

ИСПОЛЬЗОВАНИЕ МАТЕМАТИЧЕСКИХ МЕТОДОВ ДЛЯ ОБНАРУЖЕНИЯ СТАТИСТИЧЕСКИХ ОТКЛОНЕНИЙ НА ПРИМЕРЕ ПРОСТЕЙШЕЙ ИГРЫ

Волынец В.И. Хренков Д.Д.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Баркова Е.А. – канд. физ.-мат. наук

Проводится следующая игра. Каждый игрок бросает монету, если выпадает орёл – игрок победил, если выпадает решка – игрок проиграл. Одни игроки играют честно, то есть вероятность выпадения орла у них 50%. Другие игроки играют нечестно, то есть вероятность выпадения орла у них больше 50%. назовем таких игроков нарушителями. Требуется составить такой тест (k и более орлов из n бросков), чтобы раскрыть максимальное количество нарушителей, при этом обвинить ложно минимальное количество честных игроков, затратив минимальный возможный объём данных.

Рассмотрим следующую задачу:

Определить минимум 98% нарушителей, ложно обвинив менее 5% честных игроков при минимальном количестве бросков, зная, что шанс выпадения орла на монете нарушителя 55%.

Для решения задачи мы будем анализировать вероятность выпадения орла различного количества раз за определённое число бросков. Для нахождения вероятности выпадения орла используем формулу Бернулли:

$$P(k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \quad (1),$$

где n – количество бросков монеты,

k – количество положительных исходов,

$\binom{n}{k}$ – биномиальный коэффициент,

p – вероятность положительных исходов при k ,

$1 - p$ – вероятность положительного исхода каждый бросок.

Постепенно увеличивая общее количество бросков и анализируя вероятности выпадения орла, получим ответ.

В процессе решения задачи было замечено, что даже для небольших значений расчёты при помощи формулы Бернулли являются объёмными. Отсюда появляется вопрос автоматизации процесса решения по формуле Бернулли при помощи программного кода.

Для нахождения конкретного решения мы перебираем по порядку все возможные комбинации бросков и выпадений орла, вычисляя кумулятивную сумму для честного и нечестного игроков. В момент, когда сумма для честного игрока не подходит под статистические требования, мы останавливаем текущий цикл и переходим на следующую итерацию общего количества бросков. Решение считается найденным, когда две суммы удовлетворяют статистическим условиям.

В определённых случаях для решения подобных задач, в виду громоздкости использования формулы Бернулли, удобнее воспользоваться интегральной формулой Муавра-Лапласа. Основными параметрами формулы являются математическое ожидание и стандартное отклонение. Важным параметром для данной формулы является ограниченность в использовании. Так как интегральная формула Муавра-Лапласа не является абсолютно точной, перед её использованием нужно проверить рациональность её использования в рамках конкретной задачи.

Так как использование формулы Бернулли является неудобным даже для относительно больших значений, а для использования интегральной формулы Муавра-Лапласа необходимо выполнение определённых условий, мы решили сократить количество этапов подсчёта, при этом сохранив возможность использовать формулу для случаев, когда $n \cdot p \cdot q < 20$. Готовых решений, подходящих в данной конкретной ситуации найдено не было, поэтому провели преобразования самостоятельно и получили собственную формулу:

$$P(X \geq k) = p^k \binom{n}{k} (1 - p)^{n-k} \cdot 2F_1\left(1, k - n; k + 1; \frac{p}{p-1}\right) \quad (2),$$

По формуле Бернулли с помощью автоматизированного подхода получили результаты, приведённые на рисунке 1. Результатом является 715 и более выпадений орла из 1368 бросков.

Результат расчётов по собственной формуле совпал с результатом расчётов по формуле Бернулли.

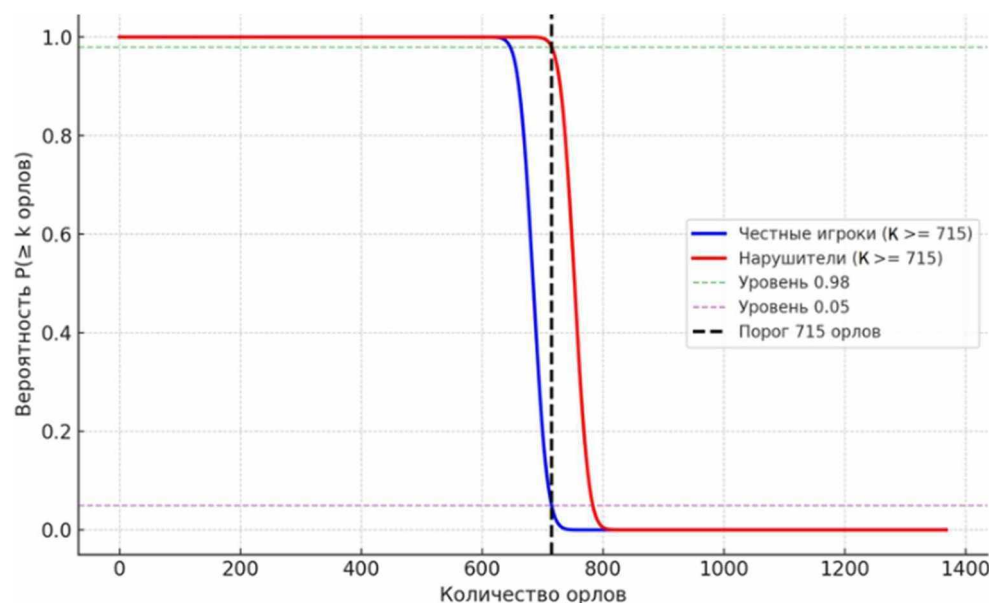


Рисунок 1 – График визуализации результатов расчетов по формуле Бернулли

Расчеты для честного игрока

Расчет математического ожидания и стандартного отклонения

$$\sqrt{npq} \sqrt{1368 \cdot 0,5 \cdot 0,5} \approx 18,49$$

Преобразование в стандартную нормальную формулу и нахождение вероятности

$$Z = \left(\frac{715 - 684}{18,49} \right) = \frac{31}{18,49} \approx 1,67$$

$$P(X \geq 715) = 1 - \Phi(1,67) = 1 - 0,9525 = 0,0475$$

Расчеты для нарушителя

Расчет математического ожидания и стандартного отклонения

$$\sqrt{npq} \sqrt{341,4} \approx 18,47$$

Преобразование в стандартную нормальную формулу и нахождение вероятности

$$Z = \left(\frac{715 - 714,4}{18,47} \right) = \frac{0,6}{18,47} \approx -2,13$$

$$P(X \geq 715) = 1 - \Phi(-2,13) = 1 - 0,0166 = 0,9834$$

Методы анализа статистических отклонений применяются в различных областях, где данный подход может быть интегрирован как компонент систем анализа, например в образовательных платформах, системах контроля качества и игровой индустрии.

Список использованных источников:

1. Информационный архив Корнеллского университета [Электронный ресурс]. Режим доступа: <https://arxiv.org/abs/1802.04888>
2. В. А. Быковский, Д. А. Фроленков, "Некоторые интегральные представления для гипергеометрической функции", Дальневост. матем. журн., 15:1 2015 – С. 38-40
3. Шундалов, Б. М. Статистика. Общая теория. Учебник. Минск 2012. – С. 14-28.
4. The American Statistician // 2019, выпуск 73, № 1-1