О счастливых билетах по-казански

М.Д. Бронштейн, Э.Ю. Лернер

И, потихоньку набирая ход,
Слегка искрящий жёлтенький трамвайчик
Бредёт туда, где неуклюжий мальчик
Опять за три копейки счастья ждёт...
Л. Сергеев

Следующая задача предложена авторами для студенческой олимпиады, посвящённой 225-летию со дня рождения Н.И. Лобачевского:

Задача 1. *Билет называется счастливым по-питерски, если сумма его цифр, стоящих на чётных местах, совпадает с суммой цифр, стоящих на нечётных местах. Назовем билет счастливым по-казански, если все его цифры можно разбить на две непересекающиеся группы так, что сумма цифр в одной совпадает с суммой цифр в другой.*
а) Пусть $P_{\Pi}(n)$ *— вероятность того, что случайно взятый* n*-значный билет является счастливым по-питерски. Найти* $\lim_{n\to\infty} P_{\Pi}(n)$.
б) Пусть $P_{\mathrm{K}}(n)$ *вероятность того, что случайно взятый* n*-значный билет является счастливым по-казански. Найти* $\lim_{n\to\infty} P_{\mathrm{K}}(n)$.

Во избежание недоразумений поясним, что номера в катушке с билетами могут начинаться с цифры “0”. Например, для $n = 3$ множество всех билетов имеет вид : $\{000, 001, 002, \dots, 999\}$. В частности, $P_{\Pi}(2) = P_{\mathrm{K}}(2) = 0.1$.

Основной целью этой заметки является обобщение этой задачи, в котором вместо цифр $\{0, 1, \dots, 9\}$ рассматривается произвольное целочисленное множество с каким угодно вероятностным распределением на нём.

Приведём сначала аккуратное решение задачи 1.

Пункт а) этой задачи совсем не оригинален: вероятности приобретения случайного билета счастливого по-питерски или по-московски посвящено много литературы (см. [1, с. 147–148], [7], [5]). Ниже мы дадим сначала решение задачи 1.а с помощью центральной предельной теоремы. Последняя не часто используется в этой задаче, хотя может быть применена и для получения точной асимптотики (см. Упражнение 1). Мы также дадим второй способ решения этой задачи. Он чуть длиннее предыдущего, но использует только элементарный математический аппарат. Этот способ обобщается на случай билетов, “цифры” которых берутся из бесконечного подмножества $\mathbb{Z}$ (см. Упражнение 4.а).

Решение задачи 1.а, 1-й способ. Пусть $(x_1, \dots, x_n)$ — цифры рассматриваемого билета, обозначим через ξ_i разности $x_{2i} - x_{2i-1}$, $i = 1, \dots, [n/2]$. Очевидно, что ξ_i — независимые случайные величины с нулевым средним и конечной дисперсией, обозначим её через V. Центральная предельная теорема (см. [2]) утверждает, что в этом случае для любого $\varepsilon > 0$ имеет место соотношение

$$\mathbb{P}(|\sum_{i=1}^{[n/2]} \xi_i| < \varepsilon\sqrt{[n/2]V}) \xrightarrow{n\to\infty} \int_{-\varepsilon}^{\varepsilon} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\, dx. \qquad (1)$$

Для чётного n билет является счастливым, если $\sum_{i=1}^{[n/2]} \xi_i = 0$. Для нечётного n билет является счастливым, если $\sum_{i=1}^{[n/2]} \xi_i = x_n$. В любом случае вероятность того, что билет является счастливым меньше

$$\mathbb{P}(|\sum_{i=1}^{[n/2]} \xi_i| \leqslant \mathrm{const}), \qquad (2)$$

где $\mathrm{const} = \max |x_n|$ (т.е. $\mathrm{const} = 9$ в данном случае).

Зафиксируем в (1) сколь угодно малое $\varepsilon > 0$. При $n > 2(\mathrm{const}/\varepsilon)^2/V$ сравнивая левую часть этого неравенства с (2) получаем, что $P_\Pi(n) \xrightarrow{n\to\infty} 0$. $\square$

Упражнение 1. *Пользуясь центральной предельной теоремой, вычислить предел* $\lim_{n\to\infty} \frac{P_\Pi(n)}{1/\sqrt{n}}$.

Решение задачи 1.а, 2-й способ. Зафиксируем произвольно большое натуральное число N и докажем, что $\lim_{n\to\infty} P_\Pi(n) \leqslant 1/N$. Рассмотрим случайные величины $Z(n) = \sum_{i=1}^{n} x_i(-1)^{i-1}$ в кольце остатков от деления на N. Обозначим через $A(n,k)$ событие: $Z(n) = k \bmod N$. Билет $(x_1,\dots,x_n)$ будет счастливым по-питерски только тогда, когда произойдет событие $A(n,0)$. Мы докажем, что вероятности событий $A(n,k)$ стремятся к $1/N$ независимо от $k = 0,\dots,N-1$ при $n\to\infty$.

Обозначим для краткости вероятность $\mathbb{P}(A(n,k))$ через $p(n,k)$. В силу формулы полной вероятности имеем

$$p(n+m,k) = \sum_{j=0}^{N-1} p(n,j)\; p\big(m,(k-j)(-1)^n\big),$$

здесь N событий $A(n,j)$ образуют полную группу событий с вероятностями $p(n,j)$, а $p\big(m,(k-j)(-1)^n\big)$ — соответствующие им условные вероятности перехода из состояния $A(n,j)$ в состояние $A(n+m,k)$.

Для дальнейшего нам важно знать лишь, что все p неотрицательны, $\sum_{j=0}^{N-1} p(\ell,j) = 1$, $\ell\in\mathbb{N}$, причем $\min_{0\leqslant j<N}\{p(m,j)\} = \varepsilon_m > 0$, при $m > [N/9]$. Пусть $\Delta(n,j) = p(n,j) - (1/N)$, заметим, что $\sum_{j=0}^{N-1}\Delta(n,j) = 0$.

Для любого $k\in\mathbb{N}$ справедливо неравенство

$$|\Delta(n+m,k)| = \left|\sum_{j=0}^{N-1} \Delta(n,j)\; p\big(m,(k-j)(-1)^n\big)\right| \leqslant \max_{0\leqslant j<N} |\Delta(n,j)|\,.$$

Зафиксируем теперь $m_N = [N/9] + 1$ и $\delta_N = \varepsilon_{m_N}$ и усилим предыдущее неравенство:

$$|\Delta(n + m_N, k)| = \left| \sum_{j=0}^{N-1} \Delta(n,j) \Big(p\big(m, (k-j)(-1)^n\big) - \delta_N \Big) \right| \leqslant$$

$$\max_{0 \leqslant j < N} |\Delta(n,j)| \sum_{j=0}^{N-1} \left| \Big(p\big(m, (k-j)(-1)^n\big) - \delta_N \Big) \right| = (1 - N\delta_N) \max_{0 \leqslant j < N} |\Delta(n,j)|\,.$$

Таким образом, $\max_j |\Delta(n,j)|$ при $n \to \infty$ стремится к нулю быстрее, чем геометрическая прогрессия со знаменателем $(1 - N\delta_N)^{8/N}$. Отсюда вытекает, что $\lim_{n\to\infty} p(n,k) = 1/N$ при всех k и, в частности, $\lim_{n\to\infty} \mathbb{P}_{\Pi}(n) \leqslant 1/N$. $\square$

Приведенное выше доказательство по сути устанавливает, что любая конечная марковская цепь с циркулянтной матрицей перехода имеет предельное распределение с равными вероятностями для всех достижимых состояний.

Решение задачи 1.б. Пусть $(x_1, \ldots, x_n)$ — цифры рассматриваемого билета, $S = \sum_{i=1}^n x_i$ — их сумма. Легко проверить, что с вероятностью $1/2$ сумма S нечётна. Очевидно, что в этом случае рассматриваемый билет не является счастливым по-казански.

Прежде чем рассматривать случай чётной суммы S оценим долю билетов, содержащих по крайней мере 8 единичек. Обозначим рассматриваемое событие через A. Для грубой оценки снизу $\mathbb{P}(A)$ разобъем набор из n цифр на непересекающиеся группы по 8 цифр. Для каждой такой группы вероятность того, что она отлична от “11111111” равна $1 - 0.1^8$, поэтому

$$\mathbb{P}(A) \geqslant 1 - \left(1 - 0.1^8\right)^{[n/8]} \xrightarrow{n\to\infty} 1.$$

Таким образом, для любого заранее заданного $\varepsilon > 0$ мы можем считать, что n настолько велико, что доля билетов, среди цифр которых не содержится 8 единичек, меньше ε. “Счастливость по-казански” таких билетов нам сложно анализировать. Зато остальные билеты удобны для дальнейшего рассмотрения.

Среди билетов с чётной суммой S возьмём только такие удобные для рассмотрения билеты. Докажем, что все они являются счастливыми по-казански. Очевидно, что нам достаточно представить алгоритм выбора цифр дающих в сумме $S/2$.

Вычеркнем сначала из рассматриваемого билета 8 единичек. Будем включать в искомый набор по очереди все невычеркнутые цифры до тех пор, пока они либо не закончатся, либо их сумма после взятия следующей цифры не станет больше $S/2$ (очевидно, что если сумма станет в точности равной $S/2$, то искомый набор найден). Если взятый набор содержит все невычеркнутые цифры, то их сумма S' меньше $S/2$, а сумма вычеркнутых единичек соответственно больше $S/2$, то есть $S/2 < 8$, тем более $S/2 - S' < 8$. Если же взятый набор содержит не все невычеркнутые цифры, то его сумма

S' отличается от $S/2$ не более, чем на 8, по построению. В обоих ситуациях дополнив взятый набор необходимым числом единиц (из тех, что были вычеркнуты вначале), мы получим искомый набор.

Таким образом, мы доказали, что для любого сколь угодно малого $\varepsilon > 0$ при всех достаточно больших n выполнено $1/2 - \varepsilon \leqslant P_{\mathrm{K}}(n) \leqslant 1/2$, значит $P_{\mathrm{K}}(n) \xrightarrow{n\to\infty} 1/2$. □

Следующая таблица иллюстрирует изменение точных значений вероятностей $P_{\Pi}(n)$ и $P_{\mathrm{K}}(n)$ с ростом n:

n	$P_{\Pi}(n)$	$P_{\mathrm{K}}(n)$
2	0.10	0.10
3	0.055	0.136
4	0.0670	0.2056
5	0.04840	0.29246
6	0.055252	0.376414
7	0.0436975	0.4366881
8	0.04816030	0.47111408
9	0.040051495	0.487875964
10	0.0432457640	0.4951921240
11	0.03715101654	0.49815780829
12	0.039581170420	0.499304300676
13	0.0347847754670	0.4997363405880
14	0.03671331273480	0.49989815235610
15	?	?
16	0.0343900019857310	0.4999832460244272
17	0.03113537578058979	0.49999282163551040
18	0.032458256583753952	0.499996822399017380
19	0.0296896918816556380	0.4999985554326500949
20	0.03081918923741896840	0.49999932964605448756
21	0.028426639840832015685	0.499999684083134646700

Обратим внимание читателей на небольшие осцилляции $P_{\Pi}(n)$, в то время как $P_{\mathrm{K}}(n)$ монотонно стремится к своему предельному значению с ростом n.

Упражнение 2. *Вычислите (написав компьютерную программу) точные значения $P_{\Pi}(15)$ и $P_{\mathrm{K}}(15)$.*

Оставшаяся часть заметки посвящена обобщению задачи 1, в котором вместо цифр билета, равномерно распределённых на множестве $\{0, \ldots, 9\}$, рассматриваются произвольные одинаково распределённые независимые случайные величины x_i. Эти величины будут принимать значения в конечном целочисленном множестве $Y = \{y_1, \ldots, y_r\}$, $r > 1$. Вероятностное распределение на этом множестве обозначим через $\mathbf{p}$: $\mathbf{p} = (p_1, \ldots, p_r), p_j > 0$; набор из n экземпляров таких независимых случайных величин x_i обозначим через $\mathbf{x}(n)$.

Теорема 1. *Пусть $P_{\Pi}(n)$ вероятность того, что для случайного набора $\mathbf{x}(n)$ выполняется равенство $\sum_{i:i \bmod 2=0} x_i = \sum_{i:i \bmod 2=1} x_i$. Пусть $P_{\mathrm{K}}(n)$*

вероятность того, что индексы $[n] = \{1, \ldots, n\}$ *элементов набора* $\mathbf{x}(n)$ *можно разбить на части* I *и* $J = [n] \setminus I$ *так, что* $\sum_{i\in I} x_i = \sum_{i\in J} x_i$ *(назовем такие наборы* $\mathbf{x}(n)$ *счастливыми по-казански). Тогда*
а) $\lim_{n\to\infty} P_{\Pi}(n) = 0$; *б)* $\lim_{n\to\infty}(P_{\mathrm{K}}(n) + P_{\mathrm{K}}(n+1))/2 = 1/2$.

Доказательство утверждения пункта а) дословно совпадает с первым решением задачи 1.а.

Наличие суммы $P_{\mathrm{K}}(n) + P_{\mathrm{K}}(n+1)$ в формулировке пункта б) отличает Теорему 1 от задачи 1.

Упражнение 3. *Приведите пример множества* Y, $|Y| = 2$, *для которого* $\lim_{n\to\infty} P_{\mathrm{K}}(n)$ *не существует.*

Очевидно, что в условии Теоремы 1 *без ограничения общности можно считать что наибольший общий делитель чисел* $y_1, \ldots, y_r$ *равен единице.* В дальнейшем изложении мы будем считать выполненным это предположение.

Подчеркнем, что рассматриваемое множество Y конечно. Представляет интерес аналог Теоремы 1 для различных распределений на счётном целочисленном множестве.

Упражнение 4. *Пусть случайные величины* x_i *принимают счётное множество значений.*
а) Докажите, что по прежнему $\lim_{n\to\infty} P_{\Pi}(n) = 0$.
б) Приведите пример распределения, для которого $\lim_{n\to\infty} P_{\mathrm{K}}(n) = 0$.

Пункт б Теоремы 1 имеет непосредственное отношение к проблеме разбиения. Напомним точную постановку этой компьютерной проблемы ([3, 6]).

Компьютерная задача 1 (проблема разбиения). *Имеется произвольный набор натуральных чисел* $(x_1, \ldots, x_n)$. *Существует ли разбиение* $[n]$ *на части* I *и* $J = [n] \setminus I$, *такое, что* $\sum_{i\in I} x_i = \sum_{i\in J} x_i$.

Это проблема относится к классу NP-полных задач, если на значения x_i не накладывается никаких условий. Если они ограничены, как в Теореме 1, то существуют квадратичные и даже линейные алгоритмы для решения задачи о разбиении (см., в частности, [6, раздел 17.3]). Ниже мы приведём простой критерий, который позволит никогда не ошибаться в ситуации, когда этот критерий выдает отрицательный ответ о существовании разбиения (см. Лемму 1). Кроме того, если натуральные числа $(x_1, \ldots, x_n)$ представляют собой реализацию случайного вектора $\mathbf{x}(n)$ из Теоремы 1, то, как мы докажем в Лемме 2, вероятность ошибки в случае прохождения критерия стремится к нулю, когда n стремится к бесконечности.

Важную роль в дальнейшем изложении играет наибольший общий делитель чисел из набора $\mathbf{x}(n)$, *обозначим его через* $d(\mathbf{x})$. Нам так же удобно ввести *обозначение* $S(\mathbf{x})$ *для суммы* $\sum_{i=1}^{n} x_i$.

Лемма 1 (критерий несуществования разбиения). *Если* $S(\mathbf{x})$ *в нечётное число раз больше* $d(\mathbf{x})$, *то набор* $\mathbf{x}(n)$ *не является счастливым по казански.*

Доказательство: Сокращение всех x_i на $d(\mathbf{x})$ сводит всё к случаю $d(\mathbf{x}) = 1$. В этом случае необходимо разбить нечётное $S(\mathbf{x})$ на две равные целые части, что невозможно. $\square$

Ключевая идея доказательства Теоремы 1 основана на следующей Лемме.

Лемма 2. *Пусть для набора $\mathbf{x}(n)$ из Теоремы 1 выполнено: "$S(\mathbf{x})$ в чётное число раз больше $d(\mathbf{x})$". Тогда для любого сколь угодно малого ε можно найти $N = N(\varepsilon, Y, \mathbf{p})$ такое, что при всех $n > N$ рассматриваемый набор $\mathbf{x}(n)$ является счастливым по-казански с вероятностью большей $1 - \varepsilon$.*

Прежде чем доказывать Лемму 2 обсудим вероятность выполнения ограничения "$S(\mathbf{x})$ в чётное число раз больше $d(\mathbf{x})$" из её условия.

Очевидно, что если все y_j, $j = 1, \dots, r$, нечётны, то событие "$S(\mathbf{x})$ в чётное число раз больше $d(\mathbf{x})$" является достоверным при чётном n и невозможным при нечётном n. Следовательно в данной ситуации, согласно Лемме 1, имеем $P_{\mathrm{K}}(2n-1) = 0$, а согласно Лемме 2, $\lim_{n\to\infty} P_{\mathrm{K}}(2n) = 1$. Сейчас мы рассмотрим более типичный случай.

Лемма 3. *Допустим, что в множестве Y содержится хотя бы одно чётное число y. Тогда вероятность того, что для случайного набора $\mathbf{x}(n)$ выполнено ограничение "$S(\mathbf{x})$ чётно", стремится к $\frac{1}{2}$ при $n \to \infty$.*

Доказательство леммы 3: Пусть a_n – вероятность рассматриваемого события для случайного набора $\mathbf{x}(n)$ и $b_n = 1 - a_n$ — вероятность противоположного события. Заметим, что $1 > a_1$, иначе единица не могла бы быть наибольшим общим делителем элементов множества Y. В силу условия леммы, $a_1 > 0$. Поэтому $0 < a_1 < 1$, тем более $|a_1 - b_1| < 1$.

Имеем

$$a_n = \sum_{\ell:\ell \bmod 2=0} \binom{n}{\ell} b_1^{\ell} a_1^{n-\ell}, \qquad b_n = \sum_{\ell:\ell \bmod 2=1} \binom{n}{\ell} b_1^{\ell} a_1^{n-\ell}.$$

Следовательно, $a_n - b_n = (a_1 - b_1)^n \xrightarrow{n\to\infty} 0$. $\square$

Заметим, что в силу конечности r со сколь угодно близкой к единице вероятностью $1 - \delta$ в случайном наборе $\mathbf{x}(n)$ будут присутствовать все значения y_j, $j = 1, \dots, r$, если только n достаточно велико. Соответственно, доля тех наборов $\mathbf{x}(n)$, для которых $d(\mathbf{x})$ отлично от единицы меньше δ. Следовательно, Лемма 3 остаётся верной при замене события $A =$"$S(\mathbf{x})$ чётно" на событие $A' =$"$S(\mathbf{x})$ в чётное число раз больше $d(\mathbf{x})$" в её формулировке. Аналогично и в Лемме 2, событие A' можно заменить на событие A, поскольку $A' \subseteq A$ и вероятность события $A \setminus A'$ составляет асимптотически нулевую долю от вероятности события A.

Доказательство леммы 2: Как только что было отмечено, в условии Леммы вместо события "$S(\mathbf{x})$ в чётное число раз больше $d(\mathbf{x})$" можно рассматривать событие "$S(\mathbf{x})$ чётно".

Расширенный алгоритм Евклида даёт возможность представления наибольшего общего делителя (единицы) в виде целочисленной линейной комбинации y_j с некими коэффициентами c_j [4]. Пусть $M = \max_j |y_j|$. Нам будут удобны наборы $\mathbf{x}(n)$, в которых найдется по крайней мере $M|c_j|$ чисел, равных y_j, $j = 1, \ldots, r$. Назовем такие наборы *удобными по-казански.*

Доказательство того, что доля удобных по-казански наборов среди всех наборов стремится к единице при n стремящемся к бесконечности легко получается тем же приёмом, что и доказательство аналогичного факта о доле билетов содержащих 8 единичек (см. решение задачи 1.б). Таким образом, нам остается предъявить явный алгоритм построения множеств I и J в ситуации, когда $S(\mathbf{x})$ чётно и рассматриваемый набор $\mathbf{x}(n)$ является удобным по-казански.

Упражнение 5. *Модифицируя решение задачи 1.б, докажите, что если набор $\mathbf{x}(n)$ является удобным по-казански и сумма его цифр $S(\mathbf{x})$ чётна, то этот набор является счастливым по-казански.*

Упражнение 5 завершает доказательство Леммы 2. $\square$

Утверждение Теоремы 1, очевидно, следует из доказанных Лемм. Попутно мы построили простой алгоритм решения задачи о разбиении. Он эффективно работает при достаточно больших n, если статистика исходных данных удовлетворяет нашим предположениям. Как и во многих других задачах, вероятностный взгляд на мир позволил дать простое решение нетривиальной проблемы "почти для всех типичных" случаев.

Список литературы

[1] *Виленкин Н.Я.* Популярная комбинаторика. М.: Наука, 1975.

[2] *Гнеденко Б.В.* Курс теории вероятностей. М.: Едиториал УРСС, 2005.

[3] *Гэри М., Джонсон Д.* Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982.

[4] *Кормен Т.Х., Лейзерсон Ч.И., Ривест Р.Л., Штайн К.* Алгоритмы. Построение и анализ. М.: Вильямс, 2012.

[5] *Ландо С.К.* Счастливые билеты // Математическое просвещение. Сер. 3. Вып. 2. М.: МЦНМО, 1998. С. 127–132.

[6] *Пападимитриу Х., Стайглиц К.* Комбинаторная оптимизация. Алгоритмы и сложность. М.: Мир, 1984.

[7] *Финк Л.М.* Ещё раз о счастливых билетах // Квант. 1976. №12. С. 68–70.

М.Д. Бронштейн, Казанский национальный исследовательский технологический университет, `bronmich@gmail.com`

Э.Ю. Лернер, Казанский (Приволжский) федеральный университет, `eduard.lerner@gmail.com`