Бросание кости до первой шестерки

Задание. На открытом уроке учительница математики моделировала геометрическое распределение[1] для иллюстрации распределения случайной величины «число попыток до достижения первого успеха». Она попросила каждого ученика бросать кубик до тех пор, пока не выпадет шестерка, и записать, сколько на это потребовалось бросков. Каждый ученик должен был повторить этот эксперимент три раза. Набор полученных чисел дал распределение случайной величины «число бросков до первой шестерки».

На уроке присутствовала другая учительница, которая решила повторить такую работу в своем классе. Она заметила, что у некоторых учеников три шестёрки выпали довольно быстро, и они бездельничали, поджидая, пока невезучие одноклассники их нагонят.

Поэтому вторая учительница усовершенствовала опыт: она попросила каждого из своих учеников бросать кубик ровно пять минут, отмечать выпадение шестерок и каждый раз записывать, через сколько бросков случилась очередная шестерка. После этого она взяла все числа, записанные школьниками, и тоже построила распределение величины «число бросков до первой шестерки».

Действительно ли два этих опыта моделируют одно и то же геометрическое распределение? Если они различаются, то каким образом и почему?

Для написания эссе на эту непростую тему потребуется довольно много экспериментировать. Для экспериментов можно использовать готовые программы для бросания игральных костей или даже обычный Excel.

Эссе Александры Нестеренко

В первом эксперименте честно моделируется распределение длин серий до первой шестерки. Если длина серии (включая конечную шестерку) k, то вероятность такой серии $g_k = pq^{k-1}$, где $p = 1/6$, $q = 1 - p = 5/6$.

Во втором эксперименте моделируется какое-то другое распределение R. Представим себе, что вторая учительница дала на эксперимент не 5 мин, а так мало времени, что хватит сделать только один бросок кости. Тогда примерно у шестой части учеников выпадет шестерка, и они сообщат, что число бросков до шестерки – единица. Все остальные ученики не сообщат ничего. Мы должны заключить, что вероятность шестерки после одного броска равна 1, после двух и больше бросков – 0.

Если времени хватит на два броска, то не равные нулю вероятности будут для чисел бросков 1 и 2, равные нулю – для всех чисел бросков, больших, чем 2. Из этого можно сделать несколько выводов.

1. Вероятность серии длины k из второго эксперимента зависит не только от k но и от числа бросков L, которое успеет сделать ученик, то есть распределение R имеет параметры p и L: $R = R(L, p)$ Обозначим функцию вероятности этого распределения $r_{L,k}$:

$$r_{L,k} = \mathrm{P}(X = k), \text{ если } X \sim R(L, p).$$

[1] Название связано с тем, что вероятности в этом распределении образуют геометрическую прогрессию с первым членом p и со знаменателем $q = 1 - p$ (p и q — традиционные обозначения вероятностей успеха и неудачи в каждом отдельном испытании).

2. Во втором эксперименте по сравнению с первым завышается вероятность коротких серий и занижается вероятность длинных серий.

3. $r_{1,1}=1$. Можно вычислить вероятность $r_{L,L}$. Если N учеников делают по L бросков, то математическое ожидание количества учеников, которые сообщат о длине серии L равно $pq^{L-1}N$ (это вероятность элементарного исхода длиной L в опыте до первого успеха, умноженная на количество опытов). В среднем за L бросков будет Lp успехов. Поэтому ученики сообщат о в среднем LpN экспериментах до первого успеха. Тогда

$$r_{L,L}=\frac{pq^{L-1}N}{LpN}=\frac{q^{L-1}}{L}.$$

Если во втором эксперименте число бросков ограничено шестью и измеряется вероятность серии до первого успеха длины 6, то вероятность серии длины 6, совпадает с геометрической:

$$r_{6,6}=\frac{qp^5N}{6pN}=pq^5=g_6.$$

Обозначим $\delta_{L,k}$ относительное отклонение $r_{L,k}$ от g_k:

$$\delta_{L,k}=\frac{r_{L,k}-g_k}{g_k}=\frac{r_{L,k}}{g_k}-1.$$

Уже выяснилось, что

$$\delta_{1,1}=\frac{r_{1,1}}{g_1}-1=\frac{1}{p}-1,\ \delta_{6,6}=0 \text{ и } \delta_{L,L}=\frac{r_{L,L}}{g_L}-1=\frac{1}{pL}-1=\frac{1/p-L}{L}.$$

В числителе выражения для $\delta_{L,L}$ заменим L переменной k и положим

$$\delta_{L,k}=\frac{1/p-k}{L}. \tag{1}$$

Формула (1) дает правильное значение относительного отклонения для случаев, когда измеряется вероятность серии до первого успеха той же длины, что разрешенное число бросков. Но (1) имеет и другие достоинства.

1. Для данной длины серии k отклонение $R(L,p)$ от геометрического распределения $G(p)$ уменьшается с ростом L. А это и ожидалось. Ведь если последовательность бросков очень длинная, уже не важен вклад одной серии до первой шестерки, которая не завершена и отброшена.

2. Формула (1) правильно показывает, что вероятность для малых длин серий больше, а для больших – меньше по сравнению с вероятностями геометрического распределения.

3. Формула (1) простая. Если бы надо было придумать выражение, которое меняет знак с ростом k и уменьшается с ростом L, то проще, чем $\dfrac{Const-k}{L}$ не получилось бы, наверное.

Примем формулу (1) в качестве рабочей гипотезы для относительного отклонения $\delta_{L,k}$. Если гипотеза (1) верна, то функция вероятности распределения $R(L,p)$ из второго эксперимента равна

$$r_{L,k} = \left(\delta_{L,k}+1\right) g_k = \frac{\frac{1}{p}+L-k}{L} pq^{k-1} = \frac{(L-k)p+1}{L} q^{k-1} \qquad (2)$$

при $k=1,2,\ldots,L$.

Используем гипотезу (1), чтобы сравнить эксперименты двух учительниц количественно. Я провела свой опыт, чтобы определить, сколько раз успею бросить кость за пять минут, если при этом записываю числа бросков до очередной шестерки. Мой результат 103 броска.

Если в классе 25 человек, то в первом эксперименте было ровно $N_1 = 25\cdot 3 = 75$ серий до первой шестерки, а во втором примерно $N_2 = 25\cdot 103/6 \approx 429$ серий. Функция вероятности во втором эксперименте равна

$$r_{103,k} = \frac{(103-k)/6+1}{103}\cdot\frac{5^{k-1}}{6^{k-1}} = \frac{109-k}{103}\cdot\frac{5^{k-1}}{6^k}.$$

Обозначим γ_k частоту серии до первой шестерки длины k, которая измеряется в первом эксперименте. Будем следить за какой-то фиксированной длиной серии k. В первом эксперименте каждый из N учеников сообщает свои три длины серии. Нас интересуют только серии длины k, поэтому можно считать сообщение о серии такой длины успехом (У), а сообщение о серии любой другой длины – неудачей (Н). Тогда последовательность из N_1 сообщений вида

ННУННННННУН...УН

будет элементарным исходом эксперимента первого типа, и

$$\gamma_k = \frac{\textit{количество успехов в элементарном исходе}}{N_1}.$$

Серии из постоянного количества испытаний описываются схемой Бернулли. Поэтому для оценки разброса γ_k можно пользоваться дисперсией частоты для схемы Бернулли $\mathrm{D}\gamma_k = \dfrac{g_k(1-g_k)}{N_1}$.

Для количественной оценки разброса результатов первого эксперимента возьмем коэффициент вариации, то есть отношение стандартного отклонения к ожидаемому количеству серий длины k_0:

$$cv(\gamma_k) = \frac{1}{g_k}\cdot\sqrt{\frac{g_k(1-g_k)}{N_1}} = \sqrt{\frac{1-g_k}{N_1 g_k}}.$$

Второй эксперимент к схеме Бернулли не сводится, так как здесь нет постоянного числа испытаний. Но для оценки, наверное, можно считать, что количество сообщений об успехе близко к $N_2 = 429$, и для приближения также использовать биномиальное распределение. Поэтому

$$cv(\rho_k) = \sqrt{\frac{1 - r_{103,k}}{N_2 r_{103,k}}},$$

где ρ_k – частота серий длины k во втором эксперименте. В таблице даны результаты вычислений для k от 1 до 10. Рассеивание во втором эксперименте меньше, чем в первом. Этого и следовало ожидать, так как N_2 больше, чем N_1 в четыре с лишним раза. Но рассеивание частот при моделировании геометрического распределения намного больше, чем разность между вероятностями распределений. Отсюда можно сделать два вывода.

Длина серии k	1	2	3	4	5	6	7	8	9	10
g_k	0,167	0,139	0,116	0,096	0,080	0,067	0,056	0,046	0,039	0,032
$r_{103,k}$	0,175	0,144	0,119	0,098	0,081	0,067	0,055	0,046	0,038	0,031
$\delta_{103,\,k}$	0,048	0,039	0,029	0,019	0,010	0	-0,01	-0,019	-0,029	-0,039
$cv(\gamma_k)$	0,26	0,29	0,32	0,35	0,39	0,43	0,47	0,52	0,58	0,63
$cv(\rho_{103,k})$	0,10	0,12	0,13	0,15	0,16	0,18	0,20	0,22	0,24	0,27
$\dfrac{\lvert g_k - r_{103,k}\rvert}{\sqrt{D\gamma_k}}$	0,194	0,128	0,088	0,045	0,020	0,001	0,031	0,021	0,034	0,064

Во втором эксперименте невозможно заметить, что моделируется распределение, отличающееся от геометрического.

«Нечестный» второй эксперимент в данных условиях может измерять геометрическое распределение даже с большим успехом, чем «честный» первый.

***

На рис.1 показаны результаты численного моделирования второго эксперимента для $N_2 = 10000$ учащихся и количества бросков $L = 100$.

Красными точками показаны частоты длин серий, полученные при моделировании второго эксперимента. Синяя линия – график вероятности $g_k = pq^{k-1}$. Вероятность в геометрическом распределении и частота $\rho_{100,k}$ серий различаются не сильно, и видна закономерность: при $k < 6$ частота $\rho_{100бk}$ выше соответствующей вероятности геометрического распределения, при $k > 6$ – ниже. Это подтверждает вывод, что в условиях второго эксперимента вероятность по сравнению с геометрическим распределением перераспределяется в пользу коротких серий за счет длинных.

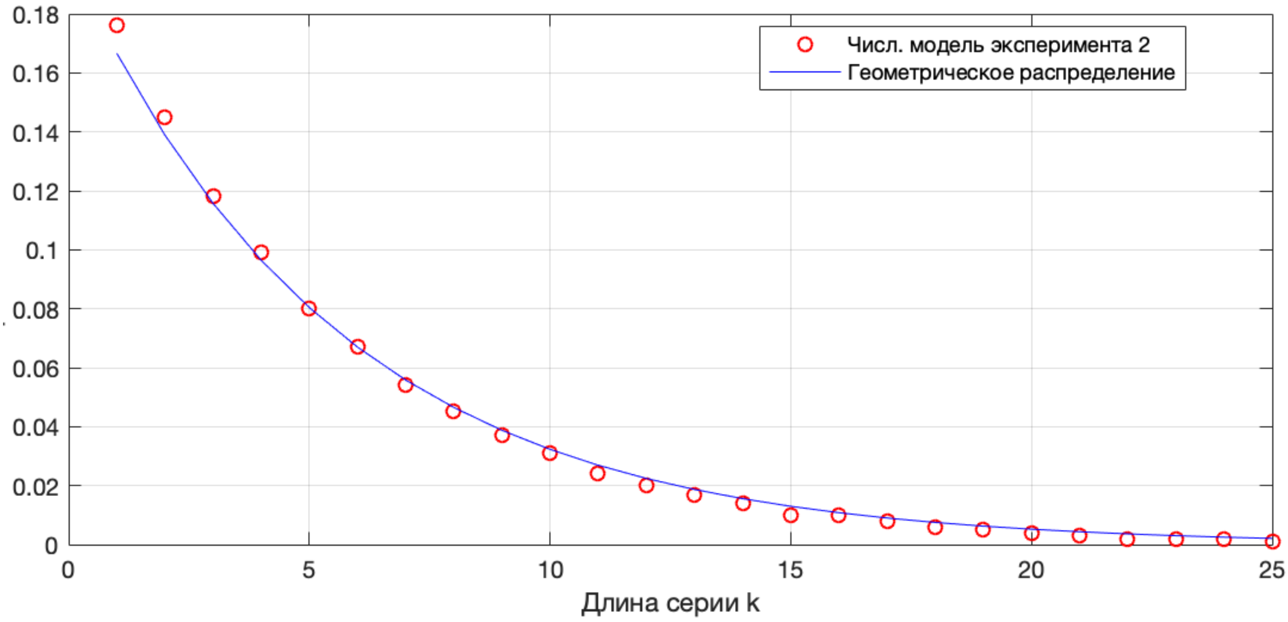

Рис.1.

Гораздо лучше различия между распределениями видны на графике относительной погрешности (рис.2).

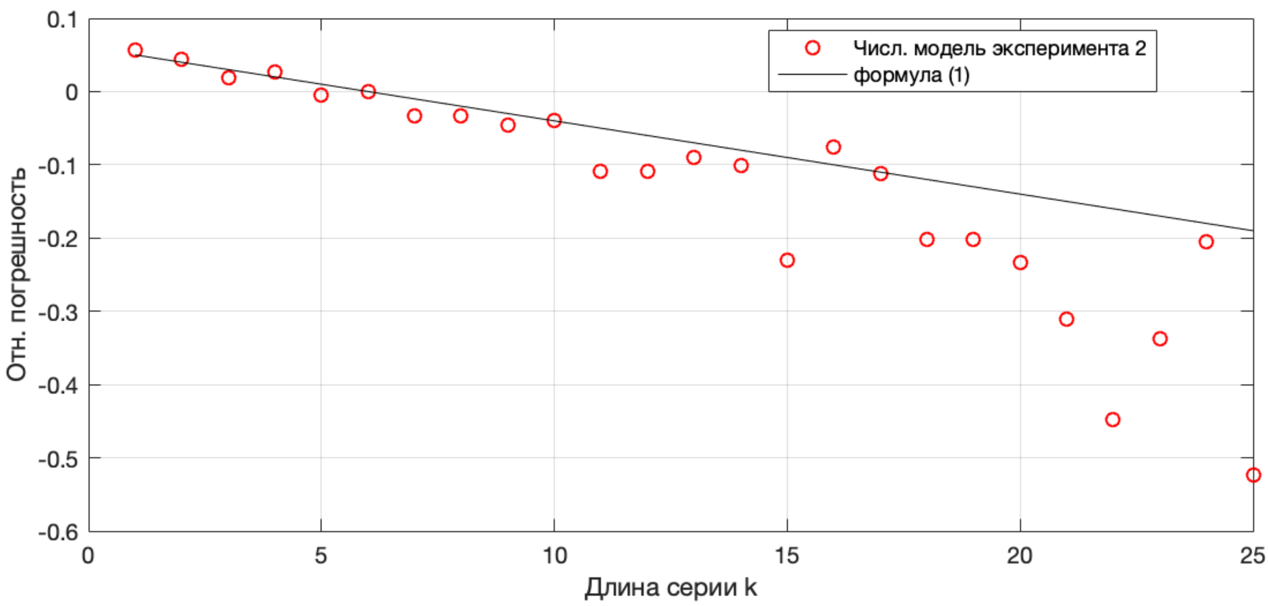

Рис.2.

Красные точки на рис. 2 – это относительное отклонение частот серий в численной модели от геометрической вероятности. Черная прямая – график гипотетического относительного отклонения, вычисленного по формуле (1). При не очень больших длинах серий (k < 15), точки лежат близко к прямой, при больших длинах серий разброс точек становится слишком большим и трудно сказать что-то определенное.

Комментарий. Когда в олимпиаду этого года я предлагал эссе о различиях двух распределений, меня вдохновляла совершенно реальная ситуация. Часто учителя, проводящие у себя в классе практическую работу, допускают именно эту ошибку: не фиксируют количество серий бросков до шестерки, а позволяют школьникам бросать кость «до свистка», отбрасывая затем конечные броски, не давшие шестерку.

При этом я не без оснований полагал, что получающееся распределение, если можно так выразиться, обрезанное геометрическое с функцией вероятности

$$r_{L,k} = \mathrm{E}\frac{X_k}{Y} = \frac{pq^{k-1}}{1-q^L} \text{ при } k = 1, 2, \ldots, L \text{ и } r_{L,k} = 0 \text{ при } k > L,$$

где L – число бросков, которое экспериментатор успел сделать до окончания отведенного времени[2], X_k – число серий длины k, окончившихся шестеркой, а Y – общее случившееся количество серий, окончившихся шестеркой, во всей последовательности из L бросков.

Когда я говорю «не без оснований», я имею в виду, что это действительно так, если мы имеем одну последовательность из L бросков, и вычисляем $r_{L,k}$ как отношение числа X_k серий, давших шестерку при k броске, к общему количеству Y серий, окончившихся шестеркой. Это можно доказать с помощью несложной комбинаторики или понять из симметрии: вопрос эквивалентен вопросу о том, какова вероятность того, что первая серия в такой последовательности будет иметь длину k.

Иными словами, если бы вторая учительница попросила каждого из своих учеников в своем индивидуальном эксперименте подсчитать отношение числа серий длины k к общему числу серий, окончившихся шестеркой, а потом усреднила бы все полученные частоты, то получила бы что-то близкое к тому, что написано выше.

Поэтому, получив эссе Александры Нестеренко, я решил, что где-то ошибка. Во всяком случае, гипотеза о том, что относительное отклонение распределения второй учительницы от геометрического равно $\delta_{L,k} = \frac{6-k}{L}$ уж точно, как мне казалось, не выдерживает критики: она не согласуется с распределением $r_{L,k} = \frac{pq^{k-1}}{1-q^L}$ и не может быть получена из сравнения одной-единственной пары известных вероятностей.

Однако, разбираясь в тексте, я понял, что не могу найти ошибку. Тогда я занялся компьютерным моделированием, которое раз за разом подтверждало Сашину правоту, а вовсе не мою.

Несколько раз проделывая одни и те же выкладки и рассуждения, я искал ошибку уже у себя, и, наконец, нашел. Ошибка оказалась поучительной. Дело в

[2] Во избежание путаницы в комментарии оставляю все буквенные обозначения из текста эссе.

том, что данные не являются частотами из повторяющихся последовательностей бросков, а частота собирается сразу по всему классу, где кости бросают ученики с номерами i от 1 до N, и каждый из них получает в результате своей деятельности Y_i серий, из которых $X_{i,k}$ имеют длину k. И тогда частотная оценка вероятности зависит от всех Y_i и $X_{i,k}$, то есть искомое распределение, имеет не параметр L «число сделанных бросков», а векторный параметр $\boldsymbol{L} = \left(L_1, L_2, ..., L_N\right)$, содержащий длины всех последовательностей бросков, выполненных классом, а, значит, еще и параметр N «число учащихся». Таким образом, получается не одно, а целое семейство распределений $R\left(N, \boldsymbol{L}, p\right)$, а оценки вероятностей какого-то из этих распределений, сделанные второй учительницей, равны

$$\hat{r}_{\boldsymbol{L},k} = \frac{\sum_{i=1}^{N} X_{i,k}}{\sum_{i=1}^{N} Y_i}.$$

Говоря о каком-то распределении, я имею в виду то, которое получается при конкретном числе учащихся N и случившемся $\boldsymbol{L}$.

Чтобы хоть как-то понять, с чем мы имеем дело, разумно найти предельное распределение этого семейства при $N \to \infty$, считая, что все случайные компоненты L_i имеют одинаковое распределение с математическим ожиданием $\mathrm{E}L$. Переходя к пределу по вероятности, получаем:

$$r_{\mathrm{E}L,k} \xleftarrow[N\to\infty]{P} \hat{r}_{\boldsymbol{L},k} = \frac{\sum_{i=1}^{N} X_{i,k}}{\sum_{i=1}^{N} Y_i} \xrightarrow[N\to\infty]{P} \frac{\mathrm{E}\sum_{i=1}^{N} X_{i,k}}{\mathrm{E}\sum_{i=1}^{N} Y_i} = \frac{\mathrm{E}X_k}{\mathrm{E}Y} = \frac{\mathrm{E}X_k}{p\mathrm{E}L},$$

Предпоследнее равенство вытекает из естественного предположения, что все величины $X_{i,k}$ имеют одно и то же распределение с математическим ожиданием $\mathrm{E}X_k$ (X_k – число серий длины k в случайной бинарной последовательности, полученной случайным школьником) и что все величины Y_i также распределены одинаково и имеют математическое ожидание $\mathrm{E}Y$ (Y – общее число серий в этой случайной бинарной последовательности).

Обозначим среднее число бросков $\mathrm{E}L$, сделанных учеником, просто буквой L. Это, конечно, нехорошо, но не вызывает путаницы и согласуется с обозначением из Сашиного эссе. Тогда предельное распределение R получает параметр L, и закон больших чисел вкупе с единственностью предела обеспечивает для предельного распределения функцию вероятности

$$r_{L,k} = \frac{\mathrm{E}X_k}{pL}.$$

Числитель правой части можно вычислить непосредственно:

$$\mathrm{E}X_k = \left(\left(L-k\right)p+1\right)pq^{k-1}, \text{ откуда } r_{k,L} = \frac{\left(L-k\right)p+1}{L}q^{k-1},$$

что совпадает с выражением (2), полученным Сашей Нестеренко на основе гипотетической формулы относительного отклонения

$$\delta_{L,k} = \frac{1/p-k}{L},$$

которую Саша сформулировала, сравнив лишь числа $r_{L,L}$ и g_L, и о достоинствах которой пишет: «*Если бы надо было придумать выражение, которое меняет знак с ростом* k *и уменьшается с ростом* L, *то проще, чем* $\dfrac{Const-k}{L}$ *не получилось бы, наверное*». Согласитесь, мы видим пример незаурядной математической интуиции.

Интересно и заключение: «*Нечестный» второй эксперимент в данных условиях может измерять геометрическое распределение даже с большим успехом, чем «честный» первый*».

Мы не хотим агитировать за нечестность. Александра, видимо, тоже. Так или иначе, Саша не повторила резон второй учительницы, хотя тоже, возможно, согласна с тем, что при такой организации опыта дисциплина крепче.

И.Высоцкий