Задача коллекционера

И. Р. Высоцкий

Введение и исторический обзор

Всякий родитель сталкивается с математической задачей, которую задаёт чадо, получившее свой первый киндер-сюрприз с игрушкой внутри и описанием всей коллекции игрушек, которую всенепременно нужно собрать полностью. В коллекции могут оказаться 12 принцесс, 10 бегемотов или персонажи мультфильма.

Естественный вопрос — сколько придётся купить шоколадных яиц, чтобы завершить коллекцию — приводит к любопытной вероятностной модели, которая, будучи простой поначалу, развивается в непропорционально и обидно сложное обобщение, как только детей становится двое и требуется собрать две коллекции на двоих.

Впервые задача появилась, вероятно, в сочинении Муавра «Измерение шансов» (De Mensura Sortis, 1712); позже упоминалась Лапласом в «Аналитической теории вероятностей» (Theorie Analytique des probabilities, 1812).

В 1930-х американская компания Dixie Cup[1)], вероятно, одной из первых в истории маркетинга использовала гениальный ход: под крышкой бутылки с напитком был спрятан купон с изображением кого-то из популярных персонажей цирка Dixies. Разумеется, предъявителю полной коллекции купонов была обещана премия. Позже в ход пошла коллекция звёзд Голливуда и игроков бейсбольных команд высшей лиги.

Благодаря этому в англоязычной литературе задача встречается не только как задача коллекционера (Coupon Collector's Problem), но и под именем Dixie Cup Problem, а её обобщение на две коллекции (или больше) часто называют Double Dixie Cup Problem.

В русскоязычной литературе по какой-то причине задача коллекционера практически не встречается (исключение составляет [4]), что стран-

[1)] Dixie, Dixy (англ.) — солдатский котелок.

но и несправедливо, поскольку эта классическая задача, бесспорно, заслуживает помещения в золотой фонд теории вероятностей на одну из почётных витрин.

В этой статье даётся классическое решение для одной, а также решение Ньюмана и Шеппа [1] для двух и более коллекций. Во второй части предлагается альтернативный подход к решению и бегло рассматриваются некоторые сопутствующие задачи.

§ 1. Классическая задача коллекционера

Формально речь идёт о длине серии независимых одинаковых испытаний, каждое из которых имеет $n \geqslant 2$ равновозможных исходов. Серия оканчивается при наступлении события «Каждый из n исходов случился хотя бы раз». В таком случае будем говорить о *завершении коллекции*, а под *коллекцией* в каждый момент времени будем понимать подпоследовательность исходов, впервые наступивших в серии. Такие исходы будем называть *экспонатами* коллекции.

Вместо сухой модели мы будем держать в голове содержательную интерпретацию, а именно — коллекцию из n бегемотов, спрятанных в шоколадных яйцах «Киндер-сюрприз». Этот образ ближе современному человеку, чем архаичные купоны Dixie Cup; кроме того, бегемоты обаятельнее безликих исходов в абстрактных испытаниях.

С точки зрения коллекционера упорядоченность экспонатов не важна, однако мы будем нумеровать экспонаты h_k в том порядке, в каком они появляются в коллекции. Под h_k мы понимаем не отдельного бегемота, а целый вид одинаковых бегемотов, представитель которого стал k-м экспонатом в коллекции.

Предположим, что в очередной попытке, т. е. из очередного яйца, с вероятностью $1/n$ может появиться один из n бегемотов независимо от результатов предыдущих попыток. Обозначим через $\xi = \xi(n)$ число яиц, купленных к моменту, когда собрана вся коллекция. Задача — найти математическое ожидание и дисперсию этой случайной величины. Указание на n почти всегда опускаем, кроме случая, где действительно возникает такая необходимость.

Задача 1. Найти $\mathrm{E}\xi$ и $\mathrm{D}\xi$.

Решение. Предположим, что в некоторый момент в коллекции уже есть $k-1$ различных экспонатов-бегемотов $(k = 1, \ldots n)$. Тогда вероятность получить k-й экспонат h_k при каждой последующей попытке равна $(n-k+1)/n$, поэтому число попыток φ_k до появления экспоната h_k имеет геометрическое распределение с параметром $(n-k+1)/n$, откуда

$$\mathrm{E}\varphi_k = \frac{n}{n-k+1}.$$

Переходя в равенстве

$$\xi = \sum_{k=1}^{n} \varphi_k \tag{1.1}$$

к математическим ожиданиям, получаем:

$$\mathrm{E}\xi = \sum_{k=1}^{n} \mathrm{E}\varphi_k = \sum_{k=1}^{n} \frac{n}{n-k+1} = n\sum_{j=1}^{n} \frac{1}{j} = nH_n, \tag{1.2}$$

где H_n обозначает n-е гармоническое число:

$$H_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}.$$

Пользуясь равенством

$$H_n = \ln n + \gamma + \frac{1}{2n} + o\left(\frac{1}{n}\right),$$

где $\gamma = 0{,}577\ldots$ — константа Эйлера — Маскерони, получаем формулу

$$\mathrm{E}\xi = n\ln n + \gamma n + \frac{1}{2} + o(1).$$

Случайные величины φ_k независимы, поэтому в равенстве (1.1) можно перейти к дисперсиям:

$$\begin{aligned}\mathrm{D}\xi = \sum_{k=1}^{n} \mathrm{D}\varphi_k = \sum_{k=1}^{n} \frac{k-1}{n} \cdot \frac{n^2}{(n-k+1)^2} = \\ = n\sum_{j=1}^{n} \frac{n-j}{j^2} = n^2\sum_{j=1}^{n} \frac{1}{j^2} - n\sum_{j=1}^{n} \frac{1}{j} = n^2 H_{2,n} - nH_n,\end{aligned} \tag{1.3}$$

где $j = n-k+1$, а $H_{2,n}$ обозначает n-е гармоническое число второго порядка:

$$H_{2,n} = 1 + \frac{1}{4} + \frac{1}{9} + \ldots + \frac{1}{n^2}.$$

Известно, что

$$H_{2,n} = \frac{\pi^2}{6} - \frac{1}{n} + \frac{1}{2n^2} + o\left(\frac{1}{n^2}\right),$$

поэтому

$$\mathrm{D}\xi = n^2\frac{\pi^2}{6} - n + \frac{1}{2} - n\ln n - \gamma n - \frac{1}{2} + o(1) = n^2\frac{\pi^2}{6} - n\ln n - (\gamma+1)n + o(1).$$

Пример. Чтобы собрать коллекцию из 10 бегемотов, в среднем потребуется купить примерно

$$10\ln 10 + 10\gamma + \frac{1}{2} \approx 29{,}3 \text{ яиц.}$$

Выбирая в качестве решающего правила три стандартных отклонения, следует рассчитывать не более чем на

$$\mathrm{E}\xi + 3\sqrt{\mathrm{D}\xi} \approx 29{,}3 + 3\sqrt{125{,}7} \approx 63 \text{ попытки.}$$

Равенства (1.2) и (1.3) допускают несколько более общий вид, а именно:

$$\mathrm{E}\zeta = nH_k, \quad \mathrm{D}\zeta = n^2 H_{2,k} - nH_k, \tag{1.4}$$

где ζ — количество попыток, которые потребуются для восполнения k недостающих бегемотов.

§ 2. Две и больше коллекций — решение Ньюмана и Шеппа

Как уже отмечалось, случай, когда требуется собрать две коллекции при тех же условиях, оказывается намного сложнее. Как мы увидим, это связано с неопределённостью *дефицита*, т. е. числа бегемотов, недостающих до завершения последней нужной коллекции в момент, когда завершена предпоследняя. В случае же одной коллекции дефицит определённо равен n.

Точное решение задачи для m коллекций дали Дональд Ньюман и Лоуренс Шепп в 1960 году [1].

Как и прежде, считаем, что каждая попытка оканчивается одним из n равновозможных бегемотов. Испытания проводятся последовательно до тех пор, пока не собраны все возможные экспонаты для m коллекций, т. е. бегемоты каждого вида не появились хотя бы m раз ($m \geqslant 2$). Количество потребовавшихся попыток обозначим через ξ_m.

Задача 2. Найти $\mathrm{E}\xi_m$.

Теорема 1 (Ньюман, Шепп).

$$\mathrm{E}\xi_m = n\int_0^\infty \left(1 - \left(1 - \sum_{k=0}^{m-1} \frac{t^k}{k!} \cdot e^{-t}\right)^n\right) dt. \tag{2.1}$$

Асимптотическое приближение даётся следующей теоремой.

Теорема 2 (Ньюман, Шепп).

$$\mathrm{E}\xi_m = n\ln n + n(m-1)\ln\ln n + c_m + o(1), \tag{2.2}$$

где c_m — некоторая константа.

Природа чисел c_m исследована Эрдёшем и Реньи в 1961 году [2].

Доказательство теоремы 1 чрезвычайно остроумно. Пусть I_i — индикатор события «i попыток оказалось недостаточно, чтобы укомплектовать m коллекций» ($i \geqslant 1$). Вероятность этого события равна

$$\mathrm{E}I_i = \mathrm{P}(I_i = 1) = \frac{W_i}{n^i},$$

где W_i — количество последовательностей, составленных из i бегемотов, в каждой из которых хотя бы один бегемот встречается менее m раз. Переходя в равенстве

$$\xi_m = 1 + \sum_{i=1}^{\infty} I_i$$

к математическим ожиданиям, получаем:

$$\mathrm{E}\xi_m = 1 + \sum_{i=1}^{\infty}\mathrm{E}I_i = 1 + \sum_{i=1}^{\infty}\frac{W_i}{n^i}.$$

Поскольку $W_0 = 1$, можно для общности включить единицу в сумму:

$$\mathrm{E}\xi_m = \sum_{i=0}^{\infty}\frac{W_i}{n^i}.$$

Введём функцию

$$f_i(x_1, \ldots, x_n) = \left\{(x_1 + x_2 + \ldots + x_n)^i\right\},$$

где фигурными скобками обозначена операция удаления из многочлена или степенного ряда всех членов, содержащих каждую из переменных $x_1 \ldots x_n$ в степени m или выше. Тогда

$$W_i = f_i\underbrace{(1, 1, \ldots, 1)}_{n\ \text{единиц}},$$

и

$$\mathrm{E}\xi_m = \sum_{i=0}^{\infty}\frac{W_i}{n^i} = \sum_{i=0}^{\infty}\frac{f_i(x_1, x_2, \ldots, x_n)}{n^i}\bigg|_{x_1=x_2=\ldots=x_n=1}. \tag{2.3}$$

Осталось удалить лишние слагаемые. Это несложно сделать для функции

$$\exp\sum_{j=1}^{n}x_j=\sum_{i=0}^{\infty}\left(\frac{1}{i!}\Big(\sum_{j=1}^{n}x_j\Big)^i\right).$$

Чтобы оставить в этом ряде только те слагаемые, куда хотя бы одно из x_j входит в степени меньше чем m, вычтем из $\exp\sum_{j=1}^{n}x_j$ произведение

$$\sum_{k=m}^{\infty}\frac{x_1^k}{k!}\cdot\sum_{k=m}^{\infty}\frac{x_2^k}{k!}\cdot\ldots\cdot\sum_{k=m}^{\infty}\frac{x_n^k}{k!}.$$

Получим

$$\exp\sum_{j=1}^{n}x_j-\prod_{j=1}^{n}\sum_{k=m}^{\infty}\frac{x_j^k}{k!}=\Big\{\exp\sum_{j=1}^{n}x_j\Big\}=\Big\{\sum_{i=0}^{\infty}\frac{1}{i!}\Big(\sum_{j=1}^{n}x_j\Big)^i\Big\}=$$
$$=\sum_{i=0}^{\infty}\Big\{\frac{1}{i!}\Big(\sum_{j=1}^{n}x_j\Big)^i\Big\}=\sum_{i=0}^{\infty}\frac{f_i(x_1,\ldots,x_n)}{i!}.$$

Полученное выражение отличается от требуемого (2.3) только знаменателем. Нужно придумать способ превратить факториалы $i!$ в степени n^i, оставив всю прочую структуру нетронутой. На помощь приходит равенство

$$\int_0^{\infty}\frac{t^i}{i!}e^{-nt}\,dt=\frac{1}{n^{i+1}}.$$

Получаем

$$\sum_{i=0}^{\infty}\frac{f_i(x_1,\ldots,x_n)}{n^i}=\sum_{i=0}^{\infty}\left(f_i(x_1,\ldots,x_n)\cdot n\int_0^{\infty}\frac{t^i}{i!}e^{-nt}\,dt\right)=$$
$$=n\int_0^{\infty}\sum_{i=0}^{\infty}\frac{f_i(x_1,\ldots,x_n)}{i!}t^ie^{-nt}\,dt=n\int_0^{\infty}\sum_{i=0}^{\infty}\frac{f_i(x_1t,x_2t,\ldots,x_nt)}{i!}e^{-nt}\,dt=$$
$$=n\int_0^{\infty}\left(\exp\sum_{j=1}^{n}x_jt-\prod_{j=1}^{n}\sum_{k=m}^{\infty}\frac{(x_jt)^k}{k!}\right)e^{-nt}\,dt.$$

Тогда

$$\mathrm{E}\xi_m=n\int_0^{\infty}\left(\exp\sum_{j=1}^{n}t-\Big(\sum_{k=m}^{\infty}\frac{t^k}{k!}\Big)^n\right)e^{-nt}\,dt=$$
$$=n\int_0^{\infty}\left(e^{nt}-\Big(e^t-\sum_{k=0}^{m-1}\frac{t^k}{k!}\Big)^n\right)e^{-nt}\,dt=n\int_0^{\infty}\left(1-\Big(1-\sum_{k=0}^{m-1}\frac{t^k}{k!}\cdot e^{-t}\Big)^n\right)dt.\ \square$$

Доказательство теоремы 2. Утверждение теоремы следует из равенства

$$\frac{\mathrm{E}_m\xi(n+1)}{n+1}-\frac{\mathrm{E}_m\xi(n)}{n}=\frac{1}{n}+\frac{m-1}{n\ln n}+\lambda_n,$$

где $\sum_{n=2}^{\infty}|\lambda_n|<\infty$. Доказательство носит технический характер и приведено в [1]. □

Ньюман и Шепп в этот момент пишут: «Если первая коллекция „стоит“ $n\ln n$, то вторая и каждая последующая стоят $n\ln\ln n$». Под стоимостью следует понимать трудозатраты коллекционера, либо — буквально — стоимость элементов коллекции в денежных единицах.

Если теорема 1, являясь образцом искромётного остроумия, не даёт приемлемого метода вычисления, то теорема 2, напротив, даёт метод оценки, но не открывает природы выражения $\ln\ln n$. Может ли этот повторный логарифм появиться из более естественных соображений? Если этот вопрос интересен читателю, то он может пробежать глазами следующий параграф.

§ 3. Дефицит коллекции

Рассмотрим момент, когда $(m-1)$-я коллекция собрана $(m\geqslant 1)$. Для завершения m-й коллекции в эту минуту не хватает какого-то количества бегемотов. Теоретически этот *дефицит* бегемотов δ_m может принимать любое значение от 1 до n.

Теорема 3. *Для всех $m \geqslant 2$ случайные величины δ_m распределены одинаково, причём*

$$\mathrm{E}\delta_m = H_n, \tag{3.1}$$

$$\mathrm{D}\delta_m = H_n^2 - 3H_n + 2H_{2,n}. \tag{3.2}$$

Первое равенство поможет «раскрыть тайну» повторного логарифма Ньюмана — Шеппа в (2.2). Равенство (3.2) говорит, что с ростом n диапазон вероятных значений дефицита растёт примерно с той же скоростью, что и математическое ожидание, а это значительно затрудняет поиск хорошей асимптотики для $\mathrm{E}\xi_m$.

Доказательство. Пронумеруем экспонаты в том порядке, в каком они появлялись в ходе собирания $(m-1)$-й коллекции. Например, бегемот вида h_1 — это первый бегемот $(m-1)$-й коллекции, через какое-то количество попыток за бегемотом h_1 явился бегемот вида h_2, какого ещё не было в $(m-1)$-й коллекции, но бегемоты этого вида уже появлялись прежде, чтобы пополнить предыдущие коллекции, если таковые есть.

Обозначим через A_{jk}, где $j \leqslant k$, событие «при комплектовании $(m-1)$-й коллекции между экспонатами h_k и h_{k+1} не появился ни один бегемот вида h_j». Здесь $k = 1, \ldots, n-1$. Событие A_{jk} состоит в том, что после экспоната h_k, но до экспоната h_{k+1} последовало несколько (может быть, ни одного) уже появлявшихся прежде (старых) бегемотов, но среди них не было h_j. Старых бегемотов всего k видов $(h_1, \ldots, h_k)$, а без бегемота h_j их всего $k-1$ вид. Новым экспонатом h_{k+1} может оказаться бегемот любого из $n-k$ видов, не встречавшихся ещё в $(m-1)$-й коллекции.

Поэтому вероятность события A_{jk} равна

$$\mathrm{P}(A_{jk}) = \frac{n-k}{n} + \frac{k-1}{n} \cdot \frac{n-k}{n} + \left(\frac{k-1}{n}\right)^2 \cdot \frac{n-k}{n} + \ldots = \frac{n-k}{n-k+1}. \tag{3.3}$$

Если к моменту завершения $(m-1)$-й коллекции не хватает бегемота некоторого вида для следующей, m-й коллекции, то будем называть такого бегемота *дефицитным*. Вероятность события A_j «бегемот h_j дефицитный» равна

$$\mathrm{P}(A_j) = \prod_{k=j}^{n-1} \mathrm{P}(A_{jk}) = \frac{n-j}{n-j+1} \cdot \frac{n-j-1}{n-j} \cdot \ldots \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{n-j+1}.$$

В частности, бегемот h_1 дефицитен с вероятностью $1/n$, а бегемот h_n дефицитен с вероятностью 1, что неудивительно, ибо $(m-1)$-я коллекция как раз и завершается появлением экземпляра именно этого вида.

Событие $\delta_m = k$ выражается алгебраически через события A_j. Оно равно объединению всевозможных пересечений k различных событий вида A_j и дополнений к остальным $n-k$ событиям A_j. Вероятности событий A_j не зависят от m, следовательно, вероятность события $\delta_m = k$ также не зависит от m.

Найдём математическое ожидание δ_m:

$$\delta_m = \sum_{i=1}^{n} I_i,$$

где I_i — индикатор события A_i, и

$$\mathrm{E}\delta_m = \sum_{i=1}^{n} \mathrm{E}I_i = \sum_{i=1}^{n} \mathrm{P}(A_i) = \sum_{i=1}^{n} \frac{1}{n-i+1} = H_n.$$

Чтобы найти дисперсию дефицита δ_m, нужны математические ожидания произведений $I_i I_j$ $(i < j)$, т. е. индикаторов событий

B_{ij} «бегемоты h_i и h_j оба дефицитные».

Рассуждения здесь аналогичны тем, что мы только что провели. При $i \leqslant k < j$ согласно (3.3) бегемот h_i не появится между бегемотами h_k и h_{k+1} с вероятностью

$$\mathrm{P}(A_{ik}) = \frac{n-k}{n-k+1}.$$

При $i < j \leqslant k$ вероятность события B_{ijk} «между бегемотами h_k и h_{k+1} нет бегемотов h_i и h_j» равна

$$\mathrm{P}(B_{ijk}) = \frac{n-k}{n} + \frac{k-2}{n} \cdot \frac{n-k}{n} + \left(\frac{k-2}{n}\right)^2 \cdot \frac{n-k}{n} + \ldots = \frac{n-k}{n-k+2}.$$

Следовательно,

$$\mathrm{P}(B_{ij}) = \prod_{k=i}^{j-1} \mathrm{P}(A_{ik}) \cdot \prod_{k=j}^{n-1} \mathrm{P}(B_{ijk}) = \prod_{k=i}^{j-1} \frac{n-k}{n-k+1} \cdot \prod_{k=j}^{n-1} \frac{n-k}{n-k+2} =$$

$$= \frac{(n-i)!(n-j+1)!2!}{(n-i+1)!(n-j+2)!} = \frac{2}{(n-i+1)(n-j+2)}.$$

Теперь в равенстве

$$\delta_m^2 = \sum_{i=1}^{n} I_i^2 + 2 \sum_{1 \leqslant i < j \leqslant n} I_i I_j = \sum_{i=1}^{n} I_i + 2 \sum_{1 \leqslant i < j \leqslant n} I_i I_j$$

переход к математическим ожиданиям даёт:

$$\mathrm{E}\delta_m^2 = \sum_{i=1}^{n} \mathrm{P}(A_i) + 2\sum_{1\leqslant i<j\leqslant n} \mathrm{P}(B_{ij}) = \sum_{i=1}^{n} \frac{1}{n-i+1} + 2\sum_{1\leqslant i<j\leqslant n} \frac{2}{(n-i+1)(n-j+2)}. \tag{3.4}$$

Преобразуем сумму во втором слагаемом:

$$\sum_{1\leqslant i<j\leqslant n} \frac{2}{(n-i+1)(n-j+2)} = \sum_{0\leqslant i\leqslant j\leqslant n-2} \frac{2}{(n-i)(n-j)} =$$

$$= \sum_{0\leqslant i\leqslant j\leqslant n-1} \frac{2}{(n-i)(n-j)} - 2\sum_{i=0}^{n-1} \frac{1}{n-i} =$$

$$= \sum_{i,j=0}^{n-1} \frac{1}{(n-i)(n-j)} + \sum_{i=0}^{n-1} \frac{1}{(n-i)^2} - 2\sum_{i=0}^{n-1} \frac{1}{n-i} = H_n^2 + H_{2,n} - 2H_n.$$

Теперь из (3.4) получаем

$$\mathrm{E}\delta_m^2 = H_n + 2H_n^2 + 2H_{2,n} - 4H_n = 2H_n^2 - 3H_n + 2H_{2,n}.$$

Следовательно,

$$\mathrm{D}\delta_m = \mathrm{E}\delta_m^2 - \mathrm{E}^2\delta_m = 2H_n^2 - 3H_n + 2H_{2,n} - H_n^2 = H_n^2 - 3H_n + 2H_{2,n}. \qquad \square$$

* * *

Чтобы сделать следующий шаг, нужно определить гармоническое число H_x для произвольного неотрицательного x или хотя бы для $x \geqslant 1$. Это можно сделать разными способами. Обычно полагают

$$H_x = \int_0^1 \frac{1-t^x}{1-t}\,dt. \tag{3.5}$$

Очевидно, при $x = n \in N$ это определение согласуется с привычным равенством

$$H_n = \sum_{i=1}^{n} \frac{1}{i}.$$

Легко проверить, что

$$\frac{d^2H_x}{dx^2} < 0,$$

поэтому функция $y = -H_x$ выпукла на промежутке $[1; +\infty)$: её надграфик $\{(x; y) \mid x \geqslant 0,\ y \geqslant -H_x\}$ является выпуклой фигурой.

Возможны и другие естественные продолжения последовательности H_n до непрерывной функции. Например, кусочно-линейное продолжение

$$H_x = H_{[x]} + \frac{\{x\}}{[x]+1}. \tag{3.6}$$

При таком определении функция $y = -H_x$ также выпукла.

* * *

Вернёмся к центральному вопросу о $\mathrm{E}\xi_m$.

Теорема 4. *При* $m \geqslant 2$

$$\mathrm{E}\xi_m < nH_n + n(m-1)H_{H_n}. \tag{3.7}$$

Доказательство. Обозначим через ζ_j случайную величину, равную количеству попыток, которые потребуются, чтобы восполнить дефицит j-й коллекции δ_j. В силу (1.4)

$$\mathrm{E}\zeta_j = n\sum_{k=1}^{n} H_k \mathrm{P}(\delta_j = k).$$

Следовательно,

$$\begin{aligned}\mathrm{E}\xi_m &= \mathrm{E}(\xi + \zeta_2 + \zeta_3 + \ldots + \zeta_m) = \\ &= \mathrm{E}\xi + \sum_{j=2}^{m} \mathrm{E}\zeta_j = \mathrm{E}\xi + n\sum_{j=2}^{m}\sum_{k=1}^{n} H_k \mathrm{P}(\delta_j = k).\end{aligned}$$

Все величины δ_j распределены одинаково. Поэтому, говоря о числовых характеристиках любой из величин δ_j, можно заменить δ_j абстрактной случайной величиной δ, имеющей то же распределение. Тогда

$$\mathrm{E}\xi_m = nH_n + n(m-1)\sum_{k=1}^{n} H_k \mathrm{P}(\delta = k).$$

Рассмотрим отдельно сумму $\sum_k H_k \mathrm{P}(\delta = k)$. Функция $y = -H_x$ выпукла (здесь в качестве H_x можно взять любое подходящее продолжение последовательности H_n, например (3.5) или (3.6)). Поэтому в силу неравенства Йенсена

$$\sum_{k=1}^{n}(-H_k \mathrm{P}(\delta = k)) > -H_{\sum_k k\mathrm{P}(\delta=k)} = -H_{\mathrm{E}\delta} = -H_{H_n}.$$

Значит,

$$\mathrm{E}\xi_m < nH_n + n(m-1)H_{H_n}. \qquad \square$$

Полученная оценка проливает свет на появление повторного логарифма в (2.2), а ход доказательства выявляет общую природу слагаемых nH_n и nH_{H_n}.

Можно получить нижнюю оценку

$$\mathrm{E}\xi_m > nH_n\Big(1+\mathrm{P}\Big(\delta \geqslant \frac{n}{H_n}\Big)\Big) \tag{3.8}$$

(см. задачу 6 в § 5). Однако эта оценка не очень полезна ввиду трудоёмкости вычисления вероятностей в распределении дефицита.

Следующий параграф посвящён распределению величины δ и иному подходу к решению задачи коллекционера, совершенно отличному от изложенных выше.

§ 4. Коллекционирование как случайный процесс

Граф комплектования двух коллекций

Напомним, что через δ_m обозначен дефицит m-й коллекции, т. е. количество экспонатов, недостающих в m-й коллекции в момент завершения $(m-1)$-й коллекции, а ζ_m — число попыток, нужных, чтобы восполнить дефицит δ_m.

Очевидно, $\delta_1 = n$; мы видели, что при $m \geqslant 2$ все величины δ_m распределены одинаково, а, следовательно, все ζ_m при $m \geqslant 2$ также имеют одинаковые распределения. Значит, чтобы решить задачу коллекционера для произвольного числа коллекций, достаточно понять, как она решается в случае двух коллекций: $m = 2$.

Будем искать только $\mathrm{E}\xi_2$, поскольку при $m > 2$, как следует из доказанного выше,

$$\mathrm{E}\xi_m = (m-1)\mathrm{E}\xi_2 - (m-2)\mathrm{E}\xi, \tag{4.1}$$

где, как и прежде, $\xi = \xi_1$ — число попыток, нужных для завершения первой коллекции.

* * *

Обозначим через $s_{i,j}$ состояние, когда первая коллекция содержит уже i экспонатов, а вторая j экспонатов $(0 \leqslant j \leqslant i \leqslant n)$. Эти состояния в совокупности образуют случайный процесс S с начальным состоянием $s_{0,0}$ и конечным состоянием $s_{n,n}$. Переходы между состояниями имеют вероятности, зависящие от i и j. Изобразим весь процесс с помощью ориентированного графа $\mathfrak{S}$, подписав вероятности переходов около соответствующих стрелок (рис. 1).

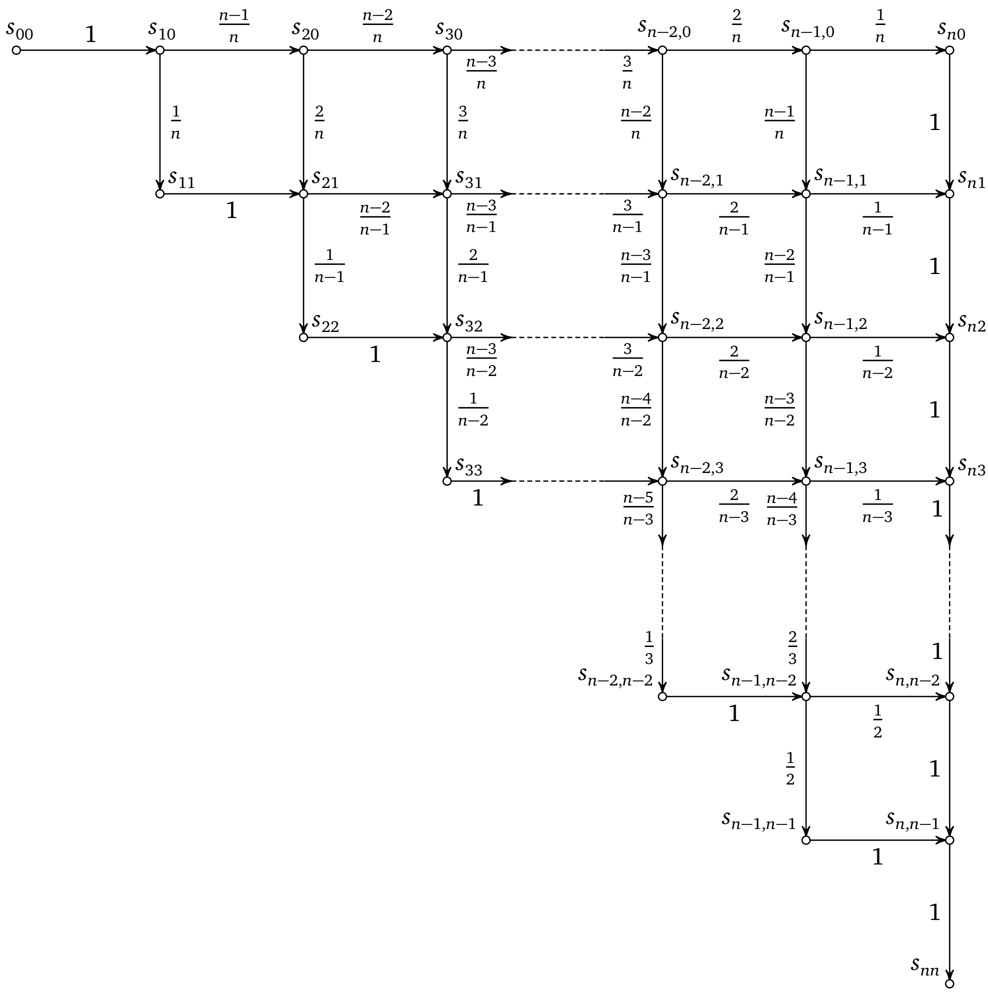

Рис. 1. Граф $\mathfrak{S}$

Каждое состояние может появиться с вероятностью $q_{i,j} = \mathrm{P}(s_{i,j})$. Положим $q_{0,0} = 1$, так как начальное состояние появится наверняка. Для общности положим $q_{0,j} = 0$ для $j > 0$ и $q_{i,-1} = 0$ для $1 \leqslant i \leqslant n$. Тогда вероятности всех прочих состояний при $1 \leqslant j \leqslant i \leqslant n$ вычисляются по формуле

$$q_{i,j} = \frac{n-i+1}{n-j} q_{i-1,j} + \frac{i-j+1}{n-j+1} q_{i,j-1}. \tag{4.2}$$

Каждый способ собрать 2 коллекции в графе $\mathfrak{S}$ изображается одной из C_n цепочек[2)] длины $2n$, ведущих из $s_{0,0}$ в $s_{n,n}$. Каждый переход в цепоч-

[2)] C_n здесь обозначает n-е число Каталана $\dfrac{(2n)!}{n!(n+1)!}$.

ке из состояния $s_{i,j}$ в одно из следующих возможных состояний $s_{i+1,j}$ или $s_{i,j+1}$ осуществляется за некоторое случайное число шагов $\varphi_{i,j}$.

Общее число попыток при комплектовании двух коллекций равно

$$\xi_2 = \sum_{\mathfrak{S}\setminus\{s_{nn}\}} I_{i,j}\varphi_{i,j},$$

где $I_{i,j}$ — индикатор события «процесс S пришёл в состояние $s_{i,j}$». Суммирование происходит по всем состояниям процесса, кроме конечного состояния $s_{n,n}$, из которого нет выхода, поскольку, попав в это состояние, коллекционер находит обе коллекции завершёнными.

Случайная величина $\varphi_{i,j}$ принадлежит геометрическому распределению $G\left(\frac{n-j}{n}\right)$. Действительно, при каждой попытке выход из состояния $s_{i,j}$ происходит тогда, когда появился новый экспонат для первой или второй коллекции, т. е. бегемот одного из $n-j$ видов, которых ещё нет во второй коллекции.

Поэтому

$$\mathrm{E}\xi_2 = \sum_{\mathfrak{S}\setminus\{s_{n,n}\}} \mathrm{E}I_{i,j}\mathrm{E}\varphi_{i,j} = \sum_{\mathfrak{S}\setminus\{s_{n,n}\}} q_{i,j}\cdot\frac{n}{n-j} = \sum_{j=0}^{n-1}\left(\frac{n}{n-j}\sum_{i=j}^{n} q_{i,j}\right). \tag{4.3}$$

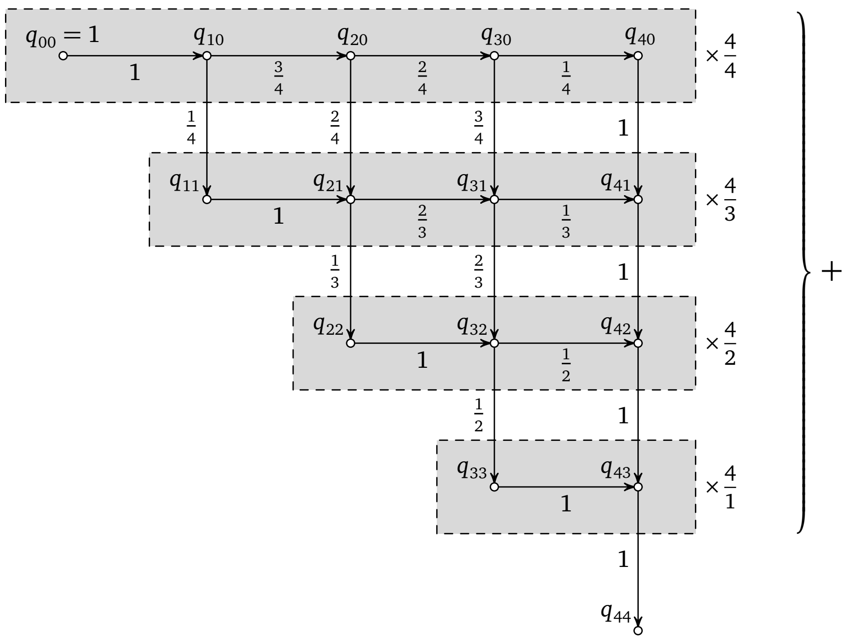

Рис. 2. Суммы вероятностей $q_{i,j}$ (кроме $q_{n,n}$) умножаются на соответствующий множитель $\frac{n}{n}$, $\frac{n}{n-1}$, …, $\frac{n}{1}$. Произведения суммируются

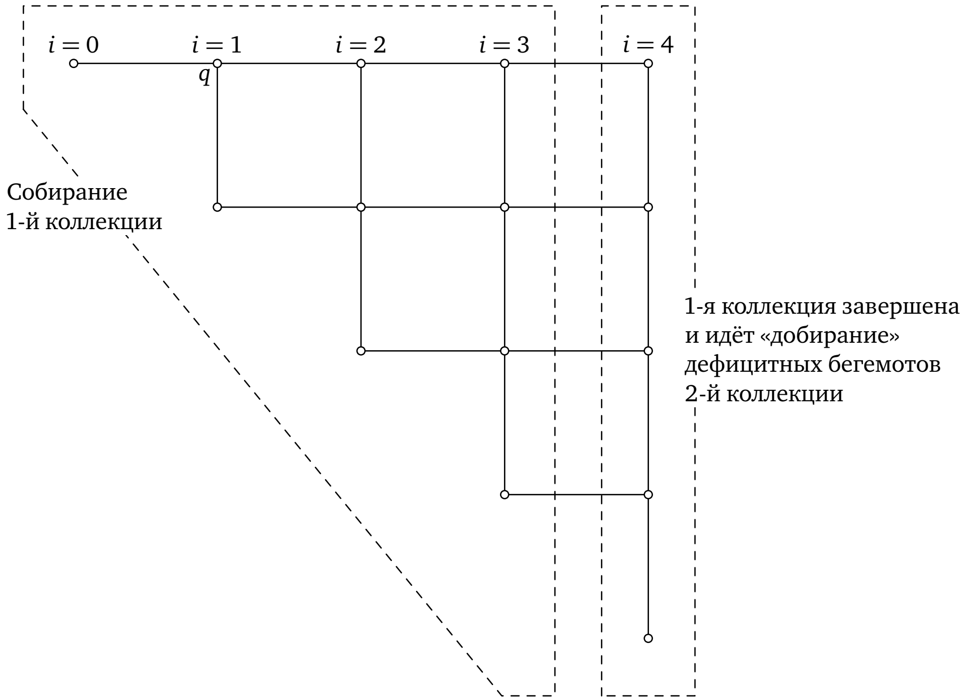

Рис. 3

Полученную сумму можно изобразить наглядно на графе $\mathfrak{S}$ (рис. 2). Для простоты и обозримости возьмём случай $n = 4$. Состояния $s_{i,j}$ будем подписывать их вероятностями $q_{i,j} = \mathrm{P}(s_{i,j})$.

Все вертикальные цепи в графе $\mathfrak{S}$, кроме самой правой, в совокупности содержат состояния $s_{i,j}$, где $i < n$, когда первая коллекция ещё не собрана. Завершение первой коллекции означает переход на последнюю вертикальную цепь, где $i = n$ (рис. 3).

Поэтому в сумме (4.3) слагаемые с $i < n$ в сумме дают $\mathrm{E}\xi$, т. е. nH_n:

$$\mathrm{E}\xi_2 = \sum_{j=0}^{n-1} \frac{n}{n-j} \sum_{i=j}^{n-1} q_{i,j} + \sum_{j=0}^{n-1} \frac{n}{n-j} q_{n,j} = nH_n + n \sum_{j=0}^{n-1} \frac{q_{n,j}}{n-j}. \tag{4.4}$$

Формулы (4.3) или (4.4) совместно с (4.1) дают точное решение задачи коллекционера для двух, а значит, и для произвольного числа коллекций.

Матрица переходов

Общее количество состояний в графе $\mathfrak{S}$ равно C_{n+2}^2. Упорядочим их построчно: s_{00}, s_{10}, …, s_{n0}, s_{11}, s_{21}, …, s_{n1}, …, $s_{n-1,n-1}$, $s_{n,n-1}$, $s_{n,n}$ и построим матрицу переходов S для графа $\mathfrak{S}$. При $n=3$ матрица имеет размер 10×10 и выглядит следующим образом (сверху и справа указаны состояния):

$$S=\begin{array}{cc}
\begin{matrix} s_{00} & s_{10} & s_{20} & s_{30} & s_{11} & s_{21} & s_{31} & s_{22} & s_{32} & s_{33} \end{matrix} & \\
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\hdashline
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\hdashline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\hdashline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} &
\begin{matrix} s_{00} \\ s_{10} \\ s_{20} \\ s_{30} \\ s_{11} \\ s_{21} \\ s_{31} \\ s_{22} \\ s_{32} \\ s_{33} \end{matrix}
\end{array}.$$

Элемент в строке $s_{i,j}$ и в столбце $s_{k,l}$ равен вероятности перехода из $s_{i,j}$ в $s_{k,l}$. Матрица естественным образом разделяется на горизонтальные полосы. В первой полосе $n+1$ строк от $s_{0,0}$ до $s_{n,0}$, во второй n строк от $s_{1,1}$ до $s_{n,1}$ и так далее.

Матрица S не является стохастической в полном смысле слова, поскольку процесс конечен и поэтому последняя строка $s_{n,n}$ нулевая.

Теорема 5. *Элементы первой строки матрицы* $(E-S)^{-1}$ *равны* $q_{i,j}$, *т. е. вероятностям соответствующих состояний процесса* $\mathfrak{S}$.

Доказательство. В первой строке матрицы S^k в столбце, соответствующем состоянию s_{ij}, стоит число, равное вероятности попадания из состояния $s_{0,0}$ в состояние $s_{i,j}$ ровно за k переходов ($k=1,\ldots,2n$). Единичная матрица E обладает тем же свойством: левая верхняя единица может рассматриваться как $q_{0,0}$, т. е. вероятность того, что процесс начинается с состояния $s_{0,0}$. Следовательно, вероятности попадания из $s_{0,0}$ в $s_{i,j}$ за некоторое число переходов образуют матрицу

$$E+S^1+S^2+\ldots+S^{2n}=\sum_{k=10}^{2n}S^k.$$

Матрица $E-S$ верхнетреугольная с единицами на главной диагонали. Поэтому обратная матрица $(E-S)^{-1}$ существует, и, суммируя геометрическую прогрессию, получаем

$$\sum_{k=0}^{2n} S^k = (E-S)^{-1}(E-S^{2n+1}) = (E-S)^{-1}.$$

Последнее равенство верно, поскольку матрица S^{2n+1} нулевая. □

Обозначим через $\boldsymbol{g}$ вектор-столбец

$$\Bigg(\underbrace{\frac{n}{n}, \frac{n}{n}, \ldots, \frac{n}{n}}_{n+1}, \underbrace{\frac{n}{n-1}, \frac{n}{n-1}, \ldots, \frac{n}{n-1}}_{n}, \ldots, \underbrace{\frac{n}{1}, \frac{n}{1}}_{2}, 0\Bigg)^T$$

и заменим в матрице $E-S$ первый столбец вектором $\boldsymbol{g}$. Полученную матрицу назовём Q.

Теорема 6.

$$\mathrm{E}\xi_2 = \det Q. \tag{4.5}$$

Доказательство. Из теоремы 5 и равенства (4.3) следует, что первый элемент вектора

$$\boldsymbol{x} = (E-S)^{-1}\boldsymbol{g} \tag{4.6}$$

равен

$$x_1 = \sum_{j=0}^{n-1}\left(\frac{n}{n-j}\cdot\sum_{i=j}^{n} q_{i,j}\right) = \mathrm{E}\xi_2.$$

Умножая слева обе части (4.6) на $E-S$, получим $(E-S)\boldsymbol{x} = \boldsymbol{g}$. Методом Крамера из этого уравнения находим:

$$x_1 = \frac{\det Q}{\det(E-S)} = \det Q.$$ □

Для произвольного количества коллекций $m \geqslant 2$ из (4.1) теперь получаем:

$$\mathrm{E}\xi_m = (m-1)\det Q - (m-2)nH_n.$$

Подобный подход к решению задачи коллекционера с помощью случайного процесса описан Марко Ферранте и Моникой Сальталамаккья (Marco Ferrante, Monica Saltamacchia) в [3]. Однако помимо существенных переходов между состояниями Ферранте и Сальталамаккья рассматривают и переходы вида $s_{i,j} \to s_{i,j}$, которые в нашем случае изображались бы петлями в графе $\mathfrak{S}$. Матрица получается стохастической, приводит к линейной системе, дающей решение задачи, однако необратимость матрицы, аналогичной $E-S$ (в наших обозначениях), затрудняет получение результата, подобного теореме 6.

* * *

Пример для $n = 3$. Построим матрицу Q:

$$Q = \begin{array}{cc}
\begin{array}{cccccccccc} s_{00} & s_{10} & s_{20} & s_{30} & s_{11} & s_{21} & s_{31} & s_{22} & s_{32} & s_{33} \end{array} & \\
\begin{pmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -\frac{2}{3} & 0 & -\frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & -\frac{1}{3} & 0 & -\frac{2}{3} & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
\frac{3}{2} & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
\frac{3}{2} & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 \\
\frac{3}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
0 & *0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} &
\begin{array}{c} s_{00} \\ s_{10} \\ s_{20} \\ s_{30} \\ s_{11} \\ s_{21} \\ s_{31} \\ s_{22} \\ s_{32} \\ s_{33} \end{array}
\end{array}$$

и найдём её определитель: $\mathrm{E}\xi_2 = \det Q \approx 9{,}639$. Тогда

$$\mathrm{E}\xi_3 = 2\mathrm{E}\xi_2 - 3H_3 \approx 13{,}778, \qquad \mathrm{E}\xi_4 = 3\mathrm{E}\xi_2 - 2\cdot 3H_3 = 17{,}917, \quad \text{и т. д.}$$

В заключение — таблица значений $\mathrm{E}\xi_2$, приближений по формуле (2.2) без учёта констант, верхних оценок[3] (3.7) и нижних (5.1) (усиление оценки (3.8)) для $n = 2, \ldots, 10$.

n	$\mathrm{E}\xi_2$	Оценка (2.2)	Верхняя оценка (3.7)	Нижняя оценка (5.1)
2	5,5	0,6533	5,5	5,5
3	9,6389	3,5780	9,75	9,6385
4	14,1887	6,8517	16,4444	14,0830
5	19,0414	10,4266	21,8889	18,6790
6	24,1339	14,2497	27,6	23,3342
7	29,4247	18,2815	33,5333	28,0141
8	34,8847	22,4923	39,6571	32,6474
9	40,4919	26,8598	45,9477	37,4319
10	46,2296	31,3662	52,3862	42,2063

[3] Для вычисления H_{H_n} использовалась формула (3.6).

§ 5. Сопутствующие задачи

1. Докажите:
а) $\mathrm{P}(\delta = k) = \frac{1}{k} q_{n-1,n-k}$;
б) $\mathrm{P}(\delta = k) = q_{n,n-k} - q_{n,n-k-1}$ (полагая для общности $q_{n,-1} = 0$);
в) $\mathrm{P}(\delta \geqslant k) = q_{n,n-k}$.
2. Докажите:

$$\sum_{j=0}^{i} q_{i,j} \frac{n-i}{n-j} = 1$$

(в частности, при $i = n-1$ получаем сумму вероятностей $\mathrm{P}(\delta = k)$ из задачи 1а).

3. Найдите математическое ожидание «промежуточного дефицита бегемотов» $\delta^{(i)}$, т. е. числа экспонатов, которых недостаёт во второй коллекции в момент, когда первая коллекция пополняется $(i+1)$-м экспонатом $(i = 0, \ldots, n-1)$.

4. Докажите:

$$\sum_{j=0}^{i} q_{i,j} = 1 + H_n - H_{n-i}$$

(в частности, $\sum_{j=0}^{n-1} q_{n-1,j} = H_n = \mathrm{E}\delta$).

5. а) Вероятности дефицита одного и двух бегемотов равны:

$$\mathrm{P}(\delta = 1) = \mathrm{P}(\delta = 2).$$

б) Вероятность того, что дефицитных бегемотов ровно k $(1 < k < n)$, больше вероятности того, что их ровно $k+1$:

$$\mathrm{P}(\delta = 2) > \mathrm{P}(\delta = 3) > \ldots > \mathrm{P}(\delta = n).$$

6. Докажите:

$$\mathrm{E}\xi_m > nH_n\Big(1 + \mathrm{P}\Big(\delta \geqslant \frac{n}{H_n}\Big)\Big).$$

7. Найдите математическое ожидание количества лишних бегемотов каждого вида h_k, образовавшегося в ходе комплектования первой коллекции.

8. Найдите вероятность того, что для завершения первой коллекции нужно купить ровно k бегемотов.

Решения задач

1. а) $q_{n-1,n-k}$ — вероятность того, что в ходе коллекционирования случилось состояние, когда в первой коллекции $n-1$ экспонат, а во второй $n-k$ экспонатов. Тогда $q_{n-1,n-k}\cdot\frac{1}{k}$ — вероятность того, что состоится переход $q_{n-1,n-k}\to q_{n,n-k}$, т. е. первая коллекция будет завершена в момент, когда во второй не хватает ровно k экспонатов:

$$\frac{1}{k}\cdot q_{n-1,n-k}=\mathrm{P}(\delta=k).$$

б) Из (4.2) получаем:

$$q_{n,n-k}=\frac{1}{k}q_{n-1,n-k}+q_{n,n-k-1},$$

откуда

$$q_{n,n-k}-q_{n,n-k-1}=\frac{1}{k}q_{n-1,n-k}=\mathrm{P}(\delta=k).$$

в) Утверждение является очевидным следствием утверждения п. б).

2. $\frac{n-i}{n-j}q_{i,j}$ — вероятность того, что в процессе S состоялся переход

$$q_{i,j}\to q_{i+1,j},$$

т. е. событие $C_{i,j}$ «в момент, когда произошло пополнение первой коллекции экспонатом h_{i+1}, во второй коллекции было ровно j экспонатов». События $C_{i,j}$ несовместны и при $j=0,\ldots,i$ исчерпывают все возможности (образуют полную группу элементарных событий). Поэтому

$$\sum_{j=0}^{i}\frac{n-i}{n-j}q_{ij}=\sum_{j=0}^{i}\mathrm{P}(C_{i,j})=1.$$

3. $\mathrm{E}\delta^{(i)}$ можно найти тем же способом, какой использован при выводе формулы (3.1):

$$\mathrm{E}\delta^{(i)}=(n-i)\Bigl(\frac{1}{n}+\frac{1}{n-1}+\ldots+\frac{1}{n-i+1}+\underbrace{\frac{n}{n-1},\frac{n}{n-1},\ldots,\frac{n}{n-1}}_{n-i\text{ слагаемых}}\Bigr)=$$
$$=(n-i)(H_n-H_{n-i}+1).$$

4. Пусть $i<n$. События $C_{i,j}$ из решения задачи 2 можно понимать иначе: $C_{i,j}=\{$к моменту появления в первой коллекции $(i+1)$-го экспоната во второй коллекции не хватает $n-j$ бегемотов$\}$. Значит, $\mathrm{E}\delta^{(i)}$ (см. задачу 3) равно

$$\sum_{j=0}^{n-1}(n-j)\mathrm{P}(C_{i,j})=\sum_{j=0}^{i}(n-j)\frac{n-i}{n-j}q_{i,j}=(n-i)\sum_{j=0}^{i}q_{i,j},$$

поскольку $\mathrm{P}(C_{i,j}) = 0$ при $j > i$. Утверждение задачи следует из уравнивания полученного выражения с выражением для $\mathrm{E}\delta^{(i)}$, найденным в задаче 3.

Пусть $i = n$. Утверждение теперь следует из того, что $q_{n,n-j} = \mathrm{P}(\delta \geqslant j)$ при $\leqslant j \leqslant n$:

$$\sum_{j=0}^{n} q_{n,j} = \sum_{j=0}^{n} q_{n,n-j} = \sum_{j=0}^{n} \mathrm{P}(\delta \geqslant j) = \sum_{j=0}^{n} (j+1)\mathrm{P}(\delta = j) = \mathrm{E}\delta + 1 = H_n + 1.$$

5. а) Из задачи 1а) получаем:

$$\mathrm{P}(\delta = 1) = q_{n-1,n-1} = \frac{1}{2} q_{n-1,n-2} = \mathrm{P}(\delta = 2).$$

б) Из задачи 1а) и формулы (4.2) при $k \geqslant 2$ находим:

$$\mathrm{P}(\delta = k) = \frac{1}{k} q_{n-1,n-k} = \frac{1}{k}\left(\frac{2}{k} q_{n-2,n-k} + \frac{k}{k+1} q_{n-1,n-k-1}\right) > \\ > \frac{1}{k+1} q_{n-1,n-k-1} = \mathrm{P}(\delta = k+1).$$

6. Воспользуемся равенством (4.4):

$$\mathrm{E}\xi_2 = nH_n + n\sum_{j=0}^{n-1} \frac{1}{n-j} q_{n,j} = nH_n + nH_n \sum_{j=0}^{n-1} \frac{1}{H_n(n-j)} q_{n,j}.$$

Рассмотрим последнюю сумму отдельно. Из задач 1б) и 5б) следует, что кусочно-линейное продолжение $y = q_{n,x}$ последовательности вероятностей $q_{n,j}$ — выпуклая нелинейная возрастающая функция. При этом

$$\sum_{j=0}^{n-1} \frac{1}{H_n(n-j)} = \frac{H_n}{H_n} = 1.$$

В силу неравенства Йенсена

$$\sum_{j=0}^{n-1} \frac{1}{H_n(n-j)} q_{n,j} > q_{n,f(n)},$$

где

$$f(n) = \sum_{j=0}^{n-1} \frac{j}{H_n(n-j)} = \sum_{j=0}^{n-1} \frac{j-n}{H_n(n-j)} + \frac{n}{H_n} \sum_{j=0}^{n} \frac{1}{n-j} = n - \frac{n}{H_n}.$$

Отсюда

$$\mathrm{E}\xi_2 > nH_n + nH_n q_{n,n-\frac{n}{H_n}}. \tag{5.1}$$

Осталось заметить, что $q_{n,n-k} = \mathrm{P}(\delta \geqslant k)$ (см. задачу 1в)). Поэтому

$$q_{n,n-\frac{n}{H_n}} \geqslant \mathrm{P}\Big(\delta \geqslant \frac{n}{H_n}\Big).$$

Отсюда следует утверждение задачи.

7. Начиная с момента, когда в первой коллекции появился экспонат h_k, и до появления экспоната h_{k+1} в среднем появляется (см. с. 217, решение задачи 1)

$$\mathrm{E}\varphi_{k+1} - 1 = \frac{n}{n-k} - 1 = \frac{k}{n-k}$$

лишних экземпляров. Из этих лишних $\frac{k}{n-k}$ бегемотов в среднем $\frac{1}{n-k}$ бегемотов принадлежит виду h_k. Аналогично между экспонатами h_{k+1} и h_{k+2} появляется в среднем

$$\mathrm{E}\varphi_{k+2} - 1 = \frac{k+1}{n-k-1}$$

бегемотов, из них в среднем $\frac{1}{n-k-1}$ вида h_k и так далее. Математическое ожидание числа лишних бегемотов вида h_k равно

$$\frac{1}{n-k} + \frac{1}{n-k-1} + \ldots + 1 = H_{n-k}.$$

8. Обозначим через A_k событие «k бегемотов окажется недостаточно, чтобы собрать полную коллекцию». Пусть событие B_j состоит в том, что «после покупки k-го бегемота бегемот h_j дефицитный». Тогда $A_k = \bigcup_{j=1}^{n} B_j$.

Применяя формулу включения-исключения, получаем:

$$\mathrm{P}(A_k) = n\mathrm{P}(B_1) - C_n^2\mathrm{P}(B_1 \cap B_2) + \\ + C_n^3\mathrm{P}(B_1 \cap B_2 \cap B_3) - \ldots + (-1)^{n-1}\mathrm{P}(B_1 \cap B_2 \cap \ldots \cap B_n).$$

Вероятности событий B_j и их всевозможных пересечений найти несложно:

$$\mathrm{P}(B_1 \cap B_2 \cap \ldots \cap B_l) = \Big(\frac{n-l}{n}\Big)^k,$$

в частности $\mathrm{P}(B_1 \cap B_2 \cap \ldots \cap B_n) = 0$. Включая в сумму для общности последнее нулевое слагаемое, получаем:

$$\mathrm{P}(A_k) = \sum_{j=1}^{n} (-1)^{j+1} C_n^{j+1} \Big(\frac{n-j}{j}\Big)^k.$$

Значит, искомая вероятность равна

$$\mathrm{P}(A_{k-1}) - \mathrm{P}(A_k) = \sum_{j=1}^{n} (-1)^{j+1} C_n^{j+1} \bigg(\Big(\frac{n-j}{n}\Big)^{k-1} - \Big(\frac{n-j}{n}\Big)^k\bigg) = \\ = \sum_{j=1}^{n} (-1)^{j+1} C_n^{j+1} \frac{j(n-j)^{k-1}}{n^k}.$$

Список литературы

[1] *Newman D. J., Shepp L.* The double Dixie Cup problem // Amer. Math. Monthly. 1960. Vol. 67, № 1. P. 58–61.

[2] *Erdős P., Renyi A.* On a classical problem of probability theory // Magyar. Tud. Akad. Mat. Kutatoм Int. Kőzl. 1961. Vol. 6. P. 215–220.

[3] *Ferrante M., Saltalamacchia M.* The Coupon Collector's Problem // Mater. Mat. 2014. Vol. 2. P. 1–35.

[4] *Феллер В.* Введение в теорию вероятностей и её приложения. Т. 1. М.: Мир, 1963. С. 229–230.

Иван Ростиславович Высоцкий, МЦНМО

i_r_vysotsky@hotmail.com