УДК 004.93

Формула Ликкена

Летков Юрий Владимирович

Полиграфолог, разработчик СППРП «Сокол», г. Вологда, Россия

math-polygraph@yandex.ru

Аннотация. В статье дан вывод распределения вероятности общего суммарного ранга релевантного вопроса в тестах Ликкена. Показано, что это распределение совпадает с распределением в таблице Ликкена для 10 тестов знания виновного с точностью до ошибок округления. Так же показано, что комбинаторное распределение Бозе-Эйнштейна (задача о шарах и ящиках), которое лежит в основе квантового статистического распределения Бозе-Эйнштейна, с распределением общего суммарного ранга в тестах Ликкена, и тестовых рядов МВСИ – с формулой Ликкена связано условно.

Ключевые слова: алгоритм классификации полиграмм, ранжирование реакций, тест на знания виновного, диагностика значимости релевантного стимула, распределение Бозе-Эйнштейна.

Letkov Yuri Vladimirovich

Polygraph examiner, developer of "Sokol", Vologda, Russia

math-polygraph@yandex.ru

Annotation. The article gives the conclusion of the probability distribution of the total t rank of the relevant question in the Likken tests. It is shown that this distribution coincides with the distribution in the Likken table for 10 tests of knowledge of the culprit up to rounding errors. It is also shown that the combinatorial Bose–Einstein distribution (the problem of balls and boxes), which underlies the quantum statistical Bose–Einstein distribution, is not related to the distribution of the total rank in the Likken tests, and the test series of the MVSI, with the Likken formula.

Keywords: polygram classification algorithm, reaction ranking, culprit knowledge test, relevant stimulus significance diagnostics, Bose–Einstein distribution.

Введение

Для оценки своих тестов на знания виновного американский психофизиолог Дэвид Ликкен создал систему, которую назвал «методом средних рангов».

«Метод оценки GKT, учитывающий такие возможности, — это метод средних рангов. Сначала ранжируются реакции на каждый вопрос, а затем усредняются ранги на релевантные вопросы. Вероятность того, что невиновный подозреваемый получит определенный средний ранговый балл на релевантный вопрос, можно определить из таблицы, подобной табл. 21.1.» [1, стр. 302] (перевод С.В. Поповичева [2]). Ниже изображена таблица из книги Ликкена «Дрожь в крови».

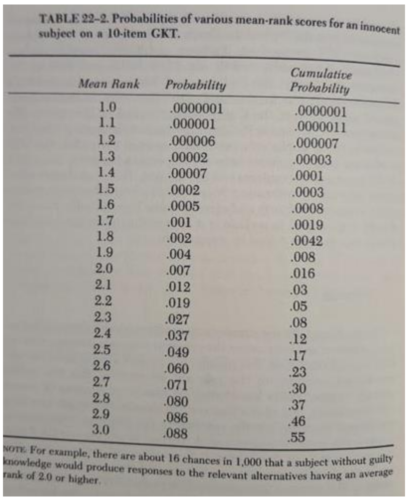


TABLE 22–2. Probabilities of various mean-rank scores for an innocent subject on a 10-item GKT.

Mean Rank	Probability	Cumulative Probability
1.0	.0000001	.0000001
1.1	.000001	.0000011
1.2	.000006	.000007
1.3	.00002	.00003
1.4	.00007	.0001
1.5	.0002	.0003
1.6	.0005	.0008
1.7	.001	.0019
1.8	.002	.0042
1.9	.004	.008
2.0	.007	.016
2.1	.012	.03
2.2	.019	.05
2.3	.027	.08
2.4	.037	.12
2.5	.049	.17
2.6	.060	.23
2.7	.071	.30
2.8	.080	.37
2.9	.086	.46
3.0	.088	.55

NOTE: For example, there are about 16 chances in 1,000 that a subject without guilty knowledge would produce responses to the relevant alternatives having an average rank of 2.0 or higher.

Рис.1 Таблица средних рангов Ликкена

Следуя тексту Ликкена, чтобы получить суммарный ранг, необходимо произвести ранжирование реакций и вычислить общий суммарный ранг. Так же, чтобы получить распределение вероятностей как на таблице Ликкена необходимо сначала научиться вычислять количество случаев $\boldsymbol{N}$ в результате, которых может быть получен общий суммарный ранг как функцию от количества $\boldsymbol{m}$ – стимулов теста, $\boldsymbol{k}$ – количества используемых физиологических признаков реакции, $\boldsymbol{l}$ – количество повторов теста. Сам Ликкен в своих тестах использовал только амплитуду КГР, мы получим формулу для $\boldsymbol{N}$ для произвольного $\boldsymbol{k}$.

Для решения поставленной задачи нам необходимо обратиться к уже известной и решенной задаче. Это задача о размещении $\boldsymbol{m}$ неразличимых шаров в $\boldsymbol{n}$ ящиках. Эта комбинаторная задача, которая лежит в основе получения квантомеханического статистического распределения бозонов по энергетическим уровням. Поэтому и комбинаторике формулу, получаемую в результате решения этой задачи, называют иногда статистикой Бозе-Эйнштейна. В этой задаче нет никаких ограничений на распределение шаров по ящикам. В какой-либо ящик может не попасть ни одного шара, а могут все шары попасть в один ящик. Различными считаются распределения шаров по ящика, отличающиеся только числом попавших в каждый ящик шаров [2, стр. 34]. Задача состоит в том, чтобы найти $\boldsymbol{F}$ – количество всех возможных, различных распределений шаров по ящикам. Решение даётся формулой:

$$\boldsymbol{F} = \boldsymbol{C}_{\boldsymbol{m}+\boldsymbol{n}-\boldsymbol{1}}^{\boldsymbol{n}-\boldsymbol{1}} = \frac{(\boldsymbol{m}+\boldsymbol{n}-\boldsymbol{1})!}{\boldsymbol{m}!(\boldsymbol{n}-\boldsymbol{1})!} \tag{1}$$

Вывод формулы Ликкена

Рассмотрим таблицу с рангами для ТЗВ с пятью стимулами, тремя физиологическими признаками, и двумя повторами. Ранг, полученный стимулом, обозначим не числом, а количеством шариков равным рангу.

	ВДЛД	КГР	ФПГ	ВДЛД	КГР	ФПГ	ОСР
	1	2	3	4	5	6	*r*
С1	●●	●●	●●	●	●●●●	●●●●●	16
С2	●●●	●●●●	●●●●●	●●●●	●	●●	19
R	●	●	●	●●●	●●	●	9
С3	●●●●●	●●●	●●●	●●●●●	●●●	●●●	22
С4	●●●●	●●●●●	●●●●	●●	●●●●●	●●●●	24
Сумма	15	15	15	15	15	15	90

Рис.2 Таблица с рангами теста ТЗВ

Ячейки, соответствующие релевантному вопросу, можно представить в виде ящиков, а ранги в виде соответствующего количества шариков. И сразу же определимся с ограничениями этой модели. Обратим внимание на эти ограничения:

1. Ни один из ящиков не может быть пуст. В нём должен быть хотя бы один шарик, так как не может быть нулевого ранга;
2. В каждом ящике не может быть более 5-ти шаров, так как максимальный ранг равен 5-ти.

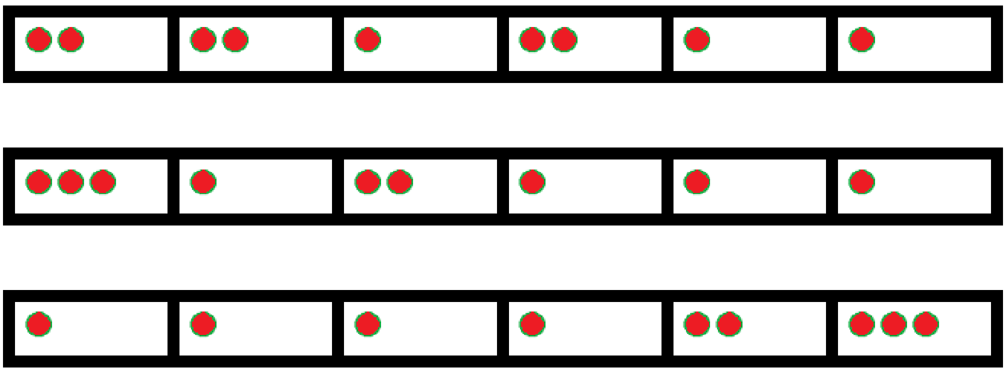
Если релевантный стимул является незначимым для опрашиваемого лица, и все стимулы теста гомогенны, то опрашиваемый никак его не выделяет на фоне других. Это справедливо и каждого из вопросов ряда. Тогда в любом из 6-ти ящиков может оказаться любое количество (от 1 до 5) шариков с одинаковой вероятностью. Это значит, что любая комбинация шаров в ящике равновозможная и равновероятна.

	ВДЛД	КГР	ФПГ	ВДЛД	КГР	ФПГ	ОСР
	1	2	3	4	5	6	
R	●	●	●	●●●	●●	●	9

В данном случае у нас имеется 6-ть ящиков. Это число равно произведению количеству признаков ($\boldsymbol{k}$ = 3) на количество повторов теста ($\boldsymbol{l}$ = 2). $6 = 3 \cdot 2$. Для произвольного случая обозначим количество ящиков как $\boldsymbol{n}$, тогда:

$$\boldsymbol{n} = \boldsymbol{k} \cdot \boldsymbol{l}$$

Зададимся вопросом: сколькими способами можно разместить $\boldsymbol{r}$ шаров в $\boldsymbol{n}$ ящиках с теми двумя ограничениями, которые мы перечислили выше? Например, для рассматриваемого случая эта задача сведётся к более простой: сколькими способами можно разместить 3-и шара в 6-ти ящиках. Могут быть такие 3-и варианта.



А сколько их всего для $\boldsymbol{r}$ = 9? А для произвольного $\boldsymbol{r}$? Шары у нас неразличимы и по одному мы должны оставлять в каждом ящике. Таким образом мы можем перекладывать по ящикам только три свободных шара. Ответом на такую задачу является формула (смотри задачу и формулу (1)):

$$\boldsymbol{F} = \boldsymbol{C}_{\boldsymbol{r-n+n-1}}^{\boldsymbol{r-n}} = \boldsymbol{C}_{\boldsymbol{r-1}}^{\boldsymbol{r-n}} = \frac{(\boldsymbol{r-1})!}{(\boldsymbol{r-n})!\cdot(\boldsymbol{n-1})!} \qquad (2)$$

Для нашего случая имеем $\frac{8!}{3!\cdot 5!} = 56$ способов. А сколько всего размещений шаров может быть в ящиках? В 1-ом ящике может быть 5 вариантов размещения шаров (от 1 до 5). При этом для каждого варианта размещения шаров в первом ящике во втором ящике будет так же пять вариантов. Тогда для двух ящиков будет $5 \cdot 5 = 5^2 = 25$ вариантов размещения шаров. А для шести ящиков будет $\boldsymbol{N} = 5^6 = 15\,625$ размещений шаров. Из этих $15\,625$ вариантов в 56 случаях будет общее количество шаров равное 9-ти. Это значит, что если релевантный стимул не значим, то вероятность, что общий суммарный ранг буде равен 9-ти равна $\frac{56}{15\,625} \approx 0{,}004$. Найти вероятность для ОСР равной 9-ти оказалось достаточно просто потому, что нам не пришлось учитывать второе ограничение. Произошло это потому, что свободных шаров оказалось всего 3-и и второе ограничение выполнялось автоматически. Как бы мы не размещали свободные шары в любом ящике будет меньше пяти шаров. Задача существенно усложниться, если ОСР будет, например, 19. Свободных шаров будет 13 и при подсчёте необходимо исключать варианты размещения шаров, когда их в ящике окажется больше 5-ти. Для этого необходимо использовать принцип включений-исключений. Для начала сосчитаем количество размещений 19-то шаров в 6-ти ящиках с учётом ограничения 1, и не учитывая ограничения 2. Для этого мы можем воспользоваться формулой (2)

$$\boldsymbol{F} = \boldsymbol{C}_{\boldsymbol{18}}^{\boldsymbol{13}} = \frac{18!}{5!\cdot 13!} = 8568 \qquad (3)$$

Но так как мы не учитывали ограничение 2 мы таким образом сосчитали и те случаи, когда в ящиках оказывалось более 5-ти шаров. А это случаи невозможны. Поэтому надо вычесть из этого числа те случаи, когда в ящике оказалось не менее 6-ти шаров. Пусть в одном ящике оказалось 6-ть шаров. Тогда осталось свободных (с учётом первого ограничения) 19 – 6 – 5 = 8. Теперь эти шары мы можем разместить в любой из 6-ти ящиков. Это можно сделать $C_{13}^{8} = \frac{13!}{5!8!} = 1287$ способами. Необходимо ещё учесть, что 6-ть шаров может оказаться в любом из 6-ти ящиков. Таким образом из (3) необходимо вычесть число $C_6^1 \cdot C_{13}^8 = 6 \cdot 1287 = 7722$. Но это еще не всё! Так как у нас 13 свободных шаров, то по 6-ть шаров может оказаться в двух ящиках. Тогда свободных шаров останется всего 3, и ситуация, когда в трех ящиках окажется по 6-ть шаров уже невозможна. Вычитая, случаи, когда в одном из ящиков оказалось 6-ть шаров мы тем самым дважды вычли те случаи, когда в ящиках оказалось по 6-ть шаров. Поэтому теперь нам надо сосчитать число случаев, когда в 2-х ящиках оказалось по 6-ть шаров и прибавить для компенсации двойного вычитания. Два ящика из шести можно выбрать $C_6^2 = \frac{6!}{2!4!} = 15$ способами. А разместить 3 свободных шара по шести ящикам можно $C_8^3 = \frac{8!}{3!5!} = 56$ способами. Таком образом мы должны прибавить $C_6^2 \cdot C_8^3 = 15 \cdot 56 = 840$. И окончательно

$$\boldsymbol{K} = \boldsymbol{C_{18}^{13}} - \boldsymbol{C_6^1} \cdot \boldsymbol{C_{13}^8} + \boldsymbol{C_6^2} \cdot \boldsymbol{C_8^3} = 8568 - 7722 + 840 = 1686$$

Теперь можно вычислить вероятность ранга равного 19-ти

$$\mathbf{P(r = 19, n = 6, m = 5)} = \frac{\mathbf{K(r, n, m)}}{\mathbf{N(m, n)}} = \frac{\mathbf{C_{18}^{13} - C_6^1 \cdot C_{13}^8 + C_6^2 \cdot C_8^3}}{\mathbf{m^n}} = \frac{1686}{15625} \approx 0{,}11$$

Теперь запишем формулу Ликкена для произвольного случая. Для этого получит формулу числа комбинаций рангов показателей физиологических признаков для определённого суммарного ранга r. Обозначим это число как ***F (r, n, m).*** Положим в каждый ящик по одному предмету. У нас останется ***(r – n)*** предметов, которые необходимо будет разместить по **n** ящикам. Без учёта ограничения по количеству предметов в ящике это можно сделать $\mathrm{C}_{\boldsymbol{r-1}}^{\boldsymbol{n-1}}$ способом. Пусть ***j*** – количество ящиков, в которых уже помещено ***m + 1*** предмет. Осуществить их выбор можно $\boldsymbol{C_n^j}$ количеством способов. Оставшиеся ***(r - n - m)*** предметов можно разместить по n ящикам $\boldsymbol{C_{r-mj-1}^{n-1}}$ способами. Таким образом для числа комбинаций рангов показателей физиологических признаков можно записать:

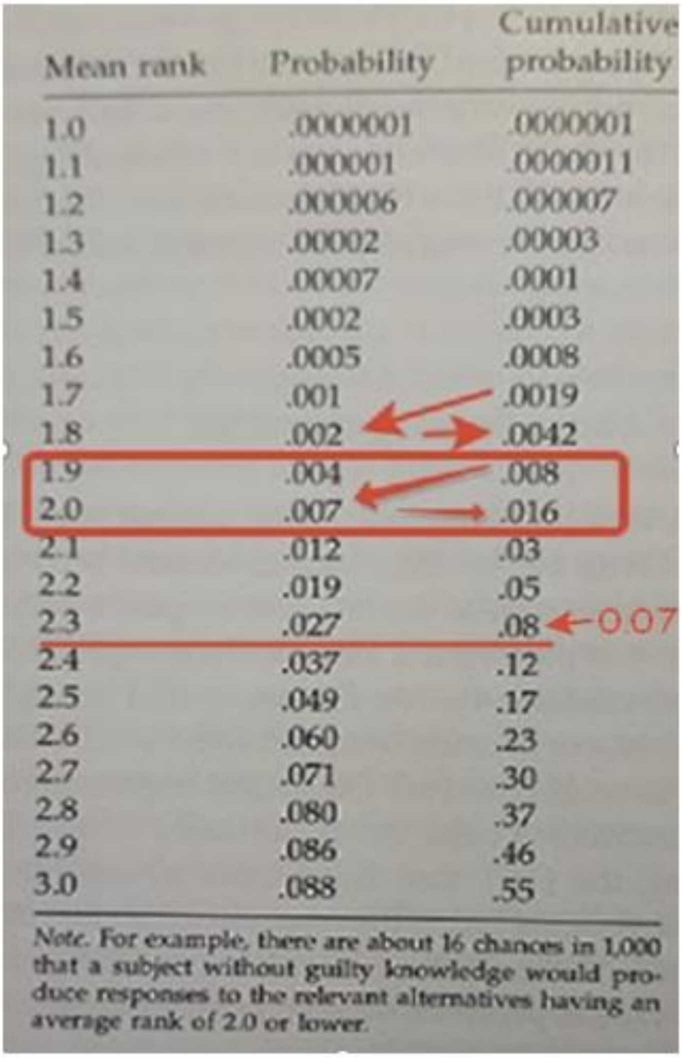
$$\boldsymbol{F}(\boldsymbol{r},\boldsymbol{n},\boldsymbol{m}) = \sum_{j=0}^{p}(-\mathbf{1})^{j}\, \boldsymbol{C_n^j C_{r-mj-1}^{n-1}} \text{ где, } \boldsymbol{p} = [(\boldsymbol{r}-\boldsymbol{n})/\boldsymbol{m}] \qquad (4)$$

Теперь запишем формулу Ликкена — это формула для вероятности ОСР для незначимого стимула.

$$\boldsymbol{P}(\boldsymbol{r},\boldsymbol{n},\boldsymbol{m}) = \boldsymbol{m}^{-\boldsymbol{n}} \sum_{j=0}^{p}(-\mathbf{1})^{j}\, \boldsymbol{C_n^j C_{r-mj-1}^{n-1}} \qquad (5)$$

По этой формуле была рассчитана таблица аналогичная таблице Ликкена. Для этой таблицы ***m = 5, n = 10***.

ОСР	Вероятность	Суммарная вероятность
10	0,0000001	0,0000001
11	0,0000010	0,0000011
12	0,0000056	0,0000068
13	0,0000225	0,0000293
14	0,0000732	0,0001025
15	0,0002040	0,0003065
16	0,0005023	0,0008088
17	0,0011151	0,0019239
18	0,0022641	0,004188
19	0,0042465	0,0084345
20	0,0074141	0,0158486
21	0,0121201	0,0279686
22	0,0186373	0,0466059
23	0,0270561	0,0736621
24	0,0371881	0,1108502
25	0,0485063	0,1593564
26	0,0601498	0,2195062
27	0,0710093	0,2905155
28	0,0798899	0,3704054
29	0,0857190	0,4561244
30	0,0877512	0,5438756



Mean rank	Probability	Cumulative probability
1.0	.0000001	.0000001
1.1	.000001	.0000011
1.2	.000006	.000007
1.3	.00002	.00003
1.4	.00007	.0001
1.5	.0002	.0003
1.6	.0005	.0008
1.7	.001	.0019
1.8	.002	.0042
1.9	.004	.008
2.0	.007	.016
2.1	.012	.03
2.2	.019	.05
2.3	.027	.08 ← 0.07
2.4	.037	.12
2.5	.049	.17
2.6	.060	.23
2.7	.071	.30
2.8	.080	.37
2.9	.086	.46
3.0	.088	.55

Note. For example, there are about 16 chances in 1,000 that a subject without guilty knowledge would produce responses to the relevant alternatives having an average rank of 2.0 or lower.

Рис.3 Таблицы распределения вероятностей рангов

Обратите внимание на второй столбец – «вероятность» в обеих таблицах. Если в левой таблице производить округление до тех же разрядов, что и в таблице Ликкена, то столбцы полностью совпадут. Третий же столбец это кумулятивная или суммарная вероятность. Значения кумулятивной вероятности получаются суммирование значений во втором столбце. Таким образом, произведя округления до тех же разрядов, мы должны в левой таблице получить те же значения в третьем столбце, что и в таблице Ликкена. Но с ранга 2.3 (23) имеются расхождения. Объяснить их можно только одним – ошибками округления, когда производили сложение чисел, округленных до различных разрядом.

Таблицы представляют закон распределения вероятностей ОСР. Его можно представить в виде гистограммы.

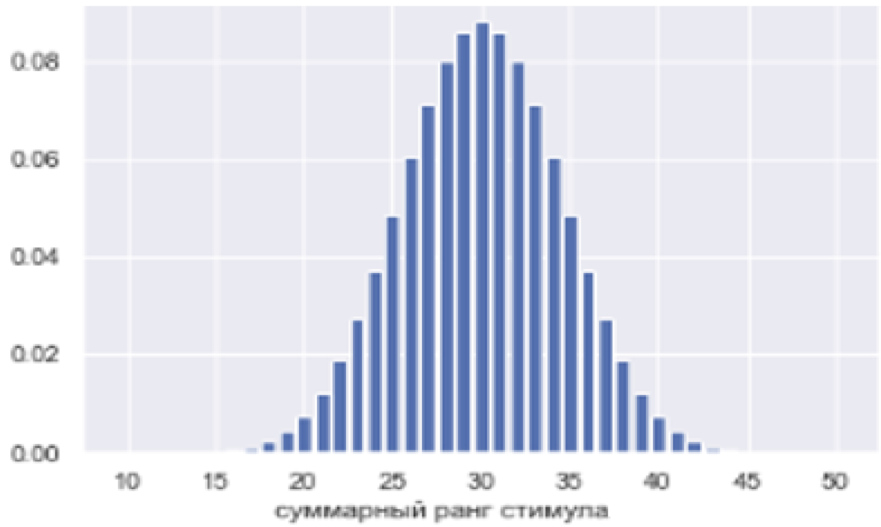


Рис.4 Гистограмма распределения вероятностей ОСР

Имея закон распределения вероятностей несложно вычислить достигаемый уровень значимости (***p-value***) – вероятность того, что ОСР примет наблюдаемое значение или ещё большее (в общем случае - экстремальное). Сделать это можно воспользовавшись вычисленной таблицей. Значение достигаемого уровня значимости можно получить, просуммировав вероятности, начиная от полученного ОСР до максимально возможного ОСР. Для таблицы Ликкена максимальная ОСР равна 50-ти. А таблица только до среднего ранга равного 3-м. В этом случае можно воспользоваться кумулятивной вероятностью – ***p-cumulative.*** Они связаны между собой формулой ***p-value = 1- p-cumulative.*** Например, для ОСР = 30 ***p-value*** = *1- 0,54* = *0,46*. Для ОСР 40

p-value (40) = p-cumulative (20) *= 0,16.* Тут мы воспользовались симметричностью распределения ранга ОСР. Достигаемый уровень значимости можно использовать для диагностики значимости релевантного вопроса. Так как ОСР значимого стимула должен быть в среднем меньше, чем ОСР незначимого стимула, то, чем меньше достигаемый уровень значимости, тем вероятнее, что стимул значим. Достигаемый уровень значимости можно посчитать, воспользовавшись формулой Ликкена (5). Для этого надо определить как вычислить максимальный ранг в произвольном случае. Сделать это не сложно – $\text{ОСР}_{max} = m \cdot n$. Для таблицы Ликкена $\text{ОСР}_{max} = 50$. Если наблюдаемый ОСР = ***r*** будем иметь.

$$\boldsymbol{p - value(r, m, n) = m^{-n} \cdot \sum_{i=r}^{m \cdot n} \sum_{j=0}^{p} (-1)^j \, C_n^j C_{i-mj-1}^{n-1}} = \qquad (6)$$

$$= \boldsymbol{m^{-n}} \cdot \sum_{\boldsymbol{i=r}}^{\boldsymbol{m \cdot n}} \sum_{\boldsymbol{j=0}}^{\boldsymbol{p}} (-\boldsymbol{1})^{\boldsymbol{j}} \frac{\boldsymbol{n}!}{\boldsymbol{j}!\,(\boldsymbol{n}-\boldsymbol{j})!} \frac{(\boldsymbol{i}-\boldsymbol{mj}-\boldsymbol{1})!}{(\boldsymbol{i}-\boldsymbol{mj}-\boldsymbol{n})!\,(\boldsymbol{n}-\boldsymbol{1})!}$$

Обсуждение

Как было сказано выше, формулу 6 можно использовать для диагностики значимости стимула в тестовых рядах. Но как показало изучение такого способа точность получается весьма посредственная [4], если не использовать дополнительные приёмы, которые могут значительно поднять точность алгоритма классификации, основного на (6). Такой алгоритм разработан и включён в СППРП «Сокол» [5]. Он получил название Combi Calc v.2, что является сокращением от Combination Calculator, по-русски – «Вычислитель Комбинаций». Речь идёт о вычислении с помощью формулы (5) количества возможных комбинаций размещений с ограничениями неразличимых шаров по ящикам. Нет никаких сомнений, что полученная с помощью (5) таблица с точностью до ошибок округлений в третьем столбце, совпадает с таблицей Ликкена. Поэтому формула (5) и была названа формулой Ликкена.

Решенная задача получения формулы с помощью, которой можно вычислить закон распределения ОСР для незначимого стимула, совпадает с комбинаторной задачей «о размещении неразличимых шаров в ящиках», лежащей в основе вычисления квантомеханического распределения Бозе-Эйнштейна, только используемой моделью «шаров и ящиков». Отличается от неё наложенными ограничениями на количество шаров в ящиках. Таких ограничений на поведение бозонов в квантовой механике нет. Кроме того, квантовая статистика имеет дело с гигантскими ансамблями квантовых частиц. В психофизиологии речь идёт о рангах не превышающих нескольких десятков. Не обсуждая физических подробностей, можно сказать одно – последнее обстоятельство приводит к тому, что физическое распределение Бозе-Эйнштейна представляет из себя экспоненциальную непрерывную функцию:

$$\langle \boldsymbol{n} \rangle = \frac{\mathbf{1}}{exp\left(\frac{E-\mu}{kT}\right)-\mathbf{1}} \qquad (7)$$

Именно, выражение (7) называется в физике распределением Бозе-Эйнштейна [6 стр. 338]. С формулой Ликкена оно связано только мысленной моделью «шаров и ящиков», которая лежат в основе их получения. Причём формула Ликкена может быть получена и другим способом – с помощью производящей функции. В этом случае в мысленной конструкции в виде «шаров и ящиков» нет необходимости, и вся кажущаяся связь с квантовой физикой исчезает.

Список литературы

1. Lykken D. T. A Tremor in the Blood. New York: McGraw-Hill, 1981. p.118.
2. Формула Ликкена//Полиграф – форум, URL: http://ld.eposgroup.ru/forum/viewtopic.php?p=94408#94408
3. Бочаров П. П., Печинкин А. В. Теория вероятностей. Математическая статистика. - 2-е изд. - М.: ФИЗМАТЛИТ, 2005. - 296 с.
4. Летков Ю. В. 2023. Оценка тестов методики скрываемой информации методом ранжирования реакций. PREPRINTS.RU. https://doi.org/10.24108/preprints-3112672

5. СППРП «Сокол»: [сайт] /2020. – URL: http://www.skl-ol.ru/ (Дата обращения 21.11.2023)
6. Квантовая физика: Учебное пособие. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2004. — 496 е.: ил. (Физика в техническом университете / Под ред. Л.К. Мартинсона, А.Н. Морозова).