Решения заданий III интернет-олимпиады по теории вероятностей и статистике

Задания-эссе

1. **Закон Ньюкомба-Бенфорда (задача-эссе. 9-11).** В 1881 году Саймон Ньюкомб заметил интересную закономерность, получившую впоследствие название “закон Бенфорда” (в честь переоткрывателя Фрэнка Бенфорда). Закон гласит, что в любом большом массиве однородных числовых данных (площади стран, высоты гор, длины рек, курсы валют, значения физических величин и т.п.*) цифра 1 встречается на первом месте примерно в шесть раз чаще, чем цифра 9. Или, в общем случае, вероятность встретить на первом месте цифру k равна $\lg\frac{k+1}{k}$ **, где $k=1,...,9$.

Обсуждение закона можно найти в статье академика В.И.Арнольда «Антинаучная революция и математика»

(http://www.mccme.ru/edu/index.php?ikey=viarn_vatikan)

или «Статистика первых цифр степеней двойки и передел мира» (В.И.Арнольд. «Жесткие» и «мягкие» математические модели. М., МЦНМО, 2008). Напишите небольшое эссе на тему закона Бенфорда, в котором

а) проверьте, выполняется ли этот закон для населения субъектов Российской Федерации (см. файл data1.xls);

б) опишите любой другой набор однотипных величин, у которых на первом месте может стоять любая цифра, и проверьте согласованность этих данных с законом Бенфорда.

*Важно, чтобы величины в наборе могли начинаться на любую цифру.

** $\lg x$ логарифм по основанию 10 (десятичный).

Вариант эссе а). Построим таблицу по данным из файла data1.xls. В первом столбце укажем цифры от 1 до 9. Во втором – частоту, с которой каждая из цифр стояла на первом месте в наших данных о населении (с точностью до двух цифр после запятой). В последнем столбце поместим теоретические значения вероятностей встретить каждую из цифр на первом месте, вычисленные по закону Бенфорда. Например, для цифры 3 в последнем столбце будет стоять $\lg\left(\frac{3+1}{3}\right)=\lg\frac{4}{3}\approx 0{,}12$.

1	2	3	4	5	6	7	8	9
0,36	0,18	0,08	0,1	0,06	0,05	0,05	0,06	0,07
0,3	0,18	0,12	0,1	0,08	0,07	0,06	0,05	0,05

Построим график, где по оси OX цифры от 1 до 9, а по оси OY процент их появления на первом месте в теории и на практике. Судя по полученному графику, можно сделать вывод, что данные довольно хорошо согласуются с законом Бенфорда. Кроме цифры 3: в реальных данных она встречается намного реже, чем в теории. И цифры 1: она, наоборот, встречается чаще в реальных данных. Однако, возможно, это связано с тем, что мы рассмотрели не очень большой объем данных: всего 84 числа. В целом же закон Бенфорда, проверенный на основе данных о населении субъектов Российской Федерации, вполне правдоподобен.

2. **Прогноз погоды (задача-эссе. 6-11).** Считается, что лучший прогноз погоды на завтра «Завтра погода будет такая же, как сегодня». Метеорологи часто говорят, что точность этого прогноза около 80%, то есть вероятность того, что погода завтра будет похожа на сегодняшнюю равна примерно 0,8. Зайдите на сайт www.rp5.ru или используйте любой другой источник с архивом погоды за несколько месяцев. Выберите свой город и проверьте, насколько правы метеорологи, которые так говорят. Выполняя это задание, составьте таблицы погоды, учитывая температуру воздуха, облачность и направление ветра за месяц.

3. **Мобильные телефоны и женское образование (задача-эссе. 6-11).** В таблице из файла data2.xls во втором столбце указано число грамотных (умеющих читать и писать) женщин $X = \{x_1, \ldots, x_{40}\}$, а в третьем количество мобильных телефонов $Y = \{y_1, \ldots, y_{40}\}$ по странам Европы. Постройте диаграмму рассеивания для пар $\{X, Y\} = \{(x_1, y_1), \ldots, (x_{40}, y_{40})\}$. Пользуясь диаграммой, напишите небольшое эссе, посвященное следующим вопросам.

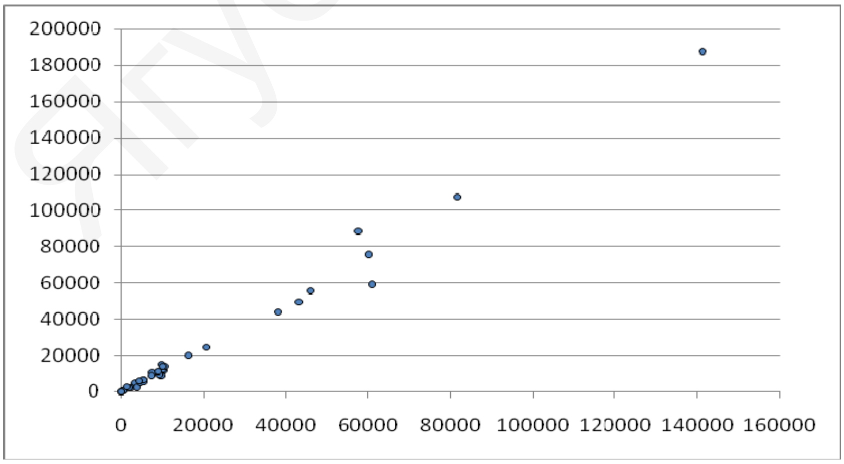
а) Можно ли считать, что чем больше в стране грамотных женщин, тем больше сотовых телефонов или наоборот?

б) Предположим, что, изучив европейский опыт, правительство некоторой развивающейся страны начало интенсивную программу по ликвидации безграмотности среди женщин, и через три года доля грамотных женщин в стране заметно выросла. Следует ли ожидать, что число мобильных телефонов за это время тоже пропорционально изменится?

в) Предположим, что правительство развивающейся страны закупило и раздало населению большое число мобильных телефонов. Следует ли рассчитывать на то, что число грамотных женщин в этой стране быстро увеличится?

г) если на диаграмме прослеживается связь между числом грамотных женщин и числом мобильных телефонов, то можно ли объяснить связь?

Возможная идея эссе. Построим диаграмму рассеивания.



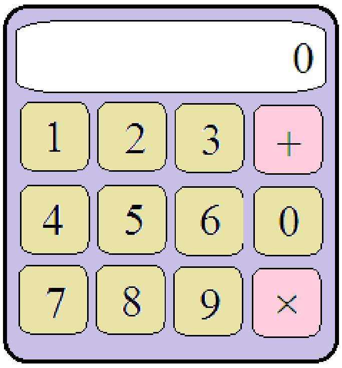
Почти все точки укладываются в одну линию. Такой вид диаграммы рассеивания должен говорить о наличии положительной взаимосвязи между количеством грамотных женщин и количеством мобильных телефонов в стране. Однако это не говорит о том, что чем больше телефонов, тем больше образованных женщин или наоборот. Рассмотрим еще одну величину N – численность населения в стране. При примерно равных условиях, характерных для большинства стран Европы, можно говорить о наличии почти линейной зависимости между X и N, поскольку чем больше население страны, тем больше найдется в этой стране грамотных женщин: $X = a_1N + b_1$. Похожая зависимость имеет место для количества сотовых телефонов: $Y = a_2N + b_2$. Значит, имеется почти линейная связь между X и Y:

$$Y = \frac{a_2}{a_1}X + \left(b_2 - b_1\frac{a_2}{a_1}\right)$$

Получаем, что величины X и Y, скорее всего, связаны через третью величину, но не сами по себе.

Задачи, требующие решения

4. Калькулятор (6-9).



а) (1 б.) На клавиатуре калькулятора есть цифры от 0 до 9 и знаки двух действий (см. рисунок). Вначале на дисплее написано число 0. Можно нажимать любые клавиши. Калькулятор выполняет действия в последовательности нажатий. Если знак действия нажать подряд несколько раз, то калькулятор запомнит только последнее нажатие. Кнопка со знаком умножения <×> сломалась и не работает. Рассеянный Учёный нажал несколько кнопок в случайной последовательности. Какой результат получившейся цепочки действий более вероятен — четное число или нечетное?

б) (2 б.) Решите предыдущую задачу, если кнопку со знаком умножения починили.

Решение. а) Заметим, что хотя бы одно действие сложения выполнено, даже если Учёный набрал только одно число – тем самым он прибавил это число к нулю.

Обозначим A событие «результат нечетный». Четность результата определяется последним слагаемым. Поясним это подробнее.

Пусть с вероятностью p предпоследнее число нечетное. Тогда результат останется нечетным, только если прибавить четное число. Значит, вероятность нечетного результата равна $p \cdot \frac{1}{2}$.

Предпоследнее число было четным с вероятностью $1-p$. Результат будет нечетным, только если последнее слагаемое нечетно. Значит, в этом случае событие A имеет вероятность $(1-p)\cdot\frac{1}{2}$.

Складывая вероятности этих несовместных событий, получим:

$$\mathrm{P}(A)=p\cdot\frac{1}{2}+(1-p)\cdot\frac{1}{2}=\frac{1}{2}.$$

Таким образом, если разрешено только сложение, четный и нечетный результаты имеют равные шансы.

б) Рассмотрим последнее сложение в цепочке действий. После него получился либо четный, либо нечетный результат, причем вероятности этих событий равны, как мы видели при решении задачи а).

Если цепочка оканчивается ровно одним умножением, то оно дает нечетный результат, только если оба последних числа нечетные. Вероятность $\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$.

Если цепочка оканчивается двумя умножениями, то нечетный результат получится, только если все три сомножителя нечетные. Вероятность $\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$ и так далее. Если цепочка оканчивается k действиями умножения, то при этом условии вероятность нечетного результата равна $\frac{1}{2^{k+1}}$.

Если обозначить p_k вероятность того, что цепочка действий оканчивается ровно k действиями умножения, то

$$\mathrm{P}(A)=p_0\cdot\frac{1}{2}+p_1\cdot\frac{1}{4}+p_2\cdot\frac{1}{8}+\ldots+p_k\cdot\frac{1}{2^{k+1}}+\ldots+p_n\cdot\frac{1}{2^{n+1}},$$

где n — общее число действий в цепочке.

Вынесем $\frac{1}{2}$ за скобки:

$$\mathrm{P}(A)=\frac{1}{2}\left(p_0+p_1\cdot\frac{1}{2}+p_2\cdot\frac{1}{4}+\ldots+p_n\cdot\frac{1}{2^n}\right)<\frac{1}{2}\left(p_0+p_1+\ldots+p_n\right)=\frac{1}{2}\cdot 1=\frac{1}{2}$$

,

поскольку сумма всех p_k равна единице — ведь каким-то образом цепочка из n действий все же оканчивается.

Значит, если разрешено умножение, четный результат имеет больше шансов. Очевидно, что вероятность четного результата намного выше, чем нечетного, если умножений в конце много — оценка сверху, которую мы сделали при переходе к неравенству, довольно грубая.

5. **Гирлянда (1 б. 6-11)**. На новогоднюю елку повесили 100 лампочек в ряд. Затем лампочки стали переключаться по следующему алгоритму: зажглись все, через секунду погасла каждая вторая лампочка, еще через секунду каждая третья лампочка переключилась: если горела, то погасла и наоборот. Через секунду каждая четвертая лампочка переключилась, еще через секунду - каждая пятая и так далее. Через 100 секунд все закончилось. Найдите вероятность того, что случайно выбранная после этого лампочка горит (лампочки не перегорают и не бьются).

Решение. Очевидно, лампочка с номером n останется гореть, только если ее переключили нечетное число раз, то есть если число n имеет нечетное количество натуральных делителей. Ясно, что этому условию удовлетворяют только квадраты: $n = 1, 4, 9, ..., 100$. Таким образом, останется гореть 10 лампочек из 100. Поэтому вероятность случайно выбрать горящую лампочку равна $0{,}1$.

6. **Вероятностное голосование (6-9).** В финал конкурса спектаклей к 8 марта вышли два спектакля. В первом играли n учеников 5 класса А, а во втором - n учеников 5 класса Б. На спектакле присутствовали $2n$ мам всех $2n$ учеников. Лучший спектакль выбирается голосованием мам. Известно, что каждая мама с вероятностью $\frac{1}{2}$ голосует за лучший спектакль и с вероятностью $\frac{1}{2}$ – за спектакль, в котором участвует её ребенок.

а) (1 б.) Найдите вероятность того, что лучший спектакль победит с перевесом голосов.

б) (1 б.) Тот же вопрос, если в финал вышло больше двух классов.

Решение.

а) Назовем маму уверенной, если ее ребенок играет в лучшем спектакле. Уверенная мама с вероятностью 1 проголосует за лучший спектакль. Уверенных мам ровно n, поэтому лучший спектакль наберет не меньше половины голосов – он получит, по крайней мере, голоса всех уверенных мам. Единственный случай, когда лучший спектакль не победит с перевесом голосов – ничья, когда худший спектакль тоже получит голоса всех неуверенных мам. Это возможно только если все неуверенные мамы проголосуют за своих детей. Вероятность этого события равна

$p=\left(\frac{1}{2}\right)^n$. Значит, вероятность того, что за лучший спектакль будет подано больше голосов, равна $1-\left(\frac{1}{2}\right)^n$.

б) Пусть в финал вышло $s>2$ классов с одинаковым числом n участников спектакля в каждом из них. Лучший спектакль получит, по крайней мере, n голосов – по одному голосу от каждой уверенной мамы. Каждый из остальных спектаклей получит не более чем n голосов всех мам. Если хотя бы одна неуверенная мама проголосовала за лучший спектакль, он побеждает. Значит, лучший спектакль не победит с перевесом голосов, только если все неуверенные мамы (их ровно $(s-1)n$) проголосовали за своих детей. Вероятность этого события равна $\frac{1}{2^{sn-n}}$. Значит, вероятность того, что лучший спектакль будет признан лучшим, равна $1-\frac{1}{2^{ns-n}}$.

7. **Голосование мам (9-11).** В финал конкурса спектаклей к 8 марта вышли два спектакля. В первом играли n учеников 5 класса А, а во втором - n учеников 5 класса Б. На спектакле присутствовали $2n$ мам всех $2n$ учеников. Лучший спектакль выбирается голосованием мам. Известно, что ровно половина мам честно голосует за лучший спектакль, а другая половина в любом случае голосует за спектакль, в котором участвует её ребенок.

а) (2 б.) Найдите вероятность того, что лучший спектакль победит с перевесом голосов.

б) (1 б.) Тот же вопрос, если в финал вышло больше двух спектаклей.

Решение

а) Назовем маму честной, если она в любом случае голосует за лучший спектакль. Если мама голосует за своего ребенка, будем считать ее нечестной. Известно, что n честных мам будет голосовать за лучший спектакль, поэтому он заведомо не проиграет (половина голосов точно за него). Единственный случай, когда лучший спектакль не победит – ничья, когда худший спектакль также наберет половину голосов. Это возможно только если все дети нечестных мам играют в худшем спектакле. Найдем вероятность этого события.

Всего существует C_{2n}^{n} случаев распределения n нечестных мам среди общего числа $2n$ мам. И только в одном из этих случаев все нечестные мамы голосуют за худший спектакль. Таким образом, вероятность ничьей равна $p = \frac{1}{C_{2n}^{n}} = \frac{(n!)^2}{(2n)!}$. Следовательно, вероятность события "лучший спектакль наберет больше голосов" равна $1 - p = \frac{(2n)! - (n!)^2}{(2n)!}$.

Замечание. Если считать, что n велико, то факториал $n!$ можно вычислить по приближенной формуле Стирлинга: $n! \approx \sqrt{2\pi n} \cdot n^n e^{-n}$. Посмотрим, что получится, если приблизительно вычислить вероятность ничьей с помощью этой формулы:

$$p = \frac{(n!)^2}{(2n)!} \approx \frac{2\pi n \cdot n^{2n} e^{-2n}}{\sqrt{2\pi \cdot 2n}\,(2n)^{2n}\, e^{-2n}} = \frac{\sqrt{\pi n}}{4^n}.$$

Например, если в каждом спектакле участвовало по 20 актеров, то подсчет по приближенной формуле будет уже оправдан: $p \approx \frac{\sqrt{20\pi}}{4^{20}} \approx 7{,}2 \cdot 10^{-12}$. Как видим, в этом случае рассчитывать на ничью не приходится.

б) Пусть спектаклей $s > 2$. В этом случае лучший спектакль побеждает наверняка – он получает, по крайней мере, половину голосов по одному от каждой честной мамы. Голоса остальных мам делятся между прочими спектаклями, каждый из которых получает не более, чем $\frac{1}{s} < \frac{1}{2}$ всех голосов.

8. **Отбор матросов на подлодку**. **(1 б. 6-9)** Служить на подводной лодке может матрос, рост которого не превышает 168 см. Есть четыре команды А, Б, В и Г. Все матросы в этих командах хотят служить на подводной лодке и прошли строгий отбор. Остался последний отбор – по росту.

В команде А средний рост матросов равен 166 см

В команде Б медиана роста матросов равна 167 см

В команде В самый высокий матрос имеет рост 169 см

В команде Г мода роста матросов равна 167 см.

В какой команде по крайней мере половина матросов точно может служить на подводной лодке?

Решение.

Пример команды А: три матроса с ростом 160, 169 и 169 см. Средний рост 166 см, но двое из трех не годны к службе на подводной лодке. Значит, в команде А не обязательно половина матросов пригодна.

Пример команды В: два матроса с ростом 169 см. Служить на лодке не может ни один. Команда В тоже не дает ответа на вопрос.

Пример команды Г: пять матросов с ростом 167, 167, 169, 170 и 171 см. Мода равна 167, но трое из пяти не могут служить на лодке.

Рассмотрим команду Б. По определению медианы, не меньше половины матросов имеют рост не больше, чем 167 см. Эти матросы могут служить на подводной лодке. А, возможно, есть еще такие, чей рост ровно 168. Таким образом, заведомо не менее половины матросов команды Б может служить на подлодке.

Ответ: Б.

9. **Прямоугольный треугольник (7-11)**. Длина гипотенузы прямоугольного треугольника равна 3.

а) (1б.) Рассеянный Учёный вычислил дисперсию длин сторон этого треугольника и нашел, что она равняется 2. Не ошибся ли он в расчетах?

б) (2 б.) Какое наименьшее стандартное отклонение сторон может иметь такой прямоугольный треугольник? Какие у него при этом катеты?

Решение. а) Пусть a и b – катеты треугольника. Тогда $a^2+b^2=9$. Найдем дисперсию набора длин сторон. Для этого воспользуемся формулой $S^2=\overline{x^2}-\overline{x}^2$ – **средний квадрат без квадрата среднего**:

$$S^2=\frac{a^2+b^2+9}{3}-\left(\frac{a+b+3}{2}\right)^2=\frac{18}{3}-\left(\frac{a+b+3}{2}\right)^2=6-\left(\frac{a+b+3}{3}\right)^2. \quad (1)$$

Из неравенства треугольника следует, что $a+b>3$. Значит.

$$S^2<6-\left(\frac{6}{3}\right)^2=6-4=2.$$

Ответ: **ошибся**.

б) Дисперсия любого набора неотрицательна. Вычислим отдельно дисперсию двух катетов: $S^2_{катетов}=\frac{a^2+b^2}{2}-\left(\frac{a+b}{2}\right)^2$. Но ведь дисперсия неотрицательна. Следовательно,

$$\frac{a^2+b^2}{2}-\left(\frac{a+b}{2}\right)^2\geq 0\text{, откуда }\frac{a+b}{2}\leq\sqrt{\frac{a^2+b^2}{2}}\qquad(2).$$

Полученное неравенство (частный случай неравенства Коши-Буняковского) является одним из неравенств о средних: «**Среднее арифметическое любых чисел не меньше их среднего квадратичного, причем равенство достигается, только если все числа равны между собой**». Это неравенство можно было доказать самыми разными способами, мы выбрали необычный способ – через дисперсию.

Из (2) следует: $a+b\leq 2\sqrt{\frac{a^2+b^2}{2}}=2\frac{3}{\sqrt{2}}=3\sqrt{2}$. Подставляя это в (1), получим оценку S^2 снизу:

$$S^2\geq 6-\left(\frac{3\sqrt{2}+3}{3}\right)^2=6-\left(\sqrt{2}+1\right)^2=6-\left(2+2\sqrt{2}+1\right)=2-2\sqrt{2}+1=\left(\sqrt{2}-1\right)^2.$$

Следовательно, стандартное отклонение S длин трех сторон не превосходит

$$\sqrt{2}-1=0{,}41421356...$$

Это наименьшее значение достигается, только если дисперсия катетов равна нулю, то есть если $a=b$. Тогда $a=b=\frac{3\sqrt{2}}{2}$.

Ответ: наименьшее стандартное отклонение $\sqrt{2}-1$ у равнобедренного прямоугольного треугольника с катетами $\frac{3\sqrt{2}}{2}$.

10. **Жюри конкурса спектаклей (3 б. 10-11).** Учительница математики предложила изменить схему голосования на конкурсе спектаклей (см. задачу 7 а). По её мнению, нужно из всех $2n$ мам выбрать случайным образом жюри из $2m$ человек ($2m\leq n$). Найдите вероятность того, что лучший спектакль победит при таких условиях голосования.

Решение. Как и прежде, будем считать маму честной, если она голосует за лучший спектакль и нечестной, если она в любом случае голосует за спектакль, где играет ее ребенок.

Поскольку всего нечестных мам n и столько же честных, вероятность того, что в жюри попало q нечестных мам и $2m-q$ честных, равна $\frac{C_n^q C_n^{2m-q}}{C_{2n}^{2m}}$.

Если это событие осуществилось, то судьба конкурса зависит от распределения голосов нечестных мам. Если t детей нечестных мам играет в худшем спектакле, то этот худший спектакль получит ровно t голосов, тогда как все прочие голоса числом $2m-t$ будут отданы лучшему спектаклю. Таким образом, лучший спектакль побеждает с перевесом голосов, если $2m-t>t$, то есть если $t<m$.

Вероятность того, что худший спектакль получит ровно t голосов, равна вероятности того, что ровно t детей нечестных мам из жюри играет в этом спектакле: $\frac{C_q^t C_{2n-q}^{n-t}}{C_{2n}^n}$ – именно такова вероятность того, что в худший спектакль попадут t детей нечестных мам из жюри и еще $n-t$ детей честных мам или детей нечестных мам, не вошедших в жюри.

Следовательно, вероятность комбинации, при которой в жюри собирается q нечестных мам, и худший спектакль получает меньше голосов, чем лучший, равна $\frac{C_n^q C_n^{2m-q}}{C_{2n}^{2m}} \cdot \sum\limits_{\substack{t=0,\dots,q \\ t<m}} \frac{C_q^t C_{2n-q}^{n-t}}{C_{2n}^n}$.

Суммируя эти вероятности по всем возможным распределениям нечестных мам, получаем, что искомая вероятность равна

$$\sum_{q=0}^{2m}\left(\frac{C_n^q C_n^{2m-q}}{C_{2n}^{2m}} \cdot \sum_{t=0,\dots,q,\, t<m} \frac{C_q^t C_{2n-q}^{n-t}}{C_{2n}^n}\right) = \frac{1}{C_{2n}^n C_{2n}^{2m}} \sum_{q=0}^{2m}\left(C_n^q C_n^{2m-q} \cdot \sum_{t=0,\dots,q,\, t<m} C_q^t C_{2n-q}^{n-t}\right).$$

Замечание. Расчеты по этой формуле можно немного «укоротить», если заметить, что при $q<m$ у худшего спектакля нет шансов и поэтому (а также из чисто комбинаторных соображений) $\sum\limits_{t=0,\dots,q<m} \frac{C_q^t C_{2n-q}^{n-t}}{C_{2n}^n} = 1$.

11. **Функция рассеивания (2 б. 7-11).** Дан числовой набор $x_1, \dots, x_n$.

Рассмотрим функцию $d(t) = \dfrac{\min\limits_{i=1,\dots,n} |x_i - t| + \max\limits_{i=1,\dots,n} |x_i - t|}{2}$.

а) Верно ли, что функция $d(t)$ принимает наименьшее значение в единственной точке, каков бы ни был набор чисел $x_1, \dots, x_n$?

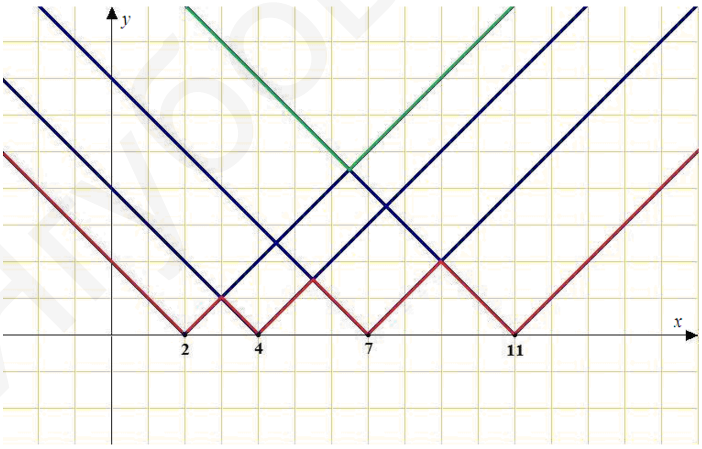
б) Сравните значения $d(c)$ и $d(m)$, где $c = \dfrac{\min\limits_{i=1,\dots,n} x_i + \max\limits_{i=1,\dots,n} x_i}{2}$, а m – его медиана.

Решение

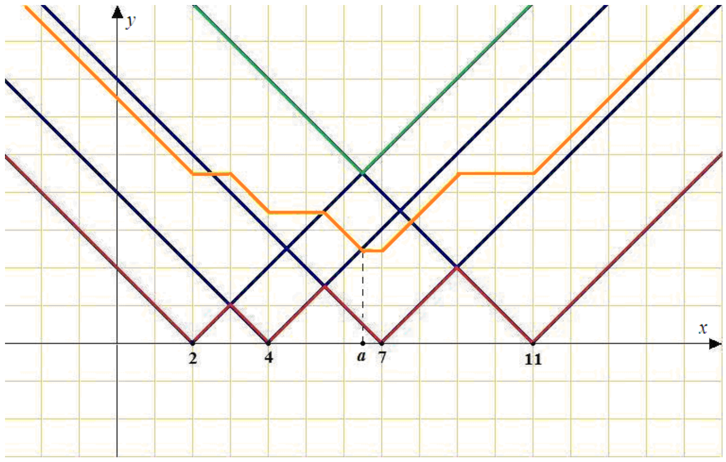
а) Рассмотрим набор $\{2, 4, 7, 11\}$ Построим функцию $d(t)$ для этого набора. Сначала построим графики функций

$$y = |x-2|,\ y = |x-4|,\ y = |x-7|,\ y = |x-11|.$$

Теперь построим графики функции $y = \min\limits_{i=1,\dots,n} |x_i - t|$ (красный) и $y = \max\limits_{i=1,\dots,n} |x_i - t|$ (зеленый).



Изобразим график функции $y = d(t)$ оранжевым цветом. Это легко сделать, если в каждой точке взять середину вертикального отрезка, соединяющего зеленый и красный графики. Этот график имеет изломы во всех точках, где имеет изломы график функции $y = \min\limits_{i=1,\dots,n} |x_i - t|$. При этом функция $y = d(t)$, очевидно, не может возрастать там, где убывает функция $y = \max\limits_{i=1,\dots,n} |x_i - t|$ и не может убывать там, где эта функция возрастает.



Видно, что функция $y = d(t)$ принимает наименьшее значение не в одной точке, а в любой точке из отрезка $[a, 7]$

б) Докажем, что $d(c) \le d(m)$. Из проведенного рассуждения следует, что функция $y = d(t)$ не может возрастать левее точки c и не может убывать правее этой точки. Значит, $d(c) \le d(t)$ для любого t, в частности для $t = m$.

12. **Задача по мотивам задачи А.Д.Сахарова (7-11)**. Вася в ярости режет прямоугольный лист бумаги ножницами. Каждую секунду он разрезает первый попавшийся кусок случайным прямолинейным разрезом на две части.

а) (1 б.) Найдите математическое ожидание числа сторон многоугольника, который случайно попадется Васе через час такой работы.

б) (2 б.) Решите эту же задачу, если вначале лист бумаги имел форму произвольного многоугольника.

Примечание. Эту задачу в другой формулировке и в более общем виде придумал и решил академик Андрей Дмитриевич Сахаров. Сахаров писал, что задача возникла, когда он рубил капусту.

Решение. Решим задачу б). Пусть в результате Васиной деятельности на столе оказалось p_3 треугольников, p_4 четырехугольников и так далее – число m-угольников равно p_m. Здесь m - наибольшее возможное число сторон у получившегося многоугольника, причем нам даже не важно, чему m равно. Тогда математическое ожидание числа сторон выбранного многоугольника равно

$$\mu = 3 \cdot \frac{p_3}{k+1} + 4\frac{p_4}{k+1} + \ldots + m\frac{p_m}{k+1} = \frac{3p_3 + 4p_4 + \ldots + mp_m}{k+1}.$$

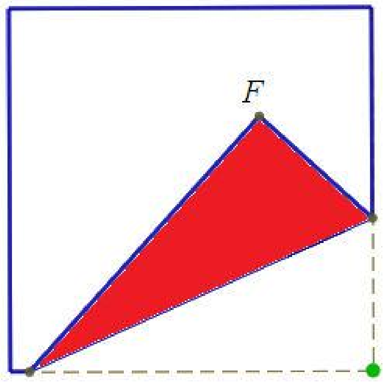
Заметим, что в числителе находится общее число сторон всех полученных многоугольников. Нам нужно найти именно эту величину.

Предположим, что у Васи был n-угольник. Вероятность сделать случайный разрез точно через какую-нибудь вершину равна нулю. Значит, каждый разрез делит какие-то две стороны разрезаемого многоугольника на две части каждую, и еще две новые стороны получаются вдоль разреза. Таким образом, каждый разрез добавляет 4 новые стороны. После первого разреза общее число сторон будет $n+4$, после второго – $n+8$ и так далее. После k разрезов общее число сторон будет равно $n+4k$.

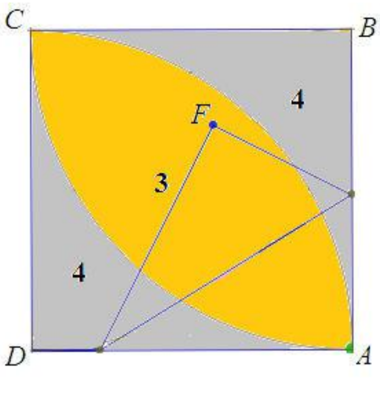
Таким образом, $\mu = \dfrac{n+4k}{k+1}$.

Заметим, что если вначале у Васи был четырехугольник, то $\mu = \dfrac{4+4k}{k+1} = 4$, что дает решение задачи а).

Если же исходный многоугольник имел другое число сторон n, то $\lim\limits_{k\to+\infty} \mu = 4$ и поэтому для достаточно больших k можно считать, что среднее число сторон практически равно 4.



13. **Оригами**. **Красный многоугольник. (2 б. 7-11).** Верхняя сторона бумажного квадрата белая, а нижняя – красная. В квадрате случайным образом выбирается точка F. Затем квадрат сгибают так, чтобы одна случайно выбранная вершина наложилась на точку F. Найдите математическое ожидание числа сторон появившегося красного многоугольника.



Решение. Безразлично, какую именно вершину мы будем накладывать на выбранную точку F. Пусть это будет вершина A. Нужно изучить геометрический вопрос: где должна находиться точка F, чтобы получился треугольник, четырехугольник, пятиугольник и т.д.

Решение этой геометрической задачи большого труда не представляет. Если точка F принадлежит оранжевому двуугольнику, то в результате сложения получится треугольник, а если точка F вне двуугольника, в серой области, то будет четырехугольник. Докажите это самостоятельно – нужно рассмотреть граничный случай, когда линия сгиба проходит через вершину B или D.

Обозначим вероятности попадания точки F в серую и оранжевую области P_g и P_o соответственно. Если считать, что площадь квадрата равна 1, то вероятности попадания точки F в каждую область равна площади этой области.

Вероятность попадания в серую область:

$$P_g = 2\cdot\left(1-\frac{\pi}{4}\right) = 2-\frac{\pi}{2}.$$

Тогда вероятность попадания в оранжевую область равна

$$P_o = 1-2+\frac{\pi}{2} = \frac{\pi}{2}-1.$$

Осталось найти математическое ожидание:

$$3\cdot P_o + 4\cdot P_g = 3\left(\frac{\pi}{2}-1\right)+4\cdot\left(2-\frac{\pi}{2}\right) = 5-\frac{\pi}{2} \approx 3,429.$$

14. **Почтальон (2 б. 7-11)**. На улице n домов. Каждый день почтальон идёт на почту, берет там письма для жителей одного дома и разносит их. Затем он возвращается на почту, берет письма для жителей другого дома и снова их разносит. И так далее он обходит все дома. В каком месте нужно построить почту, чтобы почтальону пришлось проходить наименьшее расстояние? Улицу можно считать отрезком прямой.

а) Решите задачу для $n=5$.

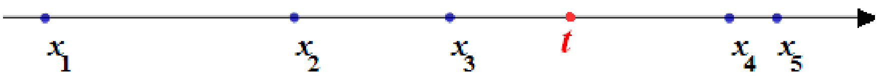
б) Решите задачу для $n=6$.

в) Решите задачу для произвольного n.

Решение.

а) Введем координатную прямую, на которой расположим дома по улице Пушкина с координатами x_1, x_2, x_3, x_4, x_5 упорядоченными по возрастанию.

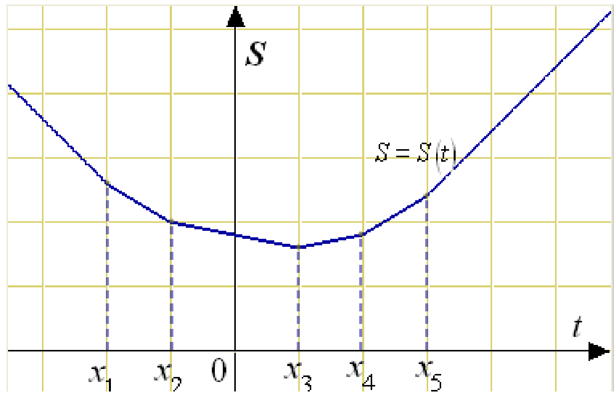
Пусть t - координата почтового отделения на числовой прямой.



Почтальон проходит от почты до каждого дома дважды: туда и обратно, то есть общее расстояние, пройденное почтальоном равно

$$2S = 2|x_1-t|+2|x_2-t|+2|x_3-t|+2|x_4-t|+2|x_5-t|.$$

Нужно сделать $2S$, а значит и S, как можно меньше. Найдем S при разных t. Рассмотрим функцию $S = S(t)$. Графиком является ломаная, точки излома в каждой из точек $x_1, x_2, \ldots, x_5$ (см. рисунок).



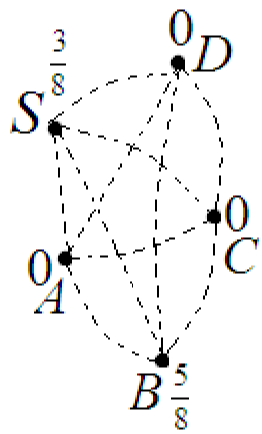
У самого левого звена угловой коэффициент -5, а если двигать t вправо, то при переходе через каждую точку x_i угловой коэффициент увеличивается на 2. Таким образом, до точки x_3 функция $S = S(t)$ убывает, а после этой точки возрастает (на рисунке масштаб по оси ординат уменьшен для удобства).

Значит, наименьшее значение достигается в точке $t = x_3$. Здание почты нужно строить в этой точке.

б) Рассуждая так же, получим, что здание почты нужно строить в x_3 или x_4 или в любом месте между этими точками. Схему графика постройте самостоятельно.

в) Для произвольного n тем же рассуждением найдем, что t должно быть медианой набора чисел $x_1, x_2, \ldots, x_n$. То есть при нечетных n положим $t = x_{\frac{n+1}{2}}$, а при четных n нужно взять $x_{\frac{n}{2}} \le t \le x_{\frac{n+2}{2}}$.

15. **Пять озер (3 б. 6-11)**. В Долине Пяти Озер есть пять одинаковых озер, некоторые из которых соединены ручьями. Маленькие караси появляются на свет только в озере S. Пока карась взрослеет, он ровно четыре раз переходит из одного озера в другое по какому-нибудь ручью (карась выбирает ручей наудачу), а затем остается жить в том озере, в котором оказался. Из каждой тысячи карасей в среднем 375 остается жить в озере S, а остальные остаются жить в озере B, в других озерах не остается жить никто. Определите, сколько ручьев в Долине Пяти Озер.



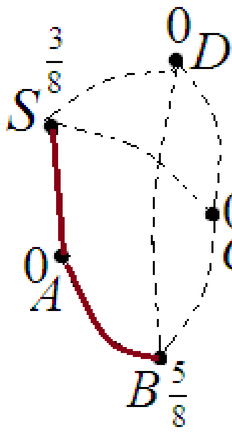
Решение. Переход из озера в озеро будем называть маршрутом длины n, если он проходит через n ручьев. Сформулируем решение в виде последовательности утверждений.

1) Из S нет маршрута длины 2 ни в одно из озер A, C и D.

Доказательство. Предположим, что из озера S можно проплыть в озеро A через одно промежуточное озеро (скажем C). Тогда был бы возможен маршрут $S-C-A-C-A$. В этом случае, была бы ненулевая вероятность попасть в озеро A за четыре перехода. Значит из S в A маршрута длины 2 нет. Точно так же, нет маршрутов длины 2 из S в C и в D.

2) Озеро S связано с озером B маршрутом длины 2.

Доказательство. Очевидно, S и B связаны. Предположим, они связаны общим ручьем. Тогда из B нет других выходов, иначе не выполнено условие (1). Значит, в B можно прийти только из S на первом или третьем ходу. Но тогда получается, что после четвертого перехода попасть в B карась не сможет. Противоречие.

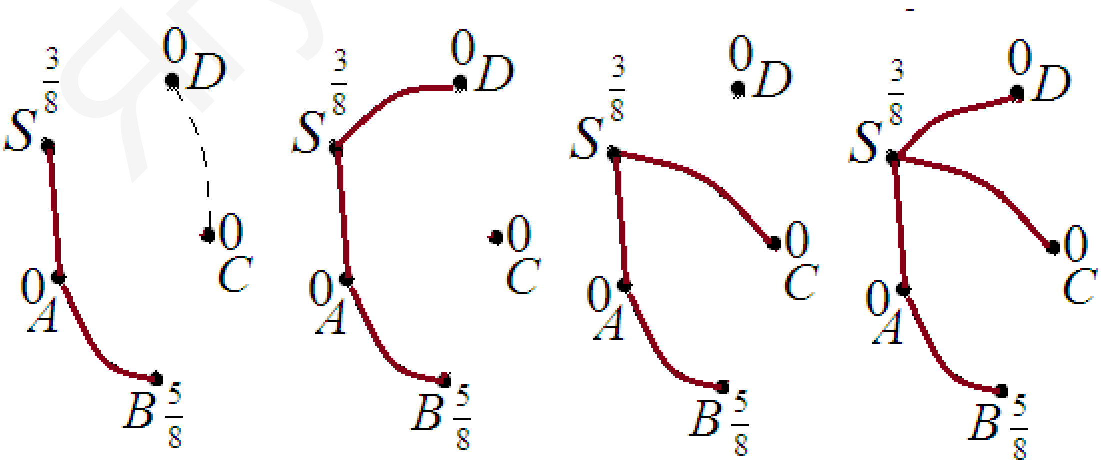


Предположим, что S и B связаны маршрутом длины 3 или 4. В этом случае нарушено условие (1) – есть маршрут длины 2 из S в одну из комнат A, C или D, а его быть не может.

Значит, из S в B есть маршрут, проходящий через одно промежуточное озеро A, C или D, причем одно ничем не хуже другого. Можно для определенности считать, что есть маршрут $S-A-B$. Из этого и утверждения (1) следует, что нет маршрутов $A-C$ и $A-D$.

3) Есть хотя бы один из маршрутов $B-C$ или $B-D$.

Доказательство. Предположим, что нет ни маршрута $B-C$, ни $B-D$ Получаем один из следующих рисунков.

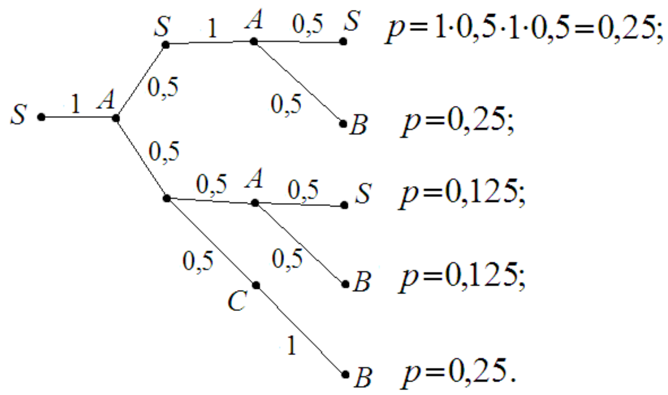


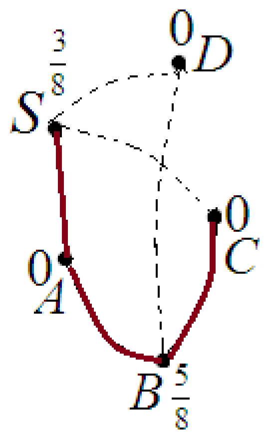
В первом случае карась первый раз переходит в озеро A, и у него остается три перехода. Озера S и B одинаково расположены по отношению к A, поэтому их шансы должны быть равны. Легко видеть, что в трех других случаях вероятность остаться в озере S больше, чем в B.

Следовательно, один из маршрутов $B-C$ или $B-D$ присутствует. Отсюда, кстати, следует, что маршрута $C-D$ нет.

4) Построенная конфигурация из трех ручьев удовлетворяет условию.

Доказательство. Для построенной конфигурации возможны следующие маршруты длины 4:





Над каждым переходом указана вероятность этого перехода, а справа – вероятность всего маршрута, полученная произведением вероятностей переходов между озерами.

Вероятность того, что карась, сделав четыре перехода, окажется в озере S, равна $\mathrm{P}(S) = 0,25 + 0,125 = \frac{3}{8}$ – как раз то, что надо.

Тогда вероятность того, что B – конечная точка маршрута, равна $\mathrm{P}(B) = 1 - \frac{3}{8} = \frac{5}{8}$.

5) Больше чем уже есть ручьев быть не может.

Доказательство. Добавляя любые ручьи $S-C$, $S-D$, $B-D$ порознь или вместе, мы изменим вероятности. В этом можно убедиться непосредственно, сделав еще семь рисунков, соответствующих возможным конфигурациям и рассчитав вероятности для каждой из полученных схем.

Таким образом, с точностью до обозначения вершин, схема, показанная на последнем рисунке – единственно возможная, а общее число ручьев равно трем.

16. **Нестрогие нарды (3 б. 9-11)**. Правильная игральная кость бросается много раз. Найдите математическое ожидание числа бросков, сделанных до того момента, когда сумма всех выпавших очков достигнет 2010.

Решение. Пусть X_n – число бросков, которые пришлось сделать, чтобы сумма очков достигла n. Введем случайную величину

$$I_k = \begin{cases} 1, \textit{если первый бросок дал } k \textit{ очков}, \\ 0 \textit{ в противном случае.} \end{cases}$$

Очевидно, $\mathrm{E} I_k = 0 \cdot \frac{5}{5} + 1 \cdot \frac{1}{6} = \frac{1}{6}$.

Если обозначить X_{n-k} число бросков, начиная со второго, которые сделаны, чтобы сумма достигла $n-k$, то

$$X_n = X_{n-1}I_1 + X_{n-2}I_2 + \ldots + X_{n-6}I_6 + 1,$$

Величина X_{n-k} не зависит от результата первого броска. Значит, величины X_{n-k} и I_k независимы. Переходя к математическим ожиданиям, получим:

$$\mathrm{E}\,X_n = \sum_{k=1}^{6} \mathrm{E}\,X_{n-k}\,\mathrm{E}\,I_k = \frac{1}{6}\sum_{k=1}^{6} \mathrm{E}\,X_{n-k} + 1.$$

Обозначая для краткости $\mathrm{E}\,X_k = e_k$, запишем рекуррентную формулу:

$$e_n = \frac{1}{6}\sum\nolimits_{k=1}^{6} e_{n-k} + 1. \quad (1)$$

В рассуждениях нигде не была использована положительность чисел $n-2$, $n-3$ и т.д. Значит, полученная формула справедлива для всех $n \ge 1$, если положить $\mathrm{E}\,X_k = 0$ при всех $k \le 0$.

Из (1) можно получить другую формулу. Запишем равенство (1) для e_{n-1}:

$$e_{n-1} = \frac{1}{6}\sum\nolimits_{k=1}^{6} e_{n-k-1} + 1.$$

Теперь вычитая второе равенство из первого почленно, получим:

$$e_n - e_{n-1} = \frac{e_{n-1} - e_{n-7}}{6}, \text{ откуда } e_n = \frac{7}{6}e_{n-1} - \frac{1}{6}e_{n-7} \qquad (2).$$

При выводе этой формулы уже предполагалось, что $e_{n-1} > 0$, поэтому по ней можно считать e_n, только начиная с $n = 2$.

По формулам (1) и (2) последовательно вычисляем:

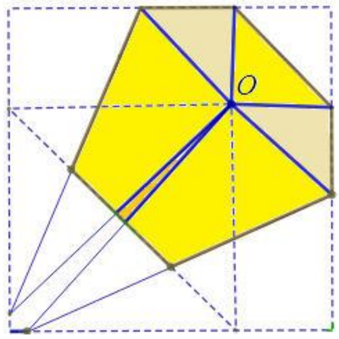
$$e_1 = 1 + \frac{0+0+0+0+0+0}{6} = 1; \qquad e_2 = \frac{7}{6}\cdot 1 - \frac{1}{6}\cdot 0 = \frac{7}{6};$$

$$e_3 = \frac{7}{6}\cdot\frac{7}{6} - \frac{1}{6}\cdot 0 = \frac{49}{36}; \qquad e_4 = \frac{7}{6}\cdot\frac{49}{36} - \frac{1}{6}\cdot 0 = \frac{343}{216}; \ \ldots$$

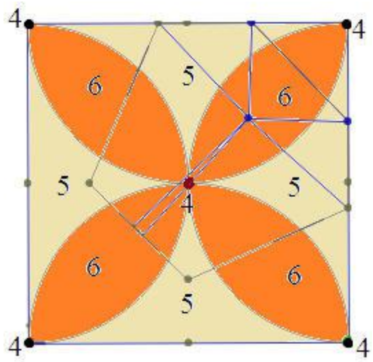
Замечание. Создается впечатление, что математические ожидания образуют геометрическую прогрессию со знаменателем $\frac{7}{6}$. Это никак не согласуется с тем интуитивно очевидным фактом, что число бросаний должно расти примерно как арифметическая прогрессия с разностью $\frac{2}{7}$ – ровно 2 броска в среднем нужно сделать, чтобы увеличить сумму на 7. Конечно, впечатление ложное – геометрическое возрастание уже после суммы 6 (как только e_{n-7} становится положительным числом) сменяется более медленным ростом и все приходит к ожидаемым значениям.

Разумеется, чтобы посчитать e_{2010} приближенно, можно воспользоваться найденной рекуррентной формулой и каким-нибудь расчетным средством. Например, с помощью Excel моментально находим: $e_{2010} = 574,761904$.

Если принять, что искомая величина растет примерно как арифметическая прогрессия с разностью $\frac{2}{7}$, можно найти $e_{2010} \approx 2010 \cdot \frac{2}{7} \approx 574,286$. Как видим, погрешность небольшая.



17. **Оригами. Лепестковый многоугольник. (3 б. 9-11)** В бумажном квадрате случайным образом выбирается точка O. Затем квадрат сгибают так, чтобы каждая вершина наложилась на точку O. На рисунке показана одна из возможных схем складывания. Найдите математическое ожидание числа сторон появившегося многоугольника.



Решение. Небольшое геометрическое исследование показывает, что в результате может получиться шестиугольник, пятиугольник и четырехугольник. Четырехугольник получается, только если точка O попала точно в центр или в вершину квадрата. Вероятность этого равна нулю. Шестиугольник получается, если точка O внутри фигуры, граница которой образована четырьмя полуокружностями (оранжевый четырехлепестковый цветок на рисунке). Если же точка O в любом другое месте (серая область), то получается пятиугольник.

Если считать, что площадь квадрата равна 1, то вероятность того, что точка O попадает в оранжевую область, равна площади этой области:

$$\mathrm{P}_o = S_o = 8 \cdot \left(\frac{\pi}{16} - \frac{1}{8} \right) = \frac{\pi}{2} - 1.$$

Вероятность того, что точка O попадает в серую область, равна

$$\mathrm{P}_g = 1 - \mathrm{P}_o = 2 - \frac{\pi}{2}.$$

Таким образом, математическое ожидание числа сторон многоугольника равно

$$5\cdot\left(2-\frac{\pi}{2}\right)+6\cdot\left(\frac{\pi}{2}-1\right)=4+\frac{\pi}{2}\approx 5{,}571.$$

18. **Рассеянный Учёный в автобусе (4 б. 6-11)**. В автобусе n мест, и все билеты проданы n пассажирам. Первым в автобус заходит Рассеянный Учёный и, не посмотрев на билет, занимает первое попавшееся место. Далее пассажиры входят по одному. Если вошедший видит, что его место свободно, он занимает свое место. Если же место занято, то вошедший занимает первое попавшееся свободное место. Найдите вероятность того, что пассажир, вошедший последним, займет место согласно своему билету?

Решение. Первый способ. Пронумеруем всех пассажиров, начиная с Учёного, в том порядке, в каком они заходили в автобус. Последний пассажир имеет номер n. Для простоты места пронумеруем так же. Пусть все, кроме последнего пассажира, уже вошли и заняли места. Осталось одно свободное место. Если бы это было второе место, то второй пассажир уже занял бы его. Точно так же – если бы это было место с одним из номеров $3, 4, 5, \ldots n-1$, то оно было бы уже занято. Однако, это место свободно. Значит, это место принадлежит либо последнему пассажиру, либо Рассеянному Учёному.

Таким образом, все зависит от того, кому принадлежит последнее место. Ясно, что оно с равными шансами может принадлежать как первому, так и последнему. В первом случае последний пассажир сядет не на свое место, а во втором – на свое. Значит, вероятности обоих событий равны $\frac{1}{2}$.

Второй способ. Воспользуемся методом математической индукции.

1. **Базис индукции**: $n=2$. Очевидно, что с вероятностью $\frac{1}{2}$ учёный займет своё место, а значит, второй (он же последний) пассажир займет своё место с такой же вероятностью $\frac{1}{2}$.

2. **Предположение индукции**. Пусть для $n \le k$ доказано, что искомая вероятность равна $\frac{1}{2}$.

3. **Шаг индукции**. Докажем, что искомая вероятность равна $\frac{1}{2}$ для $n = k+1$. С вероятностью $\frac{1}{k+1}$ учёный займет свое место. В этом случае последний пассажир занимает свое место с вероятностью 1.

С вероятностью $\frac{1}{k+1}$ Учёный займет место последнего пассажира, и тогда тот займет свое место с вероятностью 0.

С вероятностью $\frac{k-1}{k+1}$ Учёный займет место пассажира с номером i, где $1 < i \le k$. Тогда i-ый пассажир играет роль «нового Рассеянного Учёного». По предположению индукции в этом случае последний пассажир занимает свое место с вероятностью $\frac{1}{2}$.

Таким образом, по формуле полной вероятности получаем, что вероятность события «последний пассажир занял свое место» равна

$$p = \frac{1}{k+1} \cdot 1 + \frac{1}{k+1} \cdot 0 + \frac{k-1}{k+1} \cdot \frac{1}{2} = \frac{1}{2}.$$

Утверждение доказано.

19. **Строгие нарды (5 б. 10-11)**. Правильная игральная кость бросается много раз. Известно, что в какой-то момент сумма очков стала равна ровно 2010. Найдите математическое ожидание числа бросков, сделанных к этому моменту.

Решение. Первый способ. Пусть событие A_n «сумма очков равна n», а X_n – число сделанных при этом бросков. Событие A_n мы считаем уже осуществившимся, и нас интересует ожидание $\mathrm{E}\,X_n$. Пусть, как и прежде

$$I_k = \begin{cases} 1, \textit{если первый бросок дал } k \textit{ очков}, \\ 0 \textit{ в противном случае.} \end{cases}$$

Теперь вероятность события $(I_k = 1)$ условная

$$\mathrm{P}(I_k = 1) = \mathrm{P}(B_k | A_n) = \frac{\mathrm{P}(B_k \cap A_n)}{\mathrm{P}(A_n)} = \frac{\mathrm{P}(B_k \cap A_{n-k})}{\mathrm{P}(A_n)},$$

где событие B_k «первый бросок дал k очков», и событие A_{n-k} «сумма очков равна $n-k$» (при всех бросках, кроме первого).

События B_k и A_{n-k} относятся к разным броскам и поэтому независимы:

$$\mathrm{P}(I_k = 1) = \frac{\mathrm{P}(B_k)\cdot \mathrm{P}(A_{n-k})}{\mathrm{P}(A_n)} = \frac{\mathrm{P}(A_{n-k})}{6\mathrm{P}(A_n)}.$$

Пусть величина X_{n-k} – число бросков, считая со второго, давших сумму $n-k$. Как и раньше

$$X_n = \sum_{k=1}^{6} X_{n-k} I_k + 1.$$

Случайные величины X_{n-k} и I_k по-прежнему независимы, поскольку относятся к разным сериям бросков. Поэтому, переходя к ожиданиям, получим:

$$\mathrm{E}\,X_n = \sum_{k=1}^{6} \mathrm{E}\,X_{n-k}\cdot \mathrm{E}\,I_k + 1 = \sum_{k=1}^{6} \frac{\mathrm{E}\,X_{n-k}\,\mathrm{P}(A_{n-k})}{6\mathrm{P}(A_n)} + 1.$$

Вводя краткие обозначения $EX_k = e_k$ и $\mathrm{P}(A_{n-k}) = p_k$, получаем:

$$e_n = \sum_{k=1}^{6} \frac{e_{n-k}\,p_{n-k}}{6p_n} + 1. \quad (1)$$

Осталось воспользоваться тем, что $6p_n = \sum_{k=1}^{6} p_{n-k}$. Это следует, например, из того, что сумма вероятностей $\mathrm{P}(I_k = 1)$ равна единице. Это факт можно взять из решения задачи 1-й заочной олимпиады 2008 года. Тогда

$$e_n = \frac{\sum\limits_{k=1..6} e_{n-k}\cdot p_{n-k}}{\sum\limits_{k=1..6} p_{n-k}} + 1.$$

Начальные значения: $p_{-5} = p_{-4} = ... = p_{-1} = 0;\ p_0 = 1;\quad e_{-5} = e_{-4} = ... = e_0 = 0$. Последовательное вычисление дает

$$e_1 = 1,\quad p_1 = \frac{1}{6};$$

$$e_2 = \frac{1\cdot\frac{1}{6} + 0\cdot 1 + 0\cdot 0 + 0\cdot 0 + 0\cdot 0 + 0\cdot 0}{\frac{1}{6} + 1 + 0 + 0 + 0 + 0} + 1 = \frac{8}{7},\quad p_2 = \frac{\frac{1}{6}+1}{6} = \frac{7}{36};$$

$$e_3 = \frac{\frac{8}{7}\cdot\frac{7}{36} + 1\cdot\frac{1}{6} + 0\cdot 1 + 0\cdot 0 + 0\cdot 0 + 0\cdot 0}{\frac{7}{36} + \frac{1}{6} + 1 + 0 + 0 + 0} + 1 = \frac{9}{7} \quad \textit{и т.д.}$$

Применив любое расчетное средство, например, Excel, без труда получаем: $e_{2010} = 574{,}5238095...$

Второй способ. Известно, что сумма очков сравнялась с n. Пусть наступило событие A_k «в какой-то момент сумма равна k». Тогда последующие броски должны привести к событию A_{n-k}. Значит.

$$\mathrm{P}\left(A_k \middle| A_n\right) = \frac{\mathrm{P}\left(A_k \cap A_n\right)}{\mathrm{P}\left(A_n\right)} = \frac{\mathrm{P}\left(A_k \cap A_{n-k}\right)}{\mathrm{P}\left(A_n\right)}.$$

События A_{n-k} и A_k относятся к разным сериям бросков и поэтому независимы: $\mathrm{P}\left(A_k \middle| A_n\right) = \dfrac{\mathrm{P}\left(A_k\right) \cdot \mathrm{P}\left(A_{n-k}\right)}{\mathrm{P}\left(A_n\right)}$.

Рассмотрим величину $I_k = \begin{cases} 1, \textit{ если сумма в какой} - \textit{то момент равна } k, \\ 0, \textit{ в противном случае.} \end{cases}$

Очевидно, случайная величина «число полученных сумм»

$$X_n = I_1 + I_2 + I_3 + \ldots + I_n$$

равна числу сделанных бросков.

Введем краткие обозначения $\mathrm{E}\, X_k = e_k$; $\mathrm{P}\left(A_k\right) = p_k$. Математическое ожидание $\mathrm{E}\, I_k$ равно $0 \cdot \left(1 - \mathrm{P}\left(A_k \middle| A_n\right)\right) + 1 \cdot \mathrm{P}\left(A_k \middle| A_n\right) = \mathrm{P}\left(A_k \middle| A_n\right) = \dfrac{p_k p_{n-k}}{p_n}$.

Перейдем к математическому ожиданию суммы:

$$\mathrm{E}\, X_n = e_n = \sum_{k=1}^{n} \mathrm{E}\, I_i = \sum_{k=1}^{n} \frac{p_k p_{n-k}}{p_n} = \sum_{k=1}^{n-1} \frac{p_k p_{n-k}}{p_n} + 1.$$

Пользуясь соотношением $p_{n-k} = \dfrac{1}{6} \sum_{i=1}^{6} p_{n-k-i}$, запишем ожидание иначе:

$$e_n = \sum_{k=1}^{n-1} \frac{\sum_{i=1}^{6} p_k p_{n-k-i}}{6 p_n} + 1 = \sum_{k=1}^{n-1} \frac{\sum_{i=1}^{6} p_k p_{n-k-i}}{6 p_n} + 1 = \sum_{i=1}^{6} \frac{p_{n-i} \sum_{k=1}^{n-i} \frac{p_k p_{n-k-i}}{p_{n-i}}}{6 p_n} + 1 = \sum_{i=1}^{6} \frac{e_{n-i} p_{n-i}}{6 p_n}.$$

Полученное соотношение совпадает с (1).