Бросание кости до первой шестёрки

И. Р. Высоцкий

Нижеприведённая задача была предложена в разделе «Темы для эссе» на заочной олимпиаде по теории вероятностей в 2020/2021 учебном году. В ответ оргкомитет получил несколько сочинений на эту тему, среди которых выделялось эссе, написанное Александрой Нестеренко, на тот момент ученицей 8 класса школы 1278 Москвы. Публикуем его с сокращениями.

Задача

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 шестёрка, 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 шестёрки».

На уроке присутствовала другая учительница, которая решила повторить такую работу в своём 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 шестёрок и каждый раз записывать, через сколько бросков случилась очередная шестёрка. После этого она взяла все числа, записанные школьниками, и тоже построила распределение величины «число бросков до первой шестёрки».

[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).

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 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 шестёрки. Если длина серии (включая конечную шестёрку) k, то вероятность такой серии $g_k = pq$, 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 шестёрка и они сообщат, что число бросков до шестёрки — единица. Все остальные ученики не сообщат ничего. Мы должны заключить, что вероятность шестёрки 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, т. е. распределение R имеет параметры p и L: $R = R(L, p)$ Обозначим функцию вероятности этого распределения $r_{L,k}$:

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

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 $pq^{L-1}N$. 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 веро-

ятность серии длины 6, совпадает с геометрической:

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

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

$$\delta_{1,1} = \frac{r_{1,1}}{g_1} - 1 = \frac{1}{p} - 1, \quad \delta_{6,6} = 0$$

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) даёт правильное значение относительного отклонения для случаев, когда измеряется вероятность серии до первого успеха той же длины, что разрешённое 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 шестёрки, 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 $(\mathrm{const} - k)/L$, 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

$$r_{L,k} = (\delta_{L,k} + 1) g_k = \frac{1/p + L - k}{L} pq^{k-1} = \frac{(L-k)p + 1}{L} q, \quad k = 1, 2, \ldots, L. \tag{2}$$

* * *

На рис. 1 показаны результаты численного моделирования второго эксперимента для $N_2 = 10\,000$ учащихся и количества бросков $L = 100$.

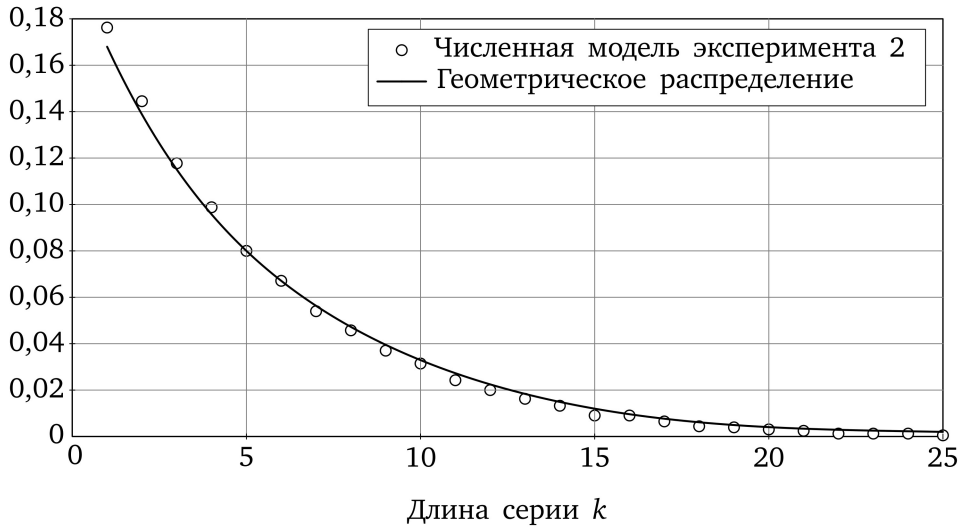

Рис. 1

Отмеченными точками показаны частоты длин серий, полученные при моделировании второго эксперимента. Сплошная 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 $\rho_{100,k}$ 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 счёт длинных. Но рассеивание частот при моделировании геометрического распределения намного больше, чем разность между вероятностями двух распределений. Отсюда можно сделать два вывода.

1. Во втором эксперименте невозможно заметить, что моделируется распределение, отличающееся от геометрического.
2. «Нечестный» второй эксперимент в данных условиях может измерять геометрическое распределение даже с большим успехом, чем «честный» первый.

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

Отмеченные точки на рис. 2 — это относительное отклонение частот серий в численной модели от геометрической вероятности. Чёрная прямая — график гипотетического относительного отклонения, вычисленного по формуле (1). При не очень больших длинах серий

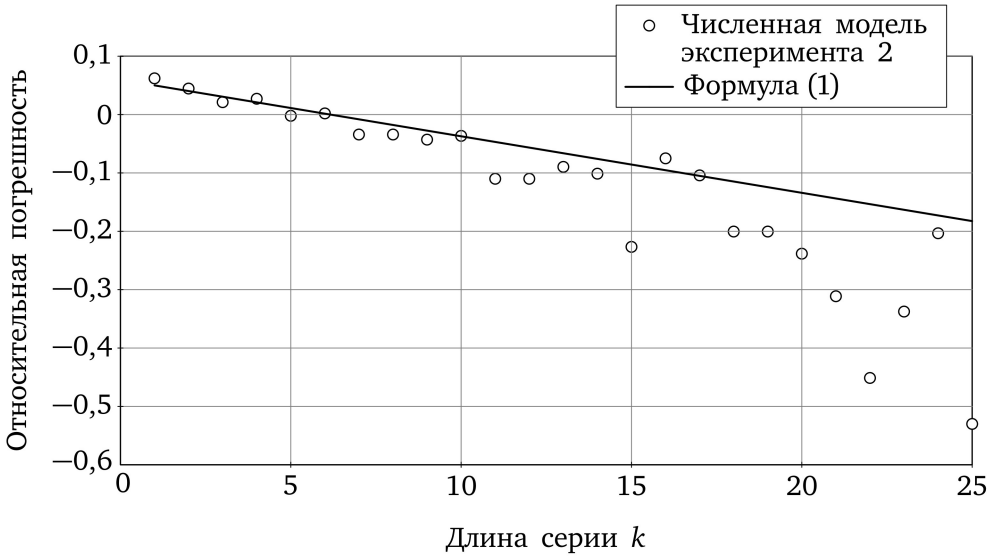

Рис. 2

($k < 15$) точки лежат близко к прямой, при больших длинах серий разброс точек становится слишком большим и трудно сказать что-то определённое.

Комментарий

Когда 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 шестёрки, а позволяют школьникам бросать кость «до свистка», отбрасывая затем конечные броски, не давшие шестёрку.

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

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

и

$$r_{L,k} = 0 \quad \text{при } k > L,$$

где L — число бросков, которое экспериментатор успел сделать до окончания отведённого времени[2)], X_k — число серий длины k, окон-

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

чившихся шестёркой, а Y — общее случившееся количество серий, окончившихся шестёркой, во всей последовательности из L бросков.

Говоря «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 шестёрку при k-м броске, к общему количеству Y серий, окончившихся шестёркой. Это можно показать с помощью несложной комбинаторики или с помощью рассуждений, привлекающих симметрию: вопрос эквивалентен вопросу о том, какова вероятность того, что первая серия в такой последовательности будет иметь длину ровно k.

Иными словами, если бы вторая учительница попросила каждого из своих учеников в своём индивидуальном эксперименте подсчитать отношение числа серий длины k к общему числу серий, окончившихся шестёркой, 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 $\delta_{L,k} = (6-k)/L$, уж точно, как мне казалось, не выдерживает критики: она не согласуется с распределением

$$r_{L,k} = \frac{pq^{k-1}}{(1-q)}$$

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. Теперь я искал ошибку уже у себя, и, наконец, нашёл.

Ошибка оказалась поучительной. Дело в том, что данные не являются частотами из повторяющихся последовательностей бросков, а частота собирается 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. Тогда частотная оценка вероятности зависит от всех Y_i и $X_{i,k}$, т. е. искомое распределение имеет не параметр L «число сделанных бросков», а векторный параметр $\boldsymbol{L} = (L_1, L_2, \ldots, L_N)$, содержащий длины всех последовательностей бросков, выполненных классом, а, значит, ещё и параметр N «число учащихся». Таким образом, получается не одно распределение, а целое семейство распределений $R(N, \boldsymbol{L}, p)$, а оценки вероятностей какого-то

из этих распределений, сделанные второй учительницей, равны

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

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

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:

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

Предпоследнее равенство вытекает из естественного предположения, что все величины $X_{i,k}$ имеют одно и то же распределение с математическим ожиданием $\mathrm{E}X_k$, где X_k — число серий длины k в случайной последовательности бросков до первой шестёрки, 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 последовательности).

Обозначим среднее число бросков $\mathrm{E}L$, сделанных учеником, просто L. Это, конечно, нехорошо, но не вызывает путаницы и согласуется с обозначением из Сашиного эссе. В интерпретации с бросанием костей в классе такой переход можно оправдать предположением, что все бросают кости с одинаковой скоростью, например, раз в две секунды, повинуясь ритмичным взмахам учительских бровей.

Тогда предельное распределение R получает параметр L, и закон больших чисел вкупе с единственностью предела обеспечивает для предельного распределения функцию вероятности

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

Найдём числитель правой части. Каждая серия длины k начинается с какого-то броска. Пусть I_j — индикатор события «серия длины k началась с броска номер j» и, следовательно, закончилась, броском номер $j+k-1$.

$$\mathrm{E}I_j = P(I_j = 1) = \begin{cases} q^{k-1}p & \text{при } j = 1, \\ p \cdot q^{k-1}p = q^{k-1}p^2 & \text{при } 1 < j \leqslant L-k+1, \\ 0 & \text{при } j > L-k+1. \end{cases}$$

Тогда, переходя к математическим ожиданиям в равенстве $X_k = \sum_{j \geqslant 1} I_j$, получаем:

$$\mathrm{E}X_k = q^{k-1}p + (L-k+1-1)q^{k-1}p^2 = ((L-k)p+1)q^{k-1}p,$$

откуда

$$r_{L,k} = \frac{(L-k)p+1}{L} q^{k-1},$$

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

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

которую Саша вывела из весьма эфемерных предположений и сравнения чисел $r_{L,L}$ 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* $(\mathrm{const}-k)/L$ *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. Саша, видимо, тоже. Так или иначе, она не повторила вслух резон второй учительницы, хотя тоже, возможно, согласна с тем, что при такой организации опыта дисциплина крепче.

Иван Ростиславович Высоцкий, МЦНМО
i_r_vysotsky@hotmail.com