{GpI_II_III_IVRcmD}) = 0.164$	$\text{corr}(\text{GpI_II_III_IVD}, \text{GpI_II_III_IVRcmD}) = 0.992$

Рис. 9.13. Продолжение

Среднее значение модулей коэффициентов корреляции первого столбца на рис. 9.13 (продолжение на с. 166) составляет значение 0.299, т.е. критерий GrI_II может эффективно использоваться в большинстве сочетаний с другими критериями рассматриваемой группы. То же самое распространяется и на два других критерия двух следующих столбцов на рис. 9.13 (продолжение на с. 166).

```

corr(GrI_R, GrI_II_III_IVR) = 0.376
corr(GrI_R, GrI_G) = 0.962
corr(GrI_R, GrI_II_III_IVG) = 0.422
corr(GrI_R, GrI_II_III_IVGD) = 0.077
corr(GrI_R, GrI_Rcm) = 0.996
corr(GrI_R, GrI_II_III_IVRcm) = 0.202
corr(GrI_R, GrI_II_III_IVRcmD) = 0.03
corr(GrI_R, GrI_II_III_IVGcm) = 0.04
corr(GrI_R, GrI_II_III_IVGcmD) = 0.037
corr(GrI_II_III_IVR, GrI_G) = 0.368
corr(GrI_II_III_IVR, GrI_II_III_IVG) = 0.887
corr(GrI_II_III_IVR, GrI_II_III_IVGD) = -0.078
corr(GrI_II_III_IVR, GrI_Rcm) = 0.375
corr(GrI_II_III_IVR, GrI_II_III_IVRcm) = 0.777
corr(GrI_II_III_IVR, GrI_II_III_IVRcmD) = -0.066
corr(GrI_II_III_IVR, GrI_II_III_IVGcm) = 0.603
corr(GrI_II_III_IVR, GrI_II_III_IVGcmD) = -0.067
corr(GrI_G, GrI_II_III_IVG) = 0.451
corr(GrI_G, GrI_II_III_IVGD) = 0.076
corr(GrI_G, GrI_Rcm) = 0.947
corr(GrI_G, GrI_II_III_IVRcm) = 0.198
corr(GrI_G, GrI_II_III_IVRcmD) = 0.028
corr(GrI_G, GrI_II_III_IVGcm) = 0.059
corr(GrI_G, GrI_II_III_IVGcmD) = 0.034

```

Рис. 9.13. Продолжение

Среднее значение модулей коэффициентов корреляции первого столбца на рис. 9.13 (см. продолжение на с. 166) составляет значение 0.349. Два пары критериев оказались сильно коррелированы, и эти пары бесперспективны. Примерно столько же бесперспективных пар содержат и два следующих столбца на рис. 9.13 (см. продолжение на с. 166).

```

corr(GrI_II_III_IVG, GrI_II_III_IVGD) = -0.042
corr(GrI_II_III_IVG, GrI_Rcm) = 0.415
corr(GrI_II_III_IVG, GrI_II_III_IVRcm) = 0.564
corr(GrI_II_III_IVG, GrI_II_III_IVRcmD) = -0.042
corr(GrI_II_III_IVG, GrI_II_III_IVGcm) = 0.488
corr(GrI_II_III_IVG, GrI_II_III_IVGcmD) = -0.04
corr(GrI_II_III_IVRcm, GrI_II_III_IVRcmD) = -0.099
corr(GrI_II_III_IVRcm, GrI_II_III_IVGcm) = 0.883
corr(GrI_II_III_IVRcm, GrI_II_III_IVGcmD) = -0.113
corr(GrI_II_III_IVGD, GrI_Rcm) = 0.077
corr(GrI_II_III_IVGD, GrI_II_III_IVRcm) = -0.117
corr(GrI_II_III_IVGD, GrI_II_III_IVRcmD) = 0.873
corr(GrI_II_III_IVGD, GrI_II_III_IVGcm) = -0.159
corr(GrI_II_III_IVGD, GrI_II_III_IVGcmD) = 0.806
corr(GrI_Rcm, GrI_II_III_IVRcm) = 0.204
corr(GrI_Rcm, GrI_II_III_IVRcmD) = 0.031
corr(GrI_Rcm, GrI_II_III_IVGcm) = 0.039
corr(GrI_Rcm, GrI_II_III_IVGcmD) = 0.037
corr(GrI_II_III_IVGcm, GrI_II_III_IVGcmD) = -0.134
corr(GrI_II_III_IVRcmD, GrI_II_III_IVGcm) = -0.117
corr(GrI_II_III_IVRcmD, GrI_II_III_IVGcmD) = 0.925

```

Рис. 9.13. Окончание

На рис. 9.13 (окончание) из 21 пары нейрокритериев только четыре пары имеют высокий уровень корреляционной сцепленности, и, соответственно, эти пары являются бесперспективными для их использования.

КРИТЕРИИ, ПОСТРОЕННЫЕ НА ИСПОЛЬЗОВАНИИ АНАЛИЗА ПРАВИЛА БАЙЕСА С УЧЕТОМ ОТСУТСТВИЯ ЧУВСТВИТЕЛЬНОСТИ К ЗНАКУ ВОЗМОЖНОЙ КОРРЕЛЯЦИОННОЙ СЦЕПЛЕННОСТИ ДАННЫХ МАЛОЙ ВЫБОРКИ

10.1. Многомерный вариант корреляционного аналога функционалов Байеса

В середине XVII в. монах Байес сформулировал свое правило относительно вероятностного влияния двух переменных друг на друга. В XX в. активно использовались последовательно-многомерные Байесовские выводы [31]. Все вероятностные рассуждения Байеса безукоризненны с точки зрения логики их построения, однако неустойчивы (плохо обусловлены). Цепочка вероятностных рассуждений Байеса может сохранять приемлемый для практики уровень устойчивости, если между учитываемыми параметрами существуют сильные корреляционные связи.

Проще всего ситуацию с плохой устойчивостью цепочки вероятностных рассуждений можно проиллюстрировать на примере последовательных событий. Если цепочка переменных независима, то вероятности событий можно перемножать. Эта вычислительная модель устойчива. Рост корреляционных связей приводит к утрате доверия к подобным вычислениям. Тем не менее, синтез корреляционных функционалов для проверки гипотезы независимости данных возможен [32, 33].

Пример программной реализации такого типа критерия приведен на рис. 10.1.

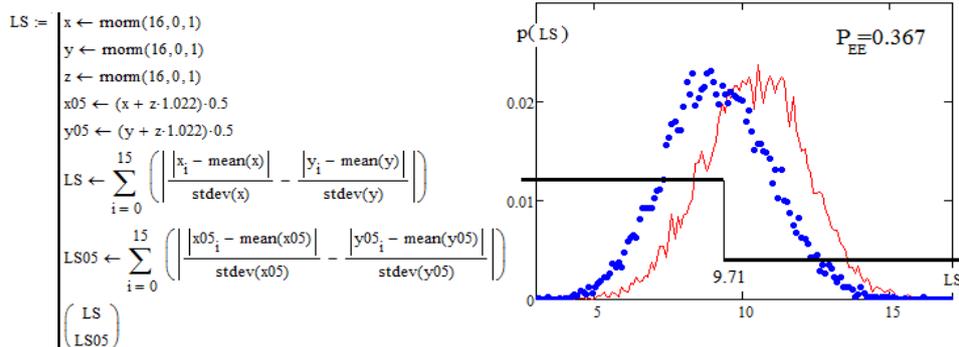


Рис. 10.1. Корреляционный функционал для многомерных устойчивых Байесовских проверок гипотезы независимости

Численный эксперимент показал, что критерии этого типа нечувствительны к знаку коэффициента корреляции данных зависимой выборки, что приводит к некоторому росту вероятностей ошибок первого и второго рода при разделении малых выборок.

Принципиально важным является то, что вычисления, на которых построен данный функционал, абсолютно устойчивы. Рост числа входов у нейрокритерия – LS всегда приводит к монотонному росту качества принимаемых им статистических решений. Исчезла низкая устойчивость многомерных Байесовских рассуждений [31].

10.2. Корреляционный функционал Байеса с накоплением данных в квадратичном пространстве

Мощность предыдущего корреляционного функционала может быть увеличена примерно на 2,5 %, если выполнять накопление данных в квадратичном пространстве. Программная реализация такого типа квадратичного нейрокритерия приведена на рис. 10.2.

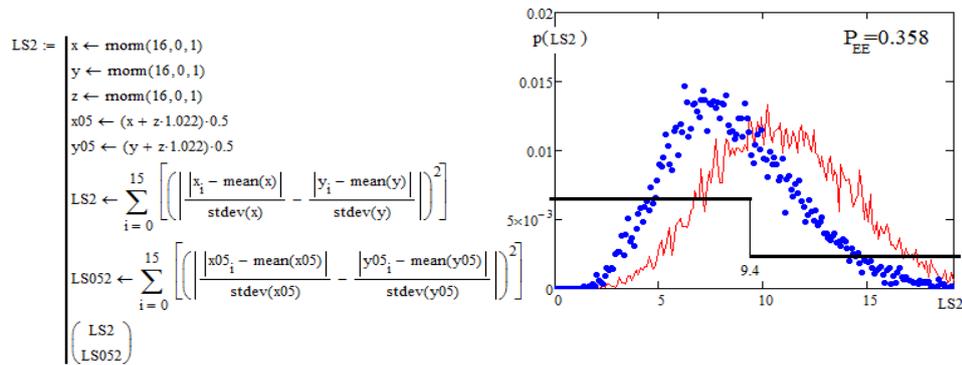


Рис. 10.2. Корреляционный функционал Байеса, выполняющий накопление данных в квадратичном пространстве

По сложности своей программные реализации двух последних нейрокритериев практически одинаковы. В этом контексте всегда выгоднее применять критерий с более высокой мощностью. Коэффициент корреляции откликов двух последних критериев значителен: $\text{corr}(LS, LS2) = 0.909$.

10.3. Корреляционный функционал Байеса, использующий сравнение координат точек с их минимальными значениями

Оставаясь в рамках функционалов с накоплением модулей, удастся снизить вероятности ошибок первого и второго рода примерно на 85.5 %, если заменить в них сравнение данных с математическими ожиданиями на сравнения данных с их минимальными значениями.

Пример программной реализации критерия приведен на рис. 10.3.

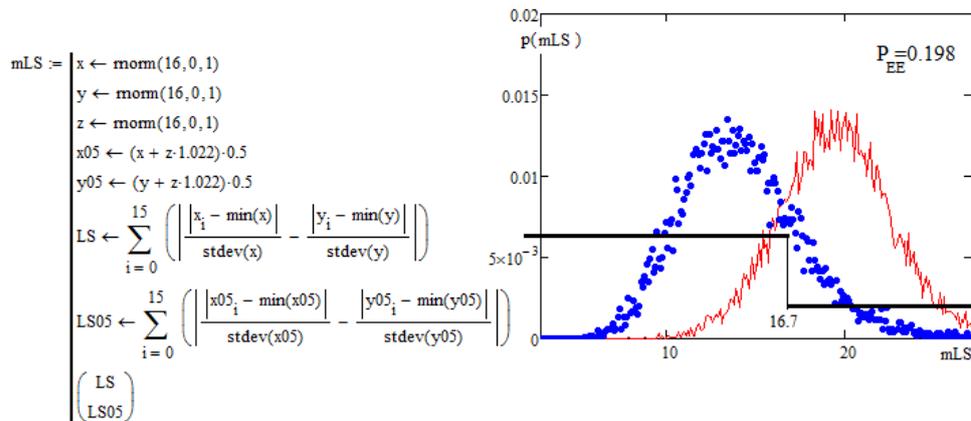


Рис. 10.3. Корреляционный функционал Байеса, усиленный сравнением данных точек малой выборки с их минимальными значениями

Важным преимуществом нового нейрокритерия является низкий уровень его корреляционной сцепленности с двумя предыдущими критериями: $\text{corr}(LS, mLS) \approx 0.166$, $\text{corr}(LS2, mLS) \approx 0.199$.

10.4. Корреляционный функционал Байеса, использующий сравнение координат точек малой выборки с их максимумами

В случае, если мы при вычислении корреляционных функционалов заменим сравнения данных малой выборки с ее центром на сравнения с ее максимальными значениями, мы опять получаем рост качества принимаемых решений. Пример программной реализации нейрокритериев этого типа приведен на рис. 10.4.

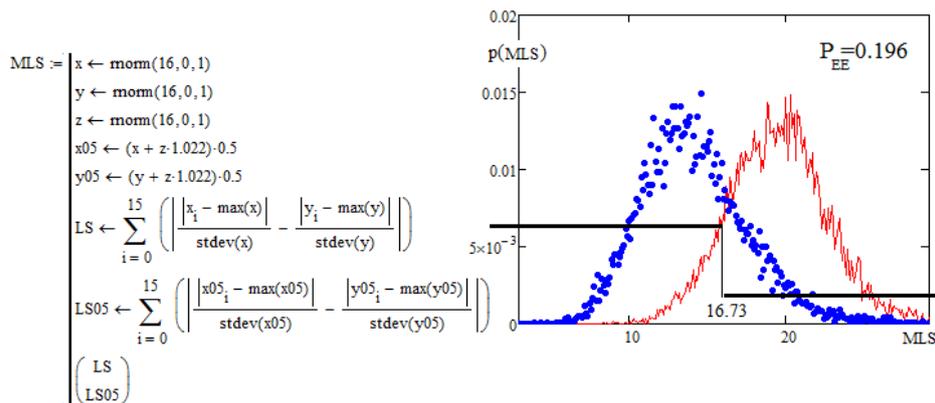


Рис. 10.4. Корреляционный функционал многомерного правила Байеса, усиленный сравнением данных точек малой выборки с их максимумами

Вероятности ошибок первого и второго рода у двух приведенных выше критериев практически совпадают. Важным преимуществом нового нейрокритерия является «низкий» уровень его корреляционной сцепленности с предыдущими критериями: $\text{corr}(mLS, MLS) \approx 0.672$, $\text{corr}(LS2, MLS) \approx 0.162$, $\text{corr}(LS2, mLS) \approx 0.198$.

10.5. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их максимумами по переменной x и минимумами по переменной y

Если данные малых выборок сравнивать с комбинацией двух предельных точек $\{\max(x), \min(y)\}$, мы получаем некоторое повышение вероятностей до значения $P_{EE} \approx 0.218$. Такой уровень вероятности появления ошибок вполне приемлем для практики.

Пример программной реализации нейрокритерия приведен на рис. 10.5. Если сравнивать новый критерий с двумя предыдущими критериями, мы получаем «средний» уровень корреляционной сцепленности: $\text{corr}(MLS, MxBLSmy) \approx 0.581$, $\text{corr}(mLS, MxBLSmy) \approx -0.554$.

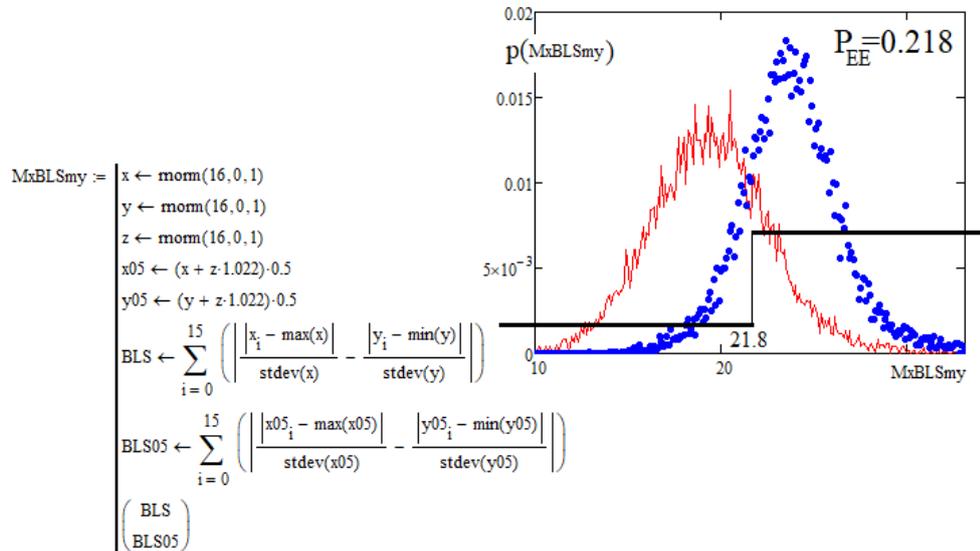


Рис. 10.5. Корреляционный функционал многомерного правила Байеса, усиленный сравнением данных точек малой выборки с $\max(x)$ и $\min(y)$

10.6. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с $\max(y)$ и $\min(x)$

Аналогичная ситуация возникает и в том случае, когда опорные точки сравнения данных малой выборки меняются на другие предельные значения $\{\max(y), \min(x)\}$ (рис. 10.6). И в этом случае наблюдается повышение вероятностей до значения $P_{EE} \approx 0.218$. Коэффициенты корреляционной сцепленности с предыдущим критерием оказываются выше «среднего»: $\text{corr}(\text{MLS}, \text{MxBLSmy}) \approx -0.523$, $\text{corr}(\text{mxLS}, \text{MxBLSmy}) \approx 0.716$.

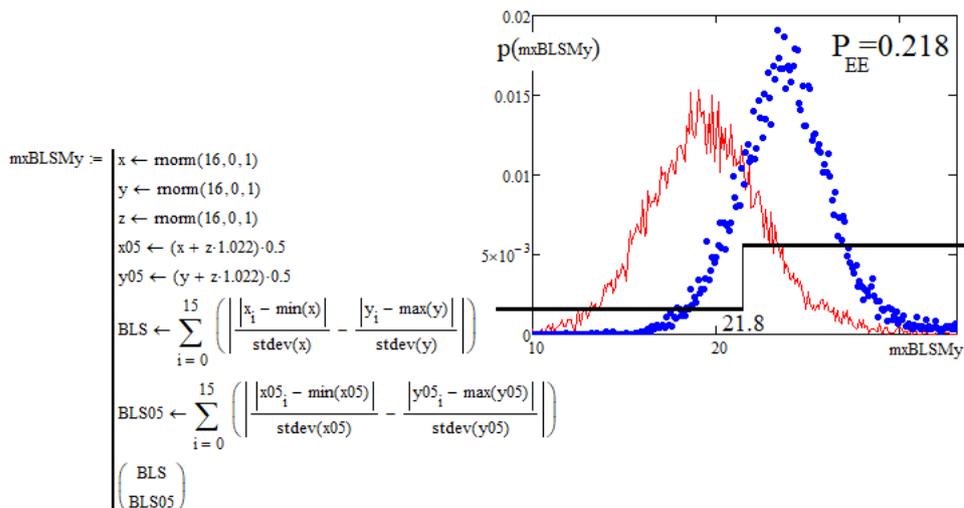


Рис. 10.6. Корреляционный функционал многомерного правила Байеса, усиленный сравнением данных точек малой выборки с $\max(y)$ и $\min(x)$

10.7. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их крайними значениями при мультипликативном свертывании промежуточных данных

Снизить вероятности ошибок первого и второго рода примерно на 20 % удастся, если выполнять сравнение данных малой выборки со всеми ее крайними значениями. При этом появляются избыточные данные, которые могут быть свернуты умножением. Программная реализация нового критерия приведена на рис. 10.7. При программировании выполнено масштабирование отклика нейрокритерия для приведения порога квантования близко к единице.

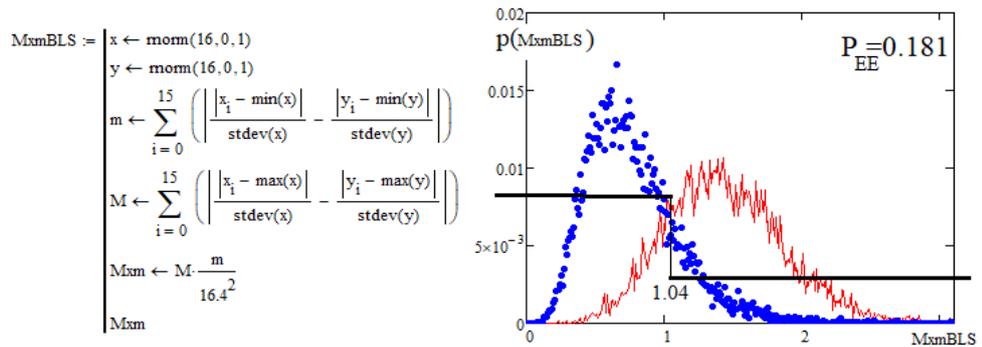


Рис. 10.7. Корреляционный функционал многомерного правила Байеса, усиленный сравнением данных точек малой выборки с их предельными значениями при мультипликативном свертывании данных промежуточных вычислений

Коэффициенты корреляционной сцепленности с предыдущими критериями близки к «средним» значениям: $\text{corr}(MxmBLS, mxLSMy) \approx -0.620$, $\text{corr}(MxmBLS, MxLSmy) \approx -0.560$, $\text{corr}(MxmBLS, MxLSmy) \approx -0.614$.

10.8. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их предельными значениями при аддитивном свертывании промежуточных данных

Кроме мультипликативного свертывания промежуточных данных, возможен иной подход по аддитивному свертыванию промежуточной избыточности. Программная реализация критерия с аддитивным свертыванием избыточности промежуточных вычислений дана на рис. 10.8. При программировании выполнено масштабирование отклика нейрокритерия для приведения порога квантования к единице.

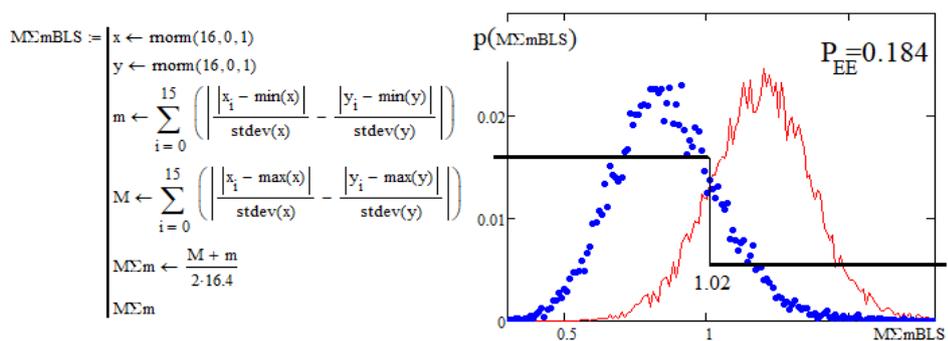


Рис. 10.8. Корреляционный функционал многомерного правила Байеса, усиленный сравнением данных точек малой выборки с их предельными значениями при аддитивном свертывании данных промежуточных вычислений

Коэффициенты корреляционной сцепленности с предыдущими критериями больше «среднего» значения:

$$\begin{aligned} \text{corr}(Mx\text{mBLS}, M\Sigma\text{mBLS}) &\approx 0.993, \text{corr}(Mx\text{LSmy}, M\Sigma\text{mBLS}) \approx -0.571, \\ \text{corr}(mx\text{LSMy}, M\Sigma\text{mBLS}) &\approx -0.620. \end{aligned}$$

К сожалению, аддитивные и мультипликативные сборки избыточности данных дают сильно коррелированные отклики.

10.9. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их предельными значениями при делении аддитивных сверток

Потенциал нейрокритерия повышается, если растет число вычисляемых промежуточных параметров. Предыдущие два критерия были построены на использовании двух сочетаний предельных точек малых выборок. Если увеличить число используемых сочетаний предельных точек до четырех, то удастся снизить вероятности ошибок примерно на 7.6 %.

Программная реализация нового критерия приведена на рис. 10.9. При реализации критерия использованы две аддитивные свертки (две суммы), далее выполнено вычисление отношения этих двух сверток. При квантовании откликов нового критерия по порогу $k = 1.57$ мы получаем высокий уровень разделимости зависимых и независимых данных. Вероятность появления ошибок нового критерия сопоставима с уровнем вероятности ошибок наиболее мощного классического критерия Эджуорта – Эдлтона – Пирсона.

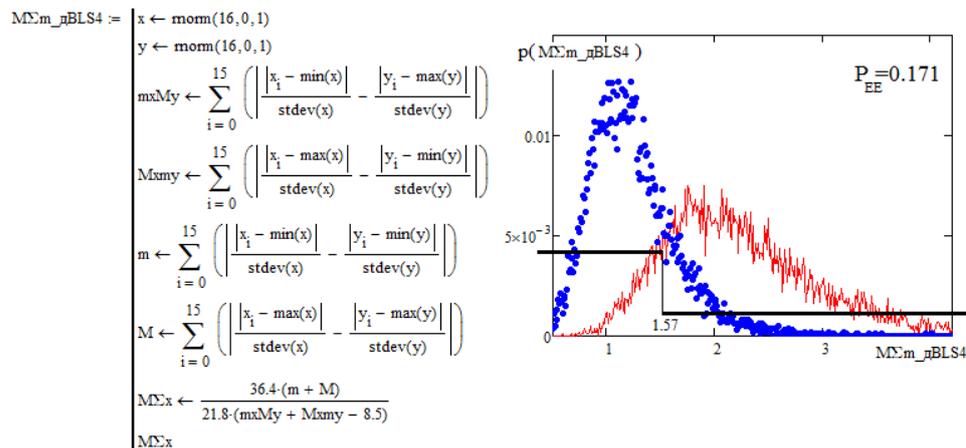


Рис. 10.9. Корреляционный функционал многомерного правила Байеса, усиленный сравнением данных точек малой выборки с их предельными значениями при делении аддитивных сверток

К сожалению, два последних нейрокритерия сильно коррелированы: $\text{corr}(\text{M}\Sigma\text{m_дBLS4}, \text{M}\Sigma\text{mBLS}) \approx 0.960$. Однако существуют другие нейрокритерии, с которыми уровень коррелированности низкий: $\text{corr}(\text{M}\Sigma\text{m_дBLS4}, \text{LS2}) \approx -0.072$.

10.10. Корреляционный функционал многомерного правила Байеса, использующий сравнение координат точек малой выборки с их предельными значениями при делении аддитивной свертки на значение мультипликативной свертки

Еще один вариант предшествующего критерия получается, если четыре компонента промежуточных вычислений свертывать через деление аддитивной свертки двух чисел на мультипликативную свертку двух других чисел. Программная реализация критерия приведена на рис. 10.10.

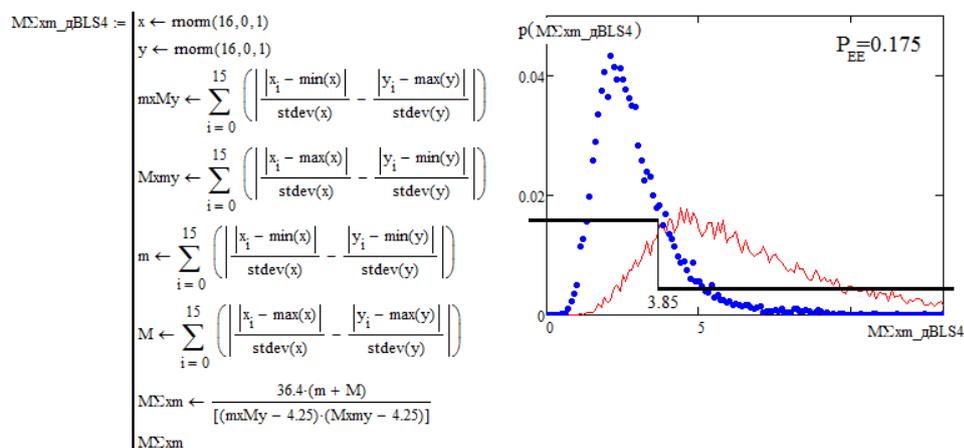


Рис. 10.10. Корреляционный функционал многомерного правила Байеса, усиленный сравнением данных точек малой выборки с их предельными значениями при делении аддитивной свертки на мультипликативную свертку

10.11. Корреляционная сцепленность группы модуль-функционалов, построенных на использовании разных вариантов многомерного правила Байеса, не способных чувствовать знак корреляционной сцепленности анализируемых данных

Объединение нейрокритериев в одну группу с исправлением ошибок эффективно только при малом уровне модулей коэффициентов корреляции (рис. 10.11 содержит их значения).

$\text{corr}(r, LS) = -0.021$		
$\text{corr}(r, LS2) = -0.008$	$\text{corr}(LS, LS2) = 0.909$	
$\text{corr}(r, mL) = -0.762$	$\text{corr}(LS, mL) = 0.211$	$\text{corr}(LS2, mL) = 0.225$
$\text{corr}(r, ML) = -0.753$	$\text{corr}(LS, ML) = 0.216$	$\text{corr}(LS2, ML) = 0.234$
$\text{corr}(r, mxBLMy) = 0.748$	$\text{corr}(LS, mxBLMy) = 0.154$	$\text{corr}(LS2, mxBLMy) = 0.186$
$\text{corr}(r, MxBLMy) = 0.772$	$\text{corr}(LS, MxBLMy) = 0.154$	$\text{corr}(LS2, MxBLMy) = 0.165$
$\text{corr}(r, MxmBL) = -0.796$	$\text{corr}(LS, MxmBL) = 0.202$	$\text{corr}(LS2, MxmBL) = 0.226$
$\text{corr}(r, M\Sigma mBL) = -0.813$	$\text{corr}(LS, M\Sigma mBL) = 0.229$	$\text{corr}(LS2, M\Sigma mBL) = 0.247$
$\text{corr}(r, M\Sigma m_дBL4) = -0.832$	$\text{corr}(LS, M\Sigma m_дBL4) = -0.071$	$\text{corr}(LS2, M\Sigma m_дBL4) = -0.072$
$\text{corr}(r, M\Sigma xm_дBL4) = -0.649$	$\text{corr}(LS, M\Sigma xm_дBL4) = -0.151$	$\text{corr}(LS2, M\Sigma xm_дBL4) = -0.152$

Рис. 10.11. Коэффициенты корреляции группы критериев, построенных на использовании вариантов многомерных правил Байеса (начало)

Первый столбец начала рис. 10.11 содержит модули коэффициентов корреляции, расположенные в интервале $\{0.021, \dots, 0.832\}$, т.е. группа рассматриваемых критериев должна хорошо дополнять классические критерии, созданные в прошлом веке. Объединять вместе критерии из этой группы далеко не всегда целесообразно, так как между ними высок уровень корреляционной сцепленности.

$\text{corr}(mL, ML) = 0.681$		
$\text{corr}(mL, mxBLMy) = -0.560$	$\text{corr}(ML, mxBLMy) = -0.517$	
$\text{corr}(mL, MxBLMy) = -0.549$	$\text{corr}(ML, MxBLMy) = -0.533$	$\text{corr}(mxBLMy, MxBLMy) = 0.696$
$\text{corr}(mL, MxmBL) = 0.912$	$\text{corr}(ML, MxmBL) = 0.910$	$\text{corr}(mxBLMy, MxmBL) = -0.585$
$\text{corr}(mL, M\Sigma mBL) = 0.916$	$\text{corr}(ML, M\Sigma mBL) = 0.917$	$\text{corr}(mxBLMy, M\Sigma mBL) = -0.587$
$\text{corr}(mL, M\Sigma m_дBL4) = 0.759$	$\text{corr}(ML, M\Sigma m_дBL4) = 0.740$	$\text{corr}(mxBLMy, M\Sigma m_дBL4) = -0.822$
$\text{corr}(mL, M\Sigma xm_дBL4) = 0.628$	$\text{corr}(ML, M\Sigma xm_дBL4) = 0.607$	$\text{corr}(mxBLMy, M\Sigma xm_дBL4) = -0.775$

Рис. 10.11. Продолжение

Возможность подбора пар критериев с низким уровнем корреляционной сцепленности подтверждается тем, что из 18 коэффициентов корреляции продолжения рис. 10.11 только семь коэффициентов имеют значения модулей более 0.75.

$\text{corr}(MxBLMy, MxmBL) = -0.560$		
$\text{corr}(MxBLMy, M\Sigma mBL) = -0.560$	$\text{corr}(MxmBL, M\Sigma mBL) = 0.993$	
$\text{corr}(MxBLMy, M\Sigma m_дBL4) = -0.815$	$\text{corr}(MxmBL, M\Sigma m_дBL4) = 0.836$	$\text{corr}(M\Sigma mBL, M\Sigma m_дBL4) = 0.822$
$\text{corr}(MxBLMy, M\Sigma xm_дBL4) = -0.765$	$\text{corr}(MxmBL, M\Sigma xm_дBL4) = 0.693$	$\text{corr}(M\Sigma mBL, M\Sigma xm_дBL4) = 0.668$
$\text{corr}(M\Sigma m_дBL4, M\Sigma xm_дBL4) = 0.960$		

Рис. 10.11. Окончание

Список литературы к главе 10

1. Рассел С., Норвиг П. Искусственный интеллект. Современный подход. М. ; СПб. ; Киев. 2006. 1407 с.
2. Иванов А. И., Ложников П. С., Сулавко А. Е. Оценка надежности верификации автографа на основе искусственных нейронных сетей, сетей многомерных функционалов Байеса и сетей квадратичных форм // Компьютерная оптика. 2017. Т. 41, № 5. С. 765–774.
3. Ложников П. С. Биометрическая защита гибридного документооборота. Новосибирск : Изд-во СО РАН, 2017. 130 с.

КРИТЕРИИ, ПОСТРОЕННЫЕ НА ИСПОЛЬЗОВАНИИ МНОГОМЕРНОГО ПРАВИЛА БАЙЕСА В ВАРИАНТАХ, ЧУВСТВИТЕЛЬНЫХ К ЗНАКУ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ

11.1. Разностно-корреляционный функционал Байеса, чувствительный к знаку корреляционной сцепленности

В предыдущей главе рассматривались критерии, нечувствительные к знаку корреляционной сцепленности данных. К сожалению, эти критерии изначально слабее критериев, чувствительных к знаку корреляционной сцепленности. Простейший такой критерий строится на вычислении разности данных двух переменных после их центрирования и нормирования.

Программная реализация нового критерия приведена на рис. 11.1.

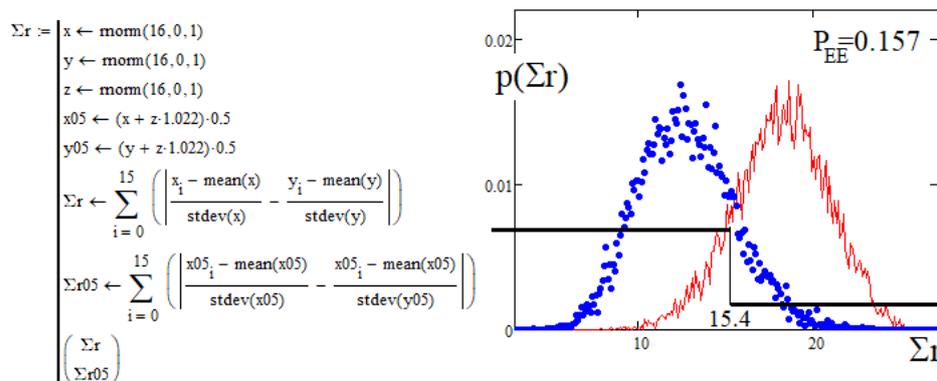


Рис. 11.1. Разностно-корреляционный функционал Байеса, хорошо разделяющий независимые данные и положительно сцепленные данные

Численный эксперимент показал, что новый нейрокритерий имеет высокий уровень разделяемости зависимых и независимых данных с вероятностью появления ошибок $P_{EE} \approx 0.157$. Этот показатель хуже всего на 9 % по сравнению с лучшим классическим критерием Эджуорта – Эдлтона – Пирсона.

Не менее важным является то, что нейрокритерии предыдущей главы, нечувствительные к знаку корреляционной сцепленности,

оказываются слабо связаны с новыми критериями этой главы:
 $\text{corr}(\Sigma r, LS) \approx -0.0201$.

11.2. Аддитивно-корреляционный функционал Байеса, чувствительный к знаку корреляционной сцепленности

Если мы в предыдущем критерии заменим операцию вычитания на операцию сложения, то мы получим новый критерий с увеличенным уровнем вероятности появления ошибок при разделении положительно зависимых и независимых данных.

Пример программной реализации нового критерия приведен на рис. 11.2.

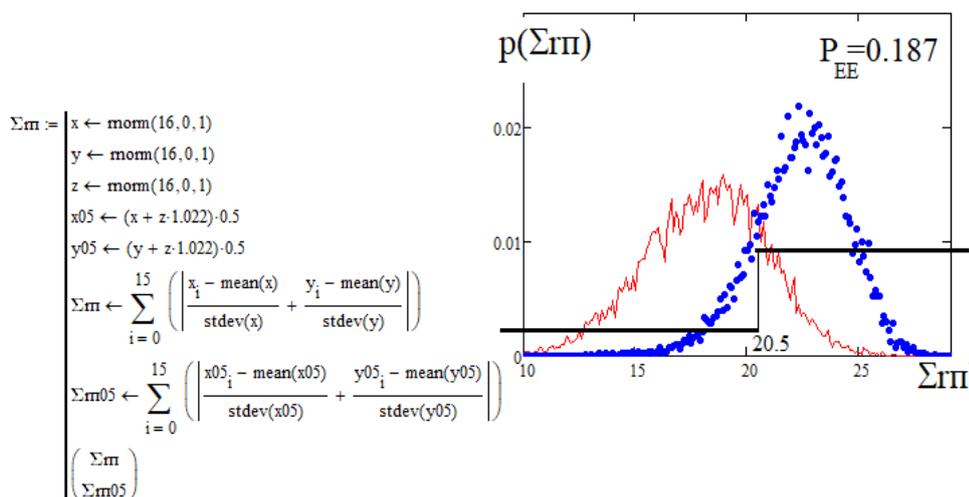


Рис. 11.2. Аддитивно-корреляционный функционал Байеса, чувствительный к знаку корреляционной сцепленности

Важно отметить, что показатели вероятностей появления ошибок первого и второго рода двух последних критериев меняются местами, если знак корреляционной сцепленности отделяемых данных изменится.

К сожалению, отклики двух последних критериев являются сильно коррелированными: $\text{corr}(\Sigma r, \Sigma \pi) \approx -0.809$, однако связи новых критериев с критериями предыдущей главы: $\text{corr}(\Sigma r, LS) \approx -0.0201$, $\text{corr}(LS, \Sigma \pi) \approx 0.121$.

11.3. Усиление разностно-корреляционных функционалов Байеса, чувствительных к знаку корреляционной сцепленности, за счет перехода к накоплению данных в квадратичном пространстве

Следует отметить, что классический корреляционный критерий Эджуорта – Эдлтона – Пирсона накапливает данные в квадратичном пространстве, а два предыдущих критерия накапливают данные в модуль-линейном пространстве. В связи с этим преобразуем разностно-корреляционный критерий, с тем чтобы он накапливал данные в квадратичном пространстве.

Программная реализация нового квадратичного критерия приведена на рис. 11.3. Численный эксперимент показал, что новый критерий оказался на 4.3 % лучше классического критерия Эджуорта – Эдлтона – Пирсона для малых выборок в 16 опытов.

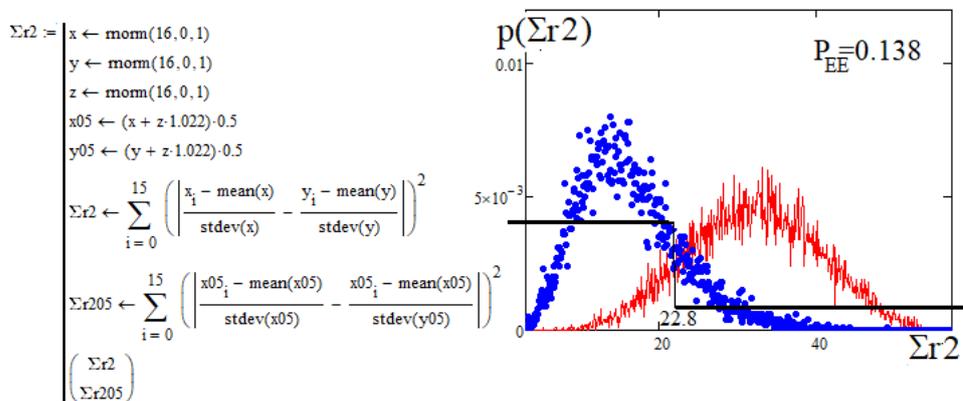


Рис. 11.3. Усиление разностно-корреляционного функционала Байеса за счет перехода к накоплению данных в квадратичном пространстве

Два последних критерия имеют сильно коррелированные отклики: $\text{corr}(\Sigma r2, \Sigma r205) \approx -0.905$. Тем не менее, последний критерий имеет несколько родственных критериев с низкой корреляционной сцепленностью: $\text{corr}(\Sigma r2, LS) \approx -0.005$.

11.4. Разностно-корреляционный функционал Байеса при замене стандартных отклонений размахом изменения данных

Известно, что стандартное отклонение в первом приближении можно заменить размахом изменения данных. Пример программной

реализации нового критерия приведен на рис. 11.4. Численный эксперимент показал, что подобная замена приводит к снижению качества принимаемых решений на 29 %.

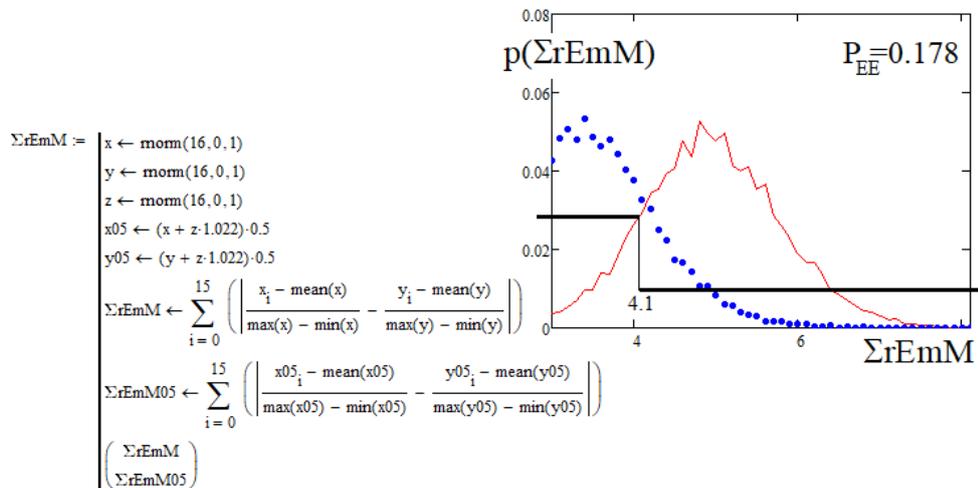


Рис. 11.4. Разностно-корреляционный функционал Байеса при замене стандартных отклонений размахом изменения данных

При этом новая модификация критерия оказывается сильно коррелирована с исходным критерием-родителем: $\text{corr}(\Sigma_r, \Sigma_rEmM) \approx 0.894$, $\text{corr}(\Sigma_r2, \Sigma_rEmM) \approx 0.780$. Сохраняется слабая корреляционная связанность критериев предыдущей и данной глав:

$$\text{corr}(LS, \Sigma_rEmM) \approx 0.021.$$

11.5. Разностно-корреляционный функционал Байеса при замене стандартных отклонений размахом данных и замене математических ожиданий минимумами данных

На малых выборках вычислить точно значения математических ожиданий сложно. Можно говорить о недостаточной устойчивости такого типа вычислений. В этом контексте при оценках коэффициентов корреляции операцию вычисления математического ожидания можно заменить операцией поиска минимума исследуемой выборки по двум переменным.

Кроме того, аналогичная ситуация возникает при оценках стандартного отклонения данных малых выборок. При замене математических ожиданий поиском минимумов и замене стандартных отклонений размахами появляется вариант разностно-корреляционного функционала с конструкцией, отраженной на рис. 11.5.

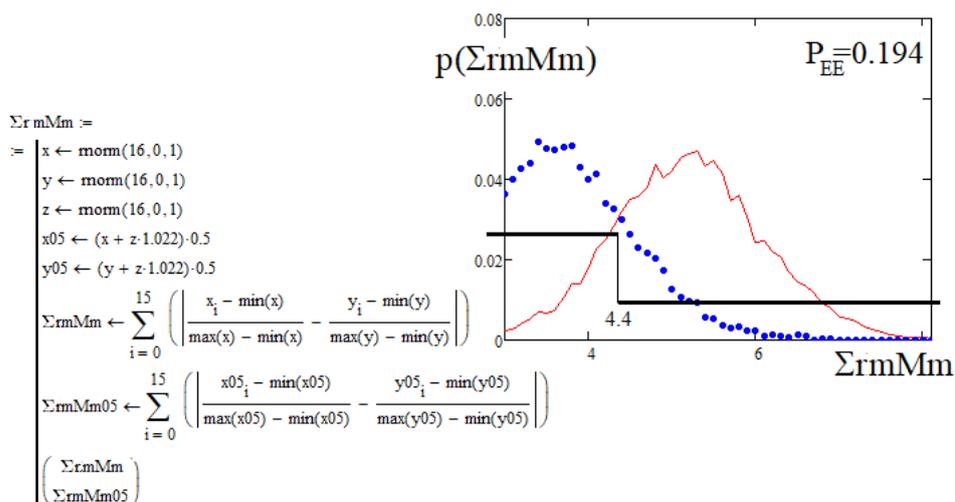


Рис. 11.5. Разностно-корреляционный функционал Байеса при замене стандартных отклонений размахами изменения данных и замене математических ожиданий минимумами

Численный эксперимент по моделированию этого нейрокритерия показал, что он имеет высокий уровень корреляционной сцепленности с предыдущим нейрокритерием: $\text{corr}(\Sigma rEmM, \Sigma rmMm) \approx 0.886$. Однако в исследуемой группе этому нейрокритерию вполне может быть найдена удачная пара с низкой корреляционной сцепленностью: $\text{corr}(LS, \Sigma rmMm) \approx 0.076$.

Очевидно, что параллельно с критерием модулей разностей (см. рис. 11.5) могут быть построены степенные варианты этого критерия (второй, четвертой, шестой степени).

11.6. Аддитивно-корреляционный функционал Байеса при замене стандартных отклонений размахом изменения данных и замене математических ожиданий максимумами данных

Одним из способов снижения корреляционной сцепленности похожих статистических критериев является замена базовой операции, на которой они созданы, на обратную операцию, т.е. вместо разностно-корреляционных функционалов Байеса мы можем воспользоваться аддитивно-корреляционными функционалами. Пример такого типа функционала приведен на рис. 11.6.

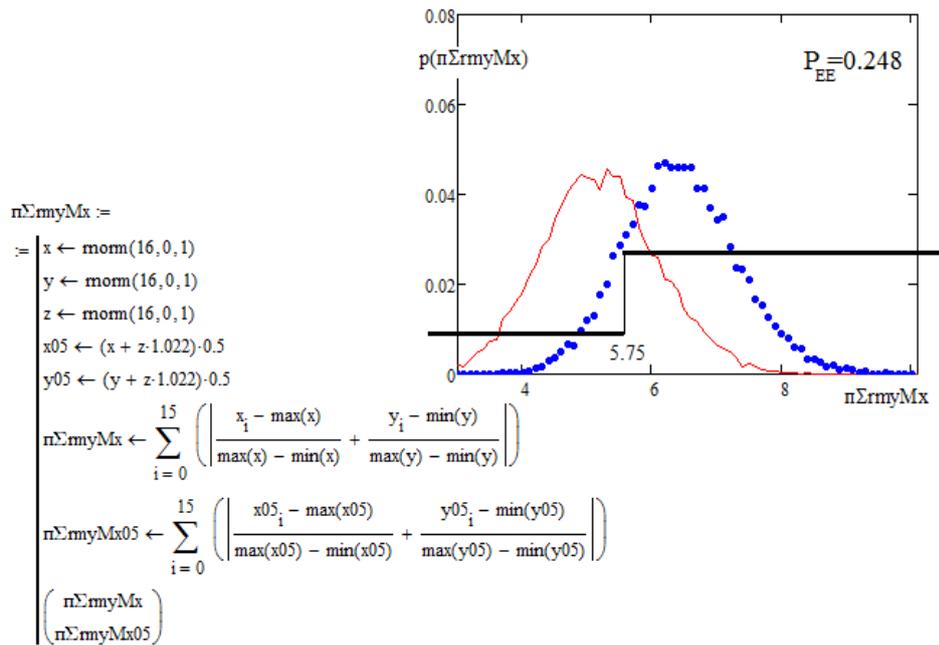


Рис. 11.6. Аддитивно-корреляционный функционал Байеса при замене стандартных отклонений размахами вариации данных и замене математических ожиданий минимумами и максимумами

Разностно-корреляционные функционалы и аддитивно-корреляционные функционалы вполне пригодны для их совместного использования, так как обладают «средним» уровнем корреляционной сцепленности: $\text{corr}(\Sigma rmyMx, \Sigma rMm) \approx -0.414$.

11.7. Аддитивно-корреляционный функционал Байеса с заменой математических ожиданий экстремальными значениями данных

Еще один вариант аддитивного критерия получается, если воспользоваться минимумом по координате x и максимумом по второй координате y . Программная реализация критерия приведена на рис. 11.7. Численный эксперимент по моделированию нового критерия дает вероятность появления ошибок $P_{EE} \approx 0.248$, что соответствует среднему уровню классических критериев и вполне приемлемо для практики.

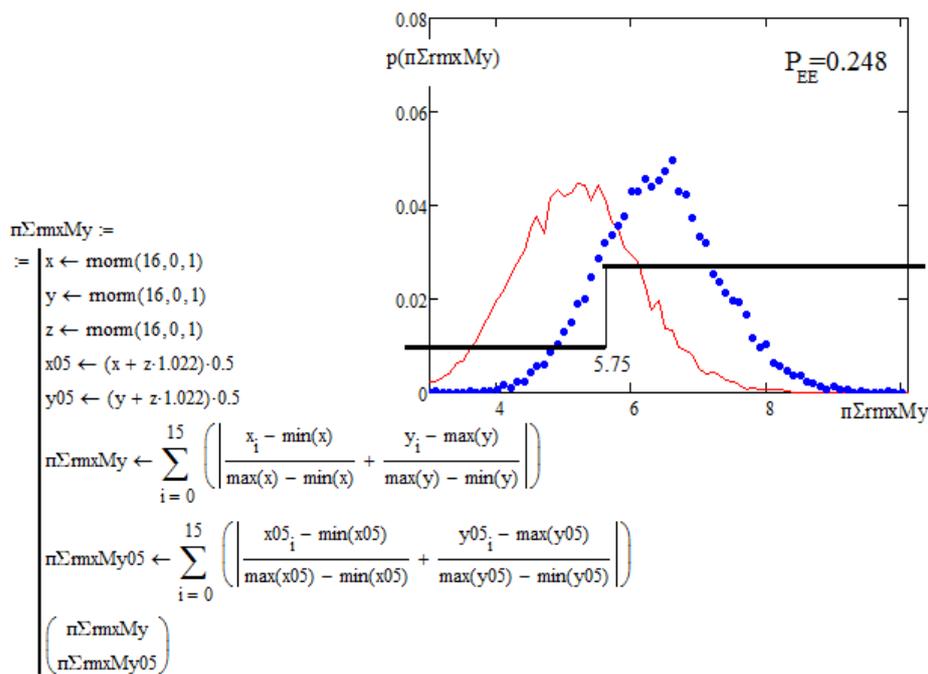


Рис. 11.7. Аддитивно-корреляционный функционал Байеса с заменой стандартных отклонений размахами изменения данных и заменой математических ожиданий экстремумами

К сожалению, отклики двух последних критериев сильно коррелированы: $\text{corr}(\text{п}\Sigma\text{ггmxMy}, \text{п}\Sigma\text{ггmyMx}) \approx 0.757$. Тем не менее, этот критерий имеет в группе хорошую пару с низкой корреляцией: $\text{corr}(\text{п}\Sigma\text{ггmxMy}, \Sigma\text{ггMm}) \approx -0.361$.

11.8. Аддитивно-корреляционный функционал Байеса с заменой математических ожиданий экстремальными значениями данных и заменой стандартных отклонений размахом выборки

Очевидно, что все приведенные в данной главе критерии могут использоваться совместно, например, при их аддитивном объединении. Пример объединения двух частных критериев приведен на рис. 11.8. Численное моделирование нового критерия показывает снижение вероятностей ошибок примерно на 26 % по сравнению с одиночными исходными компонентами.

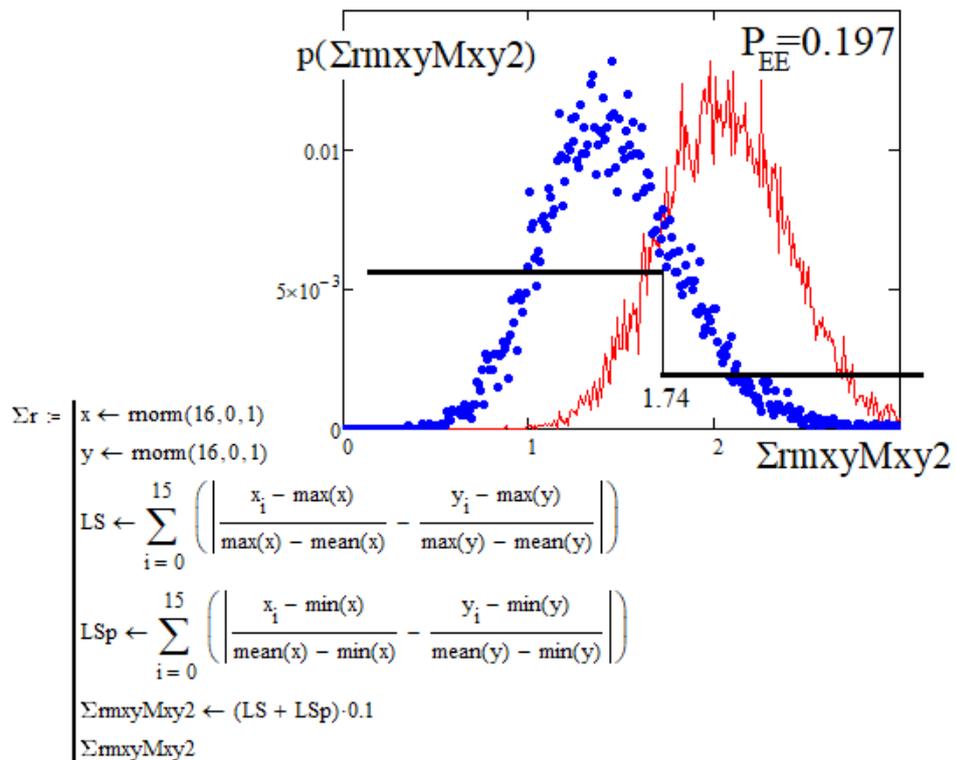


Рис. 11.8. Аддитивно-корреляционный функционал Байеса с заменой стандартных отклонений размахами изменения данных и заменой математических ожиданий экстремумами

К сожалению, аддитивное свертывание дает сильное связывание откликов критериев-потомков и критериев-родителей:

$$\text{corr}(\Sigma r M x y M x y 2, \Sigma r M x y M y) \approx -0.310, \text{corr}(\Sigma r M x y M x y 2, \Sigma r E m M) \approx 0.886.$$

11.9. Разностно-корреляционный функционал Байеса со свертыванием четырех оценок по Колмогорову – Смирнову

Ситуация с негативным наследованием сильных корреляционных связей меняется, если свертывание данных промежуточных вычислений выполняется нелинейно, например, свертывание может быть выполнено по критерию Колмогорова – Смирнова. Пример программной реализации критерия со сверткой избыточности по Колмогорову – Смирнову приведен на рис. 11.9.

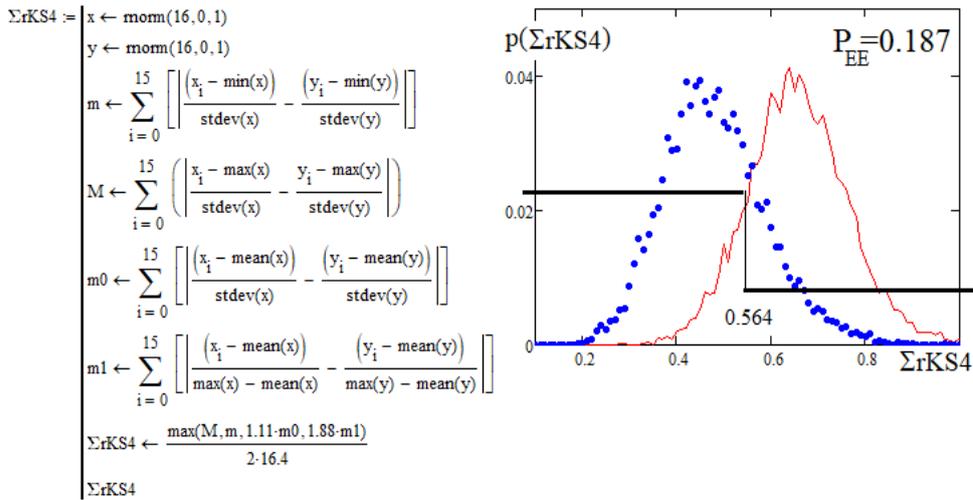


Рис. 11.9. Разностно-корреляционный функционал Байеса со свертыванием промежуточных вычислений по Колмогорову – Смирнову

Результаты численного моделирования показывают низкий уровень корреляционной сцепленности нового критерия с группой рассматриваемых в данной главе критериев:

$$\text{corr}(\Sigma rKS4, \Sigma r) \approx -0.084, \text{corr}(\Sigma rKS4, \Sigma \pi) \approx -0.066,$$

$$\text{corr}(\Sigma rKS4, \Sigma r2) \approx 0.003, \text{corr}(\Sigma rKS4, \Sigma rmyMx) \approx -0.058.$$

11.10. Корреляционная сцепленность в группе критериев Байеса, чувствительных к знаку корреляционных связей

Группы объединяемых статистических критериев должны формироваться по критерию минимума значений модулей их коэффициентов корреляции. На рис. 11.10 приведены значения коэффициентов корреляции в рассматриваемой группе нейрокритериев.

К сожалению, рассматриваемая группа критериев имеет сильные корреляционные связи с классическим критерием Эджуорта – Эдлтона – Пирсона. Об этом свидетельствуют данные первого столбца на рис. 11.10. Большинство модулей его коэффициентов корреляции больше значения 0.75.

$\text{corr}(r, LS) = 0.005$		
$\text{corr}(r, \Sigma r) = -0.901$	$\text{corr}(LS, \Sigma r) = 0.244$	
$\text{corr}(r, \Sigma m) = 0.902$	$\text{corr}(LS, \Sigma m) = 0.244$	$\text{corr}(\Sigma r, \Sigma m) = -0.806$
$\text{corr}(r, \Sigma r2) = -1$	$\text{corr}(LS, \Sigma r2) = -0.005$	$\text{corr}(\Sigma r, \Sigma r2) = 0.900$
$\text{corr}(r, \Sigma rEmM) = -0.784$	$\text{corr}(LS, \Sigma rEmM) = 0.094$	$\text{corr}(\Sigma r, \Sigma rEmM) = 0.896$
$\text{corr}(r, \Sigma mMm) = -0.751$	$\text{corr}(LS, \Sigma mMm) = 0.079$	$\text{corr}(\Sigma r, \Sigma mMm) = 0.812$
$\text{corr}(r, n\Sigma myMx) = 0.751$	$\text{corr}(LS, n\Sigma myMx) = 0.080$	$\text{corr}(\Sigma r, n\Sigma myMx) = -0.651$
$\text{corr}(r, n\Sigma mxMy) = 0.628$	$\text{corr}(LS, n\Sigma mxMy) = 0.079$	$\text{corr}(\Sigma r, n\Sigma mxMy) = -0.540$
$\text{corr}(r, \Sigma mxyMxy2) = -0.751$	$\text{corr}(LS, \Sigma mxyMxy2) = 0.079$	$\text{corr}(\Sigma r, \Sigma mxyMxy2) = 0.812$
$\text{corr}(r, \Sigma rKS4) = -0.01$	$\text{corr}(LS, \Sigma rKS4) = 0.193$	$\text{corr}(\Sigma r, \Sigma rKS4) = -0.084$

Рис. 11.10. Коэффициенты корреляции группы критериев, построенных на функционалах Байеса, чувствительных к знакам связи данных (начало)

В 25 столбцах окончания рис. 11.10 содержатся 15 коэффициентов корреляции с модулем более 0.75. Более половины коэффициентов корреляции имеют высокие значения модулей.

$\text{corr}(\Sigma m, \Sigma r2) = -0.905$		
$\text{corr}(\Sigma m, \Sigma rEmM) = -0.673$	$\text{corr}(\Sigma r2, \Sigma rEmM) = 0.789$	
$\text{corr}(\Sigma m, \Sigma mMm) = -0.655$	$\text{corr}(\Sigma r2, \Sigma mMm) = 0.760$	$\text{corr}(\Sigma rEmM, \Sigma mMm) = 0.886$
$\text{corr}(\Sigma m, n\Sigma myMx) = 0.815$	$\text{corr}(\Sigma r2, n\Sigma myMx) = -0.756$	$\text{corr}(\Sigma rEmM, n\Sigma myMx) = -0.414$
$\text{corr}(\Sigma m, n\Sigma mxMy) = 0.732$	$\text{corr}(\Sigma r2, n\Sigma mxMy) = -0.646$	$\text{corr}(\Sigma rEmM, n\Sigma mxMy) = -0.327$
$\text{corr}(\Sigma m, \Sigma mxyMxy2) = -0.655$	$\text{corr}(\Sigma r2, \Sigma mxyMxy2) = 0.760$	$\text{corr}(\Sigma rEmM, \Sigma mxyMxy2) = 0.886$
$\text{corr}(\Sigma m, \Sigma rKS4) = -0.066$	$\text{corr}(\Sigma r2, \Sigma rKS4) = 0.003$	$\text{corr}(\Sigma rEmM, \Sigma rKS4) = -0.395$
$\text{corr}(\Sigma mMm, n\Sigma myMx) = -0.448$		
$\text{corr}(\Sigma mMm, n\Sigma mxMy) = -0.316$	$\text{corr}(n\Sigma myMx, n\Sigma mxMy) = 0.757$	
$\text{corr}(\Sigma mMm, \Sigma mxyMxy2) = 1$	$\text{corr}(n\Sigma myMx, \Sigma mxyMxy2) = -0.448$	$\text{corr}(n\Sigma mxMy, \Sigma mxyMxy2) = -0.31$
$\text{corr}(\Sigma mMm, \Sigma rKS4) = -0.325$	$\text{corr}(n\Sigma myMx, \Sigma rKS4) = -0.058$	$\text{corr}(n\Sigma mxMy, \Sigma rKS4) = -0.214$

Рис. 11.10. Окончание

КРИТЕРИИ, ПОСТРОЕННЫЕ НА ФРАКТАЛЬНОМ ЗАПОЛНЕНИИ ОКРУЖНОСТЯМИ, НАКРЫВАЮЩИМИ ПО ДВЕ И ТРИ ТОЧКИ МАЛОЙ ВЫБОРКИ

12.1. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через усреднение радиусов окружностей, накрывающих половину точек выборки

В конце прошлого века активно развивались фрактальные методы обработки информации [34, 35]. Данные независимых выборок должны равномерно заполнять некоторую окружность, построенную на центре выборки и захватывающую точку с максимальным отклонением от центра. В этом случае фрактальная размерность задачи совпадает с показателем размерности площади 2. По мере увеличения корреляционной сцепленности данных размерность их статистического описания снижается, становясь дробной величиной. В предельном случае полной корреляционной сцепленности показатель размерности задачи становится единичным 1. Формально мы можем для каждой выборки оценить ее дробную или фрактальную размерность.

Простейший способ оценки дробной размерности малой выборки состоит в оценке множества радиусов, накрывающих две, три, четыре, пять точек. Окружность с усредненным радиусом накрывает примерно 30 % точек малой выборки. По этому параметру зависимые и независимые данные хорошо различимы.

Пример программной реализации критерия приведен на рис. 12.1.

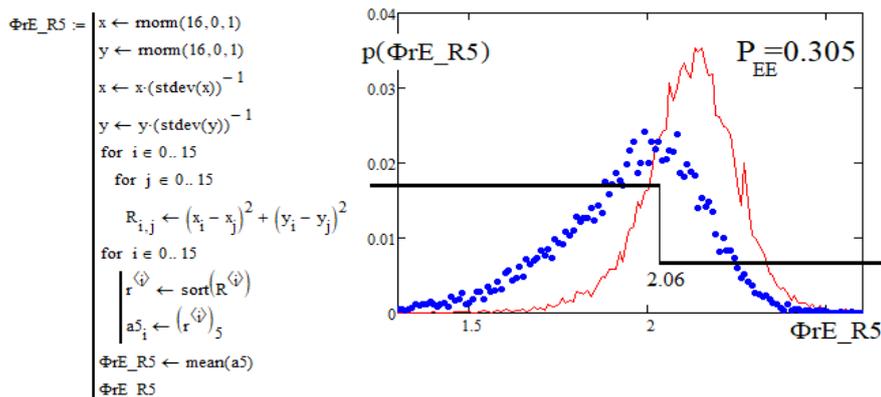


Рис. 12.1. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через усреднение радиусов окружностей

Численный эксперимент показал, что новый критерий оказался хуже примерно в два раза при сравнении с наиболее эффективным классическим критерием Эджуорта – Эдлтона – Пирсона.

12.2. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку стандартного отклонения радиусов

Фрактальная размерность данных существенно влияет на все статистические моменты распределения радиусов малой выборки. Предыдущий критерий был построен на вычислении математического ожидания, следующий статистический момент – стандартное отклонение также может быть использовано в качестве статистического критерия. Пример программной реализации критерия $\Phi\sigma_R$ приведен на рис. 12.2.

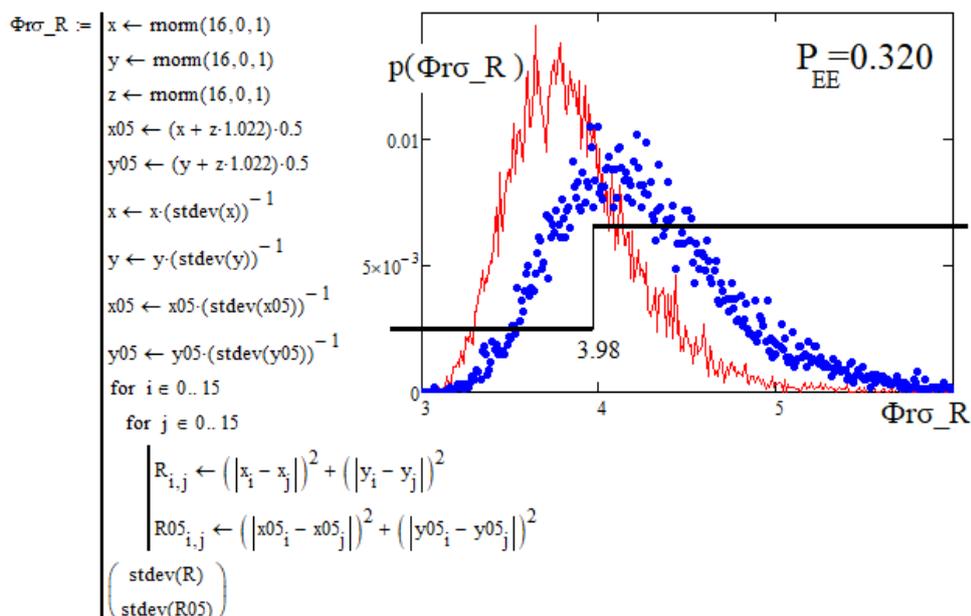


Рис. 12.2. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через вычисление стандартного отклонения радиусов окружностей

Численный эксперимент показал, что дисперсионный критерий слабее предыдущего критерия, построенного на вычислении математического ожидания радиусов окружностей, накрывающих вокруг каждой конкретной точки по четыре ближайшие точки.

12.3. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку максимального значения радиусов

Если перейти от контроля стандартного отклонения к контролю максимальных значений радиусов, мы получаем еще один критерий с относительно высоким уровнем вероятности появления ошибок первого и второго рода (рис. 12.3). Однако все эти критерии все-таки могут использоваться в одной группе, так как имеют приемлемый уровень корреляционной сцепленности: $\text{corr}(\Phi rE_R5, \Phi r\sigma_R) \approx -0.309$, $\text{corr}(\Phi rE_R5, \Phi r_MR) \approx -0.203$, $\text{corr}(\Phi r\sigma_R, \Phi r_MR) \approx 0.818$.

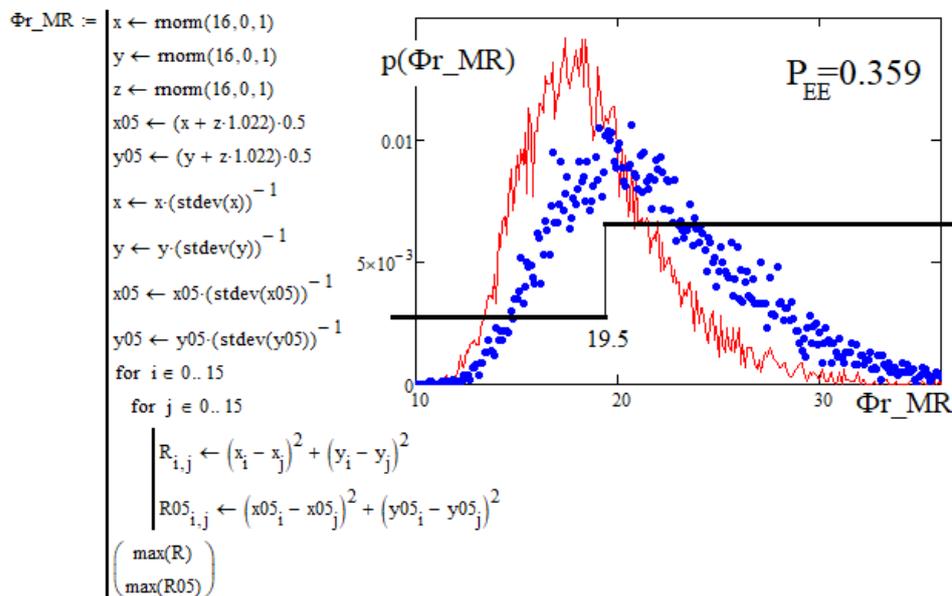


Рис. 12.3. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через вычисление максимума радиусов

12.4. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку математического ожидания радиусов и деление на стандартное отклонение радиусов

Существенного снижения уровня ошибок первого и второго рода удастся добиться, если перейти к вычислению отношения

математического ожидания радиусов окружностей, накрывающих пять рядом находящихся точек, к значению стандартного отклонения всех радиусов, накрывающих две, три, ..., 15 точек малой выборки (рис. 12.4).

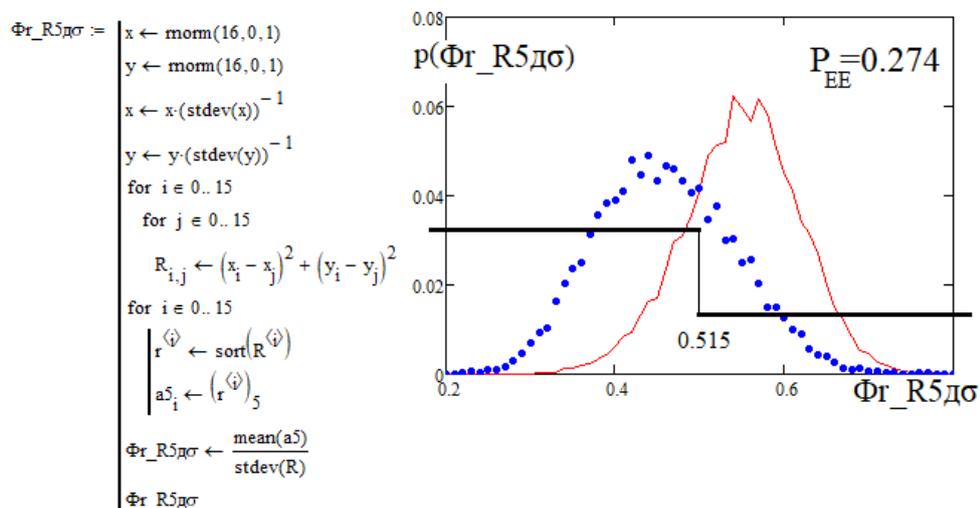


Рис. 12.4. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через вычисление математического ожидания радиусов окружностей и его деление на стандартное отклонение радиусов

12.5. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через оценку длины ломаной линии, соединяющей все точки малой выборки

Рекордного для классических статистических критериев снижения вероятностей появления ошибок первого и второго рода удастся достичь, если выполнить сортировку данных малой выборки по одной из переменных. Тогда точки второй переменной, оставшиеся случайными, могут быть соединены некоторой ломаной линией. Длина квадратов отрезков этой ломаной по порядку следования случайных точек и вторая длина квадратов колен другой ломаной линии по другому порядку соединения точек дают возможность «хорошего» разделения зависимых и независимых данных.

Программная реализация критерия приведена на рис. 12.5.

Численный эксперимент дает снижение вероятности появления ошибок до значения $P_{EE} \approx 0.078$. Это эквивалентно снижению вероятности примерно в 3.5 раза. По этому показателю рассматриваемый критерий намного лучше всех классических критериев прошлого века.

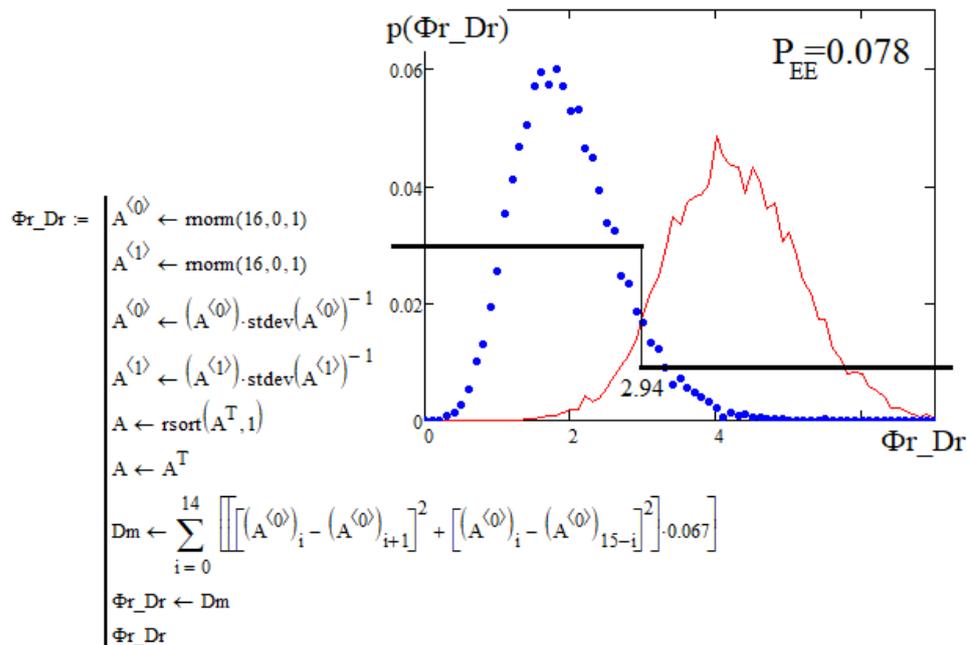


Рис. 12.5. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через вычисления среднего расстояния ломаной линии, соединяющей точки малой выборки

Очевидным является то, что упорядочивание данных по другой переменной должно приводить к появлению еще одного похожего критерия с такой же вероятностью ошибок первого и второго рода $P_{EE} \approx 0.078$. При этом оба родственных критерия будут иметь слабую корреляционную сцепленность.

12.6. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование скользящей хи-квадрат суммы с четырьмя элементами и поиск максимумов

Одной из причин высокого уровня разделимости независимых и зависимых данных малых выборок является то, что предыдущий критерий построен, опираясь на достаточно длинную траекторию обхода точек. Возникает вопрос о том, что произойдет, если мы сократим окно просматриваемых участков траектории обхода, например, до трех произвольно выбранных точек малой выборки. При этом сокращение окна просмотра можно попытаться скомпенсировать увеличением числа таких окон, например, это может быть организовано движением хи-квадрат критерия по 12 рядом расположенным окнам исследуемой выборки.

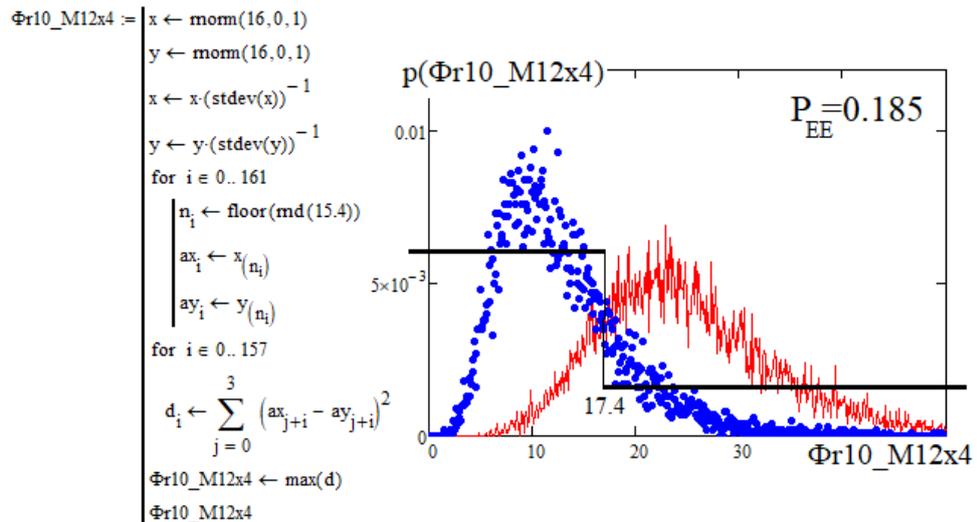


Рис. 12.6. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через 10-кратное увеличение псевдовыборки случайными перестановками точек

Суммирование квадратов координат разности двух точек должно давать хи-квадрат распределение с четырьмя степенями свободы, при этом параметры этого распределения будут существенно зависеть от корреляционной сцепленности данных малой выборки [36, 37].

Очевидно, что 10-кратное увеличение объема псевдовыборки ломаных линий, соединяющих между собой четыре точки малой выборки, приводит к существенному снижению вероятностей появления ошибок первого и второго рода. Численное моделирование рассматриваемых фрактальных функционалов было выполнено для четырех точек и 10-кратного увеличения псевдо-выборки ломаных линий, исходя из возможностей обычной на текущий момент вычислительной машины. Предположительно, что поток ошибок первого и второго рода может быть снижен, если выполнить взаимную оптимизацию перечисленных выше параметров перестановок.

Основы оптимизации опираются на статистические решения задачи «коммивояжера». В обычной постановке задачи коммивояжер пытается найти кратчайший путь обхода точек на плоскости полным перебором. Точное решение задачи «полного» перебора имеет высокую вычислительную сложность. Приближенное решение задачи по оценке средней длины случайно выбираемых путей и его стандартного отклонения является простой задачей с линейной вычислительной сложностью.

Как показано в данном параграфе, для малых выборок в 16 опытов приближенных решений задачи «коммивояжера» вполне достаточно для существенного снижения вероятностей ошибок первого и второго рода.

12.7. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через использование скользящей хи-квадрат суммы с четырьмя элементами и поиск минимумов

Предшествующий критерий построен на том, что правый «хвост» хи-квадрат распределений намного длиннее левого «хвоста». Именно по этой причине мы получили высокий уровень разделимости зависимых и независимых данных предшествующим критерием. Однако у асимметричных хи-квадрат распределений для малых выборок оказываются хорошо наблюдаемы и левые «хвосты», т.е. может быть создан критерий, построенный на поиске минимального значения выборки, увеличенной в 10 раз случайными перестановками ее точек. Программная реализация такого критерия приведена на рис. 12.7.

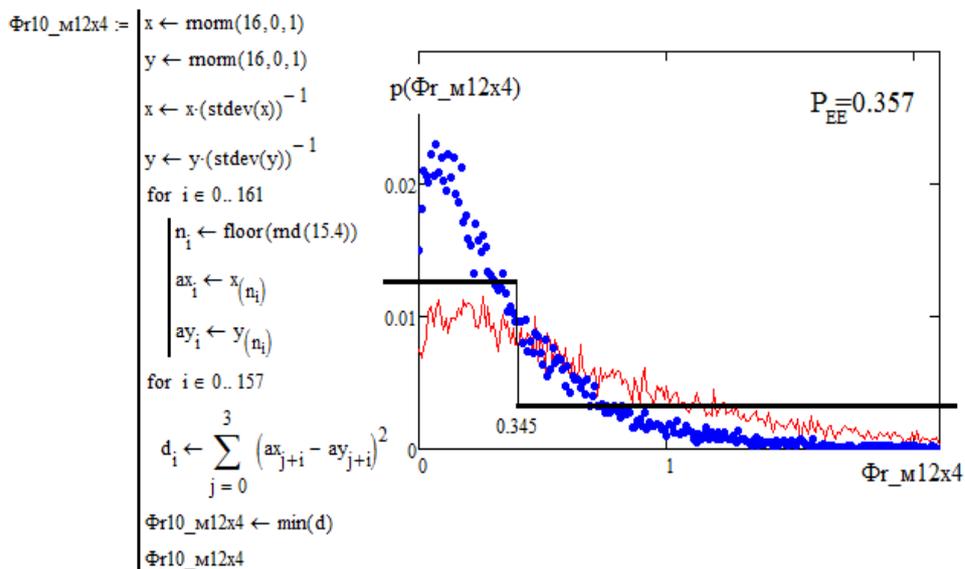


Рис. 12.7. Фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через 10-кратное увеличение псевдовыборки случайными перестановками точек и поиск минимальных значений

Проведенный численный эксперимент показал работоспособность нового критерия, однако вероятность ошибок первого и второго рода для него увеличивается примерно в два раза по сравнению с предыдущим критерием.

12.8. Дискретный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки, скользящей хи-квадрат суммы с четырьмя элементами (правый «хвост»)

Обычно непрерывный вариант того или иного критерия оказывается слабо корреляционно сцеплен с его дискретным вариантом. В связи с этим целесообразно перейти к исследованию дискретных вариантов двух предыдущих критериев. На рис. 12.8 приведена программная реализация дискретного критерия – $d\Phi_{r10_M12x4}$, являющегося аналогом непрерывного критерия – Φ_{r10_M12x4} .

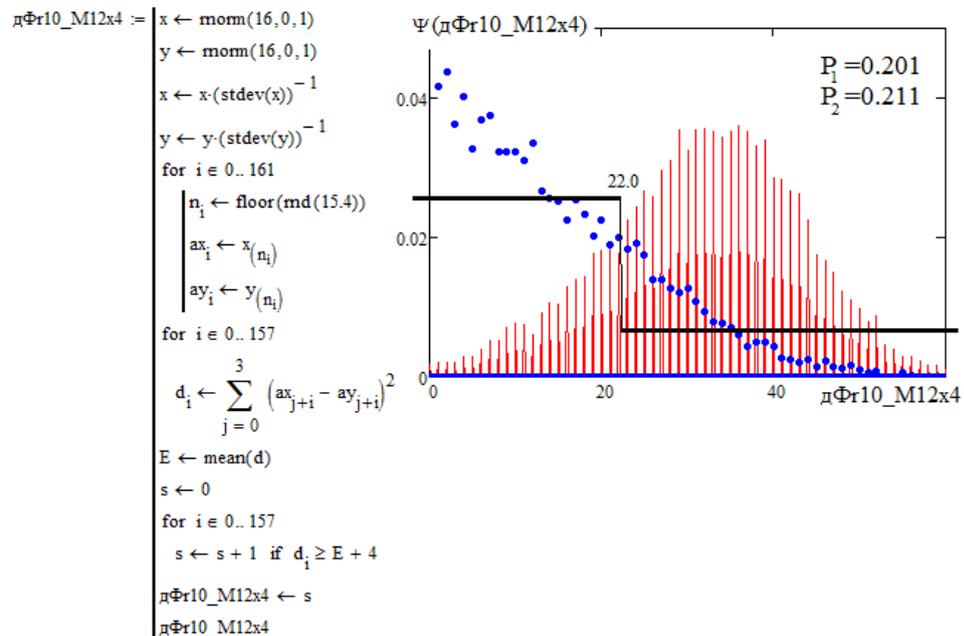


Рис. 12.8. Дискретный фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через 10-кратное увеличение псевдовыборки случайными перестановками точек (правые «хвосты» хи-квадрат распределений)

Численный эксперимент по моделированию нового критерия показал для него высокий уровень разделимости зависимых и независимых данных. Произошло незначительное повышение уровня вероятности появления ошибок всего на 8 % по отношению к его непрерывному аналогу. При этом примерно 27.3 % корреляционной связанности утрачивается: $\text{corr}(\Phi_{r10_M12x4}, d\Phi_{r10_M12x4}) \approx 0.727$. Этого уже

достаточно для совместного использования дискретной и непрерывной версий одного и того же критерия.

12.9. Дискретный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки, скользящей хи-квадрат суммы с четырьмя элементами (левый «хвост»)

Существенно иная ситуация возникает, когда мы используем критерий поиска минимальных значений хи-квадрат распределений. В этом случае разделяющая способность дискретного варианта критерия оказывается примерно в два раза выше, чем у его непрерывного аналога. Пример программной реализации нового критерия приведен на рис. 12.9.

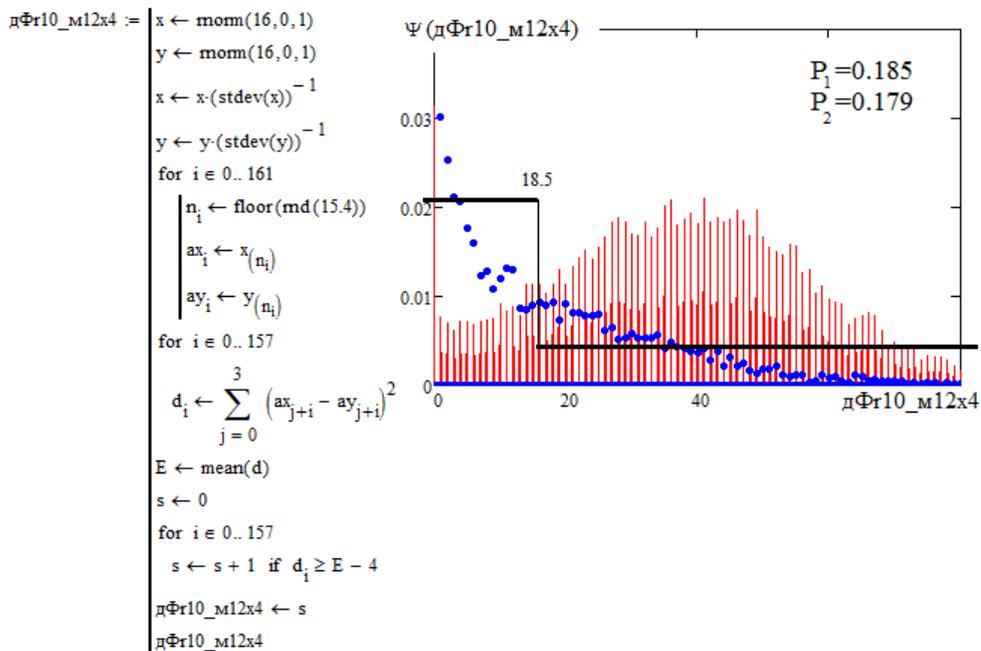


Рис. 12.9. Дискретный фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через 10-кратное увеличение псевдовыборки случайными перестановками точек (левые «хвосты» хи-квадрат распределений)

Численный эксперимент показал не только значительное увеличение делимости для нового критерия, но и значительное устранение

взаимной зависимости откликов родственных критериев:
 $\text{corr}(\Phi r10_m12x4, \text{д}\Phi r10_m12x4) \approx 0.276$.

12.10. Дискретный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки, скользящей хи-квадрат суммы с четырьмя элементами и аддитивной сверткой левого и правого «хвостов»

Очевидно, что два рассмотренных выше статистических критерия дополняют друг друга и могут быть аддитивно объединены друг с другом. Программная реализация такого объединения приведена на рис. 12.10.

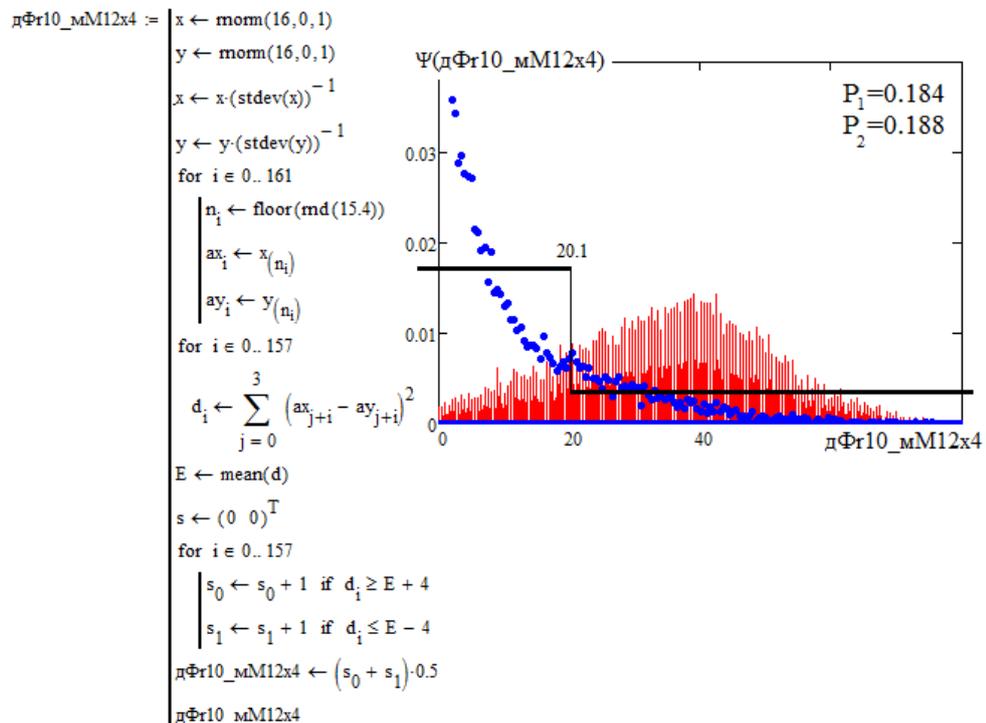


Рис. 12.10. Дискретный фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через 10-кратное увеличение псевдовыборки случайными перестановками точек с аддитивной сверткой данных по левому и правому «хвостам»

К сожалению, аддитивное объединение дает критерии-потомков, сильно зависящих от критериев-родителей:

$$\text{corr}(\Phi r10_m12x4, \text{д}\Phi r10_mM12x4) \approx 0.891,$$

$$\text{corr}(\Phi r10_M12x4, \text{д}\Phi r10_mM12x4) \approx 0.993.$$

12.11. Непрерывный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки и скользящей хи-квадрат суммы с четырьмя элементами, а также накопления статистик правого «хвоста»

Уйти от высокой корреляционной сцепленности критериев-потомков позволяет возврат к непрерывному критерию варианта анализа данных малой выборки. Вернуться удастся, если вместо подсчета числа превышений данных над порогом выполняется накапливание непрерывных значений хи-квадрат состояний с четырьмя степенями свободы. Пример программной реализации непрерывного критерия этого класса приведен на рис. 12.11.

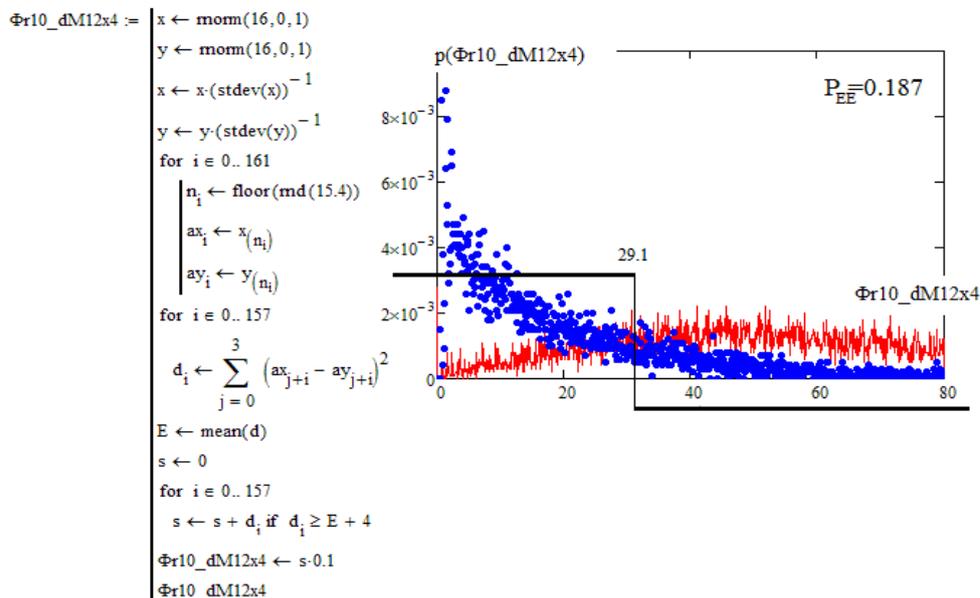


Рис. 12.11. Непрерывный фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через 10-кратное увеличение псевдовыборки случайными перестановками точек и учет статистик правого «хвоста» хи-квадрат распределения

Переход от дискретного статистического критерия к непрерывному критерию дает снижение уровня вероятностей ошибок примерно на 7.5 %, однако оба критерия имеют высокий уровень корреляционной сцепленности их откликов:

$$\text{corr}(\Phi_{r10_M12x4}, \Phi_{r10_dM12x4}) \approx 0.916.$$

12.12. Непрерывный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки и скользящей хи-квадрат суммы с четырьмя элементами и учетом статистик левого «хвоста»

Похожая ситуация возникает, когда нейросетевой критерий строится на накоплении редких событий попадания данных в левый «хвост» хи-квадрат распределений промежуточных вычислений. Программная реализация этого критерия приведена на рис. 12.12.

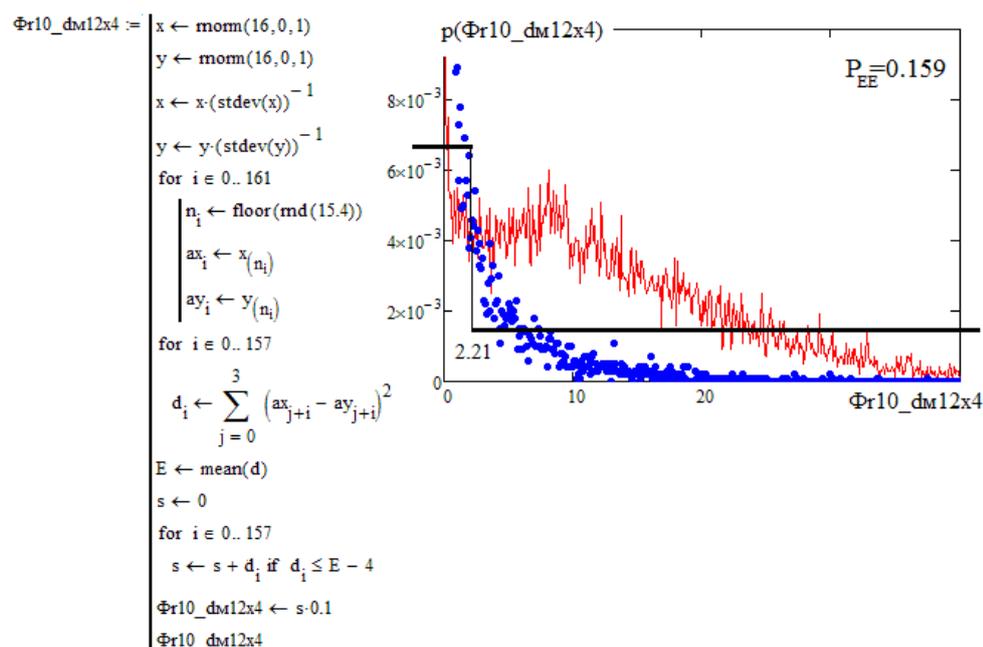


Рис. 12.12. Непрерывный фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через 10-кратное увеличение псевдовыборки случайными перестановками точек и учетом статистик левого «хвоста» хи-квадрат распределения

Численный эксперимент по моделированию этого критерия показал снижение уровня вероятности ошибок примерно на 14.5 % по сравнению с критерием-родителем. Однако между откликами критерия-потомка и критерия-родителя высок уровень корреляционной сцепленности: $\text{corr}(\Phi_{r10_dm12x4}, \Phi_{r10_dm12x4}) \approx 0.891$.

12.13. Непрерывный вариант фрактального функционала, оценивающего корреляционную сцепленность данных малой выборки через использование 10-кратного увеличения выборки, скользящей хи-квадрат суммы с четырьмя элементами и аддитивным учетом статистик двух «хвостов»

Если попытаться аддитивно свернуть статистики правого и левого «хвостов» хи-квадрат распределений промежуточных вычислений, то мы получим еще один вариант нейрокритерия для проверки гипотезы независимости. Программная реализация этого варианта критерия приведена на рис. 12.13.

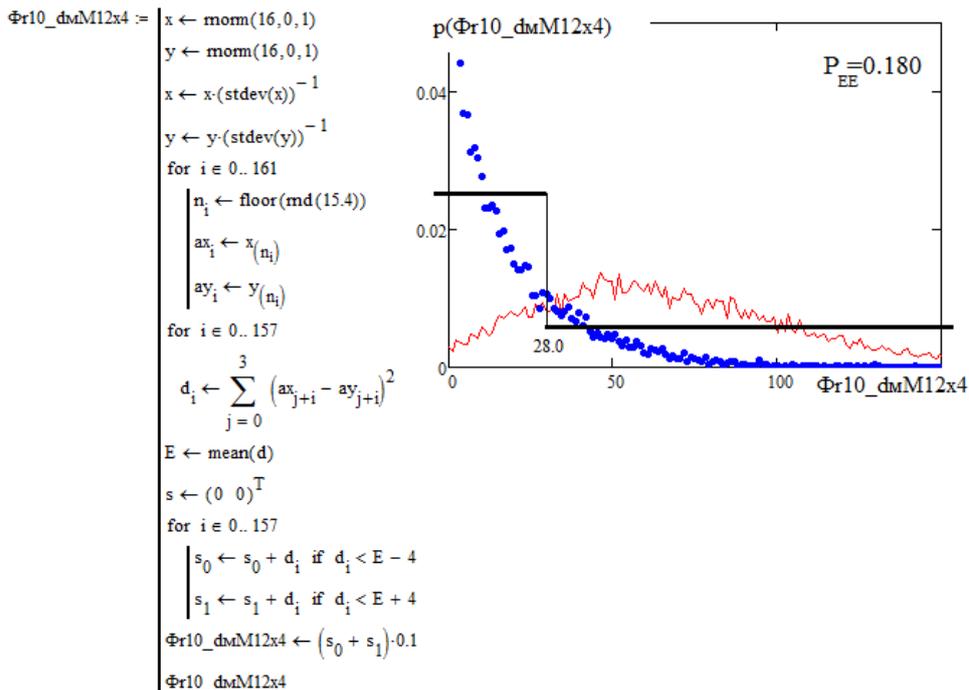


Рис. 12.13. Непрерывный фрактальный функционал, оценивающий корреляционную сцепленность данных малой выборки через 10-кратное увеличение псевдовыборки случайными перестановками точек и учетом статистик двух «хвостов» хи-квадрат распределения

К сожалению, ощутимый рост качества принимаемых решений критерием-потомком не наблюдается по сравнению с критериями-родителями.

Необходимость проверки различных вариантов похожих критериев обусловлена наличием провалов в качестве принимаемых ими решений. Так ряд критериев-потомков (GrI_II_III_IVR, GrI_II_III_IVG, ФI_IV, ФSI_IV, ...), построенных на аддитивном свертывании промежуточных данных, дает существенное снижение вероятностей появления ошибок по сравнению с критериями-родителями. Однако в нашем случае этого эффекта не наблюдается.

12.14. Корреляционная сцепленность фрактальных функционалов, оценивающих статистики ломаных линий, соединяющих точки малой выборки

Объединение новых критериев между собой в группы следует выполнять с учетом их корреляционных связей. Коэффициенты корреляции рассмотренных в данной главе нейрокритериев представлены на рис. 12.14.

$\text{corr}(r, \Phi r_{\sigma_R}) = -0.025$		
$\text{corr}(r, \Phi r_{E_R5}) = 0.020$	$\text{corr}(\Phi r_{\sigma_R}, \Phi r_{E_R5}) = -0.309$	
$\text{corr}(r, \Phi r_{MR}) = -0.029$	$\text{corr}(\Phi r_{\sigma_R}, \Phi r_{MR}) = 0.818$	$\text{corr}(\Phi r_{E_R5}, \Phi r_{MR}) = -0.203$
$\text{corr}(r, \Phi r_{R5\text{д}\sigma}) = 0.024$	$\text{corr}(\Phi r_{\sigma_R}, \Phi r_{R5\text{д}\sigma}) = -0.862$	$\text{corr}(\Phi r_{E_R5}, \Phi r_{R5\text{д}\sigma}) = 0.732$
$\text{corr}(r, \Phi r_{DR}) = 0.004$	$\text{corr}(\Phi r_{\sigma_R}, \Phi r_{DR}) = -0.17025$	$\text{corr}(\Phi r_{E_R5}, \Phi r_{DR}) = 0.043$
$\text{corr}(r, \Phi r_{10_M12x4}) = -0.343$	$\text{corr}(\Phi r_{\sigma_R}, \Phi r_{10_M12x4}) = 0.019$	$\text{corr}(\Phi r_{E_R5}, \Phi r_{10_M12x4}) = -0.006$
$\text{corr}(r, \Phi r_{10_m12x4}) = -0.3439$	$\text{corr}(\Phi r_{\sigma_R}, \Phi r_{10_m12x4}) = 0.019$	$\text{corr}(\Phi r_{E_R5}, \Phi r_{10_m12x4}) = -0.006$
$\text{corr}(r, \text{д}\Phi r_{10_M12x4}) = -0.608$	$\text{corr}(\Phi r_{\sigma_R}, \text{д}\Phi r_{10_M12x4}) = 0.113$	$\text{corr}(\Phi r_{E_R5}, \text{д}\Phi r_{10_M12x4}) = 0.014$
$\text{corr}(r, \text{д}\Phi r_{10_m12x4}) = -0.626$	$\text{corr}(\Phi r_{\sigma_R}, \text{д}\Phi r_{10_m12x4}) = 0.201$	$\text{corr}(\Phi r_{E_R5}, \text{д}\Phi r_{10_m12x4}) = 0.022$
$\text{corr}(r, \text{д}\Phi r_{10_mM12x4}) = -0.643$	$\text{corr}(\Phi r_{\sigma_R}, \text{д}\Phi r_{10_mM12x4}) = 0.176$	$\text{corr}(\Phi r_{E_R5}, \text{д}\Phi r_{10_mM12x4}) = 0.0204$
$\text{corr}(r, \text{д}\Phi r_{10_dM12x4}) = -0.702$	$\text{corr}(\Phi r_{\sigma_R}, \text{д}\Phi r_{10_dM12x4}) = 0.223$	$\text{corr}(\Phi r_{E_R5}, \text{д}\Phi r_{10_dM12x4}) = -0.041$
$\text{corr}(r, \text{д}\Phi r_{10_dm12x4}) = -0.735$	$\text{corr}(\Phi r_{\sigma_R}, \text{д}\Phi r_{10_dm12x4}) = 0.179$	$\text{corr}(\Phi r_{E_R5}, \text{д}\Phi r_{10_dm12x4}) = -0.075$
$\text{corr}(r, \text{д}\Phi r_{10_dmM12x4}) = -0.725$	$\text{corr}(\Phi r_{\sigma_R}, \text{д}\Phi r_{10_dmM12x4}) = 0.216$	$\text{corr}(\Phi r_{E_R5}, \text{д}\Phi r_{10_dmM12x4}) = -0.050$

Рис. 12.14. Коэффициенты корреляции фрактальных функционалов, полученные оценкой ломаной линии, соединяющей точки малой выборки (начало)

Из данных верхней части первого столбца на рис. 12.14 следует целесообразность объединения простейших фрактальных критериев с наиболее популярным классическим критерием Эджуорта – Эдлтона – Пирсона. Однако усложнение фрактальных критериев (нижняя часть первого столбца) приводит к существенному росту уровня корреляционной сцепленности. Это делает объединение критериев в подобные группы нецелесообразным.

Второй и третий столбцы на рис. 12.14 свидетельствуют о низком уровне корреляционной сцепленности между простейшими фрактальными критериями.

$\text{corr}(\Phi_{r_MR}, \Phi_{r_R5д\sigma}) = -0.686$		
$\text{corr}(\Phi_{r_MR}, \Phi_{r_DR}) = -0.124$	$\text{corr}(\Phi_{r_R5д\sigma}, \Phi_{r_DR}) = 0.140$	
$\text{corr}(\Phi_{r_MR}, \Phi_{r10_M12x4}) = 0.023$	$\text{corr}(\Phi_{r_R5д\sigma}, \Phi_{r10_M12x4}) = -0.017$	$\text{corr}(\Phi_{r_DR}, \Phi_{r10_M12x4}) = -0.016$
$\text{corr}(\Phi_{r_MR}, \Phi_{r10_m12x4}) = 0.023$	$\text{corr}(\Phi_{r_R5д\sigma}, \Phi_{r10_m12x4}) = -0.017$	$\text{corr}(\Phi_{r_DR}, \Phi_{r10_m12x4}) = -0.016$
$\text{corr}(\Phi_{r_MR}, д\Phi_{r10_M12x4}) = 0.101$	$\text{corr}(\Phi_{r_R5д\sigma}, д\Phi_{r10_M12x4}) = -0.082$	$\text{corr}(\Phi_{r_DR}, д\Phi_{r10_M12x4}) = -0.023$
$\text{corr}(\Phi_{r_MR}, д\Phi_{r10_m12x4}) = 0.177$	$\text{corr}(\Phi_{r_R5д\sigma}, д\Phi_{r10_m12x4}) = -0.134$	$\text{corr}(\Phi_{r_DR}, д\Phi_{r10_m12x4}) = -0.048$
$\text{corr}(\Phi_{r_MR}, д\Phi_{r10_mM12x4}) = 0.155$	$\text{corr}(\Phi_{r_R5д\sigma}, д\Phi_{r10_mM12x4}) = -0.119$	$\text{corr}(\Phi_{r_DR}, д\Phi_{r10_mM12x4}) = -0.040$
$\text{corr}(\Phi_{r_MR}, д\Phi_{r10_dM12x4}) = 0.201$	$\text{corr}(\Phi_{r_R5д\sigma}, д\Phi_{r10_dM12x4}) = -0.179$	$\text{corr}(\Phi_{r_DR}, д\Phi_{r10_dM12x4}) = -0.051$
$\text{corr}(\Phi_{r_MR}, д\Phi_{r10_dm12x4}) = 0.168$	$\text{corr}(\Phi_{r_R5д\sigma}, д\Phi_{r10_dm12x4}) = -0.161$	$\text{corr}(\Phi_{r_DR}, д\Phi_{r10_dm12x4}) = -0.041$
$\text{corr}(\Phi_{r_MR}, д\Phi_{r10_dmM12x4}) = 0.197$	$\text{corr}(\Phi_{r_R5д\sigma}, д\Phi_{r10_dmM12x4}) = -0.178$	$\text{corr}(\Phi_{r_DR}, д\Phi_{r10_dmM12x4}) = -0.05$

Рис. 12.14. Продолжение

Все столбцы продолжения рис. 12.14 свидетельствуют о том, что для большинства фрактальных критериев сохраняется низкий уровень взаимной корреляционной сцепленности.

$\text{corr}(\Phi_{r10_M12x4}, \Phi_{r10_m12x4}) = 1$		
$\text{corr}(\Phi_{r10_M12x4}, д\Phi_{r10_M12x4}) = 0.267$	$\text{corr}(\Phi_{r10_m12x4}, д\Phi_{r10_M12x4}) = 0.267$	
$\text{corr}(\Phi_{r10_M12x4}, д\Phi_{r10_m12x4}) = 0.276$	$\text{corr}(\Phi_{r10_m12x4}, д\Phi_{r10_m12x4}) = 0.276$	$\text{corr}(д\Phi_{r10_M12x4}, д\Phi_{r10_m12x4}) = 0.845$
$\text{corr}(\Phi_{r10_M12x4}, д\Phi_{r10_mM12x4}) = 0.283$	$\text{corr}(\Phi_{r10_m12x4}, д\Phi_{r10_mM12x4}) = 0.283$	$\text{corr}(д\Phi_{r10_M12x4}, д\Phi_{r10_mM12x4}) = 0.934$
$\text{corr}(\Phi_{r10_M12x4}, д\Phi_{r10_dM12x4}) = 0.327$	$\text{corr}(\Phi_{r10_m12x4}, д\Phi_{r10_dM12x4}) = 0.327$	$\text{corr}(д\Phi_{r10_M12x4}, д\Phi_{r10_dM12x4}) = 0.916$
$\text{corr}(\Phi_{r10_M12x4}, д\Phi_{r10_dm12x4}) = 0.347$	$\text{corr}(\Phi_{r10_m12x4}, д\Phi_{r10_dm12x4}) = 0.347$	$\text{corr}(д\Phi_{r10_M12x4}, д\Phi_{r10_dm12x4}) = 0.729$
$\text{corr}(\Phi_{r10_M12x4}, д\Phi_{r10_dmM12x4}) = 0.339$	$\text{corr}(\Phi_{r10_m12x4}, д\Phi_{r10_dmM12x4}) = 0.339$	$\text{corr}(д\Phi_{r10_M12x4}, д\Phi_{r10_dmM12x4}) = 0.887$
$\text{corr}(д\Phi_{r10_m12x4}, д\Phi_{r10_mM12x4}) = 0.980$		
$\text{corr}(д\Phi_{r10_m12x4}, д\Phi_{r10_dM12x4}) = 0.892$	$\text{corr}(д\Phi_{r10_mM12x4}, д\Phi_{r10_dM12x4}) = 0.935$	
$\text{corr}(д\Phi_{r10_m12x4}, д\Phi_{r10_dm12x4}) = 0.816$	$\text{corr}(д\Phi_{r10_mM12x4}, д\Phi_{r10_dm12x4}) = 0.815$	$\text{corr}(д\Phi_{r10_dM12x4}, д\Phi_{r10_dm12x4}) = 0.894$
$\text{corr}(д\Phi_{r10_m12x4}, д\Phi_{r10_dmM12x4}) = 0.891$	$\text{corr}(д\Phi_{r10_mM12x4}, д\Phi_{r10_dmM12x4}) = 0.923$	$\text{corr}(д\Phi_{r10_dM12x4}, д\Phi_{r10_dmM12x4}) = 0.993$

Рис. 12.14. Окончание

К сожалению, с усложнением фрактальных критериев уровень взаимной корреляционной сцепленности повышается.

Список литературы к главе 12

1. Мандельброт Б., Хадсон З. Л. (НЕ)послушные рынки. Фрактальная революция в финансах. М. ; СПб. ; Киев : Вильямс, 2006. 408 с.

2. Гильмутдинов А. Х. Фракталы и дробные операторы. Казань : Изд-во «Фэн» Академии наук РТ, 2010. 488 с.

3. Ахметов Б. Б., Иванов А. И. Многомерные статистики существенно зависимых биометрических данных, порождаемые нейросетевыми эмуляторами квадратичных форм : монография. Алматы : ЛЕМ, 2016. 86 с.

4. Иванов А. И., Полковникова С. А. Хи-квадрат-симметризация корреляционных связей, ориентированная на одиночные нейроны и их обучение на малых выборках // Информационно-управляющие, телекоммуникационные системы, средства поражения и их техническое обеспечение : сб. науч. ст. IV Всерос. межведомственной науч.-техн. конф. Пенза, 2022. С. 131–141.

КРИТЕРИИ, ПОСТРОЕННЫЕ НА ПОПАРНОМ СВЕРТЫВАНИИ ЛИБО СВЕРТЫВАНИИ ТРОЕК ДАННЫХ ПОСЛЕДОВАТЕЛЬНОСТИ ТОЧЕК МАЛОЙ ВЫБОРКИ ПРИ ПРОВЕРКЕ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ

13.1. Критерий накопления модулей коэффициентов ассоциаций последовательности двух точек малых выборок

В случае, когда можно выбрать из данных малой выборки в форме минимальной связанности в матрице 2×2 , мы можем воспользоваться коэффициентами ассоциаций [38], т.е. мы можем двигаться по парам отсчетов точек выборки, каждый раз вычисляя для них коэффициенты ассоциаций. Пример реализации программы для такого ассоциативного критерия приведен на рис. 13.1.

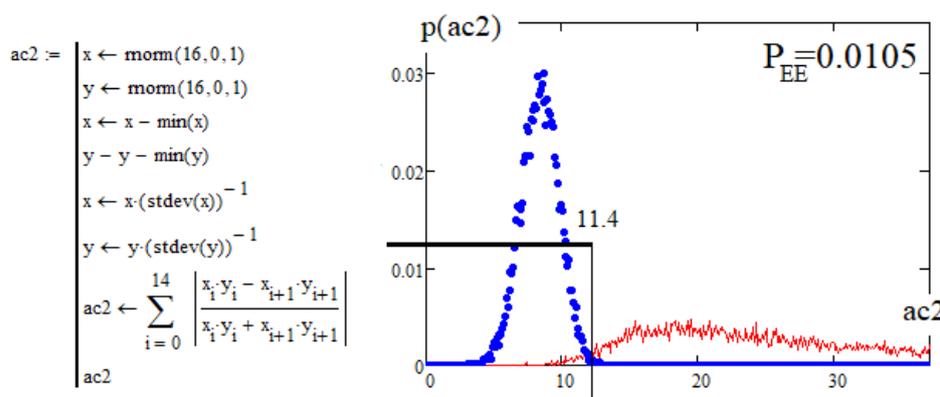


Рис. 13.1. Один из наиболее сильных статистических критериев для проверки гипотезы независимости данных малой выборки

Численный эксперимент по оценке состояний нового статистического критерия показывает достижение вероятности ошибок первого и второго рода на уровне $P_{EE} \approx 0.0105$, что вполне приемлемо для многих практически важных приложений. Столь высокий показатель выше показателя наиболее распространенного классического критерия Эджуорта – Эдлтона – Пирсона примерно в 14 раз.

13.2. Критерий накапливания модулей коэффициентов ассоциаций последовательности трех точек малых выборок

Очевидно, что по аналогии с классическим коэффициентом ассоциаций размерности $[2 \times 2]$ можно попытаться повысить размерность задачи до $[2 \times 3]$. Пример критерия, построенного на накоплении модулей коэффициентов ассоциаций $[2 \times 3]$, приведен на рис. 13.2.

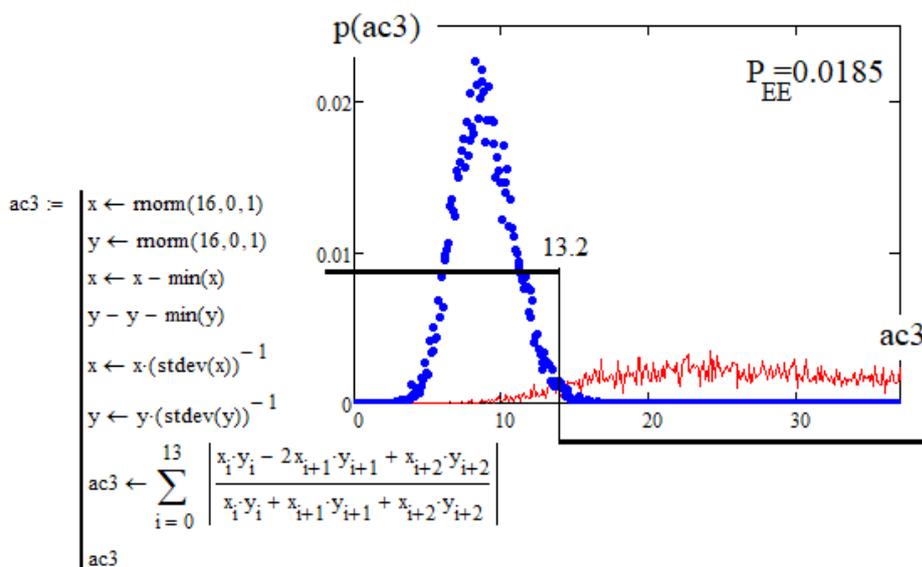


Рис. 13.2. Критерий для проверки гипотезы независимости данных малой выборки, построенный на анализе коэффициентов ассоциаций трех рядом расположенных точек

Численный эксперимент показал, что у нового статистического критерия сохраняется низкий уровень вероятности ошибок первого и второго рода при разделении зависимых и независимых данных: $P_{EE} \approx 0.0185$. Это примерно в два раза хуже предыдущего критерия, однако это не препятствует совместному использованию двух последних критериев, так как их отклики слабо коррелированы: $\text{corr}(ac2, ac3) \approx 0.351$.

13.3. Критерий накапливания модулей коэффициентов ассоциаций соседних точек малых выборок без нормирования данных при промежуточных вычислениях

Представляют интерес причины, по которым рассматриваемый класс критериев обладает высокой разделяющей способностью.

Для выяснения причин преимуществ устраним процедуру деления при вычислениях. Программная реализация этого типа критериев приведена на рис. 13.3.

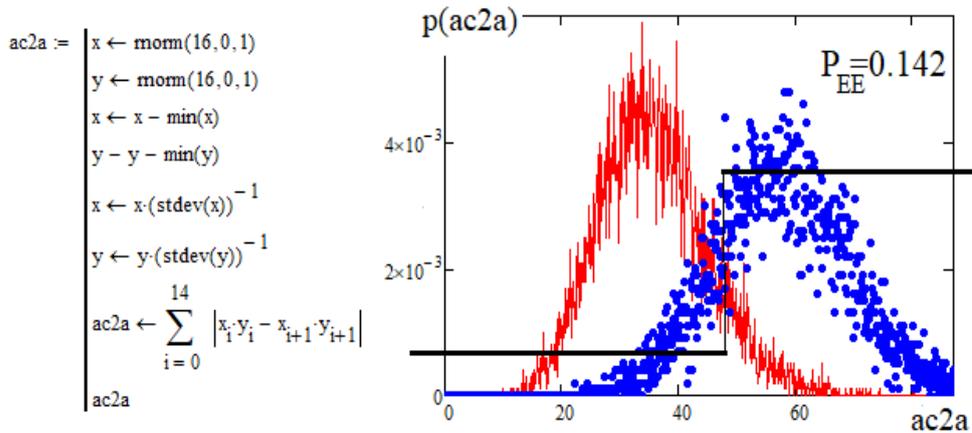


Рис. 13.3. Критерий для проверки гипотезы независимости данных малой выборки, построенный на анализе ассоциаций пары точек без нормирования данных

Численный эксперимент показал, что устранение знаменателя при вычислениях снижает качество принимаемых решений до качества классического критерия Эджуорта – Эдлттона – Пирсона (конца XVIII в.), т.е. именно нормирование данных при вычислениях классического критерия коэффициентов ассоциаций дает 14-кратное увеличение качества принятия нейросетевых решений.

Положительным является то, что новый и классический критерии имеют слабо коррелированные отклики: $\text{corr}(r, \text{ac2a}) \approx 0.023$.

13.4. Критерий объединения модулей коэффициентов коллигации Юла двух соседних точек вдоль малой выборки

Параллельно с коэффициентом ассоциаций в XX в. активно использовался коэффициент коллигаций Юла [38]. Замена коэффициентов ассоциаций критерия ac2 на коэффициент коллигации Юла дает еще один статистический критерий. Программная реализация нового критерия приведена на рис. 13.4.

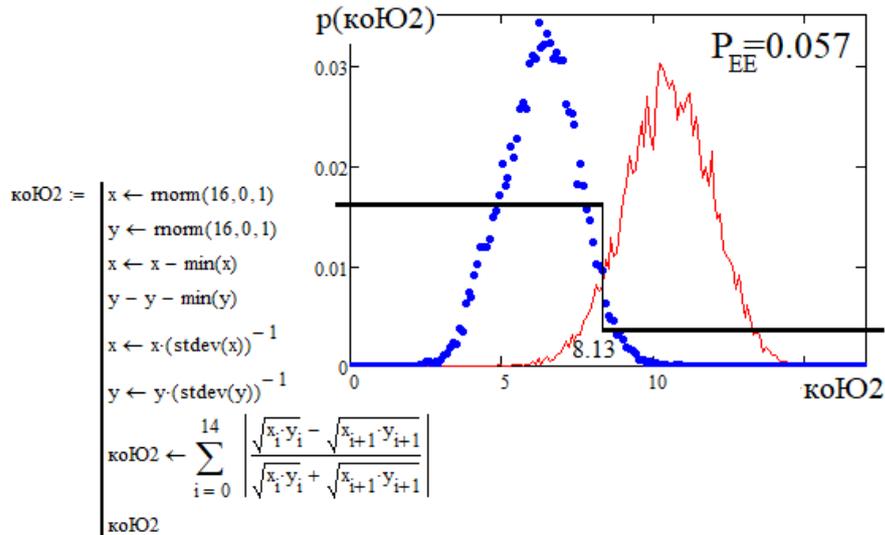


Рис. 13.4. Критерий для проверки гипотезы независимости данных малой выборки, построенный на анализе коэффициентов коллигаций Юла

Высокая разделяющая способность нового критерия для зависимых и независимых данных сохраняется. При этом корреляционная сцепленность двух критериев оказывается низкой: $\text{corr}(\text{ac2}, \text{коЮ2}) \approx 0.028$.

13.5. Критерий объединения коэффициентов сходства пары точек малых выборок

Третьим классическим критерием парного анализа точек размерности $[2 \times 2]$ является коэффициент контингенции (сходства) [38]. Накопление модулей коэффициентов контингенции дает еще один критерий для проверки гипотезы независимости данных малой выборки. Пример программной реализации нового критерия приведен на рис. 13.5.

Численный эксперимент по моделированию выходных состояний нового критерия показал, что ошибки первого и второго рода для нового критерия на уровне $P_{EE} \approx 0.0108$. Такой уровень вероятности ошибок примерно в 14 раз хуже в сравнении с классическим критерием Эджуорта – Эдлтона – Пирсона. При этом корреляционная сцепленность классического критерия и нового критерия оказывается низкой: $\text{corr}(r, \text{cx2}) \approx 0.008$.

Низкой оказывается также корреляционная сцепленность рассматриваемого критерия с двумя почти классическими критериями: $\text{corr}(\text{cx2}, \text{ac2}) \approx 0.009$, $\text{corr}(\text{cx2}, \text{коЮ2}) \approx 0.098$.

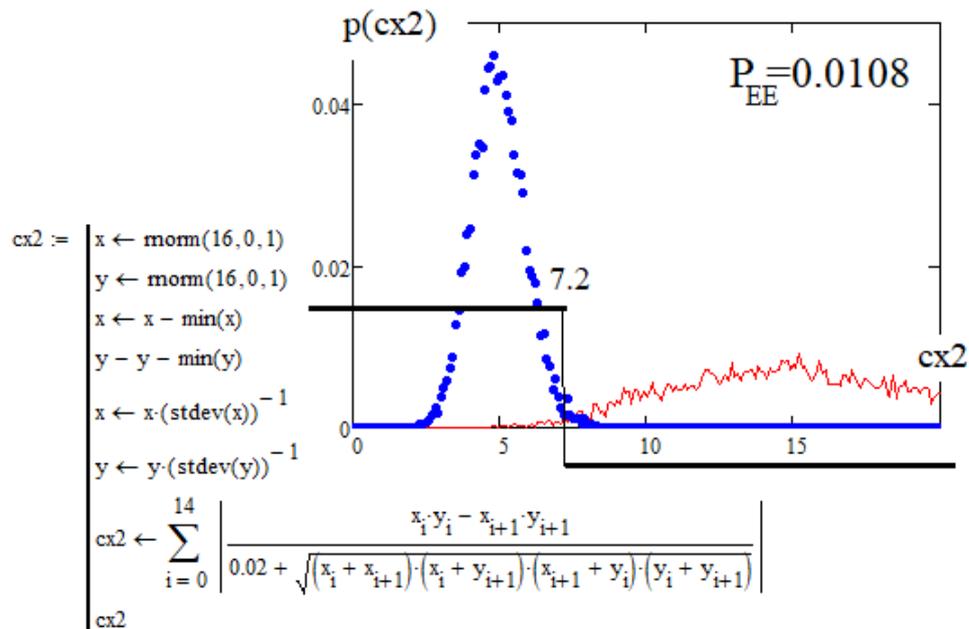


Рис. 13.5. Критерий для проверки гипотезы независимости данных малой выборки, построенный на анализе сходства

13.6. Критерий объединения коэффициентов ассоциации соседних точек малых выборок (перестановка данных в вычисляемых произведениях)

При синтезе статистических критериев в прошлом веке исследователи стремились увеличить качество принимаемых ими решений. Предполагалось использовать, синтезируемые критерии отдельно. В прошлом веке была далеко не очевидна возможность нейросетевого объединения множества уже созданных статистических критериев.

В этом веке стала очевидной возможность перехода от любого статистического критерия к эквивалентных им нейронам [27, 29], т.е. нет необходимости искать один мощный статистический критерий, который в одиночку способен решать все специфические задачи. Гораздо разумнее преобразовывать множество достаточно слабых искусственных нейронов и объединять их в одну нейросеть. В этом случае качество решений нейросети зависит только от числа объединяемых ею искусственных нейронов.

В этом плане в XXI в. необходимо пытаться создавать как можно больше статистических критериев. При этом желательно, чтобы новые критерии были либо независимы, либо хотя бы имели некоторую независимую компоненту по отношению к уже созданным критериям.

В рассматриваемом нами случае получить новые варианты статистических критериев удастся простой перестановкой перемножаемых компонент при вычислении последовательности коэффициентов ассоциаций. Программная реализация нового критерия приведена на рис. 13.6.

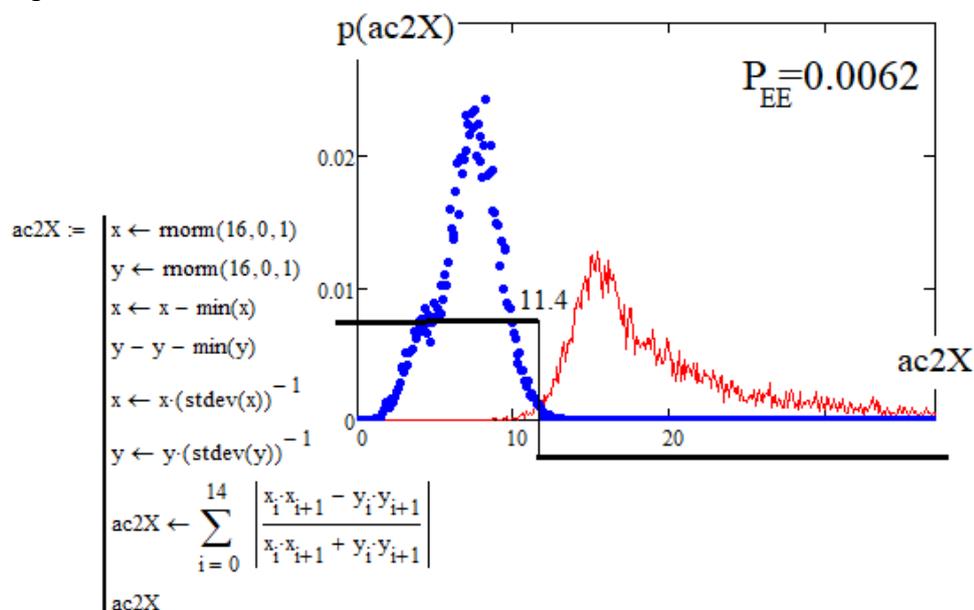


Рис. 13.6. Еще один вариант критерия для проверки гипотезы независимости данных малой выборки, построенный на коэффициентах ассоциации с учетом перестановки компонент между произведениями

Главным свойством коэффициентов ассоциаций является то, что в числителе дроби содержится разность произведений координат двух точек, а в знаменателе дроби содержится сумма произведений числителя. Принципиально важным является то, что уровень вероятности ошибок первого и второго рода нового критерия снижен по отношению к критерию-родителю. Вторым принципиально важным обстоятельством является низкий уровень корреляционной сцепленности нового и предшествующего критерия: $\text{corr}(\text{ac2}, \text{ac2X}) \approx -0.001$.

13.7. Вторая перестановка координат в произведениях при синтезе нового критерия, объединяющего коэффициенты ассоциаций координат пары точек

Возможен еще один вариант нового критерия – ac2XX , синтезированный еще одним (вторым) вариантом перестановок компонент

двух произведений. Программная реализация нового критерия приведена на рис. 13.7. Численный эксперимент показал, что по сравнению с предшествующим критерием мы наблюдаем снижение вероятностей ошибок на 5.1 %. При этом низкой оказывается корреляционная сцепленность нового критерия с близкими ему аналогами:

$$\text{corr}(\text{ac2X}, \text{ac2XX}) \approx -0.0002, \text{corr}(\text{ac2}, \text{ac2XX}) \approx -0.0007.$$

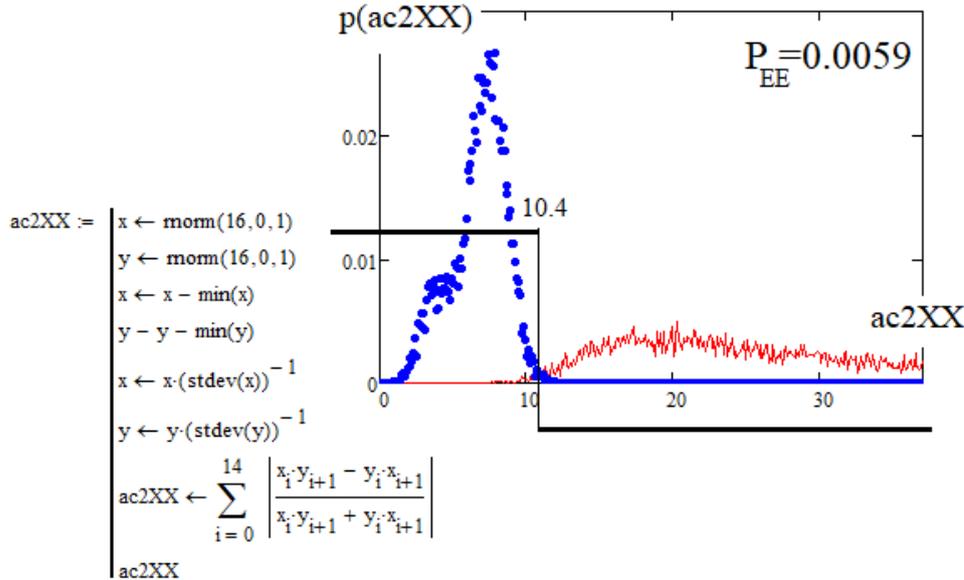


Рис. 13.7. Критерий для проверки гипотезы независимости данных малой выборки, построенный на втором варианте вычисления коэффициента ассоциации соседних точек

Очевидным преимуществом рассматриваемой группы критериев является относительная простота получения нескольких вариантов разных критериев, обладающих похожими свойствами и даже выполняющих решение исходной задачи с более высоким качеством.

13.8. Модификация критерия сборки коэффициентов Юла для проверки гипотезы независимости малых выборок

По аналогии с синтезом двух предшествующих критериев коэффициентов ассоциации могут быть построены два новых варианта критериев, построенных на использовании коэффициентов коллигации Юла [38]. Первая модификация строится перемножением одинаковых координат двух точек малой выборки. Программная реализация нового критерия приведена на рис. 13.8.

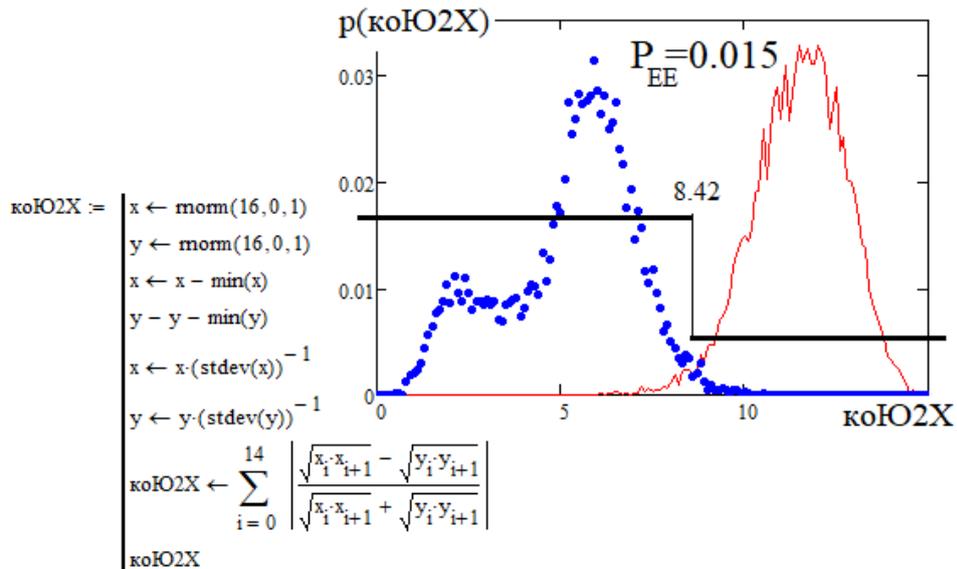


Рис. 13.8. Критерий для проверки гипотезы независимости данных малой выборки, построенный на модификации критерия коллигаций Юла

Численный эксперимент по моделированию нового критерия показал сохранение высокого уровня разделимости зависимых и независимых данных малой выборки в 16 опытов. Уровень вероятности ошибок первого и второго рода для исследуемого критерия составляет $P_{EE} \approx 0.015$. Этот уровень примерно в 10 раз ниже по сравнению с наиболее распространенным критерием Эджуорта – Эдлтона – Пирсона. Мы наблюдаем значительный рост качества принимаемых группой новых критериев решений по сравнению с группой классических статистических критериев.

13.9. Вторая модификация критерия сборки коэффициентов Юла для проверки гипотезы независимости малых выборок

Вторая модификация критерия коэффициентов Юла получается перемножением разных координат (координаты x и координаты y) двух разных точек (рис. 13.9). К сожалению, этот критерий дает рост вероятности появления ошибок примерно на 53 % по сравнению с предыдущим критерием. Эта особенность характерна для разделения положительно коррелированных данных и независимых данных. При смене знака коррелированности отделяемых данных уровень вероятностей появления ошибок изменяется. Происходит обмен значениями вероятностей ошибок между рассматриваемыми критериями.

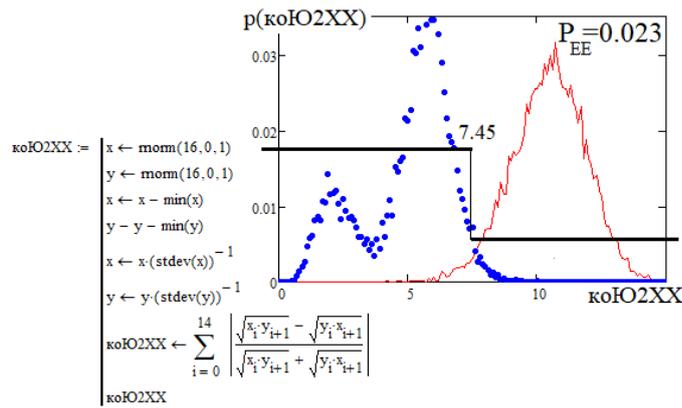


Рис. 13.9. Критерий для проверки гипотезы независимости данных малой выборки, построенный на второй модификации критерия коллигаций Юла через вторую перестановку перемножаемых координат

13.10. Модификация критерия, обобщающего коэффициенты сходства для проверки гипотезы независимости малых выборок

Формально коэффициенты сходства [38] являются аналогами коэффициентов ассоциаций и коэффициентов коллигации Юла, т.е. мы имеем право синтезировать еще две модификации новых критериев для коэффициентов сходства. Программная реализация нового критерия приведена на рис. 13.10. Новый критерий дает снижение вероятностей ошибок первого и второго рода примерно в 3.5 раза, или на 34,3 %.

При столь значительном снижении вероятностей ошибок корреляционная сцепленность откликов нового критерия с похожими критериями имеет низкий уровень:

$$\text{corr}(cx2, cx2X) \approx 0.0007, \text{corr}(cx2X, \text{коЮ2XX}) \approx 0.094.$$

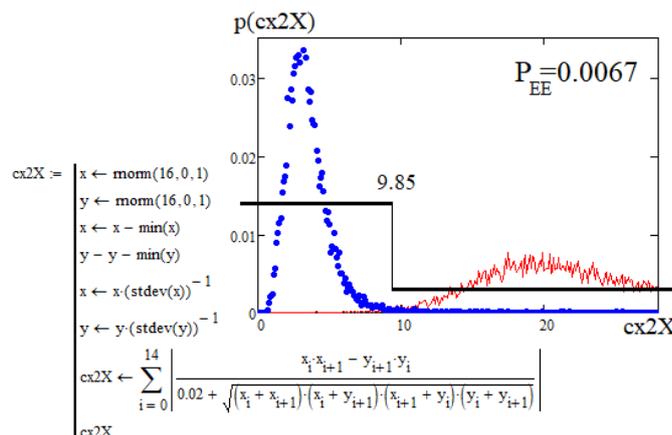


Рис. 13.10. Критерий для проверки гипотезы независимости данных малой выборки, построенный на модификации критерия сходства

При этом коэффициенты сходства все-таки имеют свою особенность. В отличие от коэффициентов ассоциаций и коллигации, числитель их дроби всегда совпадает для всех модификаций этого критерия. Модификации критериев, построенных на связывании коэффициентов сходства, будут отличаться только знаменателями вычисляемых дробей.

13.11. Вторая модификация критерия сходства для проверки гипотезы независимости малых выборок через использование еще одной перестановки перемножаемых координат соседних точек

К сожалению, следующий вариант рассматриваемой группы критериев имеет значительное повышение вероятности ошибок первого и второго рода. Пример его программной реализации приведен на рис. 13.11.

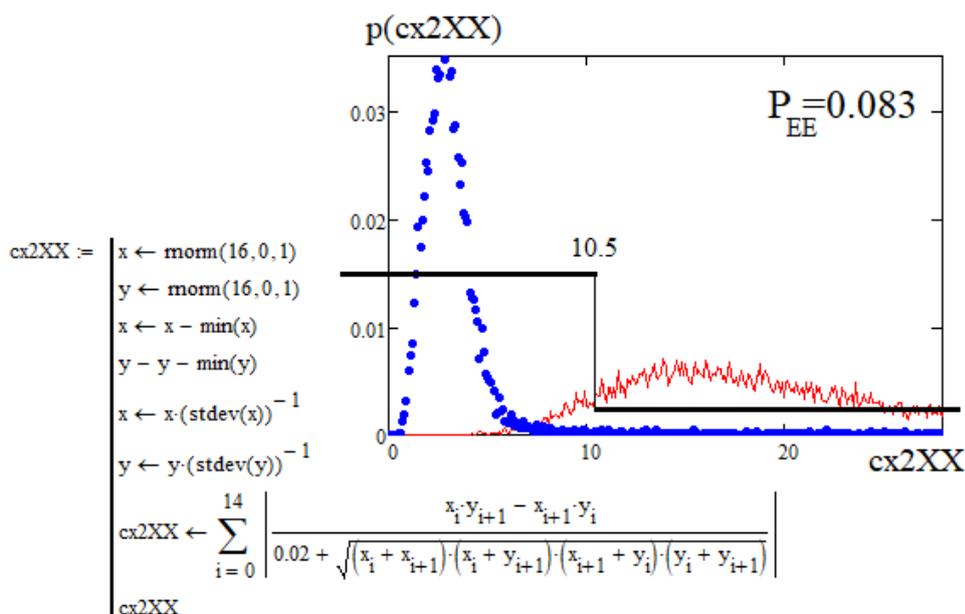


Рис. 13.11. Вторая модификация критерия для проверки гипотезы независимости данных малой выборки, построенная на еще одной перестановке перемножаемых координат соседних точек

Еще одним негативным моментом этой математической конструкции является ее высокая корреляционная сцепленность с откликами предыдущего критерия при выделении положительно коррелированных данных: $\text{corr}(cx2X, cx2XX) \approx 0.904$.

13.12. Критерий, построенный на использовании вычисления коэффициентов ассоциаций двух соседних точек как процедуры предварительной обработки данных с последующим применением ассоциативного разделения независимых и зависимых данных

В этом справочнике были использованы десятки типовых преобразований, примененных исследователями прошлого века при синтезе разных статистических критериев. Общая идея проста. Если несколькими авторами был использован тот или иной математический прием, то, скорее всего, его можно применить и для синтеза других статистических критериев.

В данной главе были использованы свертки координат двух точек размерности $[2 \times 2]$. При этом получились новые нейрокритерии с высокой разделяющей способностью. Однако примененные в данной главе преобразования не являются универсальными. Это становится очевидным, если попытаться вкладывать друг в друга матричные преобразования.

Мы можем рассматривать эти преобразования как достаточно эффективную входную предобработку данных. Если преобразования достаточно универсальны, то второй уровень подобных преобразований должен приводить к положительному результату. Программное обеспечение примера критерия с двухуровневой обработкой данных приведено на рис. 13.12.

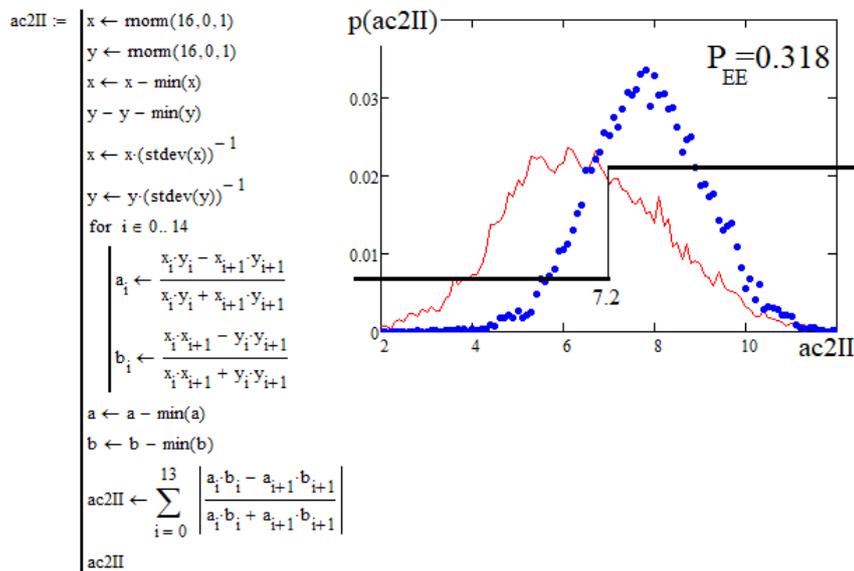


Рис. 13.12. Использование вычисления коэффициентов ассоциаций двух соседних точек как процедуры предварительной обработки данных (ассоциативные вычисления в двух уровнях)

К сожалению, мы наблюдаем «катастрофическое» повышение уровня ошибок первого и второго рода для нового статистического критерия, т.е. вкладывать друг в друга рассматриваемые преобразования нецелесообразно.

13.13. Корреляционная сцепленность группы критериев, построенных на свертывании координат двух соседних точек малой выборки

Объединять нейрокритерии в один фрагмент нейросети целесообразно, только если каждый из нейрокритериев имеет вероятности ошибок первого и второго рода, близкие к его соседям. Кроме того, желательно, чтобы нейрокритерии одной группы имели низкие по модулю корреляционные связи.

На рис. 13.13 даны коэффициенты корреляции между парами нейрокритериев, построенных на двухмерном свертывании координат двух точек.

$\text{corr}(r, \text{ac}2) = 0.004$		
$\text{corr}(r, \text{ac}3) = -0.004$	$\text{corr}(\text{ac}2, \text{ac}3) = 0.351$	
$\text{corr}(r, \text{ac}2a) = 0.023$	$\text{corr}(\text{ac}2, \text{ac}2a) = 0.033$	$\text{corr}(\text{ac}3, \text{ac}2a) = 0.012$
$\text{corr}(r, \text{коЮ}2) = 0.008$	$\text{corr}(\text{ac}2, \text{коЮ}2) = 0.028$	$\text{corr}(\text{ac}3, \text{коЮ}2) = 0.009$
$\text{corr}(r, \text{cx}2) = -0.017$	$\text{corr}(\text{ac}2, \text{cx}2) = 0.009$	$\text{corr}(\text{ac}3, \text{cx}2) = -0.001$
$\text{corr}(r, \text{ac}2X) = -0.003$	$\text{corr}(\text{ac}2, \text{ac}2X) = -0.001$	$\text{corr}(\text{ac}3, \text{ac}2X) = -0.0009$
$\text{corr}(r, \text{ac}2XX) = -0.002$	$\text{corr}(\text{ac}2, \text{ac}2XX) = -0.0007$	$\text{corr}(\text{ac}3, \text{ac}2XX) = -0.0009$
$\text{corr}(r, \text{коЮ}2X) = 0.001$	$\text{corr}(\text{ac}2, \text{коЮ}2X) = 0.036$	$\text{corr}(\text{ac}3, \text{коЮ}2X) = 0.0217$
$\text{corr}(r, \text{коЮ}2XX) = -0.004$	$\text{corr}(\text{ac}2, \text{коЮ}2XX) = 0.030$	$\text{corr}(\text{ac}3, \text{коЮ}2XX) = 0.016$
$\text{corr}(r, \text{cx}2X) = -0.022$	$\text{corr}(\text{ac}2, \text{cx}2X) = 0.015$	$\text{corr}(\text{ac}3, \text{cx}2X) = 0.001$
$\text{corr}(r, \text{cx}2XX) = 0.010$	$\text{corr}(\text{ac}2, \text{cx}2XX) = 0.008$	$\text{corr}(\text{ac}3, \text{cx}2XX) = -0.002$
$\text{corr}(r, \text{ac}2\Pi) = -0.011$	$\text{corr}(\text{ac}2, \text{ac}2\Pi) = -0.010$	$\text{corr}(\text{ac}3, \text{ac}2\Pi) = -0.001$

Рис. 13.13. Коэффициенты корреляции критериев, оценивающих связи двух и трех соседних точек малой выборки (начало)

Из данных первого столбца на рис. 13.13 следует низкий уровень корреляционных связей между классическими критериями и рассматриваемой группой новых нейрокритериев. Второй и третий столбцы подтверждают низкий уровень модулей коэффициентов корреляции между рассматриваемыми критериями и наиболее часто используемым критерием Эджуорта – Эдлтона – Пирсона.

$\text{corr}(\text{ac2a}, \text{коЮ2}) = 0.298$		
$\text{corr}(\text{ac2a}, \text{cx2}) = 0.023$	$\text{corr}(\text{коЮ2}, \text{cx2}) = 0.098$	
$\text{corr}(\text{ac2a}, \text{ac2X}) = -0.016$	$\text{corr}(\text{коЮ2}, \text{ac2X}) = -0.006$	$\text{corr}(\text{cx2}, \text{ac2X}) = 0.0007$
$\text{corr}(\text{ac2a}, \text{ac2XX}) = -0.009$	$\text{corr}(\text{коЮ2}, \text{ac2XX}) = 0.002$	$\text{corr}(\text{cx2}, \text{ac2XX}) = -0.0008$
$\text{corr}(\text{ac2a}, \text{коЮ2X}) = 0.426$	$\text{corr}(\text{коЮ2}, \text{коЮ2X}) = 0.755$	$\text{corr}(\text{cx2}, \text{коЮ2X}) = 0.0516$
$\text{corr}(\text{ac2a}, \text{коЮ2XX}) = 0.156$	$\text{corr}(\text{коЮ2}, \text{коЮ2XX}) = 0.889$	$\text{corr}(\text{cx2}, \text{коЮ2XX}) = 0.085$
$\text{corr}(\text{ac2a}, \text{cx2X}) = 0.046$	$\text{corr}(\text{коЮ2}, \text{cx2X}) = 0.096$	$\text{corr}(\text{cx2}, \text{cx2X}) = 0.986$
$\text{corr}(\text{ac2a}, \text{cx2XX}) = -0.037$	$\text{corr}(\text{коЮ2}, \text{cx2XX}) = 0.086$	$\text{corr}(\text{cx2}, \text{cx2XX}) = 0.914$
$\text{corr}(\text{ac2a}, \text{ac2II}) = -0.077$	$\text{corr}(\text{коЮ2}, \text{ac2II}) = 0.043$	$\text{corr}(\text{cx2}, \text{ac2II}) = 0.0002$

Рис. 13.13. Продолжение

Внутри группы рассматриваемых нейрокритериев уровень корреляционных связей может быть достаточно высоким, например, $\text{corr}(\text{cx}, \text{cx2XX}) = 0.914$, однако эта ситуация является скорее исключением, чем правилом.

$\text{corr}(\text{ac2X}, \text{ac2XX}) = -0.0004$		
$\text{corr}(\text{ac2X}, \text{коЮ2X}) = -0.011$	$\text{corr}(\text{ac2XX}, \text{коЮ2X}) = -0.002$	
$\text{corr}(\text{ac2X}, \text{коЮ2XX}) = -0.007$	$\text{corr}(\text{ac2XX}, \text{коЮ2XX}) = 0.005$	$\text{corr}(\text{коЮ2X}, \text{коЮ2XX}) = 0.758$
$\text{corr}(\text{ac2X}, \text{cx2X}) = -0.001$	$\text{corr}(\text{ac2XX}, \text{cx2X}) = -0.0004$	$\text{corr}(\text{коЮ2X}, \text{cx2X}) = 0.096$
$\text{corr}(\text{ac2X}, \text{cx2XX}) = -0.0002$	$\text{corr}(\text{ac2XX}, \text{cx2XX}) = 0.001$	$\text{corr}(\text{коЮ2X}, \text{cx2XX}) = 0.044$
$\text{corr}(\text{ac2X}, \text{ac2II}) = -0.007$	$\text{corr}(\text{ac2XX}, \text{ac2II}) = -0.0001$	$\text{corr}(\text{коЮ2X}, \text{ac2II}) = -0.054$
$\text{corr}(\text{коЮ2XX}, \text{cx2X}) = 0.094$		
$\text{corr}(\text{коЮ2XX}, \text{cx2XX}) = 0.102$	$\text{corr}(\text{cx2X}, \text{cx2XX}) = 0.904$	
$\text{corr}(\text{коЮ2XX}, \text{ac2II}) = 0.028$	$\text{corr}(\text{cx2X}, \text{ac2II}) = -0.042$	$\text{corr}(\text{cx2XX}, \text{ac2II}) = 0.002$

Рис. 13.13. Окончание

Принципиально важным является то, что низкий уровень корреляционной сцепленности прослеживается во всех приведенных выше частях рис. 13.13.

13.14. Непрерывная аддитивная сборка трех лучших частных критериев при проверке гипотезы независимости малых выборок в 16 опытов

В рамках данной главы были синтезированы новые статистические критерии, имеющие уровень вероятности появления ошибок 0.01

и ниже. В связи с этим целесообразно выбрать три критерия с высоким уровнем качества принимаемых решений и объединить их в непрерывной (аналоговой) форме. Программная реализация критерия с аддитивным свертыванием промежуточных вычислений приведена на рис. 13.14.

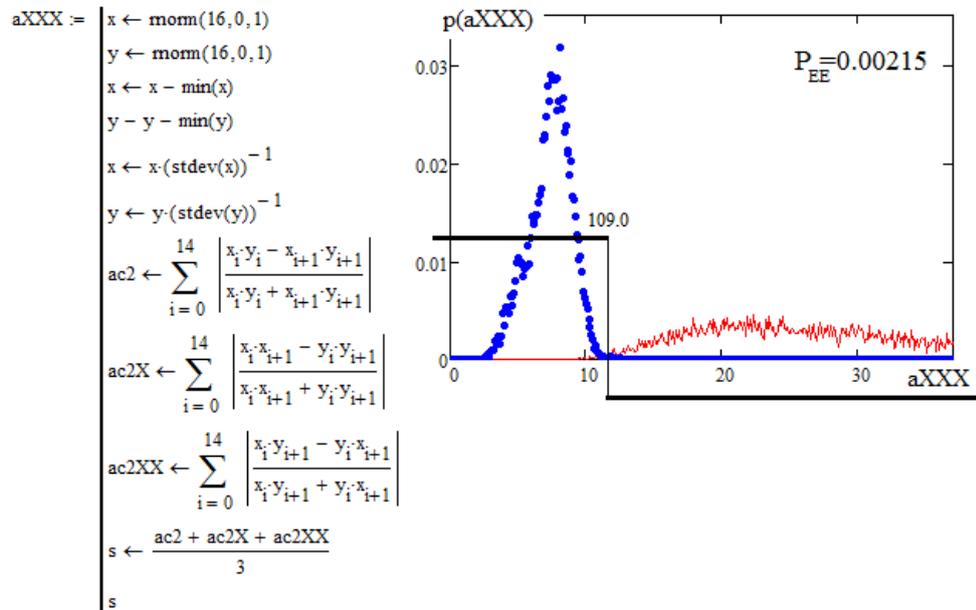


Рис. 13.14. Пример континуального свертывания трех наиболее эффективных статистических критериев

Численный эксперимент по моделированию объединения трех статистических критериев показал, что уровень вероятности ошибок первого и второго рода удастся снизить до величины $P_{EE} \approx 0.00215$, т.е. рассмотренная выше математическая конструкция дает решения с доверительной вероятностью 0.99785. Столь высокого уровня доверия вполне достаточно для многих практических приложений.

13.15. Дискретная сборка трех лучших частных критериев при проверке гипотезы независимости малых выборок в 16 опытов

Существует альтернативный вариант объединения трех критериев, например, с использованием эквивалентных нейрокритериев, в нейросеть с избыточным выходным кодом. Пример программной реализации такого объединения приведен на рис. 13.15.

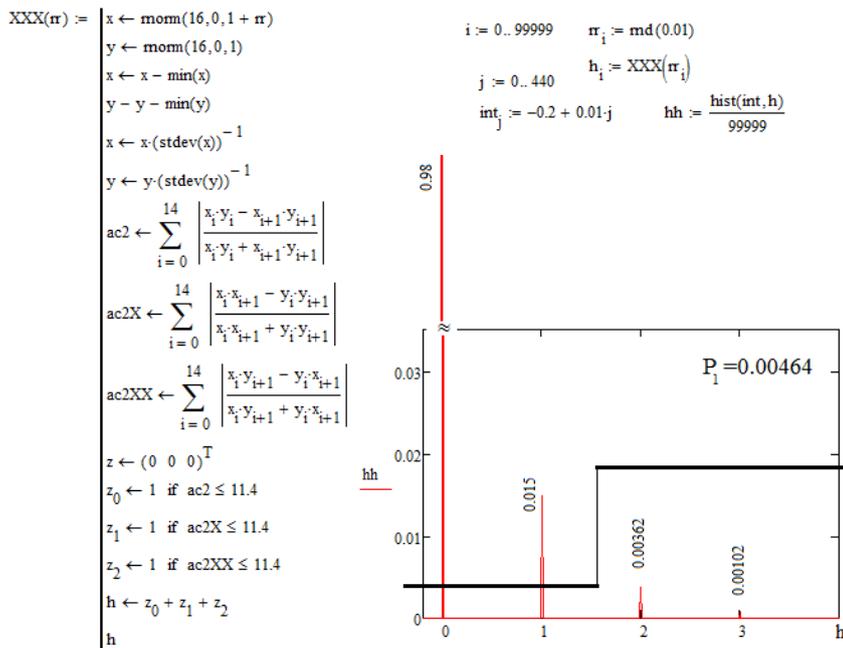


Рис. 13.15. Пример дискретного свертывания трех наиболее эффективных статистических критериев с устранением избыточности выходного кода нейросети

Программа выполняет вычисление значений трех непрерывных функционалов и сравнивает их с порогом квантования $k = 11,4$. Далее выполняется вычисление расстояния Хэмминга между выходным кодом нейросети и «идеальным» откликом «000», когда все три нейрокритерия принимают решение об обнаружении факта независимых данных.

В правой части рис. 13.15 приведен спектр из четырех линий расстояний Хэмминга. Две последние линии спектра соответствуют вероятностям появления единиц в двух и в трех разрядах кода, т.е. вероятность ошибок первого рода для этой конструкции составляет $P_{EE} \approx 0.00462$, что примерно в 2.15 раза хуже в сравнении с вероятностью ошибок предыдущего критерия.

Свертывание промежуточных (избыточных) нейровычислений выгоднее осуществлять аддитивно либо мультипликативно в непрерывных пространствах. Дискретное свертывание всегда дает результат хуже многоуровневого свертывания в непрерывных пространствах множества частных промежуточных вычислений.

Список литературы к главе 13

1. Кенделл М. Дж., Стюарт А. Статистические выводы и связи. М. : Наука : Физматлит, 1973. Т. 2. 899 с.

Справочное издание

Иванов Александр Иванович

**Нейросетевой многокритериальный
статистический анализ малых выборок.
Проверка гипотезы независимости**

Редактор *Т. Н. Судовчихина*
Технический редактор *Н. В. Иванова*
Компьютерная верстка *Н. В. Ивановой*
Дизайн обложки *И. В. Шваревой*

Подписано в печать 23.05.2023.
Формат 70×100¹/₁₆. Усл. печ. л. 17,71.
Тираж 100. Заказ № 213.

Издательство ПГУ.
440026, г. Пенза, ул. Красная, 40.
Тел.: (8412) 66-60-49, 66-67-77; e-mail: iic@pnzgu.ru