УДК 519.235:336.265

В.Ю. Дубницкий, Ю.А. Наталина

Харьковский банковский институт УАБД НБУ

СТАТИСТИЧЕСКАЯ ОЦЕНКА КОРРЕКТНОСТИ ЛОТЕРЕИ ТИПА: «УГАДАЙ M ЧИСЕЛ ИЗ N»

Предложен способ проверки корректности лотерей типа: «Угадай m чисел из n». В основе его лежит оценка совпадения фактических и расчетных вероятностей появления каждого из n чисел. Проверка совпадения вероятностей проведена по схеме сравнения наблюдаемых частот с заданным расщеплением.

лотерея, вероятностная модель, критерий хи-квадрат, расщепление выборки

Введение

Постановка задачи. Известно, что в настоящее время широко распространены лотереи, в которых от участников требуется угадать m чисел из n возможных, так называемые "генуэзские" лотереи [1, 2]. В связи с тем, что организация таких лотерей порождает большие финансовые потоки, то при аудите лотерей необходимо, прежде всего, оценить их корректность. Корректной назовем лотерею, в которой все ее участники имеют равные возможности для выигрыша.

В связи с тем, что число участников лотереи велико, а выигрышную комбинацию (ВК) формируют лототроном на глазах участников в режиме реального времени, то равновозможность выигрыша обусловлена корректным формированием набора чисел, составляющих ВК. Вариант прямого сговора организаторов лотереи с некоторыми из её участников в данном сообщении не рассмотрен, как выходящий за пределы принятой предметной области.

Корректной назовем лотерею, в которой статистическая вероятность формирования ВК совпадает с вероятностью этого же события, полученной из анализа соответствующей математической модели.

Анализ литературы. Вероятность выигрыша в генуэзской лотерее определена в работе [1]. Статистические проблемы оценки корректности такой лотереи в этой и аналогичных работах не рассмотрены. Следует отметить, что близкая по смыслу задача рассмотрена в работе [3].

Изложение результатов

Рассмотрим лотерею, в которой необходимо угадать в любой последовательности m чисел из n, m < n. Тогда количество возможных ВК

$$B = C_n^m . \quad (1)$$

Количество комбинаций, в состав которых входит число n_j, $j \in [1;n]$,

$$A = C_{n-1}^{m-1} . \quad (2)$$

Вероятность события, состоящего в том, что число n_j войдет в ВК,

$$p = \frac{A}{B} = \frac{C_{n-1}^{m-1}}{C_n^m} = \frac{(n-1)!}{(m-1)!(n-m)!} : \frac{n!}{m!(n-m)!} = \frac{m}{n} \quad (3)$$

Рассмотрим в качестве иллюстрации к приведенной схеме такой пример. Пусть в лотерее исходный набор составляют n = 5 чисел, в выигрышные комбинации входят m = 4 числа. Все допустимые комбинации ВК приведены в табл. 1

Таблица 1

Возможные выигрышные комбинации в лотерее: “Угадай четыре числа из пяти”

a	b	c	d	e
+	+	+	–	+
+	+	–	+	+
+	–	+	+	+
–	+	+	+	+
+	+	+	+	–

Примечание: “+” – число принадлежит ВК;
“–” – число не принадлежит ВК.

В соответствии с условием (3) для данной лотереи вероятность p = 4/5. Действительно, $C_5^4 = 5$, количество комбинаций $A_к$ содержащих данное число k, {k = a, d, c, d, e} определим из табл.1 непосредственным подсчетом: $A_a = A_b = A_c = A_d = A_e = 4$.

Таким образом, оценка корректности лотереи сведена к проверке утверждения о том, что частости появления любого из n_j $\{j \in [1; n]\}$чисел соответствует вероятность p = m/n.

Подобного рода задачи хорошо изучены в генетике и известны под названием задачи о сравнении наблюдаемых частот с заданным расщеплением [3].

Для проверки статистической гипотезы о том, что расщепление выборки соответствует наблюдаемому, необходимо вычислить величину

$$\chi^2{}_{факт} = \sum_{j=1}^{n} \frac{(B_j - E)^2}{E}, \quad (4)$$

где B_j – фактическая частота числа n_j (количество ВК в которые вошло число n_j).

Число E определяют по формуле:

$$E = r\frac{m}{n}, \quad (5)$$

где r – объем выборки, содержащей информацию о возможных ВК; m, n – параметры лотереи.

Если величина $\chi^2{}_{факт} < \chi^2_T$, где $\chi^2_T = \chi^2_{\alpha,(n-1)}$, то гипотеза о корректности лотереи может быть принята.

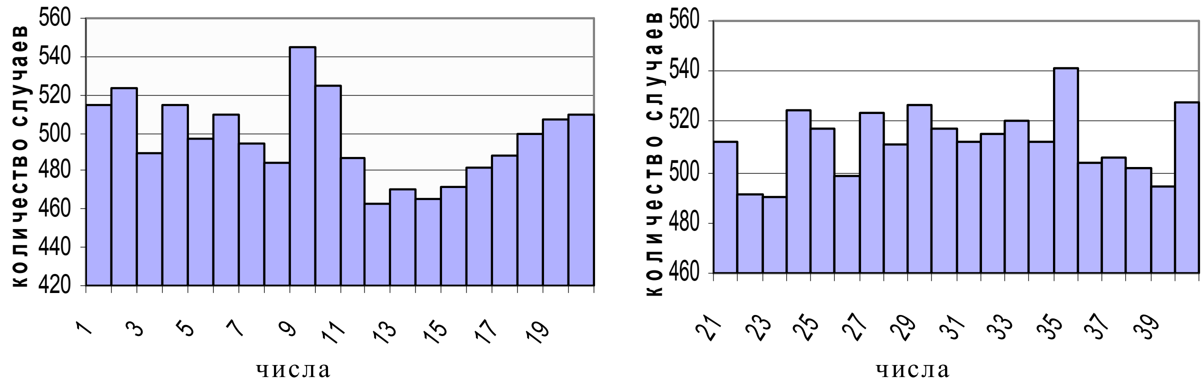
В табл. 2 приведены результаты соответствующих вычислений.

Так как для лотерей № 1 и № 2 величина $\chi^2_{факт} < \chi^2_T$, то эти лотереи можно считать корректными. Гистограмма ряда распределения для лотереи №1 показана на рис. 1, 2, а для лотереи № 2 – на рис. 3.

Таблица 2

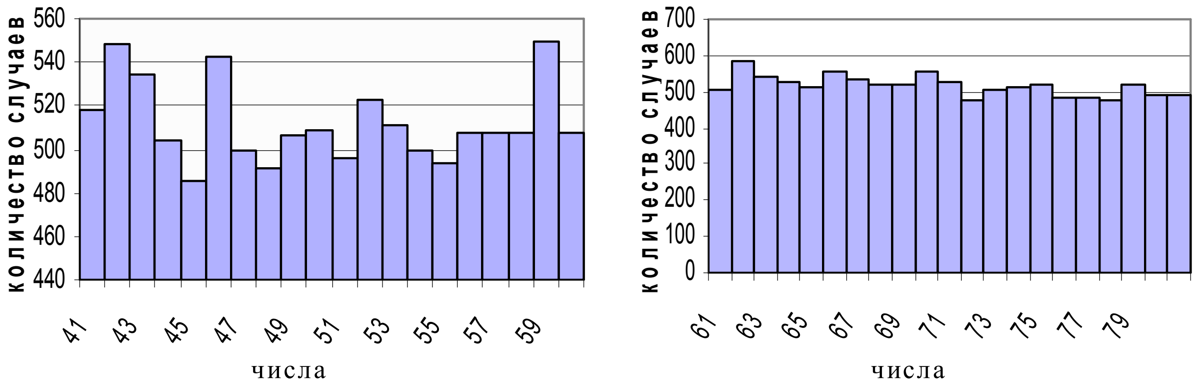
Фактические и критические значения величины χ^2

Вид лотереи	Параметры лотереи (m;n)	Объем выборки – количество тиражей (r)	Величина χ^2	Уровень значимости α	Число степеней свободы	Величина χ^2_T
№1	(20;80)	2038	76,70	0,05	79	100,74
№2	(6;49)	203	47,79	0,05	79	65,17



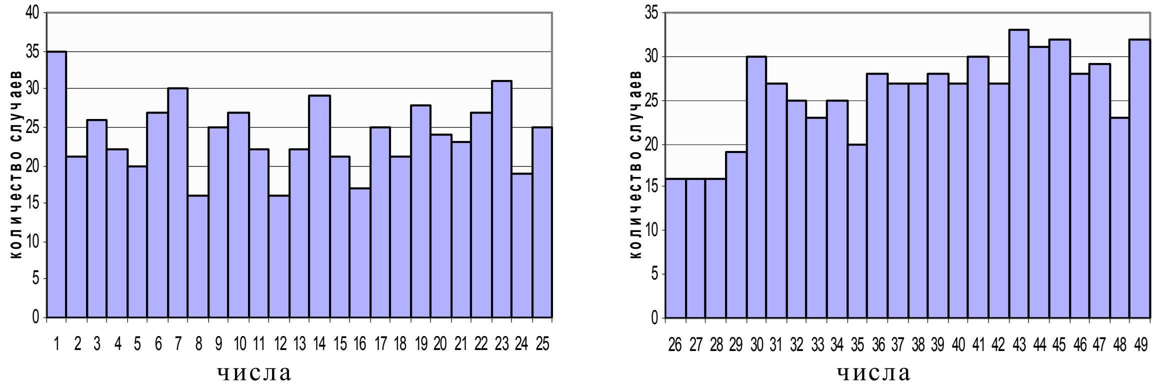
а б

Рис. 1. Гистограмма распределения цифр выигрышных комбинаций лотереи № 1: а – цифры 1-20; б – цифры 21-40



а б

Рис. 2. Гистограмма распределения цифр выигрышных комбинаций лотереи № 1: а – цифры 41-60; б – цифры 61-80



а б

Рис. 3. Гистограмма распределения цифр выигрышных комбинаций лотереи № 2: а – цифры 1-25; б – цифры 26-49

Ещё один метод контроля корректности лотереи состоит в следующем. Из работы [4] известно, что если выполнено неравенство:

$$z = \left(\left| p - \pi \right| - \frac{1}{2n} \right) \Bigg/ \sqrt{\frac{\pi(1-\pi)}{n}} < z_\alpha , \qquad (6)$$

где p – фактическая частота интересующего нас события; π – его вероятность; n – количество испытаний; z_α – критическое значение, то принято считать, что различие между частотой события и его вероятностью несущественно на уровне значимости α..

В табл. 3 приведены итоговые результаты соответствующих вычислений, выполненных для каждой из цифр, участвующей в игре.

Таким образом, фактическое различие частоты появления цифры в лотерее и вероятностью находится в пределах допустимой погрешности.

Таблица 3

Оценка существенности различия частоты появления цифры и её вероятностью

Вид лотереи	Параметры лотереи (m;n)	Количество цифр	Величина α	Уровень значимости α	Различие несущественно (количество цифр)	Различие существенно (количество цифр)
№1	(20;80)	80	1,96	0,05	77	3
№2	(6;49)	49	1,96	0,05	48	1

Выводы

1. Предложена методика оценки корректности лотереи типа: «Угадай m чисел из n возможных”.

2. Показана корректность двух типов лотерей проводимых в Украине.

Список литературы

1. *Виленкин Н.Я. Комбинаторика. – М.: Наука, 1968. – 338 с.*

2. *Закс. Л. Статистическое оценивание. – М.: Статистика, 1976. – 596 с.*

3. *Большев Л.Н. О случайных числах М. Кадырова в статистических таблицах Я.Янко // Теория вероятностей и её применение. – 1964. – № 9. – С. 152-154.*

4. *Флейс Дж. Статистические методы для изучения таблиц долей и пропорций. – М.: ‘Финансы и статистика, 1989. – 318 с.*

Поступила в редколлегию 22.12.2006

Рецензент: д-р техн. наук, проф. Е.А. Артеменко, Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков.