ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 31 МАТЕМАТИКИ Выпуск 1
2024

В. И. Круглов (Москва, МИ РАН им. В.А. Стеклова). **Сравнение математических ожиданий чисел совпадающих s-цепочек в последовательностях независимых и зависимых случайных величин.**

УДК 519.233.3 DOI https://doi.org/10.52513/08698325_2024_31_1_1

Резюме: Рассматриваются состоящие из нулей и единиц последовательности, полученные в результате извлечения из урны черных и белых шаров. Для схем извлечения шаров с возвращениями и без возвращений получены формулы для математического ожидания числа совпадений значений у подпоследовательностей заданной длины в таких последовательностях. Приводятся численные значения математического ожидания при различных количествах шаров в урне и длинах подпоследовательностей.

Ключевые слова: s-цепочки, урновая схема, последовательности зависимых случайных величин.

1. Постановка задачи и полученные равенства

Рассмотрим урну, в которой находится a белых и b черных шаров. Пусть эти шары последовательно и без возвращения извлекаются из урны, результатом эксперимента является последовательность $\overrightarrow{X} = (X_1, X_2, \ldots, X_{a+b})$, элементы которой принимают значения, равные 1 (белый шар) или 0 (черный шар). Каждая такая последовательность содержит a единиц и b нулей, будем считать, что все такие последовательности равновероятны.

Для каждого s, где $s < a$ и $s < b$, и каждого i такого, что $1 \leqslant i \leqslant a+b-s+1$, рассмотрим состоящую из s элементов подпоследовательность $Y_i = (X_i, \ldots, X_{i+s-1})$, которую будем называть s-цепочкой.

Определим событие $E_{ij} = \bigl\{Y_i = Y_j\bigr\} = \bigl\{X_i = X_j, \ldots, X_{i+s-1} = X_{j+s-1}\bigr\}$, которое заключается в совпадении значений s-цепочек Y_i и Y_j, и индикатор этого события $I\bigl(E_{ij}\bigr)$. Рассмотрим случайную величину

$$\xi = \sum_{1\leqslant i<j\leqslant a+b-s+1} I\bigl(E_{ij}\bigr),$$

равную числу совпадений s-цепочек в последовательности $\overrightarrow{X}$.

Вероятность того, что на заданных r_a местах последовательности $\overrightarrow{X}$ находятся белые шары, а на заданных r_b местах находятся черные шары, равна

$$p_{r_a,r_b} = C^{a-r_a}_{a+b-r_a-r_b} \big/ C^a_{a+b},$$

а случайная величина ξ может быть представлена в виде суммы двух случайных величин $\xi = \xi' + \xi''$, где величина

$$\xi' = \sum_{\substack{1 \leqslant i < j \leqslant a+b-s+1 \\ i+s \leqslant j}} I(E_{ij}),$$

равна числу совпадений значений у непересекающихся s-цепочек в последовательности $\overrightarrow{X}$, а величина

$$\xi'' = \sum_{\substack{1 \leqslant i < j \leqslant a+b-s+1 \\ i+1 \leqslant j \leqslant i+s-1}} I(E_{ij}),$$

равна числу совпадений значений у пересекающихся s-цепочек.

Лемма 1. *Вероятность совпадений значений у непересекающихся s-цепочек Y_i и Y_j равна*

$$\mathbf{P}\{Y_i = Y_j\} = \sum_{r=0}^{s} C_s^r p_{2r,2s-2r}$$

и для математического ожидания случайной величины ξ' справедливо равенство

$$\mathbf{E}\xi' = C_{a+b-2s+2}^2 \sum_{r=0}^{s} C_s^r p_{2r,2s-2r}.$$

Для пересекающихся s-цепочек Y_i и Y_j, где $i < j$, введем обозначение $k = j - i$, и, разделив s на k с остатком, будем обозначать

$$s = kn + m, \quad n = \left\lfloor \frac{s}{k} \right\rfloor, \quad m = s - kn = s - k\left\lfloor \frac{s}{k} \right\rfloor.$$

Лемма 2. *Вероятность совпадения значений пересекающихся s-цепочек Y_i и Y_j является функцией от величины $k = j - i$ и равна*

$$\begin{aligned} f(k) &= \mathbf{P}\{Y_i = Y_j\} \\ &= \sum_{t_1=0}^{m} \sum_{t_2=0}^{k-m} C_m^{t_1} C_{k-m}^{t_2} p_{t_1(n+2)+t_2(n+1),(m-t_1)(n+2)+(k-m-t_2)(n+1)} \end{aligned}$$

и для математического ожидания случайной величины ξ'' справедливо равенство

$$\mathbf{E}\xi'' = \sum_{i=1}^{a+b-2s+2} \sum_{k=1}^{s-1} f(k) + \sum_{i=a+b-2s+3}^{a+b-s} \sum_{k=1}^{a+b-s+1-i} f(k).$$

Из аддитивности математического ожидания и равенств для $\mathbf{E}\xi'$ и $\mathbf{E}\xi''$ следует формула для $\mathbf{E}\xi$.

Теорема.

$$\begin{aligned} \mathbf{E}\xi = \mathbf{E}\xi' + \mathbf{E}\xi'' &= C_{a+b-2s+2}^2 \sum_{r=0}^{s} C_s^r p_{2r,2s-2r} \\ &+ \sum_{i=1}^{a+b-2s+2} \sum_{k=1}^{s-1} f(k) + \sum_{i=a+b-2s+3}^{a+b-s} \sum_{k=1}^{a+b-s+1-i} f(k). \end{aligned}$$

Рассмотрим альтернативную постановку задачи, в которой в урне также находится a белых и b черных шаров, которые последовательно $a+b$ раз извлекаются из урны, и пусть после каждого извлечения шар возвращается в урну. Результатом эксперимента является состоящая из нулей и единиц последовательность $\overrightarrow{\widetilde{X}} = (\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_{a+b})$, элементы которой $\widetilde{X}_i$ независимы в совокупности и $\mathbf{P}\{\widetilde{X}_i = 1\} = \frac{a}{a+b}$, $\mathbf{P}\{\widetilde{X}_i = 0\} = \frac{b}{a+b}$, для всех $1 \leqslant i \leqslant a+b$. В таком случае вероятность того, что на заданных местах последовательности $\overrightarrow{\widetilde{X}}$ находятся r_a белых и r_a черных шаров, равна

$$\widetilde{p}_{r_a, r_b} = \left(\frac{a}{a+b}\right)^{r_a} \left(\frac{b}{a+b}\right)^{r_b}.$$

Аналогичным образом определим s-цепочки $\widetilde{Y}_i$, индикаторы совпадения s-цепочек $I\big(\widetilde{E}_{ij}\big)$, функцию $\widetilde{f}(k)$, случайную величину $\widetilde{\xi}$, равную числу совпадений s-цепочек в последовательности $\overrightarrow{\widetilde{X}}$, и ее представление $\widetilde{\xi} = \widetilde{\xi}' + \widetilde{\xi}''$ в виде суммы случайных величин $\widetilde{\xi}'$ и $\widetilde{\xi}''$, равных соответственно числам совпадений непересекающихся и пересекающихся s-цепочек.

Тогда все равенства для $\mathbf{E}\xi$, $\mathbf{E}\xi'$, $\mathbf{E}\xi''$ после замены p_{r_a, r_b} на $\widetilde{p}_{r_a, r_b}$ и $f(k)$ на $\widetilde{f}(k)$ остаются верными для $\mathbf{E}\widetilde{\xi}$, $\mathbf{E}\widetilde{\xi}$, $\mathbf{E}\widetilde{\xi}''$.

2. Численные значения Приведем значения $\mathbf{E}\xi$, $\mathbf{E}\xi'$, $\mathbf{E}\xi''$, значения $\mathbf{E}\widetilde{\xi}$, $\mathbf{E}\widetilde{\xi}$, $\mathbf{E}\widetilde{\xi}''$ и их соотношения при различных длинах цепочек s для случая, когда число белых шаров равно числу черных шаров: пусть $a = 300, b = 300$, тогда

s	$\mathbf{E}\xi$	$\mathbf{E}\widetilde{\xi}$	$\mathbf{E}\widetilde{\xi}/\mathbf{E}\xi$	$\mathbf{E}\xi'$	$\mathbf{E}\widetilde{\xi}'$	$\mathbf{E}\widetilde{\xi}'/\mathbf{E}\xi'$
8	676.443165	685.656250	1.013620	660.763205	669.550781	1.013299
12	41.417538	42.276855	1.020748	39.917525	40.711182	1.019882
16	2.535203	2.606506	1.028125	2.410743	2.474442	1.026423
20	0.155141	0.160685	1.035732	0.145547	0.150338	1.032921
24	0.009492	0.009905	1.043543	0.008784	0.009130	1.039378

s	$\mathbf{E}\xi''$	$\mathbf{E}\widetilde{\xi}''$	$\mathbf{E}\widetilde{\xi}''/\mathbf{E}\xi''$
8	15.679960	16.105469	1.027137
12	1.500013	1.565674	1.043774
16	0.124459	0.132065	1.061108
20	0.009594	0.010346	1.078375
24	0.000707	0.000775	1.095285

Приведем аналогичные значения для случая, когда число белых шаров вдвое превосходит число черных шаров: пусть $a = 400, b = 200$, тогда

s	$\mathbf{E}\xi$	$\mathbf{E}\widetilde{\xi}$	$\mathbf{E}\widetilde{\xi}/\mathbf{E}\xi$	$\mathbf{E}\xi'$	$\mathbf{E}\widetilde{\xi}'$	$\mathbf{E}\widetilde{\xi}'/\mathbf{E}\xi'$
8	1575.19638	1616.34564	1.026123	1516.18480	1555.40483	1.025868
12	148.753796	156.742920	1.053707	137.007838	144.146332	1.052103
16	14.383812	15.738153	1.094157	12.276082	13.353559	1.087770
20	1.454279	1.677476	1.153475	1.090853	1.236571	1.133582
24	0.157897	0.196116	1.242046	0.096147	0.114463	1.190500
28	0.018840	0.025870	1.373128	0.008407	0.010591	1.259775

s	$\mathbf{E}\xi''$	$\mathbf{E}\widetilde{\xi}''$	$\mathbf{E}\widetilde{\xi}''/\mathbf{E}\xi''$
8	59.011578	60.940801	1.032692
12	11.745958	12.596589	1.072419
16	2.107730	2.384594	1.131356
20	0.363426	0.440904	1.213188
24	0.061751	0.081653	1.322304
28	0.010433	0.015279	1.464469

Приведем аналогичные значения для случая, когда число белых шаров втрое превосходит число черных шаров: пусть $a = 450, b = 150$, тогда

s	$\mathbf{E}\xi$	$\mathbf{E}\widetilde{\xi}$	$\mathbf{E}\widetilde{\xi}/\mathbf{E}\xi$	$\mathbf{E}\xi'$	$\mathbf{E}\widetilde{\xi}'$	$\mathbf{E}\widetilde{\xi}'/\mathbf{E}\xi'$
8	4043.76069	4168.563583	1.030863	3870.72832	3990.833648	1.031029
12	613.889275	654.722780	1.066516	555.099138	592.425664	1.067243
16	96.559612	107.787725	1.116282	78.568201	87.909823	1.118898
20	16.260789	19.188968	1.180076	10.976090	13.039758	1.188015
24	3.039691	3.821586	1.257229	1.513568	1.933416	1.277389
28	0.643303	0.867241	1.348107	0.206035	0.286549	1.390780
32	0.152229	0.221685	1.456254	0.027688	0.042451	1.533161
36	0.038938	0.061827	1.587819	0.003674	0.006286	1.711077

s	$\mathbf{E}\xi''$	$\mathbf{E}\widetilde{\xi}''$	$\mathbf{E}\widetilde{\xi}''/\mathbf{E}\xi''$
8	173.032366	177.729935	1.027149
12	58.790138	62.297116	1.059652
16	17.991410	19.877902	1.104855
20	5.284698	6.149210	1.163588
24	1.526122	1.888170	1.237234
28	0.437267	0.580692	1.328001
32	0.124541	0.179234	1.439156
36	0.035264	0.055541	1.574979

Автор выражает признательность А. М. Зубкову за постановку задачи и ценные замечания по ходу ее решения.

СПИСОК ЛИТЕРАТУРЫ

1. *Зубков А. М., Михайлов В. Г.* Предельные распределения случайных величин, связанных с длинными повторениями в последовательности независимых испытаний. — Теория вероятн. и ее примен., 1974, т. 19, в. 1, с. 173–181.

Поступила в редакцию
23.VI.2024

UDC 519.233.3 DOI https://doi.org/10.52513/08698325_2024_31_1_1

Kruglov V. I. (Moscow, Steklov Mathematical Institute of RAS)**. A comparison of expectations of numbers of s-chains with coinciding values in sequences of independent and dependent random variables.**

Abstract: We consider sequences of zeros and ones that are obtained by extracting black and white balls from an urn. For extraction schemes with and without returns we obtain formulas for expectation of the number of coinciding subsequences of a given length in such sequences. Numerical values of expectation are presented for different numbers of balls in the urn and lengths of the subsequences.

Keywords: s-chains, urn scheme, sequences of dependent random variables.