

Летков Юрий Владимирович, полиграфолог

г. Вологда, math-polygraph@yandex.ru

Калафати Александр Юрьевич, полиграфолог, руководитель проекта полиграф «Триумф»

г. Москва, psy.akalafati@gmail.com

## **Алгоритм классификации полиграмм «Сокол»**

### **Аннотация**

*В работе описаны математические основы работы алгоритма классификации полиграмм «Сокол», приведены результаты их проверки с помощью имитационного моделирования реагирования физиологической системы методом Монте Карло; описаны методы испытания алгоритма на рандомизированной выборке из архива подтверждённых дел Национального центра оценки достоверности информации (США) и на выборке полиграмм российских полиграфологов и приведены их результаты в сравнении другими алгоритмами.*

### **Ключевые слова**

*Тест, диагностика лжи, классификация полиграмм, чувствительность, специфичность, ошибка классификации, вероятность правильной диагностики лжи, неопределённый результат.*

### **Введение**

Задачу диагностики лживости ответа, стоящую перед полиграфологом, можно разделить на две части. В первой части полиграфолог должен в соответствии с установленной методикой провести полиграфный тест. В результате, на первом этапе, у полиграфолога появляются визуальные данные – геометрические кривые, отображающие изменение регистрируемых физиологических показателей, которые получили название полиграмм. Поведение кривых, степень их изменения, в ходе задавания вопроса и в течении определённого времени после ответа опрашиваемого, полиграфологи (и не только) называют реакциями. Ко второй части диагностики лживости ответа относится задача классификации полиграмм. Она заключается в отнесении полученной полиграммы к характерным для опрашиваемых, которые давали лживые ответы в ходе теста, либо к характерным для опрашиваемых, которые давали правдивые ответы. Причём в практике полиграфологов считается допустимым, когда вторая часть остаётся не разрешённой. Делается это с целью снижения ошибки диагностики. В виду сложного поведения кривых, отображающих физиологические показатели, реакции могут быть охарактеризованы множеством различных геометрических и временных параметров, таких как длина, амплитуда, площадь под кривой,

период, частота и т.д. На этапе классификации полиграммы важную роль играют, выбираемые для этого, указанные выше параметры. Необходимо использовать те, что наиболее адекватно отражают физиологические процессы, протекающие в организме, связанные с типом даваемых ответов. Первая часть задачи диагностики решается целиком методами психологии и психофизиологии. Для решения второй используют методы, перечисленные далее. К исторически первому можно отнести визуальный метод классификации. Он заключается в том, что полиграфолог осматривает полиграммы, при этом сравнивает реакции в парах сравниваемых вопросов, и классифицирует полиграмму на основе сложившегося внутреннего убеждения. Потребность в объективизации классификации привела сначала к появлению «галочной» – Джон Рид, а затем балльной системы обчёта полиграмм – Клив Бакстер. Существует множество таких систем, которые можно разделить на те для которых не требуется проводить каких-либо измерений – баллы выставляются на основе визуального сравнения параметров реакций в сравниваемых парах (например, ESS – [Nelson и др., 2008]), и те, которые требуют измерений для выставления баллов. В обоих случаях баллы являются дискретными случайными величинами. Дискретными, т.е. значения, которых не заполняют какой-либо непрерывный интервал на числовой оси. Баллы принимают конечный ряд значений, например, -3,-2,-1,0,1,2,3, и, как правило, являются целыми числами. Решение выносится после суммирования баллов, присвоенных реакциям, по всем записанным сравниваемым парам. Отнесение полиграммы к тому или иному классу зависит от знака суммарного балла. Условия, при которых принимается решение (граничный балл) должно определяться путём накопления статистического материала, и его анализа с точки зрения допустимой ошибки классификации и процента неопределённых результатов. Ошибка классификации в первую очередь определяется используемыми тестовыми методиками и корректностью их выполнения, состоянием психики и физиологии опрашиваемого, мастерством полиграфолога и только во-вторую очередь методикой, используемой для классификации полиграмм. Но тем не менее разработка эффективных алгоритмов классификации полиграмм является важной задачей, которая способствует увеличению чувствительности и специфичности тестовых методик, уменьшению процента ошибок при принятии решений.

Дальнейшее развитие привело к появлению алгоритмов классификации полиграмм, осуществляемых программным обеспечением полиграфов – автоматических алгоритмов. Нередко эти алгоритмы защищены коммерческой тайной. Нередко об алгоритме не известно ничего кроме его названия, либо, если известен сам алгоритм, например, классификация посредством логистической регрессии, то не известно какие именно параметры реакций используются для его осуществления. При этом, если говорить об отечественных алгоритмах,

то они остаются не исследованными, не прошедшими практическую валидизации. На текущий момент сложилась следующая ситуация, когда почти все зарубежные алгоритмы оценки полиграмм активно валидируются и сравниваются между собой (Probabilities analysis, OSS-2, OSS-3, Polyscore, ESS (автоматизированная), ESS-M (автоматизированная), за исключением, пожалуй, White Alpha) [Dollins и др., 2000; Nelson и др., 2008; Webb и др., 2008; Raskin, Kircher, 2014; Nelson, Handler, 2019], в то время как отечественные алгоритмы защищены коммерческой тайной при полном отсутствии валидизации. Некоторым исключением может являться разве, что алгоритм Quest [8], разработанный Алексеевым Л.Г., который оценили производители из компании Lafayette в начале 2000-х, но сейчас этот алгоритм уже давно не поставляется в стандартном установочном пакете программного обеспечения Lafayette. Следует отметить также, что при осуществлении автоматического алгоритма важнейшей задачей является адекватное измерение программой, используемых показателей реакций. Неадекватное измерение этих параметров, может привести к тому, что в результате работы алгоритма будет увеличиваться количество ошибок классификации. При использовании автоматического алгоритма, измерение параметров реакций на конкретной полиграмме, может не находиться под контролем полиграфолога, что в известной мере, снижает его ответственность за конечный результат.

Отдельно можно выделить Метрическую (измерительную) оценку полиграмм. По имеющейся в литературе информации она основана на принципах параметрической математической статистики (критерий Стьюдента) и на постулирование существования симптомокомплекса реакций присущего конкретному человеку [1]. Не вдаваясь в технологию осуществления метрической оценки полиграмм, не останавливаясь на параметрах реакций, которые используются для её осуществления, необходимо критически отнестись к тому, на чём она основана. В первую очередь это существования симптомокомплекса. Среди отечественных специалистов известно о наличии засекреченной диссертации Ю.И.Холодного, в которой он исследовал существование симптомокомплекса. Но опубликована она до сих пор нигде не была, и получить к ней доступ для гражданского исследователя не представляется возможным. А это значит, что для научного сообщества её пока не существует. Однако, есть широко известное в США исследование John Lacey [Lacey (1959)], в котором автор среди прочего выявил, что у большинства из 42 участников имеется очень схожий паттерн реагирования на различные стрессоры: вычисление в уме, холодный эксперимент, ожидание холодного эксперимента и подбор слов. Правда, подобный паттерн наблюдался не у всех, были люди, у которых реагирование было хаотичным и паттерн не наблюдался. Так что для американских психофизиологов идея «симптомокомплекса» – или индивидуального реагирования не является чем-то новым. Как указали в своём исследовании Сасіорро и др.

[Cacioppo и др., 1991], проблемы, которые поднял Lasey [20] на десятилетия определили исследования в этой области. Другими словами, не известны опубликованные открытые исследования, которые подтвердили бы оценку симптомокомплекса как фактора, учёт которого, уменьшает процент ошибок классификации. Во-вторую, это довольно скудное описание в литературе метрической оценки, из которой нам не удалось понять математического обоснования использования коэффициента Стьюдента, особенно в том виде как он описывается. И тем более не удалось понять обоснования вычисления вероятности значимости стимула: «*Стимул...признаётся ситуационно значимым с той вероятностью, которая была рассчитана на контрольных тестах*» [1]. Задача классификации может допускать отклонения от математической строгости, если при этом достигаются приемлемые уровни ошибок классификации, чувствительности и специфичности. Но исследования Метрической оценки полиграмм на этот предмет также не известны.

Алгоритм классификации полиграмм «Сокол», результаты разработки и испытания которого, изложены в данной работе, лишён большинства указанных выше недостатков. Он может быть реализован в ручном режиме, в полуавтоматическом и может быть полностью автоматизирован. В его основе лежат три положения. Во-первых, это то, что вегетативные реакции при ложном и правдивом ответе на вопросы теста различаются друг от друга [2]. Это положение является необходимым условием для третьего положения. Во-вторых, то, что наиболее адекватно отражают физиологические процессы, связанные с типом даваемых ответов, так называемые «параметры Кирчера» эффективность которых была оценена в исследованиях американских ученых [Kircher, Raskin (1988)] и [Kircher и др., 2005]: длина линии дыхания, амплитуда КГР (ЭДА), амплитуда подъёма манжеты: длина линии дыхания, амплитуда КГР (ЭДА), амплитуда подъёма манжеты. А также амплитуда кривой ФПГ [Kircher, Raskin, 1988; Honts et al, 2015]. В-третьих, среднестатистически реакции, сопровождающие ложный ответ на проверочный вопрос, превышают реакции на специальный вопрос, используемый для сравнения; а реакции, сопровождающие правдивый ответ, меньше реакций на вопрос сравнения, идея, которая впервые была опубликована в статье Reid [24]. Алгоритм анализирует выборки значений случайных величин, представляющих собой разность параметров реакций в сравниваемых парах вопросов. Например, для КГР это будет разность амплитуд реакции на проверочный и контрольный вопрос. В том случае, если не выполняется положение три, лежащее в основе алгоритма, полученная выборка будет иметь медиану (med) равную нулю, так как амплитуды КГР реакций на проверочный и контрольный вопрос, в этом случае будут превышать друг друга с одинаковой вероятностью – 0,5. Таким образом можно сформулировать прямую статистическую гипотезу  $H_0: med = 0$ , а альтернативную гипотезы вычислить из имеющейся выборки, и в соответствии с ней отнести

полиграмму к тому или иному классу. Если альтернативная гипотеза заключается в том, что,  $med > 0$ , то полиграмма должна быть отнесена к классу правдивых (диагностирована правдивость ответов). Иначе, если  $med < 0$ , то – к классу лживых. Аналогично (но не идентично) определяются статистические гипотезы для остальных, используемых показателей реакции. Проверить выполнение нулевой гипотезы возможно с помощью методов непараметрической статистики, а именно с помощью одновыборочного критерия Вилкоксона [3,4]. Использование задачи проверки статистической гипотезы для решения задачи классификации не является чем-то новым: «Задача отнесение наблюдения  $X$  в один из двух ранее известных классов  $j = 1, 2$  тесно связана с классической статистической задачей проверки простой гипотезы против простой альтернативы» [5, с. 47]. Анализ выборок разностей физиологических параметров по отдельным показателям реакций, может однозначно дать общую для реакции на вопрос альтернативную гипотезу только в том случае, если альтернативные гипотезы по всем выборкам совпадают. Изучение этой возможности показало, что на конкретной выборке один из трёх физиологических параметров даёт отличающуюся от двух других показателей альтернативную гипотезу примерно в 50% случаев. При этом так же установлено, что исключение показателя с противоречащей альтернативой позволяет классифицировать полиграмму и даже улучшает результат классификации, но в целом не может считаться выходом из положения. Во-первых, не ясно, что предпринимать, если анализируемых показателей чётное количество, и половина из них даёт альтернативную гипотезу, не совпадающую с альтернативой гипотезой второй половины. Не ясно также каким параметрам отдавать предпочтение при исключении их из анализа. Наблюдались случаи, когда, исключение КГР приводило к ошибке классификации. Во-вторых, при отдельном анализе выборок физиологических показателей отсутствует общий, интегральный показатель, который мог бы характеризовать результат классификации и служить ориентиром для отказа от принятия решения. Хотя это затруднение можно обойти, рассматривая анализируемые разности физиологических показателей как многомерную случайную величину и использовать для этих целей многомерный критерий Вилкоксона [6, с. 403]. Алгоритм теста многомерного критерия Вилкоксона, был включён в программу «Сокол НТФП», которая была написана для исследования алгоритма «Сокол» [7], но изучение эффективности алгоритма, основанного на многомерном критерии, не производилось.

Для осуществления классификации полиграмм в алгоритме «Сокол» используется случайная величина, являющаяся функцией элементов выборки – разностей физиологических показателей, которые полиграфолог извлекает из полиграммы. Эта случайная величина была названа ФОПК – функция оценки полиграммы (косинусов). Она позволяет анализировать измеряемые физиологические параметры в совокупности, позволяет вычислять интегральные

показатели, характеризующие результат классификации и, в том числе, опираясь на которые, можно отказываться от принятия решения по результату теста.

### Функция оценки полиграммы (косинусов)

Рассмотрим тест из двух пар сравниваемых вопросов. Обозначим как  $x_2 = d_1^1 - d_2^1$  и  $x_1 = d_1^2 - d_2^2$  разности размеров физиологических показателей реакций (длин линий дыхания, амплитуд КГР, подъёмов АД, амплитуд спазма ФПГ, длин линий ФПГ и т.п.). Для каждого измеряемого физиологического параметра составим евклидово пространство с ортонормированным базисом (декартовой системы координат). По осям системы координат отложим значения равные разности измеренных параметров в парах сравнения. Эти разности будут представлять собой координаты вектора  $\vec{a}$ , который изображён на рисунке 1, для определённости назовём его вектором ФОПК.

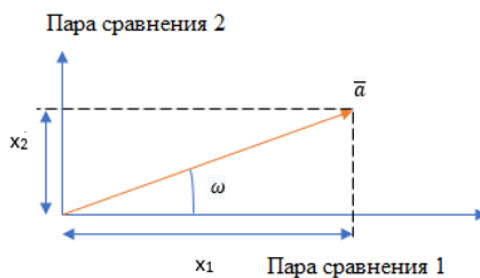


Рисунок 1. – Вектор и координаты ФОПК

Вектор ФОПК образует с осью абсцисс угол  $\omega$ . Рассмотрим косинус этого угла, как функцию разностей измеряемых показателей.

$$\cos(\omega) = \frac{x_2}{\sqrt{(x_1^2 + x_2^2)}} \quad (1)$$

Он, как функция, имеет следующие свойства:

1. Симметрична относительно нуля;
2. Ограничена, область значений  $[-1, 1]$ ;
3. Как случайная величина имеет совпадающую со случайной величиной  $x_2$  нулевую гипотезу  $H_0$  относительно медианы генеральной совокупности и среднего;
4. Имеет совпадающую альтернативную гипотезу  $H_1$  с альтернативной гипотезой  $x_2$ ;
5. Образована из  $x_2$  линейным преобразованием, которое не меняет результат вычисления  $p$  – value в статистических тестах, основанных на ранговых критериях.

Косинус угла вектора  $\vec{a}$  можно использовать в тестах одновыборочного критерия Вилкоксона и тестов из арсенала численного ресемплинга вместо разницы физиологических показателей. Свойства косинуса обеспечат идентичный результат в этих тестах, который

получается при использовании сырых параметров. Косинус угла  $\omega$  и называется ФОПК. В общем случае, когда в результате проведённых тестов, имеется  $n$  (3,4,5,6 и т.д.) сравниваемых пар, каждой сравниваемой паре, и для каждого физиологического параметра, мы можем сопоставить значение ФОПК. Формула для вычисления ФОПК в общем случае будет иметь вид:

$$\cos(\omega_i) = \frac{x_i}{\sqrt{(x_1^2+x_2^2+\dots+x_i^2+\dots+x_n^2)}} \quad (2)$$

После вычисления ФОПК для всех пар сравниваемых вопросов по всем физиологическим показателям, производится их суммирование. Вычисляются суммарные значения ФОПК как по контролируемым физиологическим показателям, так и по всему тесту. На рисунке 2 изображена таблица с вычисленными значениями ФОПК для реальной полиграммы

№	1	2	3	4	5	6	7	8	9	10	11	12	$\Sigma$	$\eta$
Дых	-0.31	0.41	0.32	-0.09	0.21	-0.37	-0.25	-0.27	0.07	0.05	-0.21	0.51	0.08	0.02
КГР...	-0.14	-0.36	-0.26	-0.26	-0.02	0.5	-0.19	-0.02	-0.48	-0.41	-0.14	-0.02	-1.82	0.53
ФПГ	0.02	0.09	0.12	0.21	-0.15	0.03	0.03	0.9	-0.11	0.17	0.02	-0.23	1.1	0.32
АД	-0.01	-0.33	-0.38	-0.63	-0.09	-0.28	-0.24	-0.2	-0.38	-0.1	0.0	-0.13	-2.76	0.8
Сумма ФОП по тесту													-3.41	

Рисунок 2. – Значения ФОПК для реальной полиграммы

Значения ФОПК не зависят от сравнительных масштабов, измеряемых физиологических показателей, также они не зависят от сравнительных масштабов разности различных физиологических показателей. Это нетрудно понять из того, что при умножении координат вектора на одно и тоже число угол  $\omega$ , а значит и косинус угла не изменится. В силу того, что выбор первой и второй пар сравниваемых вопросов осуществлён произвольно, при одинаковом знаке координат вектора, вероятности возникновения, симметричных относительно биссектрисы квадранта ( $45^\circ$ ) направлений вектора, равны. Поэтому в квадрante имеется выделенное направление – биссектрисы квадранта. В том случае если вектор будет направлен по биссектрисе, его координаты, а значит разности параметров в первой и во-второй парах сравниваемых вопросов будут одинаковыми.

ФОПК не отражает свойств физиологических реакций, связанных с масштабом, в котором производятся их измерения, и с масштабом разностей физиологических параметров. Для того что бы понять какие именно свойства физиологических реакций отражает ФОПК произведём исследование экстремальных свойств функции, являющейся суммой ФОПК по физиологическому параметру. Для простоты вычислений исследование будем проводить для двух сравниваемых пар. В случае большей размерности вычисления проводятся аналогично, но обладают большим объёмом. При этом результаты двумерного случая нетрудно обобщить для  $n$  – мерного пространства, что в конце вычислений мы и произведём.

Проводя анализ, в первую очередь отметим, что суммарное значение ФОПК (СФОПК) по всем сравниваемым парам для отдельного показателя будет близко к нулю в случае справедливости нулевой гипотезы, так как значения ФОПК будут принимать положительные и отрицательные значения с одинаковой вероятностью, и при суммировании уничтожать друг друга. Для того чтобы ответить в каком случае СФОПК будет принимать минимальное и максимальное значение необходимо использовать методы математического анализа, т.е. использовать производные и найти точки экстремума. Необходимо учесть, что на функции, которые входят в СФОПК наложено условие связи:

$$\frac{x_1^2}{x_1^2+x_2^2} + \frac{x_2^2}{x_1^2+x_2^2} - 1 = 0 \quad (3)$$

Поэтому будем искать точки условного экстремума методом неопределённых множителей Лагранжа. Для этого составим функцию:

$$f(x_1, x_2) = \frac{x_1}{\sqrt{x_1^2+x_2^2}} + \frac{x_2}{\sqrt{x_1^2+x_2^2}} + \lambda \cdot \left( \frac{x_1^2}{x_1^2+x_2^2} + \frac{x_2^2}{x_1^2+x_2^2} - 1 \right) \quad (4)$$

Теперь необходимо найти частные производные этой функции по переменным и неопределённому множителю Лагранжа и приравнять их нулю. Таким образом получим уравнения для определения точек экстремума:

$$1) x_2^2 - x_1x_2 = 0$$

$$2) x_1^2 - x_2x_1 = 0$$

Производная по множителю Лагранжа даёт уравнение:

$$3) \frac{x_1^2}{x_1^2+x_2^2} + \frac{x_2^2}{x_1^2+x_2^2} - 1 = 0$$

Нетрудно видеть, что решением системы уравнений являются все такие значения переменных, что выполняется равенство:

$$x_1 = x_2 \quad (5)$$

При  $x_1 = x_2 = 0$  значение косинусов неопределённо, кроме того, не выполняется уравнение (3), поэтому решением являются значения переменных отличные от нуля. В общем случае для  $n$  пар сравниваемых вопросов мы получим систему уравнений из квадратичных форм:

$$x_1^2 + x_2^2 + \dots + 0 \cdot x_i^2 + \dots + x_n^2 - x_i x_1 - x_i x_2 - \dots - x_i x_n = 0 \quad (6)$$

Решением которой будут является все отличные от нуля переменные, такие что:

$$x_1 = x_2 = x_3 = \dots = x_i = \dots = x_n \quad (7)$$

Надо напомнить, что переменные в данном случае являются разностями физиологических показателей реакций на соответствующие типы вопросов. Таким образом СФОПК достигнет своего максимального значения при условии постоянства разности параметров во всех парах сравниваемых вопросов. При этом не важна величина разности. Т.е.



если по дыханию разность будет постоянна и составлять 10 единиц длины, а по КГР постоянна и составлять 50 единиц длины, то в обоих случаях СФОПК достигнет одинакового максимального значения, и вклад обоих параметров в результат теста будет одинаков. Важным является то, что СФПОК будет принимать максимальное значение, когда разности показателей реакций будут иметь один знак. Это обстоятельство отличает поведение ФОПК от случая справедливости прямой гипотезы. Из вышесказанного следуют следующие выводы: во-первых, чем больше модуль ФОПК, тем менее вероятна справедливость нулевой гипотезы; во-вторых, ФОПК отражает свойство согласованности физиологических реакций в парах сравниваемых вопросов; в-третьих, при оценке полиграммы посредством ФОПК все физиологические показатели потенциально вносят одинаковый вклад в конечные результат классификации. **Под согласованностью физиологических реакций мы будем понимать близость разностей физиологических показателей реакций на соответствующие типы вопросов во всех парах сравниваемых вопросов.** Чем согласованнее реакции по отдельному физиологическому параметру, тем больший вклад он внесёт в результат теста. Кроме того, чем синхронней разности реакции в разных физиологических параметрах, тем больше будет суммарная ФОПК по всем показателям и сравниваемым парам, что в большей степени будет свидетельствовать в пользу альтернативной гипотезы. **Под синхронностью будем понимать совпадение знаков реакций по разным физиологическим параметрам.**

Вычислим максимальное значение, которое может достигнуть СФОПК (обозначим  $\lambda$ ). Для этого учитывая равенство всех переменных в точке экстремума через  $x$  запишем:

$$\lambda_{max} = \cos(\omega_1) + \cos(\omega_2) + \dots + \cos(\omega_n) = \frac{x}{\sqrt{(x^2 + \dots + x^2 + \dots + x^2)}} + \frac{x}{\sqrt{(x^2 + \dots + x^2 + \dots + x^2)}} + \dots + \frac{x}{\sqrt{(x^2 + x^2 + \dots + x^2 + \dots + x^2)}} = \frac{nx}{\sqrt{n \cdot x^2}} = \pm \sqrt{n} \quad (8)$$

Тогда значение косинуса отдельной пары сравнения будет:

$$\cos(\omega_\delta) = \frac{1}{\sqrt{n}} \quad (9)$$

где,  $\omega_\delta$  – угол между вектором СФОПК и координатой вектора СФОПК, образованной разностью параметров соответствующей пары сравниваемых вопросов, при котором наблюдается экстремум СФПОК. Таким образом, модуль экстремального значения СФОПК равен корню квадратному из количества сравниваемых пар. Для шести сравниваемых пар это с точностью до второго знака будет 2,45. И значение это не будет зависеть от измеряемых масштабов, а только от согласованности реакций. Если реакции по всем физиологическим параметрам будут полностью согласованы, то сумма ФОПК по всему тесту примет максимальное значение равное  $3\sqrt{n}$  и  $4\sqrt{n}$  для трёх и четырёх физиологических параметров соответственно. Для шести пар сравнения эти значения будут равны 7,34 и 9,78. На рисунке

2 СФПК по дыханию равна 0.08, по КГР – (- 1.82), для спазма достигает значения 1.1, для АД – (- 2.76). Общая сумма ФОПК по тесту равна (- 3.41). На рисунке 3 изображена полиграмма для которой вычислено, указанное суммарное значение ФОПК (на рисунке только сравниваемые пары вопросов, первый – вопрос сравнения).

Возвращаясь к случаю для двух сравниваемых пар, можно сказать, что СФОПК примет максимальное значение, когда вектор будет направлен по биссектрисе квадранта – единственному направлению, которое можно выделить из соображений симметрии. В этом случае разности параметров, являющиеся координатами вектора  $\vec{a}$  будут образовывать стороны квадрата и будут равны. Если в результате произведённого воздействия система отвечает строго определённым сдвигом, то это означает, что в системе отсутствуют неконтролируемые факторы, оказывающие стохастическое воздействие на реакцию системы на стимул. Тем самым ответ системы является детерминированным. Достижение максимального значения СФОПК возможно только в том случае, если реакция физиологической системы носит детерминированный характер, что в реальной ситуации является невозможным. Любая физиологическая система является стохастической, и возможно, что для разных опрашиваемых и для разных их состояний, степень влияния неконтролируемых факторов может иметь различный характер. Для оценки степени стохастичности физиологических показателей реакций можно использовать СФПОК.

Рассмотрим её формулу:

$$\begin{aligned} \lambda &= \cos(\omega_1) + \cos(\omega_2) + \dots + \cos(\omega_n) = \\ &= \frac{x_1}{\sqrt{(x_1^2+x_2^2+\dots+x_n^2)}} + \frac{x_2}{\sqrt{(x_1^2+x_2^2+\dots+x_n^2)}} + \dots + \frac{x_n}{\sqrt{(x_1^2+x_2^2+\dots+x_n^2)}} = \frac{x_1+x_2+\dots+x_n}{\sqrt{x_1^2+x_2^2+\dots+x_n^2}} = \\ &= \frac{n \frac{x_1+x_2+\dots+x_n}{n}}{\sqrt{n} \sqrt{\frac{x_1^2+x_2^2+\dots+x_n^2}{n}}} = \frac{\sqrt{n} \bar{x}}{\bar{x}_{кв}} = \frac{\lambda_{max} \bar{x}}{\bar{x}_{кв}} \end{aligned} \quad (10)$$

где  $\bar{x}$  – среднее,  $\bar{x}_{кв}$  – среднее квадратическое разниц физиологических показателей на разные типы вопросов в сравниваемых парах. Отсюда имеем:

$$\lambda = \frac{\lambda_{max} \bar{x}}{\bar{x}_{кв}} \quad (11)$$

Модуль отношение СФОП к своему максимальному значению назовём показателем детерминированности и обозначим его буквой  $\eta$

$$\eta = \left| \frac{\lambda}{\lambda_{max}} \right| = \frac{|\bar{x}|}{\bar{x}_{кв}} \quad (12)$$

Минимальное значение показателя детерминированности равно нулю. При этом значении показателя реакция физиологической системы носит чисто стохастический характер, детерминированная составляющая реакции на разные типы стимулов отсутствует. Максимальное значение показателя равно единице. При таком его значении реакции системы

носят полностью детерминированный характер, в каждой паре сравниваемых вопросов разница физиологических параметров постоянна. Вычислить значение показателя стохастичности можно как отношение модуля среднего значения переменных к их среднему квадратическому отклонению. Известно, что всегда выполняется следующее соотношение  $\bar{x} \leq \bar{x}_{кв}$  и равенство достигается только при равенстве всех переменных.

Нетрудно видеть, что физиологическая система с наибольшим показателем детерминированности будет вносить наибольший вклад в итоговый результат теста. Надо отметить, что вклад этот может как способствовать принятию решения, так и наоборот – затруднять его, в том случае, если физиологические системы с большим показателем детерминированности показывают противоположные тенденции. Для рассмотренной на рисунке 3 полиграммы и соответствующими ей значениями СФОПК, изображёнными на рисунке 2, можно видеть, что показатель детерминированности для дыхания равен всего 0,02. Если показатель детерминированности меньше 0,1 можно считать, что реагирование по данному физиологическому каналу носит чисто стохастический характер – детерминационная составляющая реакции отсутствует. Показатель детерминированности канала ФПГ равен 0,32. Показатель детерминированности от 0,1 до 0,3 говорит о сильной зашумлённости канала, когда детерминационная составляющая ответа «размывается» стохастическими процессами. Соответствующий показатель для КГР равен 0,53. Значение показателя детерминированности в диапазоне от 0,3 до 0,7 является наиболее встречающимся и может считаться свидетельствующим о средней зашумлённости канала. Причём чем ближе к 0,3, тем выше информационных шум и чем ближе к 0,7 – тем ниже. Показатель для АД равен 0,8. Как можно видеть именно этот канал внёс наибольший вклад в суммарное значение ФОПК по всему тесту. Значение показателя в диапазоне от 0,7 до 1 говорит о высокой детерминированности канала, когда реакции системы определяется ответом на внешнее воздействие. Разница параметров физиологических показателей при коэффициенте детерминированности равным единице есть «чистый» ответ системы на разницу в типе вопросов, на которую, при значении показателя меньше единицы, накладывается шум стохастических процессов.

Произведём оценку «чистого» ответа физиологической системы на разный тип ответа. Для этого запишем значение косинуса для  $i$ -ой пары сравниваемых вопросов, через среднее квадратическое отношение:

$$\cos(\omega_i) = \frac{x_i}{\sqrt{n} \sqrt{\frac{(x_1^2 + x_2^2 + \dots + x_i^2 + \dots + x_n^2)}{n}}} \quad (13)$$

$$\sqrt{n} \cos(\omega_i) = \frac{x_i}{\bar{x}_{кв}} \quad (14)$$

учитывая, что  $\cos(\omega_\delta) = \frac{1}{\sqrt{n}}$  запишем:

$$\frac{\cos(\omega_i)}{\cos(\omega_\delta)} = \frac{x_i}{\bar{x}_{KB}} \quad (15)$$

так как при  $\eta \rightarrow 1$ , выполняется  $\cos(\omega_i) \rightarrow \cos(\omega_\delta)$  и  $x_i \rightarrow x$ , то  $\bar{x}_{KB} \rightarrow x$ . Таким образом, если все разницы параметров имеют один знак, что выполняется при  $\eta \rightarrow 1$ , то их среднее квадратическое может служить оценкой «чистого» ответа физиологической системы на разницу в типе вопросов.

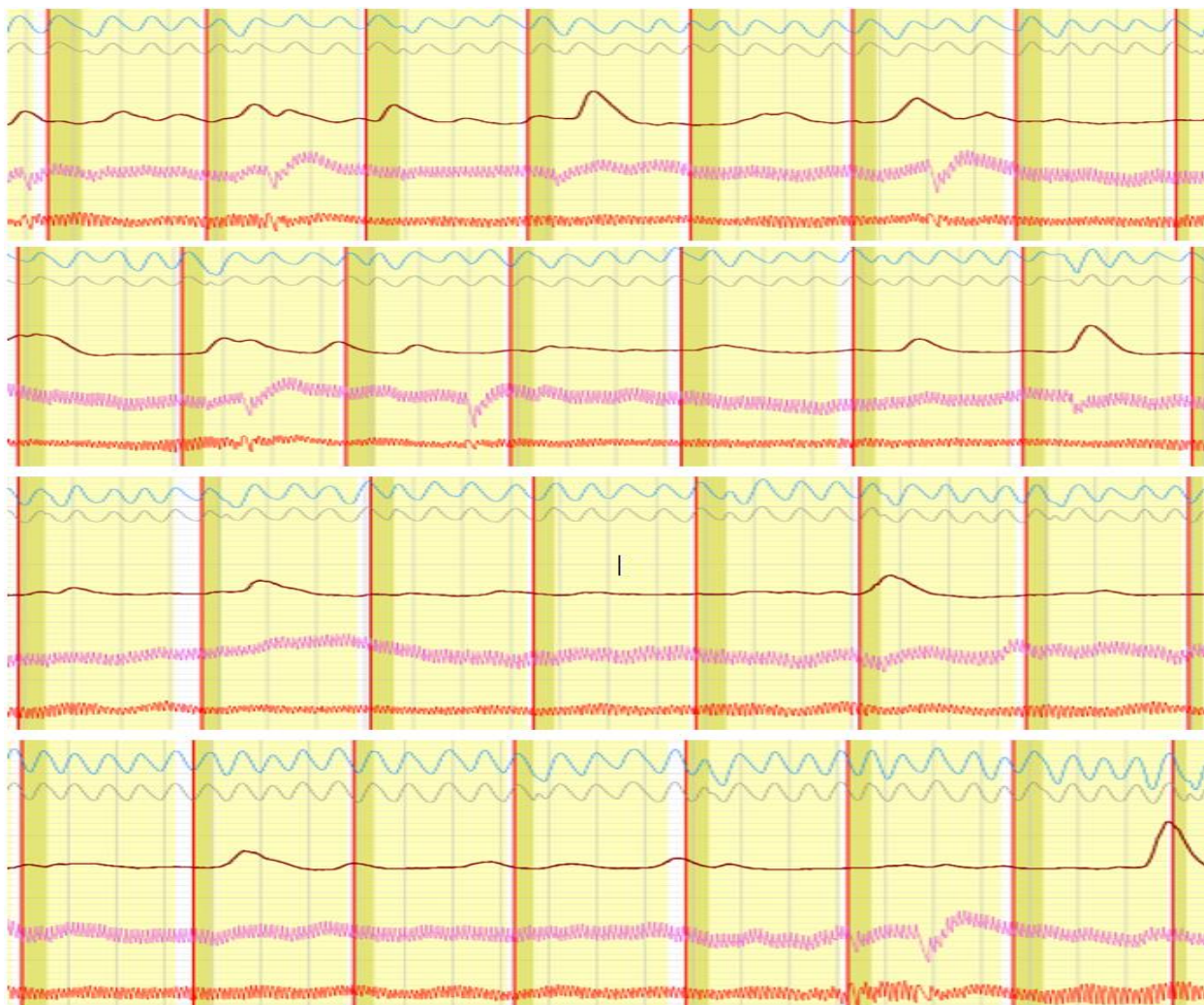


Рисунок 3. – Полиграмма случая с доказанной независимым способом лживости ответов

Кратко сформулируем установленные факты, касающиеся свойств ФОПК:

1. ФОПК отражает свойства согласованности изменений физиологического показателя, вызванных реакциями на стимулы, и свойство синхронности изменений разных физиологических параметров в ответ на стимулы, и может служить их мерой;
2. В случае истинности нулевой гипотезы значения СФОПК будут близки к нулю;
3. Чем больше модуль суммы ФОПК, тем менее вероятна нулевая гипотеза;
4. Максимальное по модулю значение СФОПК равно корню квадратному из числа сравниваемых пар;

5. Максимальное значение суммы ФОПК по всему полиграфному тесту равно произведению количества наблюдаемых параметров на корень квадратный из количества сравниваемых пар;
6. Показатель детерминированности близкий к нулю свидетельствует о зашумлённости канала, об отсутствии определяемого психологическим стимулом реагирования;
7. Показатель детерминированности может быть равный единице только в случае реакции на психологический стимул происходят без влияния сторонних стохастических процессов;
8. Оценкой составляющей реакции, детерминированной психологическим стимулом, может служить среднее квадратическое разности показателей реакций на соответствующие типы вопросов.

### **Принятие решения на основе ФОПК**

ФОП в некотором смысле является аналогом баллов, которые являются целым набором различных случайных дискретных величин. Каким бы образом они не присваивались физиологическим показателям, их распределение вероятностей зависит от количества пар вопросов сравнения, используемых в тесте, от количества повторов теста, от количества регистрируемых физиологических показателей, для которых выставляются баллы. Поэтому для достаточно строгого определения правил принятия решения – граничных баллов, через построение эмпирических распределений баллов, необходимо строить распределения для каждого из случаев индивидуально. От правил принятия решения зависят такие показатели теста как процент неопределённых исходов, ошибки DI и NDI. Для одного и того же теста с разным количеством повторов эти характеристики будут отличаться. Они будут различными и в том случае, если для одного и того же теста, с одним и теми же количеством повторов, будет использовано разное количество физиологических показателей реакций. Вышесказанное про баллы в полной мере можно отнести и к ФОП, с той лишь разницей, что она является непрерывной случайной величиной. И если определять правила принятия решения, основанные на величине и знаке ФОП, то необходимо исследовать плотности распределения вероятностей различных ФОП с учётом факторов: количества пар сравнения, повторности тестов, количеству использованных показателей, – что само по себе является объёмной задачей; для каждой комбинации сочетания различных градаций факторов необходимо иметь достаточное количество случаев, доказанных независимыми способами. Но в случае ФОП подобные трудности можно обойти, воспользовавшись методами математической статистики. Это можно сделать, если в каждом случае проверять справедливость предположения за номером три, которое лежит в основе алгоритма «Сокол»,

и о котором говорилось во введении. В случае справедливости нулевой гипотезы плотность распределения вероятностей ФОП является симметричной, поэтому нулевые гипотезы о равенстве нулю медианы и математического ожидания (среднего) равносильны. Кроме того, с помощью ФОП для всех измеряемых параметров реакций можно построить общий интегральный показатель. Дело в том, что в случае истинности нулевой гипотезы, когда снимаемые физиологические показатели представляют из себя гауссовский шум (нормальное распределение с математическим ожиданием равным нулю и произвольным стандартным отклонением), ФОПК для всех показателей имеет одинаковый закон распределения вероятностей. Обращаясь к двум сравниваемым парам и рисунку 1, нетрудно увидеть, что угол  $\omega$  в этом случае имеет одинаковое равномерное распределение для любого показателя. Координаты вектора ФОПК с одинаковой вероятностью могут иметь как положительный знак, так и отрицательный знак и поэтому сам вектор может с одинаковой вероятностью принадлежать любому из четырёх квадрантов. В случае трёх сравниваемых пар – любой из восьми долей сферы и т.д. Так как выбор координат вектора произволен, то вектор с одинаковой вероятностью может принимать одно из двух положений симметричных относительно биссектрисы квадранта. Проверить это предположение, основанное на симметрии задачи, можно произведя расчёт плотности вероятности распределения угла  $\omega$  вектора ФОПК для двух пар сравниваемых (см. рисунок 1). Этот расчёт можно увидеть в Приложении А.

Как видно в условиях справедливости нулевой гипотезы угол  $\omega$  имеет равномерное распределение, не зависящее от параметров распределений показателей физиологических реакций.

$$p(\omega) = \frac{1}{2\pi} \quad (16)$$

Аналогичные вычисления (Приложение А) для плотности функции распределения ФОПК (которая является косинусом от угла  $\omega$ ) дают следующий результат:

$$p_{\text{ФОПК}}(\lambda) = \frac{1}{\pi\sqrt{1-\lambda^2}} \quad (17)$$

Таким образом, функция распределения вероятностей ФОПК также не зависит от параметров распределений показателей физиологических реакций. График функции, которая определяется формулой (17) приведён на рисунке 5.

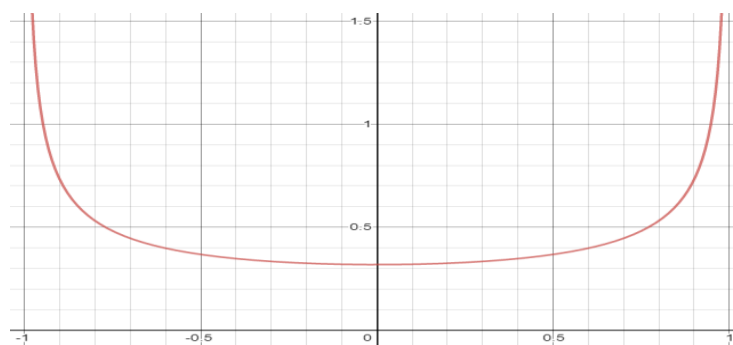


Рисунок 5. – График функции определяемой формулой (17)

Как и следовало ожидать медиана ФОПК в условиях справедливости нулевой гипотезы равна нулю. Вычисления показывают, что математическое ожидание этого распределения также равно нулю. Из вышеизложенного можно сделать вывод, что в предположении справедливости нулевой гипотезы значения ФОПК, независимо от принадлежности их различным физиологическим показателям, можно считать извлечёнными из одной генеральной совокупности. Это даёт возможность использовать их в одном непараметрическом статистическом тесте. Результатом статистического теста должна быть вычисленная из данных альтернативная гипотеза, на основе которой осуществлена классификация полиграммы, и вычислен p-value, который может служить ориентиром для принятия решения. Главное достоинство такого подхода заключается в том, что в вычисление p-value уже закладываются те условия, в которых оно производится. Происходит автоматический учёт тех факторов, о которых говорилось ранее: количестве пар сравнения в тесте, количестве повторов в тесте, количестве используемых физиологических показателей. Также автоматически учитываются пары сравнения, исключённые из анализа из-за артефактов. Статистикой называется некоторая функция от выборочных данных. Например, выборочное среднее является статистикой. Так как статистика это функция выборочных данных – случайных величин, то она сама является случайной величиной. Правило проверки статистической гипотезы при некоторой фиксированной альтернативе называется статистическим критерием. Статистика по которой судят о справедливости статистической гипотезы называется статистикой критерия и, как правило, обозначают буквой  $Z$ . Суммарное значение ФОПК может быть статистикой непараметрических тестов. Уровень значимости  $\alpha$  (alpha) – это столь маленькая вероятность, что событие с такой вероятностью принимается практически невозможным. На практике используют следующие уровни значимости: 0.001, 0.01, 0.05, 0.1. p-значением (p-value, достигаемый уровень значимости) называется вероятность того, что статистика критерия примет значение такое же как наблюдаемое  $Z$  или более "экстремальное" (большее, либо меньшее) при условии, что верна нулевая гипотеза. Если достигаемый уровень значимости оказывается меньше заранее принятой  $\alpha$ -ы, то нулевую

гипотезу отвергают. Используемые на практике уровни значимости не являются догмой и выбираются, исходя из решаемой задачи. В задаче классификации  $\alpha$  должна выбираться исходя из приемлемых ошибок классификации, чувствительности, специфичности теста и процента неопределённых исходов – когда полиграфолог отказывается отнести полиграмму к тому или иному классу. Если бы распределения суммарной ФОПК для классов были заранее известны, то классификацию можно было бы осуществлять на основе достигаемого уровня значимости для каждого из классов простым сравнением их по величине. Но в настоящее время вычислить p-value можно только для «нулевого» распределения суммарной ФОПК. На Рисунке 6 визуальным образом на условных изображениях распределений показано, что достигаемый уровень значимости «нулевого» распределения будет больше, чем соответствующий уровень значимости для распределений суммарной ФОПК для классов. Это показывает то, что, в задаче классификации мы можем выйти за пределы привычного ряда уровней значимости.

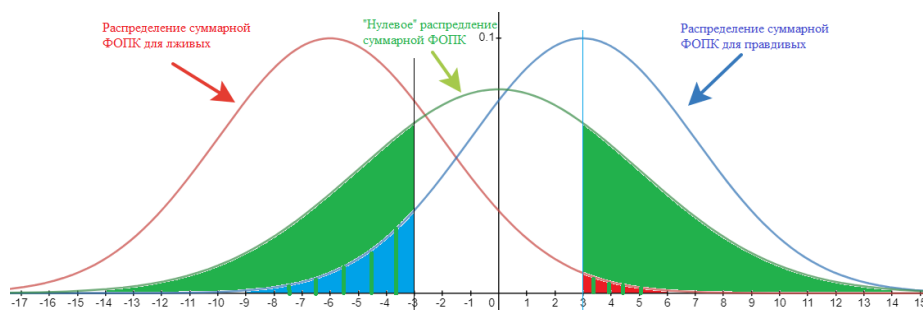


Рисунок 6. — Условные изображения распределений вероятностей СФОПК

В случае выполнения положения три (о котором говорилось во введении), что равносильно утверждению на языке статистики об истинности альтернативной гипотезы, распределение угла омега будет отличаться от равномерного; плотность распределения вероятностей ФОПК будет отличным от того, что определяется формулой (17). Рассмотрим поведение этих случайных величин на примере двух пар сравниваемых вопросов (рисунок 1). В случае правдивого ответа компоненты вектора ФОПК будут с большей вероятностью принимать положительные значения. А в силу симметричности системы центром распределения вероятностей угла омега в этом случае должен быть угол равный  $45^{\circ}$ . Центром распределения ФОПК в данном случае должно быть значение косинуса от  $45^{\circ}$  равное  $\frac{1}{\sqrt{2}}$ . В случае  $n$  – пар сравниваемых вопросов центром распределения направления вектора ФОПК будет направление, в котором вектор образует с каждой своей координатой угол косинус, которого равен  $\frac{1}{\sqrt{n}}$ . То же самое будет справедливым для лживых ответов, но с учётом знака компонент вектора.

В «Сокол»-е используется идея распределения («нулевое») ФОП соответствующего нулевой гипотезе о равенстве медианы, среднего нулю (статистически, в среднем, одинаковом



типе реагирования опрашиваемого на вопросы в парах сравнения ) Идея использовать некоторое «нулевое» распределение для классификации полиграмм (диагностики правды/лжи) не является новой. Вероятно, первым кто её предложил – это Алексеев Л.Г. Она, правда в несколько в ином виде (с использованием функции распределения вероятностей) лежит в основе алгоритма Quest [8]. Идею «нулевого» распределения можно увидеть в алгоритме Р. Нельсона, который получил название ESS-M [11]. В нём комбинаторным способом строится «нулевое» распределение баллов, исходя из предположения, что баллы получаемые с каждого из датчиков принимают значения, например: -1,0 и 1; и являются равновероятными, то есть, вероятности получить любой из них одинаковы и равны  $\frac{1}{3}$ .

## Имитационное моделирование методом Монте-Карло

Для проверки установленных свойств ФОПК, результатов вычислений, проводилось численное моделирование. Для проверки выводов о том, что ФОПК не отражает свойств физиологических реакций, связанных с масштабом, в котором производятся их измерения, и с масштабом разностей физиологических параметров, а также проверки вычислений о максимальных значениях достигаемых СФОПК в программу «Сокол НТФП» вводились значения показателей физиологических реакций моделирующие детерминированную реакцию физиологических систем на стимул. На Рисунке 7 изображён скриншот программы.

№	1	2	3	4	5	6	7	8	Σ	n
Дых	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	2.83	1.0
КГР...	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	2.83	1.0
АД	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	2.83	1.0
Сумма ФОПК по тесту									8.49	

Рисунок 7. – Главное окно программы «Сокол+ НТФП»

В полях, обведённых красной рамкой, в программу вводятся физиологические показатели реакций в сравниваемых парах вопросов. В полях, выделенных зелёной рамкой результаты расчёта соответствующих значений ФОПК. Все они одинаковы и равны  $\frac{1}{\sqrt{8}}$  (округление до второго знака). Синей рамкой отмечены значения суммарных ФОПК. Как

видно они все одинаковы независимо от сравнительных масштабов показателей, и равны 2,83. Это число является округлённым до двух знаков значением  $\sqrt{8}$ . Где 8 – это число пар сравниваемых вопросов. Показатель детерминированности для всех физиологических показателей равны единице. Увеличим показатель реакции по КГР на контрольный вопрос в одной из пар сравниваемых вопросов. На рисунке 8 во 2-й паре значение амплитуды по КГР увеличено до 330 единиц. Казалось бы, общая положительная оценка полиграммы должна увеличиться, увеличение амплитуды КГР на контрольный вопрос должно свидетельствовать в пользу отнесения полиграммы к положительному классу. Но фактически суммарная ФОПК по КГР уменьшилась до 2.32, показатель детерминированности уменьшился до 0,82. А сумма ФОПК по всему тесту стала равной 7,98. Объясняется такое поведение суммарной ФОПК тем, что причина, которая привела к увеличению разницы показателей по КГР не является откликом физиологической системы на воздействие, а заключается в некотором неконтролируемом стохастическом процессе – поведение системы перестало быть строго детерминированным. Аналогичную картину можно наблюдать на рисунке 9. При увеличении показателя реакции КГР на проверочный вопрос сумма ФОПК также уменьшается, так как система отклоняется от детерминированного поведения. Поведение ФОПК по остальным физиологическим показателям идентично.

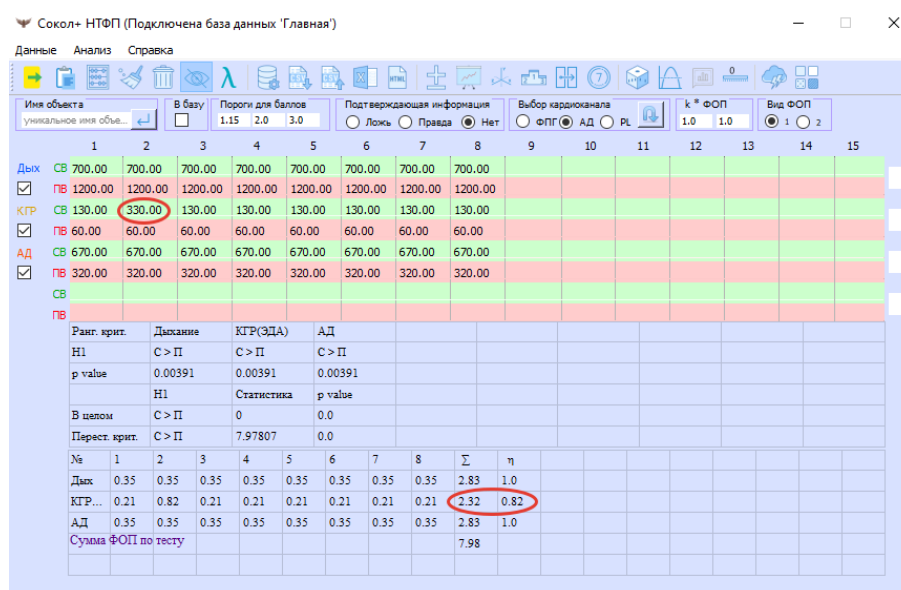


Рисунок 8. – Пример отклонение СФОПК от детерминированного поведения

Для дальнейшей проверки результатов теоретического исследования проводилось моделирование распределений вероятностей угла омега вектора ФОПК и распределение вероятностей косинуса омега – самой ФОПК для случая двух пар сравниваемых вопросов. Генерировались выборки объёмом 500 000 элементов из случайных величин, имеющих нормальное распределение с различными параметрами.

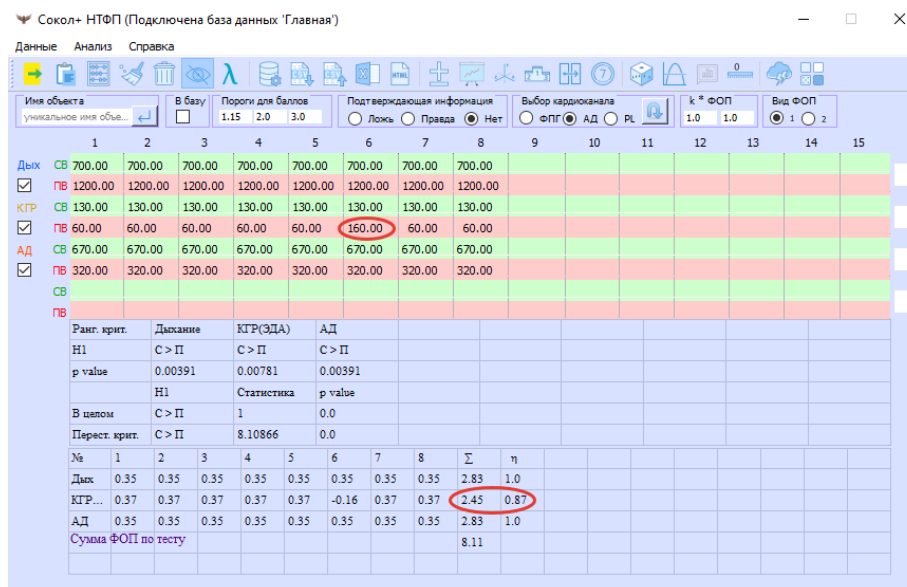


Рисунок 9. – Пример отклонение СФОПК от детерминированного поведения

Случай, соответствующий истинности нулевой гипотезы, моделировался выборкой из  $N(0,9)$ . На Рисунке 10 представлена ядерная оценка плотности распределения вероятностей, которая соответствует гауссовскому шуму – реагированию проверяемого без детерминированной стимулом составляющей. Опрашиваемый в среднем одинаково реагирует на оба типа сравниваемых вопросов. Показатель физиологической реакции на соответствующий тип вопроса с вероятностью 0,5 превышает показатель реакции на другой вопрос в паре.

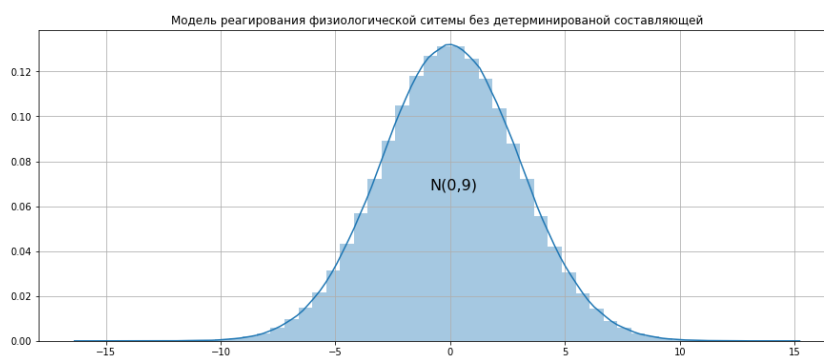


Рисунок 10. – Модель реагирования физиологической системы

На рисунке 11 размещена ядерная оценка плотности вероятностей угла  $\omega$ , полученная на выборке соответствующей гауссовскому шуму без детерминированной составляющей. Как видно оценка свидетельствует о равномерном распределении случайной величины с плотностью вероятности равной 0.16, что соответствует рассчитанному значению по формуле

$$(16) \frac{1}{2\pi} \approx \frac{1}{2 \cdot 3,14} \approx 0,16 .$$



Рисунок 11. – Ядерная оценка плотности угла  $\omega$  при реагировании без детерминированной составляющей

На рисунке 12 изображена ядерная оценка плотности распределения вероятностей ФОПК соответствующих случаю, когда реагирование физиологической системы представляет чистый гауссовский шум («нулевое» распределение ФОПК).



Рисунок 12. – Ядерная оценка плотности ФОПК при реагировании без детерминированной составляющей

Численное моделирование подтверждает теоретическое распределение ФОПК, которое представлено формулой (17) и графиком на рисунке 5. Для проверки вывода о том, что распределения вероятностей угла  $\omega$  и ФОПК при верности нулевой гипотезы не зависит от параметров гауссовского шума (стандартного отклонения) описанное выше численное моделирование проводилось для реагирования физиологической системы, описываемого случайной величиной  $N(0,25)$ . Полученные ядерные оценки плотности совпали с представленными на рисунках 11 и 12, что подтверждает утверждение о том, что в условиях нулевой гипотезы ФОПК для разных физиологических показателей можно рассматривать как выборку из одной генеральной совокупности. Не смотря на непривычный вид «нулевого» распределения ФОПК, плотность распределения вероятностей СФОПК будет иметь привычный колоколообразный вид, что подтверждается численным моделированием этого распределения, ядерную оценку плотности которого для 3 пар сравниваемых вопросов можно увидеть на рисунке 12<sup>a</sup>. Для моделирования реагирования физиологической системы правдивого опрашиваемого с детерминированной составляющей использовалась выборка объемом 500 000 элементов из случайная величина  $N(10,9)$ .



Рисунок 12<sup>а</sup>. – Ядерная оценка плотности «нулевого» распределения СФПК для 3 пар сравниваемых вопросов

На рисунке 13 представлено изображение ядерная оценка плотности распределения вероятностей ФОПК соответствующих данной модели реагирования.

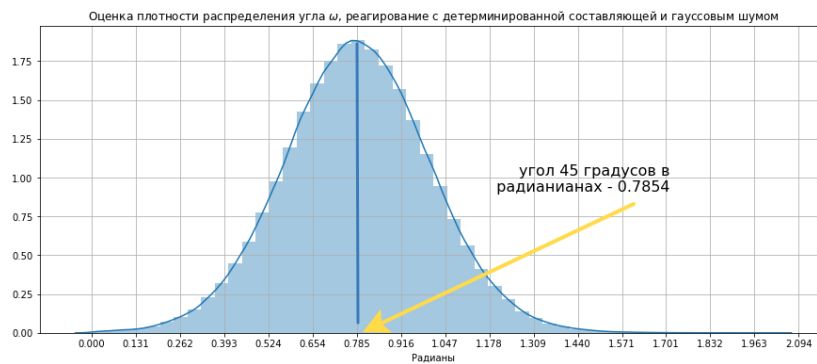


Рисунок 13. – Ядерная оценка плотности ФОПК, реагирование с детерминированной составляющей

Моделирование подтверждает вывод о том, что центром распределения вероятностей угла  $\omega$  будет угол равный  $45^0$ . Среднее выборочное равно 0,785. Медиана – 0,785. Стандартное отклонение 0,217. На рисунке 14 находится ядерная оценка плотности ФОПК для данной модели реагирования.

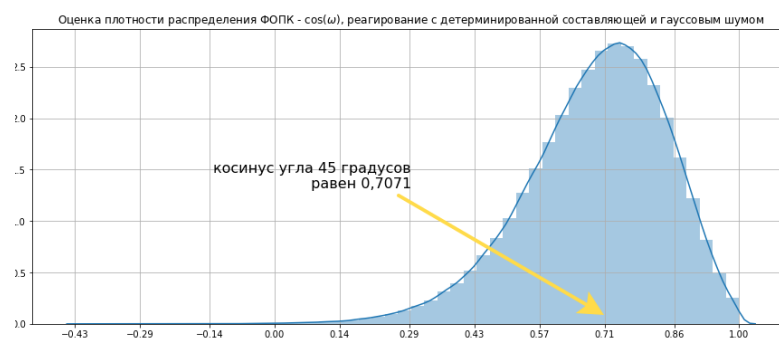


Рисунок 14. – Ядерная оценка плотности ФОПК, среднее выборочное равно 0.690, медиана – 0.707, стандартное отклонение – 0.152

Для моделирования реагирования физиологической системы лживого опрашиваемого с детерминированной составляющей использовалась случайная величина  $N(-15,16)$ . Ядерные оценки распределений угла  $\omega$  и ФОПК для данной модели реагирования находятся на

рисунках 15 и 16. Модель реагирования лживого опрашиваемого отличается от модели правдивого знаком детерминированной составляющей и стандартным отклонением. Но и в этом случае центр распределения угла  $\omega$  соответствует теоретическим представлениям.



Рисунок 15. — Ядерная оценка плотности СФОПК. Среднее выборочное равно 3,927. Медиана – 3,927. Стандартное отклонение – 0,192.

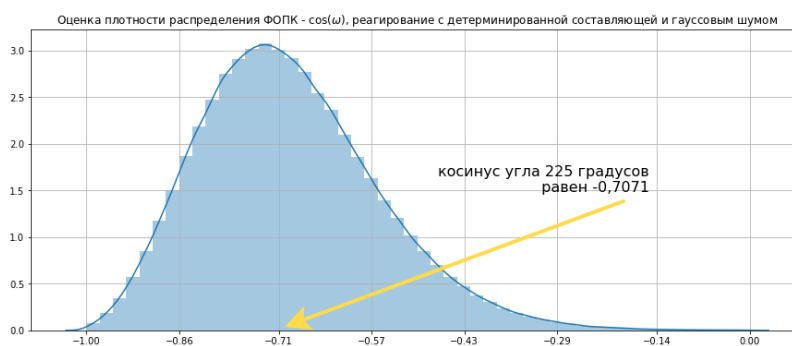


Рисунок 16. — Ядерная оценка плотности СФОПК. Среднее выборочное равно - 0,694. Медиана – 0,707. Стандартное отклонение – 0,135.

Для моделирования реагирования физиологической системы с детерминированной составляющей и малым уровнем гауссовского шума использовалась случайная величина  $N(5,0.0001)$ . Ядерные оценки распределений угла  $\omega$  и ФОПК помещены на рисунки 17 и 18.

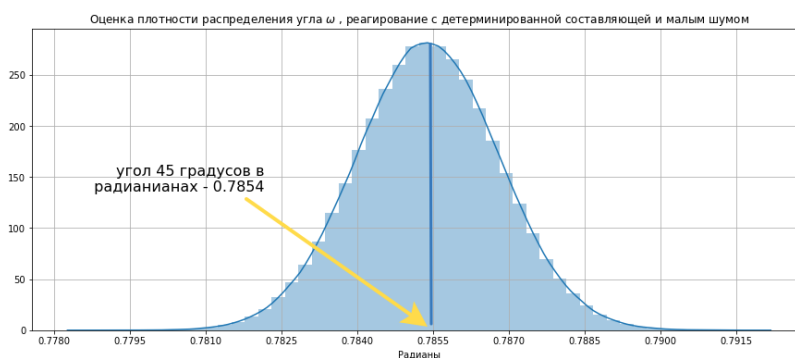


Рисунок 17. — Ядерная оценка плотности СФОПК. Среднее выборочное равно - 0,785. Медиана – 0,785. Стандартное отклонение – 0,001.

Из представленных оценок видно, что при уменьшении уровня физиологического шума уменьшается разброс (дисперсия) значений угла  $\omega$  относительно выделенных направлений (45 и 225 градусов), а значений ФОПК относительно значений косинусов этих углов.



Рисунок 18. – Ядерная оценка плотности СФОПК. Среднее выборочное равно - 0,707. Медиана – 0,707. Стандартное отклонение – 0,001

В пределе при исчезновении шума, когда останется лишь чистый детерминированный ответ физиологической системы на различие в типах вопросов, угол  $\omega$  будет строго равен 45 (225) градусам, значение ФОПК будет строго равно 0,7071 (-0,7071). Эти значения одинаковы для всех измеряемых физиологических параметров.

Подводя итоги численного моделирования, можно сказать, что оно подтвердило свойства ФОПК, полученные аналитическим путём.

### Испытание алгоритма «Сокол»

Испытание алгоритма проводилось в два этапа – на двух выборках полиграмм. Первая – рандомизированная выборка из архива подтверждённых дел Национального центра оценки достоверности информации (США). Объём выборки – сто подтверждённых случаев, из них пятьдесят тестов с правдивыми опрашиваемыми и пятьдесят со лживыми. Выборка, описанная в исследовании Krapohl, Cushman [17], а также ключи к ней, были любезно предоставлены Donald Krapohl. Выборка представляет собой тесты формата FZCT, случайным образом выбранные из архива подтвержденных дел NCCA, все тесты были записаны с помощью компьютерного полиграфа Axciton. Данная выборка использовалась в нескольких исследованиях [Krapohl, Gordon, Lombardi, 2008; Nelson, Handler, Krapohl, 2008] и использовалась для сравнения эффективности систем обсчёта Probability Analysis (в статье [22] она также упоминается как Empirical Bayesian) [Kircher, Raskin, 1988], оценки балльного обсчёта опытными специалистами, OSS-3, ESS [Nelson и др., 2008], HSS [Krapohl и др., 2008] и стала применяться в качестве эталона сравнения точности систем оценки. Полиграммы используемой выборки были размещены в файлах формата pdf. Для измерения показателей реакций снимались скриншоты полиграмм. В графическом редакторе производилось увеличение изображений так, чтобы пятисекундный интервал на полиграмме имела размер на экране монитора 1 сантиметр. Далее производилась разметка полиграммы: вертикальными и горизонтальными линиями отмечались длины кривых, расстояния между точками, которые

подлежали измерению. В настоящее время эти полиграммы находятся в сетевом сервисе «Яндекс-диск» [9]. Измерения на полиграммах производились при помощи программы «Курвиметр» (Ver. 1.4). Программа предназначена для измерения расстояния на географических картах. Измеренное расстояние показывает в километрах с точностью до второго знака. При измерении длин кривых на полиграммах, подобное масштабирование может пониматься как измерение в некоторых условных единицах. Калибровка программы производилась таким образом, чтобы пять пятисекундных интервалов на полиграмме равнялись 50 км, или, что тоже самое, пять сантиметров на экране монитора равнялись 50 км. На полиграммах производилось измерение следующих параметров: длина кривой грудного дыхания, амплитуды манжеты, амплитуды ЭДА. На рисунке 19 в качестве примера размещена полиграмма за номером 344 (только сравниваемые пары вопросов). Значения показателей заносились в программу «Сокол+ НТФП», которая производила следующие вычисления: ФОПК для каждой пары сравниваемых вопросов, суммарное значение для каждого физиологического показателя, сумму ФОПК по всему тесту. Также программа вычисляла альтернативную гипотезу, исходя из порядковых статистик, и проверяла справедливость прямой в одновыборочном критерии знаковых рангов Вилкоксона, вычисляя p-value. Кроме сказанного, программа, используя значение суммарной ФОПК по всему тесту как статистику Перестановочного критерия, исходя из её значения, вычисляла альтернативную гипотезу и проверяла прямую в тесте Перестановочного критерия. Тем самым осуществлялась классификация полиграмм двумя способами с целью их сравнительного анализа. На рисунке 20 размещён скриншот программы с результатами вычислений по данным полиграммы №344. Красным прямоугольником на скриншоте выделена область, где в верхней строчке указана альтернативная гипотеза, вычисленная на основе порядковых статистик, статистика критерия Вилкоксона и p-value, полученное в тесте Вилкоксона. Условное обозначение альтернативной гипотезы:  $\Pi > C - med < 0, C > \Pi - med > 0$ . Во второй строке указывается альтернативная гипотеза, вычисленная по значению ФОПК, статистика Перестановочного критерия и p-value. Условные обозначения альтернативной гипотезы соответствуют случаю, когда среднее меньше/больше нуля. Введённые показатели и вычисленные значения автоматически записывались в базу данных. После обчёта всех полиграмм в табличный процессор заносилась информация о каждой полиграмме, в неё входила информация об истинном классе полиграммы, о вычисленных альтернативных гипотезах и вычисленных p-value. Далее производился анализ таблицы с использованием пакета Pandas – надстройкой над языком Python. На первом шаге для каждого классификатора (критерия Вилкоксона и Перестановочного критерия, только по альтернативной гипотезе) вычислялось количество классифицированных полиграмм по классам, и в соответствии их истинным классам



результаты заносились в таблицы, представленные на рисунке 21. Для каждой категории вычислялись соответствующая чувствительность, специфичность, ошибки DI и NDI.

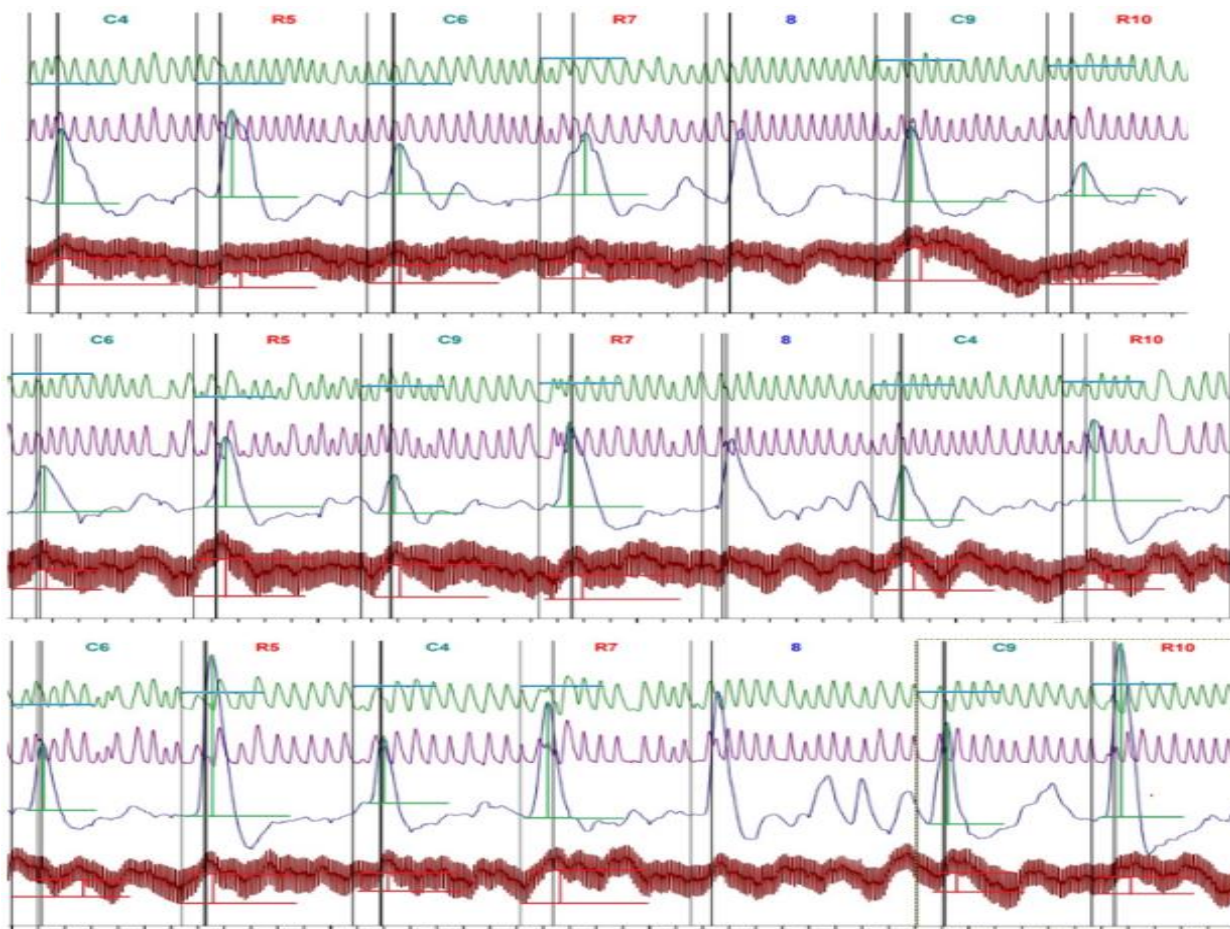


Рисунок 19. – Полиграмма № 344 (из архива подтверждённых дел, США)

Сокол+ НТФП БД 'Подтверждённые'

Данные Анализ Справка

Имя объекта	В базу	Пороги для баллов	Подтверждающая информация	Выбор кардиоканала	k * ФОП	Вид ФОП							
244	<input checked="" type="checkbox"/>	1.15 2.0 3.0	<input type="radio"/> Ложь <input type="radio"/> Правда <input checked="" type="radio"/> Нет	<input type="radio"/> ФПГ <input checked="" type="radio"/> АД <input type="radio"/> РЛ	1.0 1.0	<input checked="" type="radio"/> 1 <input type="radio"/> 2							
Ранг. крит.	Дыхание	КТР(ЭДА)	АД	Бутстреп	t-доверител... интервал	для ср. знач.							
N1	П > С	П > С	С > П	80%	(-0.19 ; -0.03)	П > С							
p value	0.08203	0.01953	0.45508	90%	(-0.22 ; -0.01)	П > С							
	N1	Статистика	p value	95%	(-0.24 ; 0.01)	Н0 не отвер...							
<b>В целом</b>	<b>П &gt; С</b>	<b>110</b>	<b>0.02885</b>										
<b>Пересч. крит.</b>	<b>П &gt; С</b>	<b>-3.15436</b>	<b>0.03382</b>										
№	1	2	3	4	5	6	7	8	9	Σ	V	V	
Дых	0	0	0	0	-1	-1	-1	0	0	-3	-68	126.95	
КТР...	-1	-1	2	-1	-2	-1	-2	-1	-1	-8	-106		
АД	1	0	3	-1	1	1	-1	-1	1	4	16		
Общий балл по тесту											-7		
№	1	2	3	4	5	6	7	8	9	Σ	η		
Дых	-0.07	0.06	0.02	-0.12	-0.46	-0.62	-0.54	0.29	-0.06	-1.5	0.5		
КТР...	-0.08	-0.08	0.28	-0.16	-0.32	-0.18	-0.64	-0.33	-0.49	-1.99	0.66		
АД	0.24	0.04	0.62	-0.42	0.21	0.28	-0.32	-0.39	0.09	0.34	0.11		
Сумма ФОП по тесту											-3.15		

Рисунок 20. – Результаты вычислений для полиграммы № 344.

Общая точность теста не отражает качества классификации в случае несбалансированных классов и может давать искажённую картину. Так как в случае классификации вводится «неопределённый результат» – фактически третий класс, то даже в случае сбалансированных выборок количество полиграмм отнесённых к классам лживых и правдивых может различаться (за счёт разного количества неопределённых исходов).

W Вилкоксона					P Перестановочный Критерий				
	Ложь	Правда	Ошибка%	Вер.		Ложь	Правда	Ошибка%	Вер.
П > С	47	4	7,8	0,92	П > С	48	2	4,0	0,96
С > П	3	46	6,1	0,94	С > П	2	48	4,0	0,96
INC					INC				
	без inc	с inc	inc % =	0,0		без inc	с inc	inc % =	0,0
Sen =	0,94	0,94	F =	0,93	Sen =	0,96	0,96	F =	0,96
Spe =	0,92	0,92			Spe =	0,96	0,96		

Рисунок 21. – Значения метрик классификаторов на выборке (из архива подтверждённых дел, США)

По этой причине для сравнения классификаторов были использованы метриками качества классификаторов из задач машинного обучения:

$$Ac = \frac{TN+TP}{TN+TP+FN+FP} \quad (18)$$

где,  $Ac$  – общая точность (*accuracy*),  $TN$  – правильно классифицированные полиграммы случаев обмана,  $TP$  – правильно классифицированные полиграммы случаев правдивых ответов,  $FN$  и  $FP$  – ошибки классификации полиграмм.  $Ac_w = 0,93$ ;  $Ac_p = 0,96$ .

$$Pr = \frac{TP}{TP+FP} \quad (19)$$

где,  $Pr$  – точность (*precision*) – доля полиграмм случаев правдивых ответов классифицированная как полиграммы с правдивыми ответами,  $Pr_w = 0,94$ ;  $Pr_p = 0,96$ .

$$Re = \frac{TP}{TP+FN} \quad (20)$$

где,  $Re$  – полнота (*recall*) – доля правильно классифицированных полиграмм с лживыми ответами,  $Re_w = 0,92$ ;  $Re_p = 0,96$ .

$$F = \frac{2 \cdot Pr \cdot Re}{Pr+Re} \quad (21)$$

где,  $F$  называется  $F$ -value. Чем больше  $F$ , тем лучше классификатор.  $F$  равно единице при  $Pr = Re = 1$ , и равен нулю, если один из этих показателей равен нулю.  $F$ -value корректно агрегирует точность и полноту в одно значение, максимизация которого означает повышение качества классификации.  $F_w = 0,93$ ;  $F_p = 0,96$ . Для сравнения качества работы алгоритмов классификации полиграмм можно сравнить только две характеристики:  $F$ -value и процент неопределённых исходов.

По всем метрикам качества классификатор на Перестановочном критерии на данной выборке оказался лучше, чем классификатор на критерии Вилкоксона. Разница в эффективности классификаторов предсказуема. Критерий Вилкоксона основанный на

ранжировании вариационного ряда выборки, теряет часть информации об элементах выборки за счёт ранжирования. Перестановочный критерий лишён этого недостатка, он использует всю информацию, которая содержится в элементах выборки. Результат алгоритма классификации полиграмм основанного на тесте Перестановочного критерия, статистически значимо отличается от случайного угадывания на достигаемом уровне значимости  $p = 6,4 \cdot 10^{-6}$ . В таблице 1 находятся 95% доверительные интервалы для чувствительности, специфичности, ошибок DI и NDI классификатора на Перестановочном критерии, которые были получены бутстреп t-методом [10].

**Таблица 1** Доверительные интервалы для параметров алгоритма Сокол

Sen	Spe	Err DI	Err NDI
0,936 < 0,960 < 0,990	0,938 < 0,960 < 0,987	0,012 < 0,04 < 0,061	0,010 < 0,04 < 0,063

Для определения уровня оптимального p-value, сбалансированного по различным показателям для принятия решения, анализировались результаты классификации выборки Перестановочным критерием с шагом p-value 0,05. На рисунке 22 находятся результаты пошагового вычисления параметров. Оптимальным уровнем p-value (на данной выборке) можно принять его величину в 0,45. Он соответствует уровням ошибки менее 5% и уровню неопределённости в 6%. Если сравнить этот результат с тем, что получен при классификации без использования неопределённого результата, то второй, за счёт отсутствия процента неопределённых результатов, выглядит предпочтительнее. Но возможно это является эффектом конкретной выборки. Определение эффективности классификатора будет более полным если сравнить его работу на конкретной выборке с другим классификатором, который является распространённым.

p value < 0.30					p value < 0.35					p value < 0.40					p value < 0.45				
	Ложь	Правда	Ошибка%	Вер.		Ложь	Правда	Ошибка%	Вер.		Ложь	Правда	Ошибка%	Вер.		Ложь	Правда	Ошибка%	Вер.
DI	40	2	4,8	0,95	DI	41	2	4,7	0,95	DI	43	2	4,4	0,96	DI	46	2	4,2	0,96
NDI	1	39	2,5	0,98	NDI	1	42	2,3	0,98	NDI	1	44	2,2	0,98	NDI	1	45	2,2	0,98
INC	9	9			INC	8	6			INC	6	4			INC	3	3		
без inc c inc inc % = 18,0					без inc c inc inc % = 14,0					без inc c inc inc % = 10,0					без inc c inc inc % = 6,0				
Sen =	0,98	0,80	F =	0,96	Sen =	0,98	0,82	F =	0,97	Sen =	0,98	0,86	F =	0,97	Sen =	0,98	0,92	F =	0,97
Spe =	0,95	0,78			Spe =	0,95	0,84			Spe =	0,96	0,88			Spe =	0,96	0,90		

Рисунок 22. – Значения метрик классификатора на выборке (из архива подтверждённых дел, США) с различными правилами принятия решения

Для сравнения было использованы результаты классификации полиграмм пятью российскими полиграфологами, осуществлявших оценку полиграмм, пользуясь системой обчёта ESS (опыт макс – 23 года, опыт мин – 0,5 года, средний опыт работы 8,9 лет). На рисунке 23 представлены усреднённые результаты классификации российскими

полиграфологами и усреднённые результаты десяти американских полиграфологов ((опыт макс – 40 лет, опыт мин – 1 год, средний опыт работы 17 лет), которые производили оценку произвольным, удобным для них способом.

Эксперты(Россия)					Эксперты(США)				
	Ложь	Правда	Ошибка%	Вер.		Ложь	Правда	Ошибка%	Вер.
DI	40,6	4	9,0	0,91	DI	39,6	5,8	12,8	0,87
NDI	3	34,8	7,9	0,92	NDI	6,1	41,2	12,9	0,87
INC	6,4	11,2			INC	4,3	3		
без inc c inc inc % = 17,6					без inc c inc inc % = 7,3				
Sen =	0,93	0,81	F =	0,91	Sen =	0,87	0,79	F =	0,87
Spe =	0,90	0,70			Spe =	0,88	0,82		

Рисунок 23. – Усреднённые результаты классификации полиграмм российскими и американскими полиграфологами

На данной выборке алгоритм классификации «Сокол» эффективнее классификатора ESS:  $F_{Skl}(0,97) > F_{ESS}(0,91)$ ;  $ErrDI_{Skl}, ErrNDI_{Skl} < 0,05$ ;  $ErrDI_{ESS}, ErrNDI_{ESS} > 0,05$ ;  $INC_{Skl}(14\%) < INC_{ESS}(17,6\%)$ . При этом значения чувствительности, специфичности, ошибок ESS не попадают в доверительный 95% интервал, рассчитанный для соответствующих параметров для «Сокол»-а, что говорит о статистически значимом различии между классификаторами, а значит о большей эффективности «Сокол»-а в общем случае. В таблице 2 приведены достигаемые уровни значимости, полученные в Хи2 тестах по проверке гипотезы об однородности двух выборок, которая позволяет выявить статистически значимую разницу между результатами алгоритма классификации полиграмм Сокол и другими алгоритмами на данной выборке. Учитывая, что выборка в 100 полиграмм является небольшой по объёму, то можно предположить, что при увеличении объёма достигаемый уровень значимости уменьшится и разница, которая заключается в эффективности работы алгоритмов, будет обнаружен на уровне 0.05 (за исключением, быть может, ESS-M). Таблицы с результатами сторонних алгоритмов, взятые из зарубежных источников [Nelson и др., 2008; Nelson, 2019], размещены в приложении Д.

**Таблица 2** Достигаемый уровень значимости и метрики алгоритмов

	Сокол	ESS-M	PA <sup>1</sup>	OSS2	OSS3	HSS
p value	1	0,66	0,06	0,01	0,12	0,07
Sensitivity	0,92	0,92	0,78	0,76	0,86	0,82
Specificity	0,90	0,80	0,76	0,72	0,72	0,84
Inc.	0,06	0,08	0,15	0,20	0,14	0,04
Err	0,03	0,07	0,08	0,06	0,07	0,13

На втором этапе испытания алгоритма использовалась выборка полиграмм Российского происхождения, полученная на полиграфах марки «Диана». В ПО «Диана»

<sup>1</sup> Probability Analysis

реализованы два алгоритма классификации полиграмм: компьютерная оценка (Юта), совместно с которой алгоритм ChanceCalc даёт (закрытый коммерческий алгоритм в ПО Диана) «вероятность значимости проверочной темы», и компьютерная оценка (OSS-2) [Крпohl, 2002], хотя изначально данный алгоритм проверялся для ограниченного круга техник, он был модернизирован в программном обеспечении CPSpro фирмы Stoelting таким образом, чтобы можно было считать множество форматов, а исследование [Raskin & Kircher, 2014] доказала эффективность такой практики. Результаты работа алгоритма Сокол сравнивалась с результатами этих алгоритмов на одном и той же выборке полиграмм. Объём выборки 138 полиграмм, из них 77 тестов с подтверждёнными правдивыми ответами и 61 тест с подтверждёнными лживыми ответами. Если полиграммы в американской выборке были получены с использованием одного теста для всей выборки, то в имеющейся российской выборке присутствуют тесты с парами сравниваемых вопросов разных видов. Исходы дел, по которым были проведены тесты и получены полиграммы, известны из признаний опрашиваемых и последующих действий (например, возвращение краденого или следственного эксперимента), с помощью дополнительных улик (например, видеозаписи преступления), следствия или решения суда. Полиграммы присылались от действующих полиграфологов, которые давали описание дела и способ подтверждения<sup>2</sup>. 79 полиграмм теста соответствовали 79-ти опрашиваемым, то есть по одной полиграмме от опрашиваемого, а другие 59 полиграмм были получены с 26-ти опрашиваемых. Из них с 20-ти опрашиваемых было получено по 2 полиграммы, с 5-ти по 3 полиграммы, и с 1-го – 4 полиграммы. В анализируемую выборку необходимо вносить не более одной полиграммы (одного теста) с опрашиваемого. Дело в том, что результаты тестов с одного опрашиваемого являются связанными с друг с другом, что неизбежно исказит результат испытаний алгоритмов. Как правило, если результат теста очевиден (идеальные полиграммы) полиграфологи дополнительные тесты проводили меньше. А по причине «плохой» физиологии или в случае противодействия, когда результат не очевиден, полиграфологи увеличивали количество тестов – оно возрастало до 6 и даже 7. Таким образом в выборке увеличивается количество полиграмм с «плохой» физиологией по отношению к количеству с «хорошей» – выборка перестаёт быть репрезентативной. Алгоритм, ошибаясь на одной полиграмме, допускает ту же ошибку на связанной полиграмме.

---

<sup>2</sup> Галушкина Мария, Давыдов Сергей, Журба Игорь, Катаев Захар, Колесников Владимир, Макаренко Ирина, Моисеева Алла, Ткаченко Евгений, Болховитина София, Ветохин Юрий, Тюлин Александр.

В выборки, на которых производились измерения, случайным образом отбирались ровно по одному тесту от опрашиваемого. Таким образом из имеющейся выборки методом случайного отбора можно получить  $2^{20} \cdot 3^5 \cdot 4 = 1\,019\,215\,872$  выборок размеров в 105 элементов, отличающихся хотя бы одним элементом, и в которую бы входило не более одной полиграммы с одного опрашиваемого. В эту выборку будут входить 62 «правдивых» полиграммы и 43 «лживых». Для уравнивания численности представителей классов из числа «правдивых» случайным образом удалялись 19 полиграмм. В итоге получалось две выборки по 43 полиграммы каждого класса. Далее для обеспечения ротации тех полиграмм, которые были получены в единственном числе, из этих выборок случайным образом удалялось по 3 полиграммы. Сгенерировать таким образом одну подвыборку и на ней провести измерение метрик алгоритмов для сравнительного анализа будет ошибочным решением. Дело в том, что в силу малого объёма выборки и ещё меньшего объёма подвыборки существует не маленькие, но примерно равные вероятности сгенерировать подвыборку, на которой алгоритм покажет как очень хороший и отличный результат, так и вероятность того, что покажет наихудший результат. В Приложении С произведён модельный расчёт, согласно которому вероятности получить подвыборку на которой алгоритм классификации покажет от 0% до 2% ошибок и от 8% до 10% ошибок классификации равна примерно 38%. Это большая вероятность. Такая вероятность нарушает принцип воспроизводства измерений. При повторной генерации подвыборки легко получить противоположный результат. С другой стороны, учитывая возможности современной вычислительной техники, не составит труда сгенерировать подвыборку, на которой классификатор покажет заранее заданный результат. Согласно модельному расчёту в среднем каждая пятая сгенерированная подвыборка будет давать либо наилучший результат, либо наихудший. Так как время генерации одной подвыборки составляет миллисекунды, то время получения подходящего результата будет маленьким. Необходимо учитывать факт того, что разные классификаторы могут ошибаться на разных полиграммах, и легко может сложиться ситуация, когда в подвыборке окажется больше «ошибочных» полиграмм для одного алгоритма, допустим для объективно лучшего, чем для другого. То есть, сравнение алгоритмов на одной подвыборке нельзя считать валидным.

Для извлечения максимума информации из имеющегося массива полиграмм и для объективизации результатов, указанным выше способом были сгенерированы 1000 подвыборок обоих классов. На каждой подвыборке для каждого классификатора вычислялись следующие значения: количество правильно определённых лжецов, количество правильно определённых правдивых опрашиваемых, количество ошибочно определённых лжецов, количества опрашиваемых, ошибочно определённых как правдивые и как лживые. А также подсчитывались результаты с неопределённым исходом. Далее производилось усреднение по

всей тысяче подвыборок и вычисление метрик алгоритмов. Таким образом из анализа был исключён элемент субъективизма, исключена возможность с большой вероятности получить не валидную для сравнения классификаторов подвыборку. Такая методика даёт возможность воспроизвести полученные результаты, так как в силу большого количества подвыборок усреднённые результаты практически не меняются раз от раза. В тысячу подвыборок для каждого алгоритма попадают все полиграммы «плохие» и «хорошие» на которых проверяется их работа, что подтверждается практической неизменностью усреднённых результатов. На рисунке 24 находятся результаты работы «OSS-2» с двумя правилами принятия решения. Первое – решение о правдивости принимается при неотрицательном балле, иначе – о лживости; второе – граничными баллами являются -6 и +6. Результаты работы «Юта» с такими же правилами принятия решения и результаты ChanceCalc находятся на рисунке 25.

Диана Компьютерная оценка (OSS-2)

(-0;+0)	Ложь	Правда	Ош%	Вер.	(-6;+6)	Ложь	Правда	Ош%	Вер.
П > С	32,94	6,07	15,55	0,84	П > С	30,28	4,83	13,75	0,86
С > П	7,06	33,93	17,22	0,83	С > П	4,66	29,12	13,80	0,86
INC	0,00	0,00			INC	5,05	6,05		
без inc c inc inc %= 0,0					без inc c inc inc %= 13,9				
Sen =	0,82	0,82	F = 0,84		Sen =	0,87	0,76	F = 0,86	
Spe =	0,85	0,85			Spe =	0,86	0,73		

Рисунок 24. – Значения метрик OSS-2 на выборке ПО Диана

Следует подчеркнуть, что значения чувствительности и специфичности при втором правиле принятия решения практически совпали с теми, что были установлены для OSS2 на выборке из архива подтверждённых дел (США), при меньшем проценте неопределённых результатов и большем проценте ошибок.

Диана Компьютерная оценка(Юта)					Диана "Оценка вероятности значимости"														
(-0;+0)	Ложь	Правда	Ош%	Вер.	(-6;+6)	Ложь	Правда	Ош%	Вер.	(0.5;0.5)	Ложь	Правда	Ош%	Вер.	(0.25;0.75)	Ложь	Правда	Ош%	Вер.
П > С	36,21	9,61	20,97	0,79	П > С	31,03	5,69	15,50	0,84	П > С	36,21	9,62	21,00	0,79	П > С	30,17	4,42	12,79	0,87
С > П	3,79	30,39	11,08	0,89	С > П	0,93	17,11	5,17	0,95	С > П	3,79	30,38	11,10	0,89	С > П	0,93	17,36	5,09	0,95
INC	0,00	0,00			INC	8,04	17,20			INC	0,00	0,00			INC	8,90	18,21		
без inc c inc inc %= 0,0					без inc c inc inc %= 31,5					без inc c inc inc %= 0,0					без inc c inc inc %= 33,9				
Sen =	0,91	0,91	F = 0,82		Sen =	0,97	0,78	F = 0,84		Sen =	0,91	0,91	F = 0,82		Sen =	0,97	0,75	F = 0,87	
Spe =	0,76	0,76			Spe =	0,75	0,43			Spe =	0,76	0,76			Spe =	0,80	0,43		

Рисунок 25. – Значения метрик ChanceCalc на выборке ПО Диана

Как видно из полученных результатов ChanceCalc превосходит OSS-2 по чувствительности, проигрывая в специфичности. OSS-2 имеет более низкий процент неопределённых результатов. На рисунке 26 находятся результаты работы алгоритма Сокол. Для Сокола приведены два правила принятия решения. Первое –  $p\text{-value} = 0.5$ , что соответствует отсутствию границ принятия решений. Второе –  $p\text{-value}$  меньше 0.237 для альтернативы соответствующей лживости ответов опрашиваемого, и  $p\text{-value}$  меньше 0.47 для альтернативы, соответствующей правдивым ответам.

Сокол									
0.5	Ложь	Правда	Ош%	Вер.	0.237/0.47	Ложь	Правда	Ош%	Вер.
П > С	39,55	4,46	10,14	0,90	П > С	35,06	2,87	7,57	0,92
С > П	0,45	35,54	1,25	0,99	С > П	0,47	35,51	1,30	0,99
	0,00	0,00				4,47	1,62		
	без inc	c inc	inc % =	0,0		без inc	c inc	inc % =	7,6
Sen =	0,99	0,99	F =	0,94	Sen =	0,99	0,88	F =	0,96
Spe =	0,89	0,89			Spe =	0,93	0,89		

Рисунок 26. – Значения метрик Сокола на выборке

Если данные в таблицах на рисунках 24,25,26 округлить, то получим данные с некоторой усреднённой (виртуальной) выборки, то можно провести тесты с целью определения статистически значимой разницы между результатами Сокола и остальных алгоритмов. Сначала это было осуществлено для результатов без порогов принятия решения (без н.р.), а затем с ними (с н.р.). Результаты размещены в таблице 3.

**Таблица 3** Достижимый уровень значимости в Хи2 тесте (Россия)

	OSS2	CCalc (Юта)	CCalc (OB3)
р (без н.р.)	0,064	0,04	0,04
р (с н.р.)	0,095	$1,5 \cdot 10^{-4}$	$6,0 \cdot 10^{-5}$

На российской выборке результат работы Сокола и OSS2 имеют статистически значимую разницу на уровне 0,1. Анализ алгоритмов, проведённый методом генерации подвыборок, показал, что на имеющемся массиве полиграмм значение всех метрик Сокола больше соответствующих метрик ChanceCalc и OSS-2, при меньшем количестве результатов, получивших статус неопределённых.

### Установка пороговых значений

В обычных задачах классификации оценка эффективности алгоритма производится по определённым характеристикам, называемых метриками. В таких задачах пользователя интересует только массовый эффект, который даёт классификация. Правильность классификации в отдельном случае при этом внимания пользователя не привлекают. При опросе на полиграфе специалист-полиграфолог, как правило, несёт определённую персональную ответственность за правильность диагностики в каждом конкретном деле. Поэтому при классификации полиграмм целесообразно введение такого результата как неопределённый исход, который заключается в том, что полиграфолог отказывается выносить решение – диагностировать лживость/правдивость ответов. Делается это в тех случаях, когда вероятность ошибки достаточно велика. Понятие это достаточности условно, но в случае работы с людьми допустимым значением ошибки считается 5%. Иначе говоря, полиграфолог должен принимать решение, когда вероятность правильного диагностирования



лживости/правдивости ответов не менее 95%. Поэтому встаёт задача определения границ принятия решения, исходя из этого условия. Нетрудно показать (приложение В), что вероятность правильного диагностирования лживого ответа в общем случае будет даваться следующей формулой:

$$P_{DI} = \frac{\alpha \cdot Sen}{\alpha \cdot Sen + (1 - \alpha)(1 - Spe - p_{2 inc})} \quad (22)$$

где  $\alpha$  – это априорная вероятность того, что опрашиваемый лжёт,

$Sen, Spe$  чувствительность и специфичность с учётом неопределённых исходов,

$p_{2 inc}$  – вероятность неопределённого исхода для правдивого опрашиваемого.

Переходя к конечным сбалансированным выборкам, будем иметь  $\alpha = \frac{1}{2}$  и можно записать:

$$P_{DI} = \frac{Sen}{1 + Sen - Spe - p_{2 inc}} \quad (23)$$

Учитывая, что в случае конечной выборки  $Sen = \frac{A}{N}$ ,  $Spe = \frac{B}{N}$ ,  $p_{2 inc} = \frac{\bar{B}}{N}$  где  $N$  – количество правдивых/лживых полиграмм (объём всей выборки  $2N$ ),  $A$  – количество правильно диагностированных случаев обмана,  $B$  – количество правильно диагностированных случаев правды,  $\bar{B}$  – количество правдивых случаев, получивших неопределённый статус, а  $\bar{B}$  – количество правдивых случаев, ошибочно диагностированных как лживые, из 32 получим простую формулу:

$$P'_{DI} = \frac{A}{A + \bar{B}} \quad (24)$$

где,  $P'_{DI}$  – оценка вероятности правильной диагностики лжи на выборке.

Из неё видно, что для увеличения вероятности правильной диагностики лживого ответа необходимо уменьшать количество правдивых случаев, ошибочно диагностированных как лживые. Это можно достигнуть, уменьшая граничное p-value для альтернативы соответствующей лживости ответов. Но при этом будет уменьшаться количество правильно диагностированных лживых ответов, они будут переводиться в разряд неопределённых. Таким образом вероятность правильной диагностики лживого ответа будет достигаться за счёт увеличения неопределённых результатов в отношении действительно лживых случаев. Для компенсации роста неопределённых результатов можно увеличивать граничный p-value для альтернативы соответствующей правдивости ответов. Как видно, значение  $P_{DI}$  зависит только от величины p-value для альтернативы соответствующей лживости ответов. Аналогичные рассуждения можно провести и для вероятности правильной диагностики правдивых ответов.

Определение наиболее оптимальных правил принятия решения и других параметров проводилась на выборке из 300 полиграмм. Выборка эта была составлена путём объединения

результатов российских полиграфологов и выборки из архива подтверждённых дел Национального центра оценки достоверности информации (США), которая использовалась для испытания алгоритма. Российские полиграммы были получены на полиграфах Диана, Триумф, Триумф 2, Эпос 10, Лафайет. Метрические данные использовались те, что определялись программным обеспечением полиграфа, с некоторой ручной коррекцией при возникновении такой необходимости. Исключением являлись полиграммы полиграфа Лафайет, данные с них снимались вручную с использованием программного обеспечения Лафайет. В выборке имеется 201 полиграмма соответствующая одному опрашиваемому и 99 полиграмм полученных с 38 опрашиваемых. При этом 24 опрашиваемым соответствовало по 2 полиграммы, семи опрашиваемым – 3, пяти – 4, двум – 5. Из выборки случайным образом отбирались по одной полиграмме так, чтобы одному опрашиваемому соответствовала только одна полиграмма. Таким образом из исходной выборки можно получить  $2^{24} \cdot 3^7 \cdot 4^5 \cdot 5^2 = 939\,309\,347\,635\,200$  выборок, отличающихся друг от друга хотя бы одной полиграммой. Каждая такая выборка будет состоять из 239 полиграмм, при этом «правдивых» полиграмм будет 130, а «лживых» – 109. Таким образом генерировалось 1000 выборок. Каждый раз для балансировки выборки случайным образом удалялись 21 «правдивая» полиграмма и получалась выборка из 109 «правдивых» и 109 «лживых». Далее из каждого класса случайным образом удалялось по девять полиграмм и получалась окончательная сбалансированная выборка из 200 полиграмм. На каждой выборке из этой 1000-и с определёнными граничными *p-value* вычислялись значения метрик, затем все результаты усреднялись. Изменяя значение *p-value*, сначала для альтернативы ложных ответов, затем то же самое для правдивых, с шагом 0.05 были получены соответствующие им значения вероятности правильной диагностики лжи/правды. Далее на этих данных для каждого класса была построена линейная регрессия вероятности правильной диагностики на *p-value*. Построенная линия регрессии для обоих классов помещена на рисунок 27.

Уравнение линейной регрессии для вероятности правильной диагностики лжи даётся формулой 25, а диагностики правды формулой 26

$$P_{DI} = a_1 - a_2 \cdot p - value \quad (25)$$

$$P_{NDI} = b_1 - b_2 \cdot p - value \quad (26)$$

Согласно уравнению регрессии, значение достигаемого уровня значимости при котором вероятность правильной диагностики лжи не меньше 95% равно 0.237.

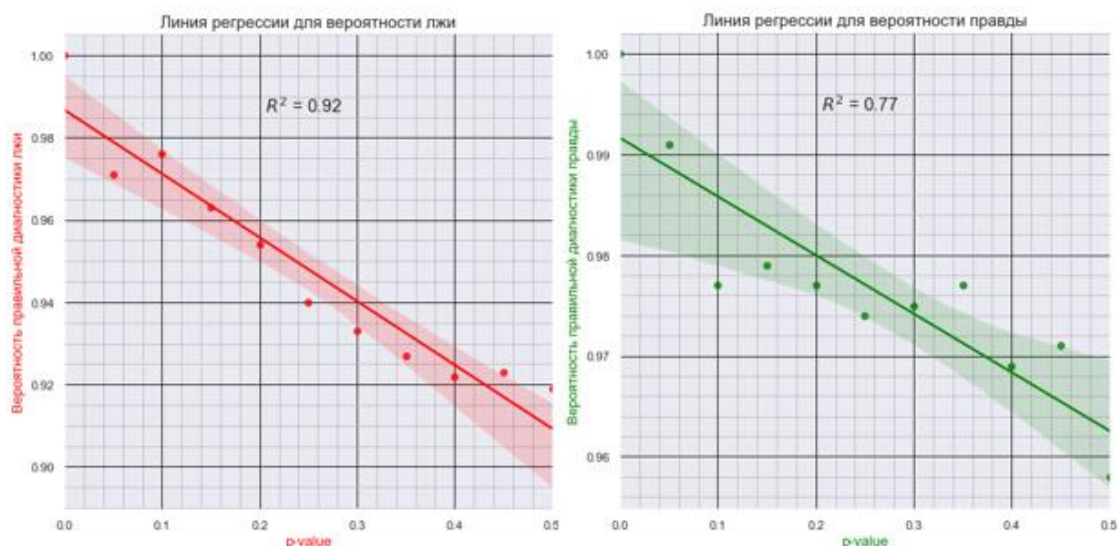


Рисунок 27. – Линии регрессии

Соответствующее значение можно определить как граничное для диагностики обмана. Меньшие значения, согласно полученного уравнения, будут соответствовать большей вероятности. Так как правильная диагностика правды на анализируемой выборке всегда больше 95%, то граничное значение можно не вводить, но всё-таки целесообразно установить хотя бы незначительный порог принятия решения, который был бы приемлем по соответствующим ему метрикам. Путём перебора вариантов лучшим решением, удовлетворяющий указанным требованиям, установлено значение достигаемого уровня значимости в 0.47. На рисунке 28 приведены соответствующие правилу 0.237/0.47 значения метрик

0,237/0,47	Ложь	Правда	Ош.%	Вер.
DI	80,761	5,115	6,0	0,940
NDI	3,054	90,736	3,3	0,967
INC	16,185	4,149		
	без inc	c inc	inc % =	10,2
Sen =	0,96	0,81	F =	0,96
Spe =	0,95	0,91		

Рисунок 28. – Значения метрик на выборке, полученные по правилу 0.237/0.47

Граничным значениям 0.237/0.47 соответствует значение INC для «лживых» полиграмм равное 16.185, что в 3.9 раза больше соответствующего значения для «правдивых» полиграмм. Для симметричного правила 0.35/0.35 наблюдается иная картина, которую можно увидеть на рисунке 29. В этом случае в зоне неопределённости «правдивых» полиграмм оказывается больше. Для первого варианта процент неопределённых результатов почти 10%, что является приемлемым с точки зрения общей эффективности, для второго варианта данная характеристика больше.

0.35/0,35	Ложь	Правда	Ош%	Вер.
П > С	86,462	6,842	7,3	0,927
С > П	1,848	78,207	2,3	0,977
INC	11,690	14,951		
	без inc	с inc	inc % =	13,3
Sen =	0,98	0,86	F =	0,95
Spe =	0,92	0,78		

Рисунок 29. – Значения метрик на выборке, полученные по правилу 0.35/0.35

Так же в первом случае значение F-метрика больше, и в отличие от второго варианта решение принимаем при вероятности правильной диагностики лживости/правдивости не меньше 95% (на выборке это значение меньше, но порог был определён по уравнению регрессии).

### Адаптация алгоритма

Использование алгоритма «Сокол» предусматривает реализацию теста Перестановочного критерия. Вычисление p-value с достаточной точностью (в третьем знаке) в данном тесте требует немалых вычислительных ресурсов. Оно не может быть реализовано вручную или с помощью табличного процессора из-за сложности алгоритма и представляет определённую проблему для слабых компьютерных систем. Но существует возможность проводить приближённые вычисления без выполнения алгоритма статистического теста, с точностью достаточной для целей классификации полиграмм. На рисунке 30 можно увидеть диаграммы рассеяния p-value на суммарной ФОПК для «лживых» и «правдивых» полиграмм.

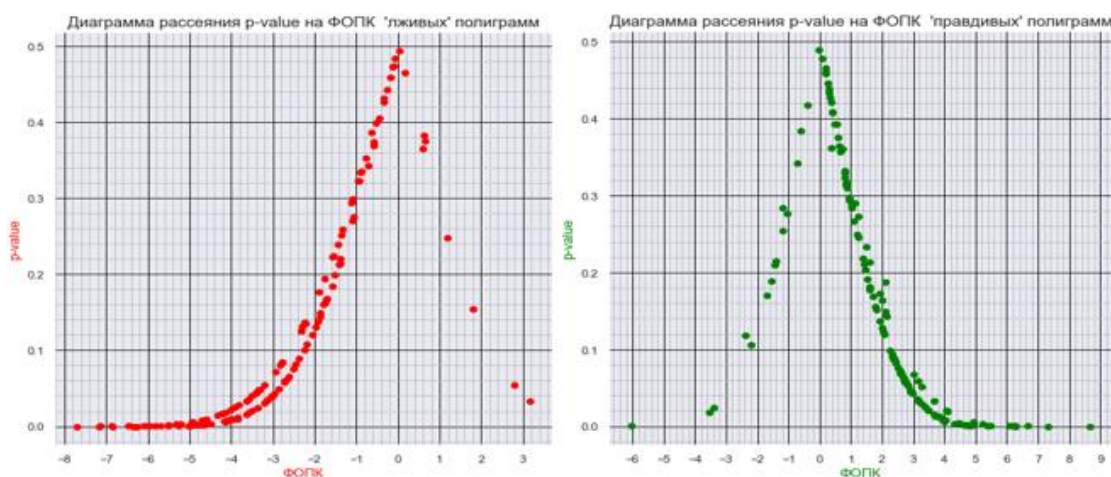


Рисунок 30. – Диаграммы рассеяния p-value на ФОПК.

Отдельно стоящие точки на диаграммах соответствуют ошибочно классифицированным полиграммам. Для «лживых» полиграмм таких точек можно насчитать восемь точек, а для «правдивых» – пятнадцать. На рисунке 31 размещена совместная

диаграмма рассеяния. Можно наблюдать высоко симметричную картину, при связи ФОПК и p-value близкой к функциональной.

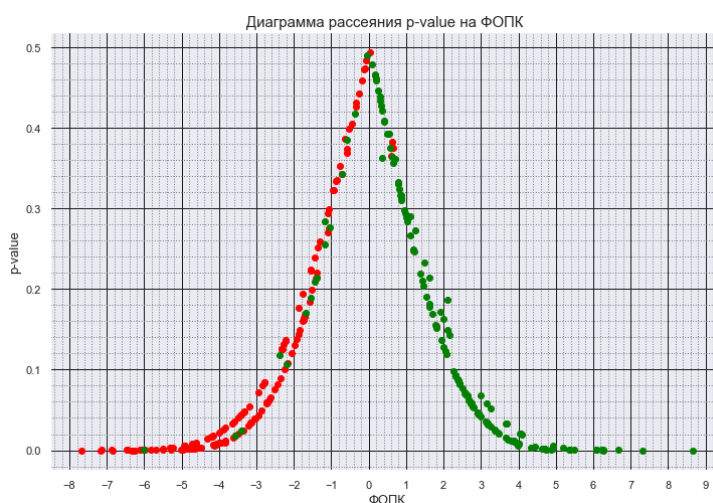


Рисунок 31. — Совместная диаграмма рассеяния p-value на ФОПК.

Такая связь позволяет с помощью полиномиальной регрессии установить достаточно точную формулу. Для этого необходимо сначала проэкспонировать значения ФОПК. На рисунке 32 изображена линия полиномиальной регрессии.

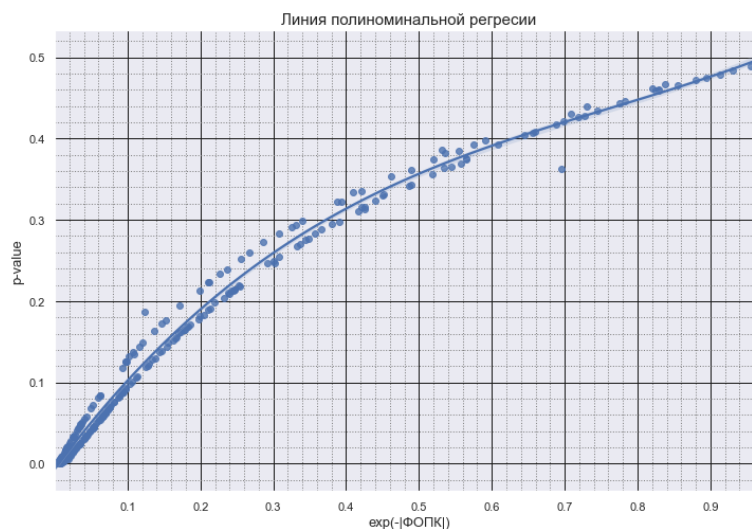


Рисунок 32. — Линия полиномиальной регрессии p-value на ФОПК.

Уравнение полиномиальной регрессии, дающее зависимость p-value от ФОПК ( $f$ ), будет иметь следующий вид:

$$p - value = b_0 + b_1 \cdot e^{(-|f|)} + b_2 \cdot e^{(-2|f|)} + b_3 \cdot e^{(-3|f|)} \quad (27)$$

Значение ФОПК для полиграммы изображённой на рисунке 19 можно увидеть на рисунке 20, там же находится значение p-value вычисленное в тесте Перестановочного критерия. Соответствующее значение p-value вычисленное с помощью формулы (27) равно 0.049 против точного значения 0.034. То есть, учитывая пороги принятия решения получается

такой же результат классификации. Вычисление вероятности правильного определения лжи по формуле (25) даст такие результаты: 0.9794 и 0.9817, что с точностью до второго знака (вполне достаточной для целей классификации) даёт одинаковый результат. Для значения ФОПК равной (-1.20751) соответствующие значения (с точностью до 3-го знака) будут: 0.259 и 0.251; вероятности: 0.947 и 0.948. В этом случае результаты так же с точностью до второго знака совпадают.

Таким образом, с помощью формулы (27), алгоритм может быть адаптирован для слабых компьютеров, или же реализован в табличном процессоре.

### Список литературы

1. Молчанов А.Ю., Оглоблин С.И. Инструментальная детекция лжи: академический курс. – Ярославль: Ньюанс, 2004. – 335 с.
2. Поповичев С.В. Ложь как психофизиологический феномен. //Детекция лжи. 2020. С.116-147
3. Ллойд Э., Ледерман У. (ред.). Справочник по прикладной статистике. Том 2. – М.: Финансы и статистика, 1990. - 526 с.
4. Холлендер М., Вульф Д. Непараметрические методы статистики. – М.: Финансы и статистика, 1983. – 518 с.
5. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д.; под ред. Айвазяна С. А. Прикладная статистика: Классификации и снижение размерности: Справ. изд. / – М.: Финансы и статистика, 1989, – 607 с.
6. Лагутин М.Б. Наглядная математическая статистика: учебное пособие. – М.: БИНОМ. Лаборатория знаний, 2009. – 472 с.
7. Сокол НТФП: [сайт] /2020. – URL: <http://mathpolygraph.tilda.ws/sokol/>. (Дата обращения 03.12.2020)
8. Алексеев Л.Г. Проблемы прикладной психофизиологии. – М.: БукиВеди, 2017. – 302с.
9. Размеченные полиграммы: [Электронный каталог] /2020. – URL: [https://yadi.sk/d/6gxZj3keSPD\\_5g?w=1](https://yadi.sk/d/6gxZj3keSPD_5g?w=1). (Дата обращения 03.12.2020)
10. Шитиков В.К. Статистические методы ресамплинга: рандомизация и бутстреп: [Электронный ресурс] //Сайт Шитикова В.К. и Розенберга Г.С /2020. – URL: <http://www.ievbras.ru/ecostat/Kiril/Article/A32/Stare.pdf>. (Дата обращения 03.12.2020)

11. Nelson R., 2017. Multinomial Reference Distributions for the Empirical Scoring System. *Polygraph & Forensic Credibility Assessment* , 46 (2), 81-115
12. Cacioppo J.T., Berntson G.G., Andersen B.L., 1991 Psychophysiological Approaches to the Evaluation of Psychotherapeutic Process and Outcome, 1991: Contributions from Social Psychophysiology. *A Journal of Consulting and Clinical Psychology*, Vol. 3. No. 3, 321-336, American Psychological Association
13. Dollins A.B., Krapohl D.J. & Dutton D.W., 2000 Computer Algorithm Comparison *Polygraph*, 29(3)
14. Honts C.R., Mark Handler, Shaw P., and Gougler M., 2015 The Vasomotor Response in the Comparison Question Test, *Polygraph*, 44(1) 62-78
15. Kircher J.C. and Raskin D.C., 1988. Human Versus Computerized Evaluations of Polygraph Data in a Laboratory Setting. *Journal of applied psychology*. 73(2) 291-302.
16. Kircher J.C., Kristjansson S.D., Gardner M.K., and Webb A. 2005 Human and Computer Decision-Making in the Psychophysiological Detection of Deception *Polygraph*, 2012 (reprint), 41(2) 77-126
17. Krapohl D.J. and Cushman, B., 2006. Comparison of evidentiary and investigative decision rules: A replication. *Polygraph*, 35(1), 55-63.
18. Krapohl D.J., 2002 Short Report Update for the Objective Scoring System. *Polygraph*, 31(4), pp 298-302
19. Krapohl D.J., Gordon N.J., and Lombardi C., 2008. Accuracy Demonstration of the Horizontal Scoring System Using Field Cases Conducted with the Federal Zone Comparison Technique. *Polygraph*, 37(4), 263-268.
20. Lacey J.I., 1959. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Padoff (Eds.), *Research in psychotherapy* (Vol. 1; pp. 160-208). Washington, DC: American Psychological Association.
21. Nelson R. and Handler M., 2019. Automated Analysis of the Marin Dataset with the ESS-M. *Polygraph & Forensic Credibility Assessment*, 48 (2), 113-123
22. Nelson R., Krapohl D.J., and Handler M., 2008. Brute-Force Comparison: A Monte Carlo Study of the Objective Scoring System version 3 (OSS-3) and Human Polygraph Scorers. *Polygraph*, 2008, 37(3), 185-215.
23. Raskin D.C., Kircher J.C., 2014 Validity of Polygraph Techniques and Decision Methods. *Credibility Assessment*, <http://dx.doi.org/10.1016/B978-0-12-394433-7.00003-8>, Elsevier Inc pp 63-127

24. Reid J.E. 1947 A Revised Questioning Technique in Lie-Detection Tests, *Journal of Criminal Law and Criminology* (1931-1951), Vol. 37, No. 6 (Mar. - Apr.,1947), pp. 542-547



## Приложение А

### Вычисление плотностей вероятностей $\omega$ и $\cos(\omega)$

Запишем угол  $\omega$  как функцию от случайных величин – разностей показателей реакций, в данном случае в двух парах сравниваемых вопросов

$$\omega = \operatorname{arctg}\left(\frac{x_2}{x_1}\right)$$

Тогда для функции распределения вероятностей угла  $\omega$  можно записать:

$$F(\omega \leq Z) = P\left(\operatorname{arctg}\left(\frac{x_2}{x_1}\right) \leq Z\right) = P((x_2, x_1) \in \pi)$$

где  $\pi$  область на Рисунке 1, определяемая уравнением  $\operatorname{arctg}\left(\frac{x_2}{x_1}\right) \leq Z$

Теперь можно записать:

$$F(\omega \leq Z) = \int_{\pi} p(x_1)p(x_2)dx_1 dx_2$$

Сделаем замену переменных под интегралом

$$x_1 = s \quad x_2 = s \cdot \operatorname{tg}(\omega)$$

И подставим выражение плотности нормального распределения

$$\begin{aligned} F(\omega \leq Z) &= \int p(s)p(s \cdot \operatorname{tg}(\omega)) \frac{s}{\cos^2(\omega)} ds d\omega = \\ &= \int_{-\frac{\pi}{2}}^Z \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{s^2}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{s^2 \operatorname{tg}^2(\omega)}{2\sigma^2}} \frac{s}{\cos^2(\omega)} ds d\omega \end{aligned}$$

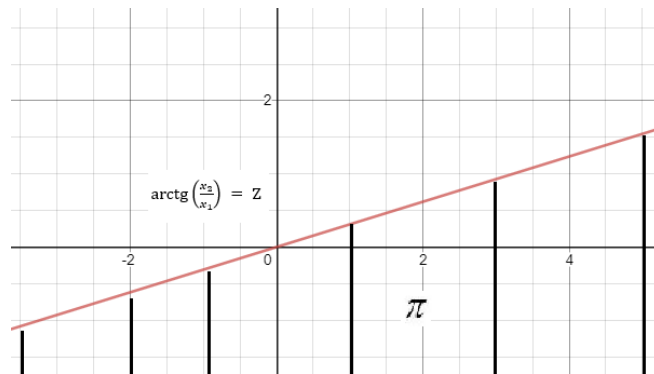


Рисунок 1. – Область, определяемая уравнением  $\operatorname{arctg}\left(\frac{x_2}{x_1}\right) \leq Z$

Таким образом получаем выражение для плотности функции распределения вероятностей угла  $\omega$ , которая будет выражаться через интеграл:

$$p(\omega) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} e^{-\frac{s^2}{2\sigma^2 \cos^2(\omega)}} \frac{s ds}{\cos^2(\omega)}$$

Вычислив интеграл, получим, что

$$p(\omega) = \frac{1}{\pi}$$

Учтя, что арктангенс принимает значения только из  $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$  окончательно получим

$$p(\omega) = \frac{1}{2\pi}$$

Запишем формулу для косинуса угла:

$$\cos(\omega) = \frac{x_2}{\sqrt{(x_1^2 + x_2^2)}}$$

Определим область интегрирования:

$$F(\omega \leq Z) = P\left(\frac{x_2}{\sqrt{(x_1^2 + x_2^2)}} \leq Z\right) = P((x_2, x_1) \in \pi)$$

Область интегрирования  $\pi$  выделена на рисунке 2 красным цветом и определяется неравенством  $\frac{x_2}{\sqrt{(x_1^2 + x_2^2)}} \leq Z$

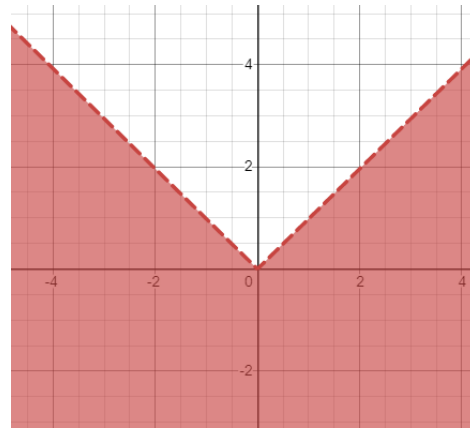


Рисунок 2. – Область, определяемая уравнением  $\frac{x_2}{\sqrt{(x_1^2 + x_2^2)}} \leq Z$

$$F(\cos(\omega) \leq Z) = \int_{\pi} p(x_1)p(x_2)dx_1 dx_2$$

Произведём замену переменных под интегралом

$$x_1 = s \quad \frac{x_2}{\sqrt{(s^2 + x_2^2)}} = t$$

$$F(\cos(\omega) \leq Z) = \int_{-\infty}^Z dt \int_{-\infty}^{+\infty} p(s)p\left(\frac{\mp t \cdot s}{\sqrt{(1-t^2)}}\right) \frac{1}{2} \frac{ds^2}{(1-t^2)^{\frac{3}{2}}}$$

Отсюда для плотности вероятности получаем

$$p(\cos \omega) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} e^{-\frac{s^2}{2\sigma^2}} e^{-\frac{t^2 s^2}{2\sigma^2(1-t^2)}} \frac{(\mp s)ds}{(1-t^2)^{\frac{3}{2}}}$$

$$p(\cos \omega) = \frac{1}{2 \cdot \pi \sigma^2 \cdot \sqrt{1-t^2}} \int_{-\infty}^{+\infty} e^{-\frac{s^2}{2\sigma^2(1-t^2)}} \frac{1}{2} \frac{ds^2}{(1-t^2)}$$

$$p(\cos \omega) = \frac{1}{2\pi\sqrt{1-t^2}} \int_{-\infty}^{+\infty} e^{-q} dq = \frac{1}{\pi\sqrt{1-t^2}} \int_0^{+\infty} e^{-q} dq = \frac{1}{\pi\sqrt{1-t^2}}$$

## Приложение Б

### Вычисление вероятности правильной диагностики лживого ответа

Введём следующие обозначения:

$D$  – событие, заключающееся в том, что опрашиваемый действительно лгал в ходе теста

$\bar{D}$  – событие противоположное  $D$  – опрашиваемый отвечал правдиво

$C$  – событие, заключающееся в том, что результат теста – опрашиваемый лгал

$\bar{C}$  – результат теста – опрашиваемый отвечал правдиво

$\bar{\bar{C}}$  – результат теста – неопределённый

$D \cdot C$  – опрашиваемый действительно лгал, а результат теста – опрашиваемый лгал

$$P(D \cdot C) = P(D)P(C|D) = P(C)P(D|C) \Rightarrow P(D|C) = \frac{P(D)P(C|D)}{P(C)}$$

$P(D|C)$  – условная вероятность, вероятность того, что опрашиваемый действительно лгал при условии того, что результат теста – опрашиваемый лгал. Это вероятность **правильной диагностики лжи** будем обозначать  $P_{DI}$ .

$P(D)$  – априорная вероятность, вероятность того, что опрашиваемый действительно лгал, обозначим её  $\alpha$ .

$P(C|D)$  – условная вероятность, вероятность того, что результатом теста будет «опрашиваемый лгал», если опрашиваемый действительно лгал. Это чувствительность теста, обозначим  $Sen$ .

Событие  $C$  произойдёт в том случае, если произойдёт одно из событий:  $C \cdot D$  или  $C \cdot \bar{D}$

$$P(C) = P(C \cdot D) + P(C \cdot \bar{D}) = P(D)P(C|D) + P(\bar{D})P(C|\bar{D}) = \alpha \cdot Sen + (1 - \alpha) \cdot P(C|\bar{D})$$

$$P(C|\bar{D}) + P(\bar{C}|\bar{D}) + P(\bar{\bar{C}}|\bar{D}) = 1 \Rightarrow P(C|\bar{D}) = 1 - P(\bar{C}|\bar{D}) - P(\bar{\bar{C}}|\bar{D})$$

$P(\bar{C}|\bar{D})$  – условная вероятность, вероятность того, что результат теста будет «опрашиваемый отвечал правдиво», при условии, что опрашиваемый отвечал действительно правдиво. Это специфичность теста, обозначим  $Spe$ .

$P(\bar{\bar{C}}|\bar{D})$  – условная вероятность, вероятность того, что результат теста будет неопределённый при условии, что опрашиваемый отвечал действительно правдиво, обозначим  $p_{2 inc}$

Имеем

$$P_{DI} = \frac{\alpha \cdot Sen}{\alpha \cdot Sen + (1 - \alpha) \cdot (1 - Spe - p_{2 inc})}$$

Для  $\alpha = 0.5$

$$P_{DI} = \frac{Sen}{Sen + (1 - Spe - p_{2 inc})}$$

Оценки на конечной выборке

$Sen = \frac{A}{N}$ ,  $Spe = \frac{B}{N}$ ,  $p_{2inc} = \frac{\bar{B}}{N}$  где  $N$  – количество правдивых/лживых полиграмм (объём всей выборки  $2N$ ),  $A$  – количество правильно диагностированных случаев обмана,  $B$  – количество правильно диагностированных случаев правды,  $\bar{B}$  – количество правдивых случаев, получивших неопределённый статус, а  $\bar{B}$  – количество правдивых случаев, ошибочно диагностированные как лживые.

$$P'_{DI} = \frac{\frac{A}{N}}{\frac{A}{N} + \left(1 - \frac{B}{N} - \frac{\bar{B}}{N}\right)} = \frac{A}{A + (N - B - \bar{B})} = \frac{A}{A + \bar{B}}$$

где,  $N = B + \bar{B} + \bar{B}$

Если выборка, на которой производятся измерения является репрезентативной, то мы получим хорошую оценку вероятности правильной диагностики лживого ответа для всей генеральной совокупности.

## Приложение В

### Вычисление вероятности получения наилучшего и наихудшего результата при генерации подвыборки

Пусть имеется выборка объёмом 100 полиграмм полученных в тестах, в которых опрашиваемые лгали, отвечая на проверочные вопросы. Так же пусть на 95 из них алгоритм классификации даёт правильные результаты, а на 5-ти ошибается. Если из этой выборки равновероятным образом отобрать подвыборку из пятидесяти полиграмм, то количество полиграмм на которых алгоритм ошибается будет подчиняться биномиальному закону. Так как в этом случае вероятность (успеха) оказаться в подвыборке будет равна 0.5, то для вероятности попадания в подвыборку нуля «ошибочных» полиграмм можно записать можно записать:

$$P(0) = \frac{5!}{0! 5!} \cdot \left(\frac{1}{2}\right)^5 = \left(\frac{1}{2}\right)^5 = 0,03125$$

И для попадания одной «ошибочной» полиграммы

$$P(1) = \frac{5!}{1! 4!} \cdot \left(\frac{1}{2}\right)^5 = 5 \cdot \left(\frac{1}{2}\right)^5 = 0,15625$$

Тогда вероятность попадания нуля или одной плохой полиграммы будет равна 0,1875 или, примерно, 19%. Ноль «ошибочных» полиграмм означает, что классификатор сделает ноль ошибок на полученной подвыборке, а одна, попавшая в подвыборку «ошибочная» полиграмма, будет означать 2% процента ошибок классификатора на полученной подвыборке. Ноль процентов и два процента – это отличный и наилучший результат, который может показать алгоритм классификации. Прделаем те же вычисления для попадания в подвыборку пять и четырёх «ошибочных» полиграмм.

$$P(5) = \frac{5!}{5! 0!} \cdot \left(\frac{1}{2}\right)^5 = \left(\frac{1}{2}\right)^5 = 0,03125$$

$$P(4) = \frac{5!}{4! 1!} \cdot \left(\frac{1}{2}\right)^5 = 5 \cdot \left(\frac{1}{2}\right)^5 = 0,15625$$

Вероятность попадания пяти или четырёх «ошибочных» полиграмм будет также равна 19%. При этом процент ошибок классификатора на подвыборке будет равен 10-ти и 8-ми соответственно. Таким образом вероятность получить экстремальный результат будет равна 38%.

Математическое ожидание количества «ошибочных» полиграмм, попавших в подвыборку, равно 2.5, что в процентном отношении будет равно 5%, как и в исходной выборке. Таким образом, если сгенерирован достаточно большое количество подвыборок и

усреднить результат, то среднее количество ошибок алгоритма будет равно 5% – как в исходной выборке.

## Приложение Г

### Результаты алгоритмов на выборке из архива доказанных дел (США)

Данные, находящиеся в нижеследующей таблице, были восстановлены по таблице 3 [21]

Таблица 1 Результаты работы алгоритмов (США)

Алгоритм	TN	TP	Corr.	FN	FP	Err.	IN	IP	INC
ESS-M	46	40	86	2	4(5)	6(7) <sup>3</sup>	2	6	8
PA	39	38	77	7	8	8	4	4	15
OSS2	38	36	74	2	4	6	10	10	20
OSS3	43	36	79	2	5	7	5	9	14
HSS	41	42	83	4	0	13	4	0	4

Таблица 2 Результаты работы алгоритма Сокол

Алгоритм	TN	TP	Corr.	FN	FP	Err.	IN	IP	INC
Сокол	46	45	91	1	2	3	3	3	6

ESS-M

	Ложь	Правда	Ошибка%	Вер.
DI	46	4	8,0	0,92
NDI	2	40	4,8	0,95
INC	2	6		
без inc		с inc	inc % = 8,0	
Sen =	0,96	0,92	F = 0,93	
Spe =	0,91	0,80		

OSS2

	Ложь	Правда	Ошибка%	Вер.
DI	38	4	9,5	0,90
NDI	2	36	5,3	0,95
INC	10	10		
без inc		с inc	inc % = 20,0	
Sen =	0,95	0,76	F = 0,92	
Spe =	0,90	0,72		

PA

	Ложь	Правда	Ошибка%	Вер.
DI	39	4	9,3	0,91
NDI	4	38	9,5	0,90
INC	7	8		
без inc		с inc	inc % = 15,0	
Sen =	0,91	0,78	F = 0,90	
Spe =	0,90	0,76		

HSS(горизонтальная система оценки)

	Ложь	Правда	Ошибка%	Вер.
DI	41	8	16,3	0,84
NDI	5	42	10,6	0,89
INC	4	0		
без inc		с inc	inc % = 4,0	
Sen =	0,89	0,82	F = 0,87	
Spe =	0,84	0,84		

OSS3

	Ложь	Правда	Ошибка%	Вер.
DI	43	5	10,4	0,90
NDI	2	36	5,3	0,95
INC	5	9		
без inc		с inc	inc % = 14,0	
Sen =	0,96	0,86	F = 0,91	
Spe =	0,88	0,72		

Рисунок 1 – Расчёт показателей алгоритмов

<sup>3</sup> По всей вероятности, в таблице 3 в работе [21] содержится ошибка, которая приводит к тому, что объём выборки должен быть равен 101. Вероятнее всего ошибка произошла за счёт увеличения FP на единицу. Поэтому для расчётов бралось значение FP равное 4