

КОНСПЕКТ ЛЕКЦИИ · ЛЕКТИКА

Теория вероятностей: от жеребьёвки к моделям случайности

Длительность: 2 ч 6 мин

Подготовлено: 2026-10-17 20:03

Кратко

Лекция и семинар вводят базовые понятия теории вероятностей. Вероятность возникла из практических задач: делёж ставки в прерванной игре, обработка записей и наблюдений, планирование экспериментов. Затем рассматриваются аксиоматика, события, независимость, серии и дерево исходов. Семинар проверяет интуицию симуляциями монеты, кубиков и плейлиста. Главный вывод: результат симуляции верен лишь в выбранной модели, а её предположения нужно проверять отдельно.

Главные выводы

- ★ Вероятность связана с моделью опыта: сначала описываем исходы, события и предположения, затем считаем
- ★ Равновероятные исходы не дают равных частот в коротких сериях; большое число повторений уточняет оценку внутри модели
- ★ Редкое событие не доказывает нарушение модели; нужно сравнивать с числом возможностей, в которых оно могло произойти
- ★ Дополнение, пересечение и объединение считаются по правилам, а несовместные события складываются
- ★ Зависимость между исходами меняет пространство исходов и вероятности
- ★ Процедура выбора (сначала альбом, затем песня) меняет шансы, поэтому слово «случайно» требует уточнения
- ★ Псевдослучайные генераторы нужно выбирать по свойствам; seed обеспечивает воспроизводимость
- ★ Модель нужно проверять отдельно: симуляция повторяет описанный опыт, а не реальный мир

Оглавление

1. Зачем считать шансы и как разделить ставку	00:00
2. Данные, измерения и планирование экспериментов	09:57
3. Аксиомы вероятности и два взгляда на неё	15:42
4. Серии, строки и дерево исходов	25:21
5. Одинаковые шансы, разные впечатления	34:22
6. Независимость, дополнение, пересечение, объединение	45:15
7. Плейлист, альбомы и генераторы случайных чисел	55:03
8. Порядок результатов, seed и выбор модели	01:06:21
9. Семинар: цели и настройка среды	01:21:05
10. Строка человека против генератора и случайное блуждание	01:23:19
11. Казино, перевес и распределение числа орлов	01:33:43
12. Серии, совпадения и проверка задач о монете	01:41:40
13. Кубики: частоты, сетка пар и суммы	01:48:39
14. Плейлист: повтор рядом, альбомы и итоги	01:58:54

Задания и сроки

- Прочитать книгу Кнута «Искусство программирования» о генераторах случайных чисел
- Запустить симуляцию из семинара с другим параметром и разобрать результат
- Поиграть с задачками семинара и попробовать разные параметры

Конспект

ГЛАВА 1

Зачем считать шансы и как разделить ставку

00:00 – 09:57

Вероятность нужна для решений в условиях неопределённости. Классическая задача Паскаля и Ферма показывает, как честно разделить банк прерванной игры.

Вероятность нужна для двух практических задач: принимать решения, когда исход неизвестен, и понимать, насколько вероятен каждый исход. Обе задачи появились задолго до формальной теории.

Жребий, игры и первые записи

Жребий помогает принять решение, когда нет оснований выбрать один вариант из нескольких. Игры известны с древности. Игроку, однако, мало знать результат жребия. Ему нужно понять, выгодны ли условия игры.

Позже люди стали записывать данные: переписи городов, наблюдения астрономов. Накопив много записей, их начали сравнивать и задаваться вопросами: почему числа получились именно такими и какой механизм их породил. Так выросли две линии, из которых позже сложились вероятность и статистика.

Задача о прерванной игре

В 1654 году Паскаль и Ферма обсуждали задачу, которая стала классической. Игра идёт до трёх побед, каждый раунд честный и независимый от других. Игру прервали при счёте 2:1, банк равен 100. Как разделить банк?

Есть три разумных варианта:

- **Поровну, 50:50.** Ставки возвращаются, игра признаётся несостоявшейся.
- **По счёту 2:1.** Банк делится по числу выигранных раундов.
- **По шансам на победу.** Оценивается, чем мог бы закончиться матч, и банк делится по этим шансам.

Из зала предложили вернуть каждому половину ставки, поскольку сыграна лишь часть партий. Мысль интересная, но доля сыгранных партий не отвечает на главный вопрос: кто ближе к победе. Кроме того, до трёх побед игра длится не больше пяти партий, а не шести, как прикидывали вначале.

Расчёт по шансам

Представим, что игра продолжается. Игроку А нужна ещё одна победа, игроку Б — две. Доиграем мысленно два раунда, чтобы все пути имели одинаковую длину. Получаем четыре равновозможных продолжения.

Задача: кто побеждает в продолжениях?

Продолжения: АА, АВ, БА, ББ. Каждое имеет шанс $\frac{1}{4}$.

- В АА, АВ и БА побеждает А: три исхода.
- В ББ побеждает Б: один исход.

Значит, $P(A) = \frac{3}{4}$, $P(B) = \frac{1}{4}$.

АА	А победил
АВ	А победил
БА	А победил
ББ	В победил

Четыре равновозможных продолжения игры и исходы

Если А выиграет первый из добавленных раундов, второй уже ничего не меняет. Мы всё равно его учитываем, чтобы сравнивать пути одинаковой длины.

Ожидание и делёж банка

Ожидаемая доля каждого игрока равна его шансу, умноженному на банк:

$$\text{доля}_A = P(A) \cdot 100 = 0,75 \cdot 100 = 75$$

$$\text{доля}_B = P(B) \cdot 100 = 0,25 \cdot 100 = 25$$

Вероятность даёт механизм, а не готовый ответ о справедливости

Делёж по шансам выглядит разумно: он учитывает и текущий счёт, и то, что могло бы произойти дальше. Но единственно справедливым он не является. Понятие «справедливо» здесь дискуссионно, и разные варианты делёжа можно обосновать по-разному. Расчёт показывает, чего можно ждать от правил игры при известных шансах, и именно это делает задачу полезной.

Данные, измерения и планирование экспериментов

09:57 – 15:42

Граунт учился сравнивать записи, Гаусс и Лаплас объединяли неточные наблюдения, Фишер предложил случайное назначение условий в опытах.

Записи о смертности: сравнение вместо отдельного случая

В 1662 году Джон Граунт опубликовал анализ лондонских записей о смертности. Судьбу отдельного человека предсказать нельзя, но в больших группах видны устойчивые доли и изменения во времени. Для этого записи объединяют в таблицы и сравнивают между собой.

Прежде чем считать

Сначала выясняем, кого и как учитывали. Пропущенные записи или разные правила учёта могут изменить вывод ещё до первой формулы.

Объединение неточных наблюдений

В начале XIX века Лежандр и Гаусс разрабатывали методы объединения неточных наблюдений. Такие задачи возникали, например, в астрономии: разные наблюдения одного объекта дают разные числа, и из них нужно получить один ответ и оценку разброса.

Пять измерений одной длины

Одну и ту же длину измерили пять раз, см: 9.8; 9.9; 10.0; 10.1; 10.2.



Пять измерений одной длины на числовой прямой

Какое значение взять за ответ? Можно выбрать то, которому больше доверяем, или использовать среднее:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Пример

$$\bar{x} = \frac{9.8 + 9.9 + 10.0 + 10.1 + 10.2}{5} = \frac{50.0}{5} = 10.0$$

Случайные отклонения в разные стороны частично взаимно гасятся, поэтому среднее устойчивее отдельного измерения.

Систематическая ошибка

Если линейка неточна, все измерения смещены в одну сторону. Усреднение уменьшает случайный разброс, но смещение сохраняется.

Ловушка

Больше повторений не исправит систематическую ошибку. Её нужно искать в самом приборе или методе измерения.

Урожайность: честное сравнение условий

Два условия, А и В, например два вида пшеницы или два варианта удобрения. На одном участке урожай выше. Причиной может быть условие, но также почва или погода. Если условие А всегда стоит на левой половине поля, а В на правой, эффект условия смешивается с различиями почвы.

Условия разделены по месту				Условия перемешаны			
А	А	В	В	А	В	А	В
А	А	В	В	В	А	В	А
А	А	В	В	А	В	В	А

Схема размещения условий А и В по участкам поля

Решение — перемешать условия по участкам. Каждый участок получает условие случайно.

Рандомизация по Фишеру

Работая в Ротамстеде, Рональд Фишер развивал планирование экспериментов. Случайное назначение условий помогает сделать группы сравнимыми: неизмеренные факторы, например свойства почвы, в среднем распределяются между группами одинаково. Это уменьшает риск спутать влияние условия с различиями участков.

Порядок действий:

1. Делим поле на участки.
2. Для каждого участка случайно выбираем А или В.
3. Сравниваем средний урожай на участках А и на участках В.

А/В-тесты

На той же идее основаны А/В-тесты: пользователей случайно распределяют по вариантам, чтобы уменьшить риск спутать влияние варианта с различиями состава групп.

ГЛАВА 3

Аксиомы вероятности и два взгляда на неё

15:42 – 25:21

Колмогоров описал вероятность через исходы, события и вероятности. Вероятность идёт от модели к данным, статистика наоборот. Частотный и байесовский взгляды дают разные ответы на вопрос, что значит «вероятность».

Три шага Колмогорова

В 1933 году Андрей Колмогоров изложил основания теории вероятностей в виде аксиом. Работа с любой задачей идёт в три шага:

1. Назвать все возможные исходы опыта. Их множество обозначают Ω .
2. Указать событие, то есть набор исходов, шанс которого нас интересует.
3. Задать вероятности.

Исход и событие

Исход — один возможный результат опыта. **Событие** — набор исходов, для которого мы считаем вероятность.

Опыт	Исходы	Событие
Монета	О или Р	выпал орёл
Два кубика	пара (a, b)	сумма не меньше 10
Плейлист	последовательность песен	есть повтор подряд

Аксиомы задают правила для вероятностей. Для любого события A выполняется

$$P(A) \geq 0, \quad P(\Omega) = 1,$$

а для несовместных событий A и B , которые не могут произойти одновременно,

$$P(A \cup B) = P(A) + P(B).$$

Отсюда следует, например, что $P(\bar{A}) = 1 - P(A)$: шанс противоположного события дополняет шанс события до единицы.

Согласованность и соответствие

Аксиомы гарантируют, что рассуждения внутри модели не противоречат друг другу. Соответствие модели реальному опыту нужно обосновать отдельно: честна ли монета, независимы ли броски, соответствует ли алгоритм перемешивания тому, что вы хотите.

Вероятность и статистика: разные направления

Вероятность идёт от модели к данным. Правило известно: шанс орла равен $1/2$. Спрашиваем, какие результаты дадут 20 бросков.

Статистика идёт наоборот. Результат уже есть: 14 орлов из 20. По нему пытаемся понять, какая модель монеты его породила.

Частота $14/20 = 0.7$ ещё не доказывает, что шанс орла равен 0.7. При честной монете такой перекося тоже возможен из-за случайного разброса. Поэтому нужен способ учитывать этот разброс.

Монета под стаканом

Монету бросили и накрыли стаканом. Сторона уже определена физически, но её никто не видит.



Монета под стаканом: результат скрыт от наблюдателя

Для тех, кто не заглядывал, ответ пока неизвестен, и шанс орла естественно считать $1/2$. Если кто-то заглянул, монета не изменилась, но его информация стала другой.

Отсюда видно, что слово «случайно» может описывать два разных факта: механизм опыта, при котором никто не влияет на бросок, или наше незнание результата.

Уточняйте смысл

Говоря «случайно», назовите, о чём речь: о механизме опыта или о том, что мы знаем о результате.

Два взгляда на вероятность

Частотный подход

Вероятность — это устойчивая доля, с которой событие происходит в длинной серии сопоставимых повторов. Если шанс орла равен $1/2$, то при многих бросках доля орлов приближается к половине.

Байесовский подход

Вероятность выражает уверенность с учётом всей доступной информации. Новые сведения могут эту уверенность изменить. Тот, кто заглянул под стакан, оценивает шанс орла иначе, чем тот, кто не заглядывал.

Оба подхода требуют ясных предположений. Число $1/2$ нужно объяснить правилами броска или тем, что нам известно о монете.

Что применяют на практике

В А/В-тестах чаще используют частотный подход. Он опирается на эмпирическое распределение: подсчитали, как клиенты ведут себя в разных вариантах, накопили данные и считаем, что они описывают клиентов за выбранный период.

Байесовский подход тоже применяется, но реже. Для байесовского расчёта нужно задать вероятностную модель и априорное распределение; подробное описание физики процесса и всех связей между признаками не обязательно. Построение модели может быть сложным как в крупных сервисах, так и в небольших компаниях.

Вероятность и статистика: исходная информация, вопрос, вывод

АСПЕКТ	ВЕРОЯТНОСТЬ	СТАТИСТИКА
Исходная информация	Модель мира: вероятности заданы	Данные: 14 орлов из 20 бросков
Вопрос	Какие данные даст модель?	Какая модель породила данные?
Направление вывода	От модели к данным	От данных к модели
Пример	Сколько орлов будет в 20 бросках?	Что говорят 14 орлов о монете?

ГЛАВА 4

Серии, строки и дерево исходов

25:21 – 34:22

Редкость серии не доказывает нечестности монеты. Дерево исходов и дополнение позволяют считать вероятности событий, например «хотя бы один орёл».

Десять орлов подряд

В пьесе Тома Стоппарда «Розенкранц и Гильденстерн мертвы» монета снова и снова падает орлом, а герои ищут объяснение странной серии. Оценим, насколько такая серия вероятна. Пусть заранее выбраны 10 честных независимых бросков (число 10 взято для примера). Шанс, что все они окажутся орлами:

$$P = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$$

Каждый бросок даёт орла с шансом $\frac{1}{2}$. Броски независимы, поэтому шансы перемножаются.

Редкое событие ещё не доказывает подвоха

Шанс $\frac{1}{1024}$ мал, но не равен нулю, значит, серия возможна. Десять орлов подряд не доказывают, что монета нечестная: честная монета тоже иногда даёт такие серии. Чтобы судить о самой монете, нужно больше бросков. По одной серии о ней можно сказать немного.

Ловушка

Вопрос «возможна ли такая серия?» и вопрос «нечестна ли монета?» — разные. Первый решается расчётом при известной модели, второй требует данных о самой монете.

Случайность как непредсказуемость

Мы называем случайным то, что нельзя предсказать. Заметив паттерн, мы считаем, что есть правило, и делаем вывод, что предположение о честной монете нарушено. Но любая конкретная последовательность из десяти бросков при честной монете имеет тот же шанс $\frac{1}{1024}$, что и серия орлов. Паттерн говорит о нашем восприятии, а не о механизме опыта.

Две строки по 24 броска

В строке А орлы и решки чередуются без длинных серий. Строка В начинается с пяти орлов подряд. Большинство участников выбирает А как «более случайную», а В считает подозрительной. Однако обе строки могли получиться честной монетой: каждая конкретная строка из 24 бросков имеет одинаковый шанс.

Событие и все исходы

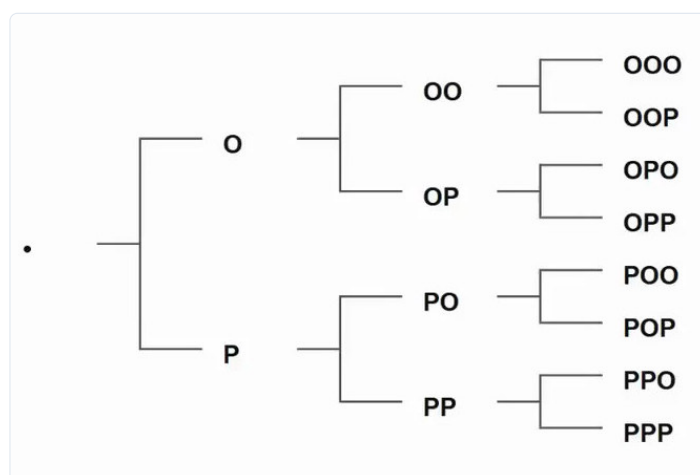
Опыт в этом разделе — один бросок. Все возможные результаты образуют пространство исходов:

$$\Omega = \{O, P\}$$

Строчной буквой ω обозначают результат, который получился. **Событие** — подмножество исходов Ω , которое нас интересует. Событие $A = \{O\}$ означает «выпал орёл».

Три броска: дерево исходов

При нескольких бросках исходов становится больше. Каждая развилка дерева добавляет орла или решку:



Дерево исходов для трёх бросков монеты

Три броска дают восемь путей: OOO, OOP, OPO, OPP, POO, POP, PPO, PPP. Путь OPO означает: сначала орёл, затем решка, затем орёл. При честных независимых бросках его шанс равен $\left(\frac{1}{2}\right)^3 = \frac{1}{8}$. Все восемь путей равновозможны, поэтому у каждого шанс $\frac{1}{8}$.

Ровно один орёл: 3/8

Чтобы найти вероятность события, выделим подходящие пути и разделим их число на общее число путей:

1. Подходят пути с одним орлом: OPP, POP, PPO.
2. Всего путей восемь.
3. Положение орла различается, но условие выполнено во всех трёх случаях.

$$P(\text{ровно один орёл}) = \frac{3}{8}$$

Когда работает подсчёт

Формула «благоприятные исходы делить на все» верна, когда все исходы равновозможны.

Хотя бы один орёл: дополнение

Перечислять семь путей с орлом можно, но проще найти единственный путь без орла: PPP. Событие «хотя бы один орёл» — дополнение к событию «ни одного орла»:

$$P(\text{хотя бы один орёл}) = 1 - P(\text{PPP}) = 1 - \frac{1}{8} = \frac{7}{8}$$

Приём

Для вопросов «хотя бы один сбой», «хотя бы один повтор» или «хотя бы одно совпадение» сначала считайте шанс противоположного события «ни одного».

ГЛАВА 5

Одинаковые шансы, разные впечатления

34:22 – 45:15

Равновероятные строки кажутся разными, потому что мы видим паттерны. Окна серий пересекаются, поэтому их шансы нельзя просто складывать.

Шесть орлов и смешанная строка

Сравним две строки из шести бросков честной монеты: 000000 и 0P0PP0. Если строка выбрана заранее, шанс каждой одинаков: $(1/2)^6 = 1/64$. Для конкретной строки нужны шесть определённых результатов в определённом порядке.

Разница появляется, когда вопрос касается не одной строки, а целого класса строк:

Событие	Строк в событии	Шанс
Шесть орлов подряд	1 из 64	1/64
Любая смешанная строка	62 из 64	62/64

Отдельная строка из шести орлов так же редка, как любая другая конкретная строка. Но смешанных строк почти все, поэтому «какая-то смешанная строка» почти наверняка выпадет.

Почему мы замечаем закономерность

Мы почти автоматически выделяем паттерны. Строку 000000 легко описать одним правилом: «всегда орёл». Смешанную строку приходится запоминать поштучно, поэтому она кажется нам «более случайной», хотя шансы совпадают.

Здесь же проявляется сжатие. Архиваторы (zip, WinRAR) ищут избыточность: если последовательность описывается короткой программой, в файле много повторов, и его можно уменьшить. Человек поступает похоже: запоминает правило, а не каждое значение. Модели машинного обучения извлекают повторяющиеся закономерности, и простое предсказание следующего токена в языковых моделях даёт сильные результаты именно потому, что заставляет находить суть.

Случайность как непредсказуемость

Случайность здесь понимается как мера непредсказуемости. Строку, которую можно предсказать или описать коротким правилом, легко запомнить, и она кажется неслучайной. Непредсказуемую строку запомнить труднее, и мы называем её случайной.

Ощущение и механизм

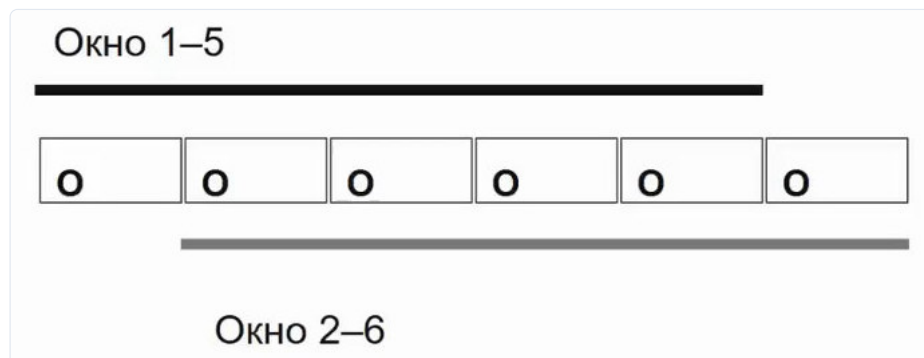
Ощущение случайности относится к нашему восприятию строки. Механизм, который её породил, может быть одинаковым: обе строки из шести бросков честной монеты равновероятны.

Пять орлов: место задано заранее или нет

Рассмотрим два события:

- **А:** первые пять бросков — орлы. Место назвали заранее, поэтому шанс равен $(1/2)^5 = 1/32 \approx 0,031$.
- **В:** среди ста бросков где-нибудь есть серия из пяти орлов подряд.

Событие В включает случаи, когда серия стоит в любом месте. Окно из пяти бросков может начинаться с 1-го по 96-й бросок, то есть окон $100 - 5 + 1 = 96$.



Одна длинная серия попадает сразу в два окна

Если сложить шансы окон, получится $96 \cdot \frac{1}{32} = 3$, что больше единицы, а вероятность не может быть больше единицы. Окна пересекаются: в шести орлах подряд есть и окно 1–5, и окно 2–6, но вся история одна. Поэтому в каждой истории серию отмечают один раз: встретилась она или нет.

Главное

Серия из пяти орлов где-нибудь в ста бросках встречается почти в 81% историй. Это гораздо чаще, чем орлы именно в первых пяти бросках (примерно 3%).

Удачный «предсказатель»

Представим, что нужно угадать все исходы турнира на выбывание. Один аккаунт почти наверняка ошибётся. Но если завести много аккаунтов и перебрать разные варианты, то какой-то из них угадает всё. Со стороны это выглядит как маловероятное предсказание, хотя результат получен процедурой, а не удачей одного прогноза.

Вопрос при удивлении

Когда что-то кажется маловероятным, спросите: сколько было возможностей его заметить? Важно не только то, что произошло, но и то, сколько попыток было сделано.

После пяти орлов

Серия из пяти орлов не меняет монету. Следующий бросок по-прежнему даёт решку с шансом $1/2$: монета не «копит» долг и не обязана исправлять счёт. Доля орлов может приближаться к половине, но не из-за серии решек, а за счёт большого числа бросков.

Ловушка «монета должна исправиться»

Серия из пяти орлов может быть поводом пересмотреть предположение о честности монеты, если её честность неизвестна. Но это именно повод, а не доказательство: длинные серии при честной монете случаются.

ГЛАВА 6

Независимость, дополнение, пересечение, объединение

45:15 – 55:03

Для честного кубика шансы граней одинаковы. Пересечение, объединение и дополнение считаются по обычным правилам. Зависимость кубиков меняет пространство исходов.

Честный кубик: шесть равных шансов

У честного шестигранного кубика каждая грань выпадает с вероятностью $\frac{1}{6}$. Исход опыта здесь одна грань, и все шесть исходов равновозможны.

Событие «выпало чётное» объединяет три исхода: 2, 4 и 6. Их шансы складываются:

$$P(\text{чётное}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Три правила, на которых держится расчёт

Любая модель опыта подчиняется трём правилам:

1. Вероятность события неотрицательна: $P(A) \geq 0$.
2. Вероятность всего пространства исходов равна 1: $P(\Omega) = 1$.
3. Для несовместных событий вероятности складываются. Несовместные события не могут произойти вместе.

Один бросок не даёт одновременно двойку и четвёрку, поэтому

$$P(2 \text{ или } 4) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

Внимание

Складывать можно только несовместные события. Если события могут произойти вместе, общие исходы засчитываются дважды. Это поправляется формулой объединения ниже.

Дополнение: шанс того, что событие не наступит

Дополнение события A — все исходы, которые в A не входят. Его обозначают A^c . Событие A и его дополнение не пересекаются и вместе покрывают все исходы, поэтому их вероятности в сумме дают 1:

$$P(A^c) = 1 - P(A)$$

Пример: шанс не выбросить шестёрку равен $1 - \frac{1}{6} = \frac{5}{6}$. Считать так проще, чем перечислять пять граней, и тот же приём работал для «хотя бы один орёл» в предыдущей части.

Пересечение: выполнены оба условия

Пусть A — выпало чётное, а B — выпало не меньше четырёх. Тогда

$$A = \{2, 4, 6\}, \quad B = \{4, 5, 6\}$$

Пересечение $A \cap B$ — исходы, которые подходят под оба условия сразу. Это грани 4 и 6:

$$A \cap B = \{4, 6\}, \quad P(A \cap B) = \frac{2}{6} = \frac{1}{3}$$

Исход 2 подходит только под A , а исход 5 — только под B . В пересечение они не попадают.

Объединение: выполнено хотя бы одно условие

Объединение $A \cup B$ — выполнено хотя бы одно из условий, включая случай, когда выполнены оба. Исходы 4 и 6 входят и в A , и в B . Если просто сложить $P(A)$ и $P(B)$, они учтутся дважды, поэтому одно пересечение вычитают:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Подставим значения:

$$P(A \cup B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6}$$

Проверка перечислением: $A \cup B = \{2, 4, 5, 6\}$, четыре исхода из шести.

Внимание

Если забыть вычесть пересечение, получится $\frac{3}{6} + \frac{3}{6} = 1$. Значит, событие «чётное или не меньше четырёх» происходило бы всегда. Но исходы 1 и 3 под него не подходят.

Совет

Прежде чем считать, выпишите исходы каждого события. Так сразу видно, какие исходы общие и нужно ли вычитать пересечение.

Неравномерный кубик: вероятности как веса

Пусть шестёрка выпадает с вероятностью $\frac{1}{2}$, а каждая из остальных граней с вероятностью $\frac{1}{10}$. Проверка: $5 \cdot \frac{1}{10} + \frac{1}{2} = 1$.

Формула $P(A) = \frac{|A|}{|\Omega|}$ здесь не годится: исходы не равновозможны, и подсчёт граней ничего не даст. Вероятности складываются как веса:

$$P(\text{чётное}) = \frac{1}{10} + \frac{1}{10} + \frac{1}{2} = \frac{7}{10}$$

Правило «число благоприятных на число всех» работает только для конечных равновозможных исходов.

Два кубика: 36 равновозможных пар

Опыт с двумя кубиками даёт упорядоченную пару (a, b) : a — результат первого кубика, b — второго. Порядок важен: пара $(4, 6)$ отличается от $(6, 4)$. Всего пар $6 \cdot 6 = 36$, и при честных независимых кубиках каждая имеет шанс $\frac{1}{36}$.

		Второй кубик					
		1	2	3	4	5	6
1-й	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Таблица сумм двух кубиков по всем 36 парам

Сумма 7 получается из шести пар: $(1, 6)$, $(2, 5)$, $(3, 4)$, $(4, 3)$, $(5, 2)$, $(6, 1)$. Поэтому

$$P(\text{сумма} = 7) = \frac{6}{36} = \frac{1}{6}$$

Разные пары дают одинаковую сумму. Сама сумма — числовая характеристика исхода. Позже такую величину назовём случайной величиной.

Зависимость меняет пространство исходов

Теперь второй кубик выставляют на то же число, что выпало на первом. Каждый кубик по отдельности остаётся честным. Но пар становится только шесть:

$$(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)$$

Сумма 7 исчезает, все суммы чётные: $a + a = 2a$. Здесь не изменилась честность граней, а изменилась связь результатов. Поэтому, прежде чем считать, нужно знать, зависят ли результаты друг от друга, потому что от этого зависит и список возможных исходов.

Внимание

Честность каждого кубика по отдельности не задаёт пространство пар. Для двух бросков нужно знать, связаны ли их результаты.

d12 или 2d6: какой кубик выбрать

Сравним два способа получить число от 1 до 12. **d12** — один двенадцатигранный кубик: каждое число выпадает с шансом $\frac{1}{12}$. **2d6** — сумма двух шестигранных, она распределена «куполлом» с пиком при 7.



Сравнение распределений результатов d12 и 2d6

Выбор зависит от порога успеха:

Условие успеха	d12	2d6	Выгоднее
результат не меньше 7	$\frac{6}{12} = 0.5$	$\frac{21}{36} \approx 0.58$	2d6
результат не меньше 10	$\frac{3}{12} = 0.25$	$\frac{6}{36} \approx 0.17$	d12

Для 2d6 при пороге 7 успешны 21 пара. Это проще посчитать дополнением: пар с суммой не больше 6 будет 15, а $36 - 15 = 21$. При пороге 10 успешны шесть пар: (4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6).

Два кубика чаще дают результат из середины, потому что к ней ведёт больше пар. При низком пороге это помогает 2d6, при высоком выгоднее плоский d12. Одинаковый максимум, 12, не означает одинаковую сложность испытания.

ГЛАВА 7

Плейлист, альбомы и генераторы случайных чисел

55:03 — 01:06:21

Результат зависит от процедуры выбора: выбор альбома, затем песни, даёт другие шансы, чем выбор из общей библиотеки. Псевдослучайные генераторы выбирают по изученным свойствам.

Три правила перемешивания

Плейлист можно собрать тремя способами. Все они используют случайность, но дают разный опыт прослушивания.

Правило	Как выбирается трек	Что может произойти
Любая песня	На каждом шаге любой трек из каталога, все равновероятны	Любимый трек может прозвучать дважды подряд
Одна перестановка	Весь каталог перемешивается и проигрывается один раз	Повторов нет: каждая песня звучит не более одного раза
Без соседнего повтора	Выбор случаен, но следующий трек не совпадает с предыдущим	Подряд повтора нет, через несколько песен он возможен

Что считать повтором

Запрет соседнего повтора не мешает вернуться к песне позже. Если раздражает только повтор подряд, достаточно правила «без соседнего повтора». Если раздражает любой повтор за вечер, нужна перестановка. Сначала нужно назвать, что именно не нравится: от этого зависят и алгоритм, и вероятность события.

20 песен и 19 переходов

Пусть каталог состоит из 10 песен, а плейлист из 20 выборов выполняется независимо, с возвращением. Нужно найти шанс хотя бы одного повтора рядом.

Удобнее считать противоположное событие: соседних повторов нет.

- Первый выбор: любая из 10 песен, шанс 10/10.
- Каждый следующий выбор должен отличаться от предыдущей песни: 9 подходящих вариантов из 10, шанс 9/10.
- Между 20 позициями 19 переходов.

Шанс, что соседних повторов нет, равен $(9/10)^{19}$. Тогда шанс хотя бы одного повтора рядом:

$$P = 1 - \left(\frac{9}{10}\right)^{19} \approx 0.865$$

Ловушка

На одном переходе шанс повтора равен всего 10%. Но переходов за вечер много, и шансы накапливаются. Повтор за вечер ожидаем, и это не признак сломанного перемешивания.

Альбомы: процедура меняет шансы

Пусть есть альбом А из двух песен (А1, А2) и альбом В из восьми песен (В1–В8). Всего 10 разных треков. Сравним два способа выбора:

1. Выбрать альбом с шансами 1/2 и 1/2, затем песню внутри него с равными шансами.
2. Выбрать сразу одну из 10 песен, не глядя на альбомы.

Альбом А: 2 песни

A1	A2
----	----

Альбом В: 8 песен

B1	B2	B3	B4	B5	B6	B7	B8
----	----	----	----	----	----	----	----

Альбом А из двух песен и альбом В из восьми песен

Для первого способа шанс каждой песни считается как произведение шансов по шагам:

- песня A1: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$;
- песня B1: $\frac{1}{2} \times \frac{1}{8} = \frac{1}{16}$.

Для второго способа каждая песня получает $\frac{1}{10}$.

Проверка: $2 \cdot \frac{1}{4} + 8 \cdot \frac{1}{16} = 1$. Все шансы в сумме дают единицу.

Почему так: альбом выбирается с одинаковым шансом независимо от размера. На короткий альбом приходится тот же вес, что на длинный, поэтому каждая его песня получает больший шанс. Таким образом, при первом способе треки короткого альбома звучат заметно чаще.



Шансы отдельных песен при выборе альбома и песни и при выборе из общей библиотеки

Описывайте процедуру

Слов «выберем случайно» недостаточно. Нужно записать шаги: что выбирается первым, с каким шансом и что выбирается затем. Только тогда видно, какие шансы получают песни.

Кнут: от жеребьёвки к алгоритму

Сначала случайность получали жеребьёвкой: бросали монету, тянули камешки или палочки разной длины. Затем появились готовые таблицы случайных цифр. С развитием компьютеров возникли две практические потребности:

- нужны длинные последовательности чисел для симуляций;
- расчёт нужно иметь возможность точно повторить, например, чтобы проверить программу.

Из этого следует задача: получить нужные свойства случайности по **воспроизводимому правилу**. Приставка «псевдо» здесь важна. Настоящую

случайность алгоритм не производит, он выдаёт последовательность, которая ведёт себя как случайная. Подробно этот путь описан Кнутом в «Искусстве программирования» (том 2, §3.1).

Генератор, который застрял

Кнут построил алгоритм К из множества запутанных преобразований и ожидал очень хорошую последовательность. Сложность сама по себе качества не гарантировала. При первом запуске алгоритм быстро пришёл к числу 6065038420 и стал повторять его:

$$\dots \rightarrow 6065038420 \rightarrow 6065038420 \rightarrow \dots$$

Алгоритм попал в неподвижную точку: из этого состояния он уже не выходит.

Из этого примера следуют несколько выводов:

- По виду результатов качество генератора не определить. Последовательность может выглядеть случайной, а при отдельных начальных значениях заикнуться или выдать строгий паттерн.
- Аналитически разобрать сложный алгоритм тоже трудно.
- Поэтому генераторы строят из операций, у которых изучены свойства и границы их поведения, а затем комбинируют.

Как выбирать генератор

Выбирайте генератор по изученным свойствам, а не по одной удачной серии чисел. Подробнее о воспроизводимости через seed — в следующей главе.

ГЛАВА 8

Порядок результатов, seed и выбор модели

01:06:21 — 01:09:44

Одинаковая доля орлов не описывает порядок бросков. Для воспроизводимых опытов фиксируют seed. Ошибочную модель не исправит никакое число повторений.

Порядок результатов несёт информацию

Возьмём две строки по двенадцать бросков, в каждой шесть орлов и шесть решек. Если смотреть только на долю орлов, строки неразличимы.

О	Р	О	Р	О	Р	О	Р	О	Р	О	Р
---	---	---	---	---	---	---	---	---	---	---	---

6 орлов, 6 решек. Что дальше?

О	О	Р	Р	Р	О	О	Р	О	Р	Р	О
---	---	---	---	---	---	---	---	---	---	---	---

Тоже 6 орлов и 6 решек

Одинаковая доля не описывает порядок.

Одинаковая доля орлов, разный порядок результатов

Верхняя строка строго чередуется, и её продолжение легко угадать. В нижней порядок нерегулярен. Доля одинаковая, но если игнорировать порядок, теряется часть информации.

Доля отвечает на вопрос «сколько орлов», порядок — на вопрос «как они выпадали». Для проверки случайности важны оба.

Seed: как повторить опыт

Генератор случайных чисел — алгоритм, а не настоящая случайность. Его начальное состояние задаёт **seed** (зерно). Один seed и один алгоритм дают одну и ту же последовательность.

```
import numpy as np

SEED = 12345
rng = np.random.default_rng(SEED) # создаём генератор один раз
flips = rng.random(10) < .5       # все броски берём из него
```

Ловушка: заново задавать один seed перед каждым броском. Тогда каждый бросок повторяет первый результат.

Записывайте seed. Если его не сохранить, результат всё равно был получен, но повторить его нельзя.

Миллион опытов не исправит неверное правило

Симуляция повторяет опыт, который вы описали. Если описание неверно, тысячи повторов лишь точнее оценят неверную модель. Пусть второй кубик копирует первый: тогда сумма равна $2x$ и всегда чётна. Сумма 7 не появится ни в одном повторе.

Большое число повторов уточняет вероятность внутри модели. Саму модель нужно проверять отдельно.

Порядок действий перед расчётом

Для нового вопроса назовите три вещи: **процедуру** (что делаем), **событие** (что считаем успехом) и **предположения** (честность, независимость, выбор с возвращением или без). Только после этого считайте формулу или запускайте симуляцию.

ГЛАВА 9

Семинар: цели и настройка среды

01:21:05 – 01:23:19

Семинар даёт базовую интуицию и умение проводить симуляции, которые нужны в курсах математической статистики и A/B-тестам. Среда фиксирована seed-ом, чтобы результаты воспроизводились.

Цель семинара — выработать базовую интуицию и умение проводить симуляции, которые понадобятся в математической статистике и A/B-тестах.

Ход занятия

Сначала разбираем монету, затем кубики и музыкальный плейлист. В каждом блоке можно менять параметры: число бросков, вероятность орла, порог суммы, длину серии. Сравнивайте свою догадку с результатом и возвращайтесь к задачам лекции.

Фиксируем среду

```
SEED = 12345
rng = np.random.default_rng(SEED)
```

`SEED` задаёт начальное состояние генератора, поэтому при одинаковом значении результаты совпадают у всех. Генератор создаётся один раз, иначе будет повторяться первый результат. Остальной служебный код подключает NumPy и matplotlib и задаёт оформление рисунков. Успешный запуск выводит строку `Среда готова. NumPy 2.1.3.`

Вспомогательные функции

- `tape(values, ax, title, highlights)` рисует ряд клеток О и Р; `highlights` обводит серии.
- `has_run(trials, length)` проверяет для каждой истории, есть ли серия из `length` орлов подряд. Орёл здесь `True`.

- `compare(labels, estimates, exact)` сравнивает симуляцию с точными значениями.
- `dice_grid(mask, title)` рисует сетку 6×6 пар кубиков и выделяет нужные клетки.

```
from numpy.lib.stride_tricks import sliding_window_view

def has_run(trials, length=5):
    x = np.asarray(trials, dtype=bool)
    if x.ndim != 2 or length < 1:
        raise ValueError('Нужен массив историй и положительная длина серии')
    if x.shape[1] < length:
        return np.zeros(x.shape[0], dtype=bool)
    return sliding_window_view(x, length, axis=1).all(axis=-1).any(axis=1)

trials = rng.random((1000, 30)) < .5
print(has_run(trials, 5).mean())
```

Литература

- «Что не так? Три парадокса теории вероятностей».
- «Важнейшая модель теории вероятностей».
- Кнут, «Искусство программирования», том 2, §3.1.

ГЛАВА 10

Строка человека против генератора и случайное блуждание

01:23:19 – 01:33:43

Признак «число переключений» показывает, насколько строка типична для честной монеты. Случайное блуждание с орлом +1 и решкой −1 показывает разницу между долей орлов и перевесом.

Переключения и серии

Переключение — это пара соседних бросков с разными результатами: орёл сменился решкой или решка орлом. **Серия** — подряд идущие одинаковые результаты.

В строке OORPPPO два переключения: между вторым и третьим бросками и между пятым и шестым. Серии здесь OO, PPP и O.

Для честной монеты следующий бросок отличается от предыдущего с шансом $1/2$. В строке из 24 бросков 23 соседние пары, поэтому в среднем ожидается $23 \cdot \frac{1}{2} = 11.5$ переключения.

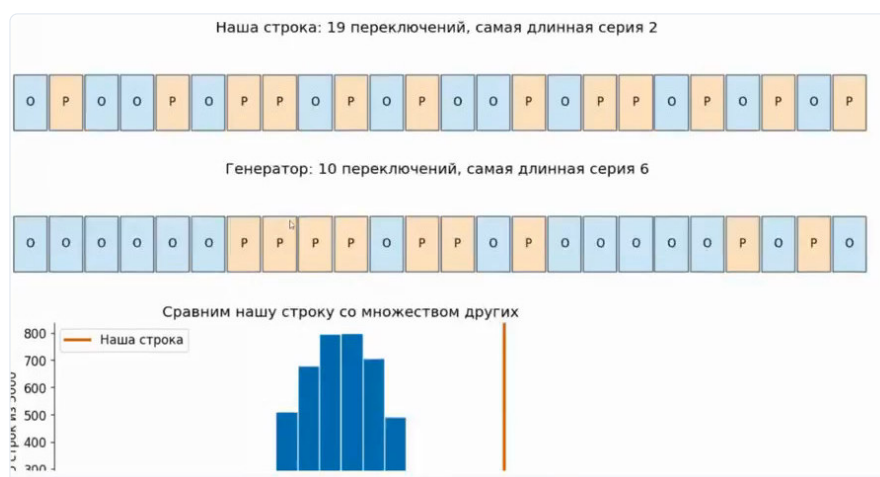
Признак выбираем до просмотра

Строку, записанную вручную из 24 бросков, сравнивают с 5000 случайными строками той же длины. Для каждой строки считают число переключений:

```
human_text = 'OP00P0P0P0P00P0P0P0P0P'
human = np.array(list(human_text)) == '0'
trials = rng.random((5000, len(human))) < .5
switch_counts = (trials[:, 1:] != trials[:, :-1]).sum(axis=1)
```

Выражение `trials[:, 1:] != trials[:, :-1]` сравнивает каждый бросок с предыдущим.

У строки «человека» 19 переключений, у генератора — 10 (при этом в его строке самая длинная серия 6). Гистограмма для 5000 строк собрана около 11-12 переключений, а значение 19 стоит на её правом краю.



Строка человека и генератор на фоне 5000 случайных строк по числу переключений

Что значит «необычно»

Строка необычна по выбранному признаку для выбранной модели. Это не доказывает, что её придумал человек, и не доказывает подделку.

Признак нужно выбрать до просмотра результата. Если перебрать много признаков и оставить тот, где строка выделилась, «необычность» найдётся почти у любой строки.

Подгонка под типичное число переключений

Чтобы проверить этот вывод, строку меняли. Сначала получили строку с 9 переключениями и самой длинной серией 4, а у генератора в той же проверке было 11 переключений и серия 5. Затем подобрали строку с 12 переключениями, как у генератора. По самой длинной серии она всё равно отличается: 4 против 5.

Вывод такой: подгоняя строку под один признак, можно добиться, чтобы она выглядела типичной по этому признаку. Но один признак ничего не доказывает. Перед сравнением назовите признак и объясните, почему он важен для вопроса.

Случайное блуждание

Пусть орёл означает шаг вправо на $+1$, а решка — шаг влево на -1 .

Накопленная сумма шагов называется **случайным блужданием**. Её значение после n бросков равно $H - T$, где H — число орлов, T — число решек. Эту величину называют перевесом.

```
n = 10000
p = .5
flips = rng.random(n) < p
steps = np.arange(1, n+1)
count = np.cumsum(flips)
frequency = count / steps
lead = 2*count - steps
```

Здесь `count` — число орлов к каждому моменту, `frequency` — доля орлов, `lead` — перевес. Поскольку $T = n - H$, получаем $H - T = 2H - n$.

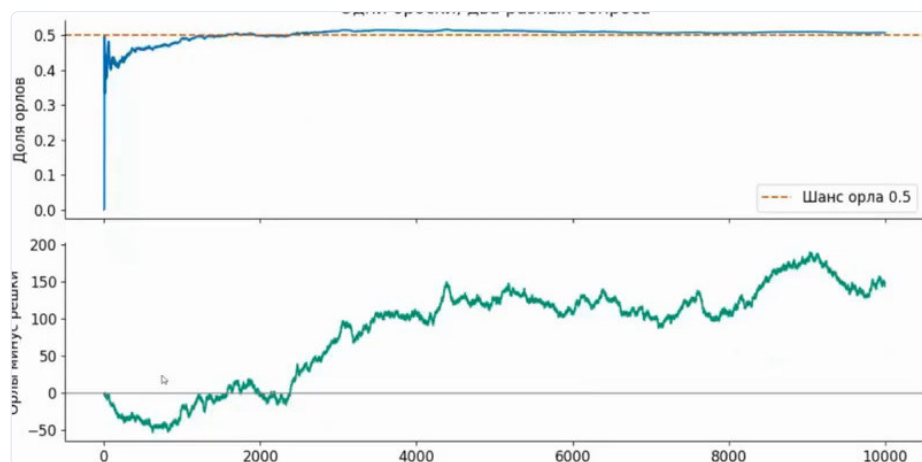
Доля и перевес

Доля орлов равна H/n . Подставим $H = \frac{(H - T) + n}{2}$ и получим точную связь:

$$\frac{H}{n} = \frac{1}{2} + \frac{H - T}{2n}$$

Доля отличается от $1/2$ на перевес, делённый на $2n$. Значит, доля может быть близка к $1/2$, а перевес при этом оставаться большим: отклонение доли равно перевесу, делённому на $2n$, и перевес не обязан уменьшаться при увеличении n .

Честная монета, 10 000 бросков:



Доля орлов и перевес орлов над решками за 10 000 бросков

Доля 0.5072, перевес +144. Проверка по формуле: $\frac{144}{2 \cdot 10000} = 0.0072$, и $0.5 + 0.0072 = 0.5072$.

Два разных вопроса

«Близка ли доля к $1/2$?» и «каков перевес?» — разные вопросы. Доля со временем сходится к шансу орла, а перевес колеблется и может уходить далеко от нуля.

Короткие истории

Для 100 бросков получили долю 0.53 и перевес +6. В начале истории доля заметно уходила от $1/2$, и лишь потом стала держаться около него. На 10 бросках доля может быть 0.7 у совершенно честной монеты (7 орлов из 10): перевес при этом всего +4.

Ловушка

Не судите о честности монеты по доле в короткой истории. Сходимость доли видна только на больших n .

Практика

Перед запуском запишите ожидание: вы смотрите на долю или на перевес, и какого поведения ждёте при большом и малом n .

ГЛАВА 11

Казино, перевес и распределение числа орлов

01:33:43 – 01:41:40

Небольшое преимущество шанса на длинной дистанции даёт устойчивый перевес. Распределение числа орлов в 20 бросках проверяется повторами эксперимента и точным подсчётом.

Случайное блуждание в игре

Пусть орёл приносит игроку +1, а решка −1. Тогда **перевес** после n бросков равен орлы − решки = $2 \cdot \text{орлы} - n$. Каждый бросок сдвигает точку вправо или влево, и такая траектория называется случайным блужданием.

Доля орлов и перевес связаны точно:

$$\frac{\text{орлы}}{n} = \frac{1}{2} + \frac{\text{перевес}}{2n}$$

Поэтому доля может приблизиться к 0.5, а перевес при этом остаться большим. Это разные величины, и отвечают они на разные вопросы.

```
flips = rng.random(n) < p
count = np.cumsum(flips)
frequency = count / steps
lead = 2*count - steps
```

Честная игра: доля стабилизируется, перевес колеблется

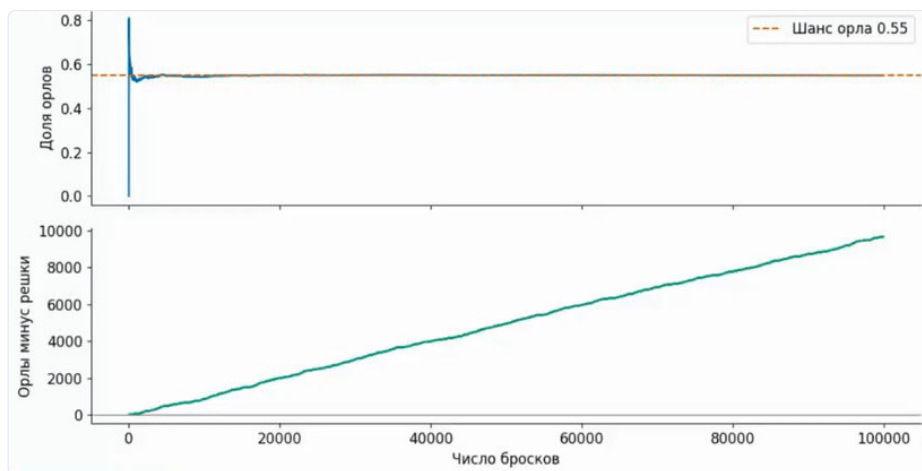
При $p = 0.5$ на 100 000 бросках получили долю 0.5005 и перевес +94. Доля орлов сходится к 0.5, а перевес на длинной дистанции не прилипает к нулю: на графике он колеблется примерно от -100 до $+250$.

Уточнение

Сходится к $\frac{1}{2}$ доля орлов, а не перевес в штуках. В честной игре перевес тоже блуждает вокруг нуля, но к нулю он не стремится.

Небольшое преимущество на длинной дистанции

При $p = 0.55$ орёл выигрывает у решки уже в самих правилах. На 100 000 бросках доля составила 0.5484, перевес +9680. Перевес растёт почти линейно, а доля уверенно держится около 0.55.



Доля орлов и перевес при шансе орла 0.55

При $p = 0.52$ получили долю 0.5173 и перевес +3464 за те же 100 000 бросков. Преимущество меньше, но на длинной дистанции оно тоже накапливается.

Ожидаемый перевес

Каждый бросок в среднем сдвигает перевес на $2p - 1$. При $p = 0.52$ за $n = 100\,000$ бросков ожидаемый перевес равен $0.04 \cdot 100\,000 = 4000$. Опыт дал 3464: для одного запуска это разумное значение.

В коротких сессиях такое преимущество почти не заметно, и результат легко принять за случайность.

Как это работает в азартных играх

Казино получает небольшое преимущество в каждой ставке. В одном вечере оно почти незаметно, но при большом числе игр и игроков оно накапливается. Поэтому долгосрочный итог определяется не удачей отдельной сессии, а шансом.

Отдельно стоят так называемые dark patterns: механики, которые вызывают зависимость и заставляют играть чаще. Они увеличивают число сессий, а значит, и число бросков, на которых работает преимущество.

Ловушка

Серия выигрышей за вечер не опровергает небольшого перевеса в пользу казино. Оценивать нужно итог на длинной дистанции.

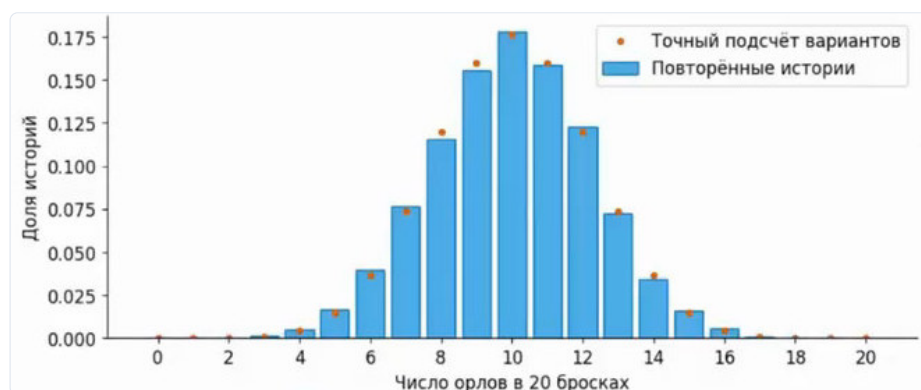
Распределение числа орлов в 20 бросках

Повторим эксперимент «20 бросков» 10 000 раз. По горизонтали отложим число орлов в одной истории, по вертикали — долю таких историй. Это другой график, чем накопленная доля орлов.

```
from math import comb

n = 20
M = 10_000
k = np.arange(n+1)
counts = (rng.random((M,n)) < .5).sum(axis=1)
observed = np.bincount(counts, minlength=n+1)/M
exact = np.array([comb(n,int(j))/2**n for j in k])
```

Ровно 10 орлов: эксперимент даёт 0.178, точный подсчёт — 0.176.



Доля историй по числу орлов в 20 бросках: повторы и точный подсчёт

Точный подсчёт опирается на число способов выбрать позиции орлов среди 20 бросков, делённое на число всех строк:

$$P(\text{ровно } k \text{ орлов}) = \frac{\binom{n}{k}}{2^n}, \quad P(\text{ровно } 10) = \frac{\binom{20}{10}}{2^{20}} = \frac{184\,756}{1\,048\,576} \approx 0.176$$

Самый частый исход — не самый вероятный в сумме

Ровно десять орлов — самый частый отдельный исход, но его шанс около 17.6%. Остальные исходы вместе дают около 82.4%. Поэтому отдельный исход почти всегда менее вероятен, чем «примерно половина орлов» целиком.

При меньшем числе повторов расхождение между опытом и точным значением становится заметнее, поэтому для проверки формулы нужно много повторов.

ГЛАВА 12

Серии, совпадения и проверка задач о монете

01:41:40 — 01:48:39

Серия длины 5 ищется по окнам, а шанс серии растёт с числом бросков и с p . Проверка задач показывает, что шансы согласуются с расчётом.

Серия длины 5 в 60 бросках

Серия — это несколько одинаковых результатов подряд. Чтобы найти серию из пяти орлов в истории из 60 бросков, проверяем все окна длины 5. Ключевая строка:

```
sliding_window_view(x, length, axis=1).all(axis=-1).any(axis=1)
```

Окно сдвигается по строке, и для каждого положения проверяется, выпали ли пять орлов подряд. Если хотя бы одно окно подошло, серия есть.



Броски монеты 1-60 с выделенными сериями орлов

В одной истории серия из пяти орлов нашлась, в другой при длине 7 её не было, а в следующем запуске семерка появилась. Ответ «есть или нет» относится к конкретной истории. Перезапуск даёт новую историю и новый ответ.

Перезапуск до нужного результата

Если запускать опыт, пока не появится нужная серия, мы подбираем историю, а не оцениваем шанс. Сначала называем событие и число бросков, потом считаем долю историй.

Окна и их шансы

В строке из 100 бросков окон длины 5 будет $100 - 5 + 1 = 96$. Соседние окна пересекаются: серия из шести орлов попадает одновременно в окна 1–5 и 2–6. Поэтому шансы окон нельзя складывать. Сумма $96 \cdot \frac{1}{32} = 3$ больше единицы, но вероятность не может быть больше 1: одна история учитывается в нескольких окнах.

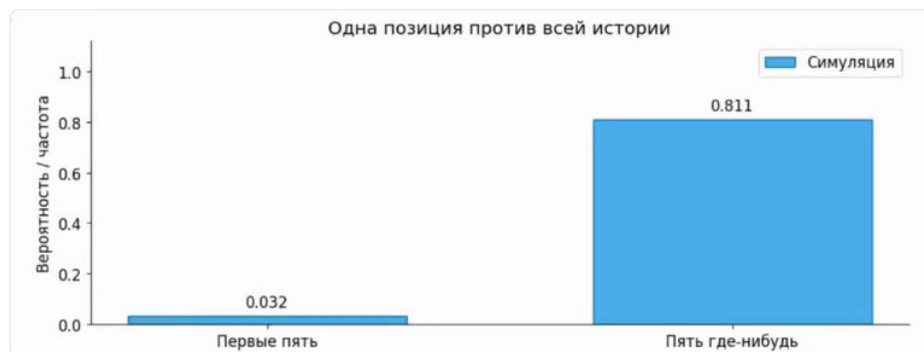
Первые пять или где-нибудь

Сравним два события для 100 бросков:

Событие	Шанс
Первые пять бросков — орлы	$\frac{1}{32} \approx 0.031$
Хотя бы одна серия из пяти орлов где-нибудь	около 0.81

Первое событие фиксирует место заранее. Второе допускает 96 позиций, поэтому его шанс заметно выше.

```
M, n, length = 30_000, 100, 5
trials = rng.random((M,n)) < .5
first = trials[:, :length].all(axis=1)
anywhere = has_run(trials, length)
```



Одна позиция против всей истории

Симуляция на 30 000 историях даёт 0.031 и около 0.81, что совпадает с расчётом.

Что меняет длина истории и шанс орла

- В 1000 бросках серия из пяти орлов почти наверняка есть: доля историй близка к 1.
- Десять орлов подряд в фиксированных первых десяти бросках имеют шанс $\left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$. Где-нибудь в 1000 бросках такая серия встречается примерно в 38% историй.

- Пусть p — шанс орла. При $p = 0.7$ доля историй с серией из пяти орлов уже около 1 при 100 бросках. При $p = 0.3$ рост идёт медленнее.

Больше бросков и больше вероятность орла дают больше мест для совпадения. Небольшие колебания графика — случайный разброс новых историй.

Проверка задач о монете

Задача 1. После пяти орлов подряд решка «нужнее»? Бросок независим от прошлых, поэтому шанс решки остаётся $\frac{1}{2}$. Симуляция: среди историй, начинающихся с пяти орлов (около 3000 из 100 000), следующая решка выпала в 0.512 случаев.

```
x = rng.random((100_000,6)) < .5
selected = x[:,5].all(axis=1)
next_tail = (~x[selected,5]).mean()
```

Задача 2. В трёх бросках что вероятнее: ровно один орёл или три орла?

- Ровно один орёл: ОРР, РОР, РРО — три из восьми равновозможных строк, шанс $\frac{3}{8} = 0.375$.
- Три орла: только ООО, шанс $\frac{1}{8} = 0.125$.

Симуляция дала 0.375 и 0.124. Результаты близки к расчёту для заданной модели честной монеты с независимыми бросками; это не проверяет честность реальной монеты.

ГЛАВА 13

Кубики: частоты, сетка пар и суммы

01:48:39 — 01:58:54

Короткие серии бросков дают заметно разные частоты граней, хотя шансы равны. Сетка пар и симуляции проверяют расчёт для суммы двух кубиков и выбора кубика.

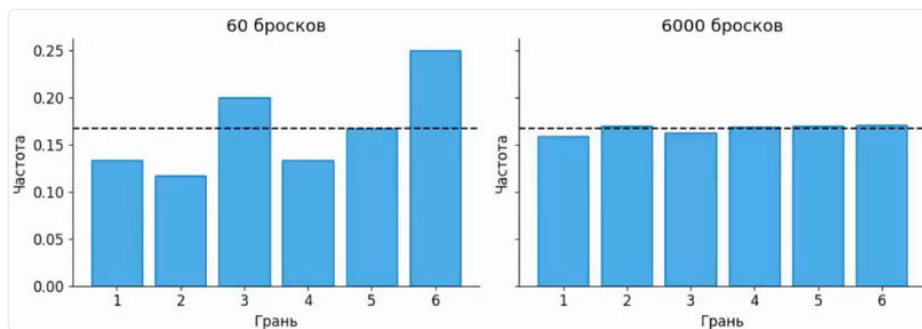
Кубик честный, если до броска все шесть граней равновозможны: шанс каждой равен $\frac{1}{6}$. Это утверждение о модели, а не обещание ровных частот в конкретной серии бросков.

Честность и частоты

Проверим на опыте: бросим честный кубик 60 раз, затем 6000 раз, и сравним доли граней с $\frac{1}{6}$.

```
for n in [60, 6000]:
    values = rng.integers(1,7,n)
    frequencies = np.bincount(values,minlength=7)[1:]/n
    # ... построение столбцов frequencies и линии 1/6
```

При 60 бросках столбцы заметно различаются: грань 2 выпала с долей около 0.23, грань 1 — около 0.10. При 6000 бросках все шесть столбцов почти совпадают с линией $\frac{1}{6}$.



Частоты граней кубика после 60 и 6000 бросков

Короткая серия показывает лишь малую часть возможных исходов, поэтому разброс ещё не доказывает нечестность кубика. Чтобы оценить такой разброс, нужна модель: сначала задаём шансы, затем смотрим, насколько типичны наблюдаемые частоты.

Сетка пар

Опыт с двумя кубиками описывается упорядоченной парой (a, b) : первый и второй кубики различимы. Для честных независимых кубиков всего пар $6 \cdot 6 = 36$, и каждая равновероятна: $\frac{1}{36}$. Удобно расположить их в таблице 6×6 : строка задаёт результат первого кубика, столбец — второго, в клетке стоит сумма.

Событие «сумма не меньше порога» выделяем булевой маской:

```
threshold = 10
first_die, second_die = np.meshgrid(np.arange(1,7), np.arange(1,7), indexing='ij')
sum_grid = first_die + second_die
event = sum_grid >= threshold
print('Точная вероятность:', event.mean())
```

`np.meshgrid` создаёт две матрицы: в первой каждая клетка хранит результат первого кубика, во второй — второго. Их сумма даёт сумму каждой пары, а `event.mean()` — долю выделенных клеток, то есть вероятность.

Порог 10 и порог 3

Сумма не меньше 10 у пар (4,6), (5,5), (6,4), (5,6), (6,5), (6,6). Это 6 из 36:

$P = \frac{6}{36} = \frac{1}{6} \approx 16.7\%$. При пороге 3 исключена только пара (1,1) с суммой 2.

Подходят 35 из 36: $P = \frac{35}{36} \approx 97.2\%$.

Пары (4,6) и (6,4) — разные исходы

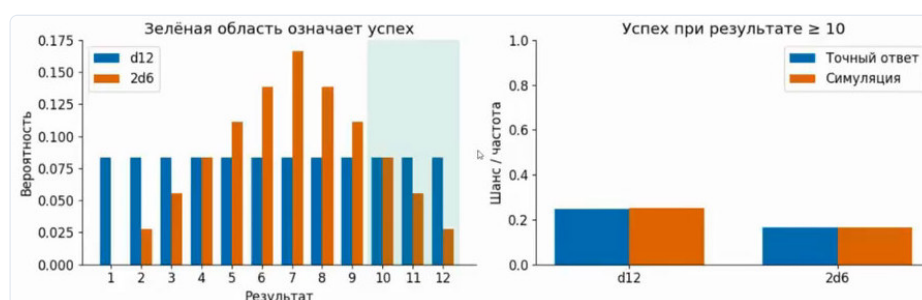
Если считать неупорядоченные пары, их будет 21, но они не равновозможны: {4, 6} получается двумя способами, а {5, 5} — одним. Поэтому шансы считают по упорядоченным парам.

d12 или 2d6

Сравним два варианта: один двенадцатигранный кубик d12 и сумму двух шестигранных 2d6. Оба дают значения от 1 до 12 (у 2d6 единица невозможна), но распределены по-разному: d12 — равномерно, у 2d6 пик приходится на 7.

Порог	d12	2d6	Выгоднее
не меньше 7	6/12 = 50%	21/36 $\approx$ 58.3%	2d6
не меньше 10	3/12 = 25%	6/36 $\approx$ 16.7%	d12
не меньше 12	1/12 $\approx$ 8.3%	1/36 $\approx$ 2.8%	d12

Симуляция подтверждает таблицу. При пороге 10 d12 дал 25.35% против точных 25.00%, 2d6 — 16.56% против 16.67%. При пороге 7 d12 дал 49.79% против 50.00%, 2d6 — 58.04% против 58.33%.



Вероятности результатов и шанс успеха при пороге 10

Причина в числе пар: к сумме 7 ведут 6 пар, а к сумме 2 или 12 — по одной. Поэтому два кубика концентрируют шансы в середине. При высоком пороге это невыгодно, при пороге около середины — выгодно. Одинаковый максимум (12) не означает одинаковую сложность испытания.

Как выбирать кубик

Посчитайте успешные исходы: для d12 — число граней, для 2d6 — число пар. Затем разделите на общее число равновозможных исходов: 12 для d12, 36 для 2d6.

Зависимость, которую не видно по одному кубику

Пусть первый кубик бросается, а второй выставляется на то же число. Заметим ли мы подмену, если смотреть только на второй кубик?

```
repeats = 12000
first = rng.integers(1,7,repeats)
independent = rng.integers(1,7,repeats)
copied = first.copy()
for second, name in [(independent, 'Независимые'), (copied, 'Второй копирует первый')]:
    print(name, ': частота суммы 7 =', np.mean(first+second==7))
```

В независимой модели сумма 7 встречается примерно в 16% случаев, во второй модели — ни разу (0.0).

Частоты граней второго кубика в обеих моделях близки к $\frac{1}{6}$, поэтому по ним подмену не заметить. Различие видно в таблице пар: независимые кубики заполняют всю сетку, а копирующая модель — только диагональ (1,1), (2,2), ..., (6,6). Там сумма всегда равна 2х и чётна, поэтому сумма 7 не появляется.

Что меняет зависимость

Каждый кубик по отдельности честный, но вместе они задают другое пространство исходов. Чтобы описать пару, нужно знать не только каждый кубик, но и связь между их результатами.

Тренажёр: кубики

1. A — «чётное», B — «не меньше 4». Найдите $P(A \cap B)$ и $P(A \cup B)$. Почему нельзя просто сложить $P(A)$ и $P(B)$?

Решение. $A \cap B = \{4, 6\}$, $P = \frac{2}{6} = \frac{1}{3}$. $A \cup B = \{2, 4, 5, 6\}$, $P = \frac{4}{6} = \frac{2}{3}$. Сумма

$P(A) + P(B) = \frac{3}{6} + \frac{3}{6} = 1$ считает 4 и 6 дважды, поэтому пересечение вычитают:
 $\frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6}$.

1. У двух независимых кубиков что вероятнее: хотя бы одна шестёрка или ровно одна?

Решение. Хотя бы одна шестёрка: $\frac{11}{36} \approx 30.6\%$. Ровно одна: $\frac{10}{36} \approx 27.8\%$. Событие «хотя бы одна» включает «ровно одну» и ещё пару (6,6). Симуляция дала 0.306 и 0.278.

1. Замените копирование на `7 - first` и подпишите вторую модель «Дополняет до семи». Предскажите вид таблицы пар и частоту суммы.

Решение. Второй кубик всегда равен 7 – first, поэтому пары лежат на побочной диагонали: (1,6), (2,5), ..., (6,1). Сумма всегда равна 7, её частота равна 1. Частоты граней второго кубика при этом остаются близкими к $\frac{1}{6}$.

Пороги 7 и 12 для выбора между d12 и 2d6 разобраны в таблице выше.

ГЛАВА 14

Плейлист: повтор рядом, альбомы и итоги

01:58:54 – 02:06:39

Процедура выбора определяет повторы и шансы песен. Симуляции подтверждают расчёт для повтора рядом. В конце предлагаются вопросы и задания для самостоятельной работы.

Повтор рядом и повтор вообще

Пусть из десяти песен независимо выбирают десять раз. Тогда «любой повтор» и «повтор рядом» — разные события.

- **Любой повтор:** какая-либо песня встречается дважды. Его не будет, только если все десять выборов различны. Шанс этого равен $\frac{10!}{10^{10}} \approx 0.0004$, поэтому повтор почти наверняка есть.
- **Повтор рядом:** две соседние позиции совпадают. На каждой паре шанс совпадения $\frac{1}{10}$, а без совпадения следующий выбор проходит с шансом $\frac{9}{10}$. Для девяти переходов получаем $1 - (9/10)^9 \approx 0.61$.

Уточнение

Повтор почти наверняка (99,96%), но не гарантирован: все десять выборов различны с шансом около 0,04%.

Три правила выбора

Правило выбора	Повтор рядом	Любой повтор
Любая песня, независимо	возможен	возможен
Случайная перестановка	нет	нет
Без соседнего повтора	нет	возможен

Перестановка использует каждую песню не больше одного раза, поэтому повторов нет вовсе. Запрет соседнего повтора не мешает вернуться к песне позже. Его можно реализовать так:

```

k, n = 10, 10
out = [int(rng.integers(k))]
for _ in range(n - 1):
    prev = out[-1]
    j = int(rng.integers(k - 1)) # выбираем из k-1 песен, без предыдущей
    out.append(j + (j >= prev)) # сдвиг индекса пропускает предыдущую песню

```



Примеры плейлистов: любая песня, перестановка, без соседнего повтора

20 выборов из 10 песен

Проверим формулу симуляцией:

```

K, n, M = 10, 20, 30_000
songs = rng.integers(0, K, (M, n))
repeat = (songs[:, 1:] == songs[:, :-1]).any(axis=1)
exact_repeat = 1 - ((K - 1) / K) ** (n - 1)

```

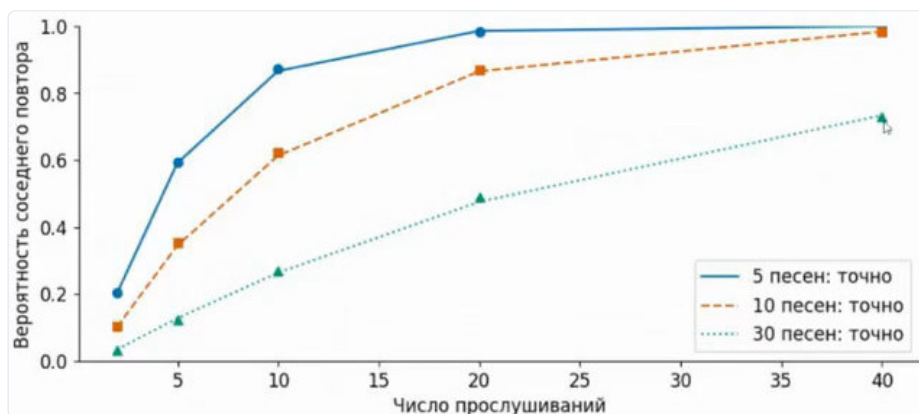
Симуляция дала 0.865, формула $1 - (9/10)^{19} \approx 0.8649$. Чтобы повтора рядом не было, каждый из 19 переходов должен пройти с шансом $\frac{9}{10}$. Искомый шанс — дополнение к этому произведению.

От числа прослушиваний и размера каталога

При n прослушиваниях и K песнях шанс соседнего повтора равен $1 - \left(1 - \frac{1}{K}\right)^{n-1}$.

Прослушиваний	5 песен	10 песен	30 песен
5	0.590	0.344	0.127
20	0.986	0.865	0.475
40	1.000	0.984	0.733

Чем длиннее плейлист, тем больше мест для совпадения. Чем больше каталог, тем меньше шанс совпасть на каждом шаге. Эти выводы относятся к независимому выбору с возвращением.



Вероятность соседнего повтора по числу прослушиваний для 5, 10 и 30 песен

Альбомы: процедура задаёт шансы

В библиотеке два альбома: А из двух песен и В из восьми. Сравним два способа выбора.

Способ	Шанс, что звучит песня из А	Шанс отдельной песни А
Сначала альбом $\frac{1}{2}$, затем песня	$\frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$
Одна из всех 10 песен	$\frac{2}{10}$	$\frac{1}{10}$

В первом способе альбом получает половину выбора независимо от размера. Поэтому песня короткого альбома получает больший вес. Во втором способе все песни равноправны.

Если сделать альбомы по 5 песен, то в первом способе каждая песня получит $\frac{1}{2} \cdot \frac{1}{5} = \frac{1}{10}$, как и во втором. Разница исчезает. Слова «выберем случайно» недостаточно: нужно описать процедуру.

Ответы на вопросы главы

1. Равные вероятности не гарантируют равных частот в короткой серии.

Короткая серия содержит случайный разброс, а частоты приближаются к $\frac{1}{6}$ только с ростом числа бросков.

2. Суммы двух честных кубиков не равновозможны. Сумма 7 получается шестью парами, а сумма 2 — только парой (1, 1): шансы $\frac{6}{36}$ и $\frac{1}{36}$.
3. Запрет соседнего повтора убирает повтор рядом по построению: выбор идёт из песен, кроме предыдущей. Перестановка убирает любой повтор за один проход.

Задания для самостоятельной работы

Перед запуском запишите ожидаемое направление изменения.

- **Длина серии 3 вместо 5.** Ответ: доля историй с серией вырастет. Короткая серия может начаться в большем числе мест.
- **Кубик с неравными вероятностями.** Ответ: частоты сойдутся к заданным весам, а не к $\frac{1}{6}$.
- **Плейлисты из 10 и 100 песен одинаковой длины (20 прослушиваний).** Ответ: для 10 песен шанс соседнего повтора около 0.865, для 100 — около $1 - 0.99^{19} \approx 0.17$. Большой каталог даёт меньше повторов.
- **Дополнительно.** Какой шанс, что звучит песня альбома А? Ответ: $\frac{1}{2}$ при выборе альбома, затем песни, и $\frac{1}{5}$ при выборе из всех десяти.

Шпаргалка

Базовые понятия

- Исход — результат одного опыта; Ω — все исходы.
- Событие A — набор исходов, например $A = \{O\}$.
- Для конечного набора равновозможных исходов: $P(A) = \frac{|A|}{|\Omega|}$.
- Дополнение: $P(A^c) = 1 - P(A)$.
- Сложение: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- Независимые броски: вероятность последовательности — произведение вероятностей бросков.

Формулы повторов и серий

- Ровно k орлов в n честных независимых бросках: $\binom{n}{k}/2^n$.
- Серия из 5 орлов в фиксированных местах: $(1/2)^5 = \frac{1}{32}$.
- Нет соседнего повтора в 20 выборах из 10: $(9/10)^{19}$; хотя бы один повтор: $1 - (9/10)^{19} \approx 0.865$.
- Доля орлов: $H/n = \frac{1}{2} + \frac{H - T}{2n}$; перевес $H - T = 2H - n$.

Частота, статистика, модель

- Вероятность: модель известна, ищем данные.
- Статистика: данные известны, ищем модель.
- Частотный подход — доля успехов в длинной серии.
- Байесовский подход — уверенность с учётом информации.
- Одна частота не доказывает точное значение вероятности.

Задачи занятия

Задача	Ответ
Ферма: счёт 2:1 до трёх побед, банк 100	75 и 25
10 орлов подряд	1/1024
Ровно 1 орёл в 3 бросках	3/8
Хотя бы 1 орёл в 3 бросках	7/8
Кубик: A чётное, $B \geq 4$	$P(A \cap B) = 2/6$, $P(A \cup B) = 4/6$
2d6: сумма ≥ 7 / d12: ≥ 7	21/36 / 6/12
2d6: сумма ≥ 10 / d12: ≥ 10	6/36 / 3/12
Альбомы из 2 и 8 песен: отдельная песня А / отдельная песня В при равновероятном выборе альбома, затем песни внутри него	1/4 / 1/16
Пять орлов где-нибудь в 100 бросках	около 81%

Программа и воспроизводимость

- `rng = np.random.default_rng(SEED)` — один генератор на эксперимент.
- `sliding_window_view(x, length, axis=1).all(axis=-1).any(axis=1)` — поиск серии.
- Переключения: `np.count_nonzero(x[1:] != x[:-1])`.
- Seed фиксируют, чтобы результат повторялся у других.

Правила выбора модели

- Сначала назвать процедуру, событие и предположения.

- Порядок результатов важен.
- Большое число повторов уточняет оценку внутри модели, но не исправляет модель.

Ключевые цитаты

«Вероятность идёт от того, что у нас есть какая-то модель мира, и она может порождать какие-то данные. Статистика идёт в обратную сторону. У нас есть данные, и мы по ним пытаемся понять, из какого распределения, из какой модели они были получены.»

17:34 Это различие задаёт, какие вопросы решает вероятность, а какие статистика.

«Частота 14 на 20, то есть 0,7, ещё не доказывает, что шанс орла равен 0,7.»

18:29 Одна наблюждённая частота не равна вероятности, поэтому нужен учёт случайного разброса.

«Когда нас что-то удивляет в жизни, сразу стоит себя спрашивать: а сколько было возможностей его заметить?»

40:36 Этот вопрос помогает отличить редкое совпадение от закономерности при большом числе попыток.

«Почему надо всегда сохранять сид? Потому что если его не сохранили, это не значит, что его не было. Он есть, просто вы его не знаете, и вы не сможете воспроизвести результаты.»

01:07:18 Воспроизводимость расчёта требует фиксации seed.

«Миллион опытов не исправит неправильные правила.»

01:08:14 Большая выборка уточняет оценку внутри модели, но не проверяет саму модель.

Глоссарий

Исход

Результат одного опыта, например выпадение орла или решки при одном броске монеты. Все возможные исходы образуют пространство исходов Ω .

Событие

Набор исходов, которые нас интересуют, например $A = \{O\}$ — «выпал орёл». Событие может объединять несколько исходов.

Вероятность для равновозможных исходов

Для конечного набора равновозможных исходов $P(A) = \frac{|A|}{|\Omega|}$: число подходящих исходов, делённое на число всех исходов.

Дополнение события

Событие A^c состоит из всех исходов, не входящих в A . Его вероятность $P(A^c) = 1 - P(A)$; этим удобно пользоваться для фраз «хотя бы один».

Пересечение и объединение

$A \cap B$ — выполнены оба условия, $A \cup B$ — выполнено хотя бы одно. Для них верно $P(A \cup B) = P(A) + P(B) - P(A \cap B)$: пересечение вычитается, потому что учтено дважды.

Независимые броски

Результат одного броска не зависит от прошлых. Для независимых бросков вероятность конкретной последовательности равна произведению вероятностей отдельных бросков, например $(1/2)^3 = 1/8$ для ОРО.

Случайная величина

Числовая характеристика исхода, например сумма очков двух кубиков. Её распределение задаётся вероятностями её значений.

Аксиоматика Колмогорова

Общий язык описания опытов: сначала задаются все возможные исходы, затем интересующее событие и вероятности. Так игры и наблюдения описываются одними правилами.

Частотный подход

Вероятность связывается с устойчивой долей успехов в длинной серии повторяемого опыта. Опирается на эмпирическое распределение данных.

Байесовский подход

Вероятность выражает уверенность с учётом всей доступной информации. Новые сведения могут изменить оценку.

Статистика

Задача, обратная вероятностной: по наблюдаемым данным пытаются понять, из какой модели или распределения они получены.

Закон больших чисел

При большом числе повторений доля успехов приближается к вероятности успеха. Например, доля орлов в честных бросках стремится к $1/2$, хотя перевес орлов над решками может оставаться большим.

Случайное блуждание

Процесс, в котором каждый бросок сдвигает положение на шаг: орёл — на $+1$, решка — на -1 . Разность орлов и решек показывает его текущее положение.

Серия

Подряд идущие одинаковые результаты, например пять орлов подряд. Чувствительность к длинным сериям влияет на наше восприятие случайности.

Переключение

Смена результата между двумя соседними бросками: орёл сменился решкой или наоборот.

Дерево исходов

Схема, в которой каждый бросок добавляет развилку. Для трёх бросков получаем восемь путей от корня до конца дерева.

Рандомизация эксперимента

Случайное назначение условий опыта, например перемешивание сортов пшеницы по участкам поля. Помогает честнее сравнивать группы и уменьшает влияние скрытых факторов.

АБ-тестирование

Сравнение двух условий, А и Б, в случайно распределённых группах. Основано на той же логике рандомизированного эксперимента.

Seed

Зерно генератора, задающее его начальное состояние. При одном и том же алгоритме и seed генератор выдаёт одну и ту же последовательность, что нужно для воспроизводимости расчётов.

Генератор псевдослучайных чисел

Алгоритм, который детерминированно выдаёт последовательность чисел, похожую на случайную. Его качество нельзя оценить по нескольким значениям: нужны изученные свойства алгоритма.

Блез Паскаль (1623-1662)

Французский математик. Вместе с Пьером Ферма в 1654 году обсуждал задачу о справедливом разделе ставки в прерванной игре.

Пьер Ферма (1607-1665)

Французский математик, партнёр Паскаля в переписке об азартных играх и разделе банка в 1654 году.

Джон Граунт (1620-1674)

Английский исследователь, который в 1662 году опубликовал анализ лондонских записей о смертности, сравнивая сведения в таблицах.

Карл Фридрих Гаусс (1777-1855)

Математик, в начале XIX века разрабатывал методы объединения неточных наблюдений, например в астрономии.

Рональд Фишер (1890-1962)

Статистик, работавший в Ротамстеде и развивавший планирование экспериментов, в том числе идею случайного назначения условий.

Андрей Колмогоров (1903-1987)

Математик, в 1933 году изложивший аксиоматические основания теории вероятностей.

Дональд Кнут (род. 1938)

Автор многотомного «Искусства программирования», где описаны генераторы псевдослучайных чисел и алгоритм К.

Вопросы для самопроверки

1. **Что такое пространство исходов и событие? Приведите пример для одного броска монеты.**

ОТВЕТ. Пространство исходов Ω — все возможные результаты опыта. Для монеты $\Omega = \{O, P\}$. Событие — подмножество исходов, например $A = \{O\}$ означает «выпал орёл».

2. **В чём разница между вероятностной задачей и статистической задачей?**

ОТВЕТ. В вероятностной задаче модель известна, и мы считаем, какие данные она даст, например, какие результаты дадут 20 бросков честной монеты. В статистической задаче данные известны, и мы пытаемся понять, какая модель их породила.

3. **Почему 14 орлов из 20 бросков не доказывают, что шанс орла равен 0.7?**

ОТВЕТ. Частота $14/20 = 0.7$ — результат случайного опыта, и её разброс может быть большим. Поэтому из одной частоты нельзя сделать вывод о точном значении вероятности, нужен учёт случайного разброса.

4. **В чём отличие частотного и байесовского взгляда на вероятность?**

ОТВЕТ. Частотный взгляд связывает вероятность с долей успехов в длинной серии повторений. Байесовский взгляд выражает уверенность с учётом доступной информации, и новые сведения могут изменить оценку.

5. **Игра до трёх побед прервана при счёте 2:1, банк 100. Как разделить банк по шансам победить?**

ОТВЕТ. Если игра продолжилась бы, А выигрывает, набрав три победы, в трёх из четырёх равновозможных продолжений: AA, AB, BA. Поэтому А получает 75, Б — 25, то есть $\frac{3}{4} \cdot 100$ и $\frac{1}{4} \cdot 100$.

6. **Чему равна вероятность выпадения десяти орлов подряд у честной монеты и что она доказывает?**

ОТВЕТ. Для заранее выбранных десяти независимых бросков $P = (1/2)^{10} = \frac{1}{1024}$. Эта серия редка, но возможна, и сама по себе не доказывает, что монета нечестная.

7. **Сколькими способами три броска монеты дают ровно один орёл, и чему равна вероятность? Чему равна вероятность хотя бы одного орла?**

ОТВЕТ. Подходят ОРР, РОР и РРО, то есть 3 из 8 равновозможных исходов: $P = \frac{3}{8}$. Хотя бы один орёл удобно считать через дополнение: $1 - \frac{1}{8} = \frac{7}{8}$.

8. **Почему шансы заранее выбранной строки из шести орлов и заранее выбранной конкретной смешанной строки одинаковы, но выглядят по-разному?**

ОТВЕТ. Шанс каждой конкретной строки равен $\frac{1}{64}$, поэтому выбранные заранее строки равновероятны. Но шесть орлов — один заметный паттерн, тогда как подходящих смешанных строк очень много, и мы не выделяем конкретную смешанную строку как особую.

9. **Для честного кубика чему равны $P(A \cap B)$ и $P(A \cup B)$, если А — «чётное», В — «не меньше четырёх»?**

ОТВЕТ. $A \cap B = \{4, 6\}$, поэтому $P(A \cap B) = \frac{2}{6}$. $P(A) = \frac{3}{6}$, $P(B) = \frac{3}{6}$, значит $P(A \cup B) = \frac{3}{6} + \frac{3}{6} - \frac{2}{6} = \frac{4}{6}$. Складывать $P(A)$ и $P(B)$ нельзя, потому что грани 4 и 6 учтены дважды.

10. **Сравните шансы успеха при пороге не меньше 10 для d12 и для суммы 2d6. Какой кубик выгоднее и почему?**

ОТВЕТ. У d12 шанс $\frac{3}{12}$, у 2d6 — $\frac{6}{36}$, поэтому выгоднее d12. Два кубика чаще дают середину суммы, и при высоком пороге этот выигрыш не помогает. При пороге не меньше 7 выгоднее 2d6: $\frac{21}{36}$ против $\frac{6}{12}$.

11. **Почему у выбора «сначала альбом, потом песня» другие шансы для песен, чем у выбора из всей библиотеки?**

ОТВЕТ. Песня короткого альбома (из двух) получает шанс $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, песня длинного (из восьми) — $\frac{1}{2} \cdot \frac{1}{8} = \frac{1}{16}$. При выборе из всех десяти песен каждая получает $\frac{1}{10}$. Процедура задаёт шансы, поэтому её нужно описать заранее.

12. **Почему в 20 выборах из 10 песен вероятность хотя бы одного соседнего повтора около 86.5%?**

ОТВЕТ. Без соседнего повтора каждый следующий выбор должен отличаться от предыдущего: 9 вариантов из 10. Вероятность отсутствия повтора равна $(9/10)^{19}$, поэтому повтор есть с вероятностью $1 - (9/10)^{19} \approx 0.865$.

13. Почему seed нужно сохранять и почему одна фиксация в начале важнее повторной установки перед каждым бросанием?

ОТВЕТ. Seed задаёт состояние генератора, поэтому тот же алгоритм с тем же seed даёт ту же последовательность, и результаты можно воспроизвести. Если задавать тот же seed перед каждым броском, будем повторять первый результат, а не получать новую последовательность.

14. Почему вероятность пяти орлов в первых пяти бросках ($1/32$) и хотя бы одной серии из пяти орлов среди ста бросков (около 81%) отличаются так сильно?

ОТВЕТ. В первом случае место серии задано заранее, поэтому вероятность $\frac{1}{32}$. Во втором случае подходит любое из 96 окон длины пять, поэтому шанс появления серии где-нибудь в истории гораздо выше. Окна пересекаются, и их вероятности нельзя просто сложить.

15. Что такое случайное блуждание в задаче с монетой и чем доля орлов отличается от перевеса орлов над решками?

ОТВЕТ. Орёл даёт шаг $+1$, решка — шаг -1 , и положение после n бросков равно разности орлов и решек. Доля орлов H/n может приближаться к $1/2$, а перевес $H - T$ при этом может оставаться заметным, например $+144$ после 10000 бросков.

Что изучить дополнительно

Байесовский подход к вероятности на практике

Лектор сказал, что он требует много информации и знания процесса генерации данных, и не раскрыл, как его применять.

Генераторы псевдослучайных чисел и алгоритм К

Лектор показал, что сложный алгоритм может заикнуться, и отправил к «Искусству программирования», том 2, §3.1.

Доверительные интервалы

Лектор упомянул, что при малом числе наблюдений интервал оценки вероятности очень широкий, но не объяснил его построение.

Законы больших чисел и случайные блуждания

Лектор сказал, что теоремы описывают поведение на больших числах, и связал блуждания с физикой, но не разобрал доказательства.

Парадоксы теории вероятностей

Лектор посоветовал статью «Что не так? Три парадокса теории вероятностей» о важности процедуры выбора.

Упоминания

Блез Паскаль (человек) — Вместе с Пьером Ферма обсуждал в 1654 году задачу о разделе ставки в прерванной игре.

Пьер Ферма (человек) — Партнёр Паскаля по переписке о задачах азартных игр, пример с игрой до трёх побед.

Джон Граунт (человек) — В 1662 году опубликовал исследование лондонских записей о смертности.

Карл Фридрих Гаусс (человек) — Разрабатывал методы объединения неточных наблюдений в начале XIX века, в том числе в астрономии.

Рональд Фишер (человек) — Работая в Ротамстеде, развивал планирование экспериментов и случайное назначение условий.

Андрей Колмогоров (человек) — В 1933 году изложил аксиоматические основания вероятности.

Дональд Кнут (человек) — Автор «Искусства программирования», где рассматриваются генераторы случайных чисел и алгоритм К.

Искусство программирования, том 2 (книга) — Дональд Кнут, §3.1: история генераторов псевдослучайных чисел и алгоритм К, рекомендовано для самостоятельного чтения.

Розенкранц и Гильденстерн мертвы (книга) — Пьеса Тома Стоппарда, в которой монета снова и снова падает орлом; использована как иллюстрация к отношению людей к случайности.

Что не так? Три парадокса теории вероятностей (книга) — Рекомендованная для чтения статья, где важна процедура выбора.

Важнейшая модель теории вероятностей (книга) — Рекомендованная статья о случайных блужданиях и связях с физикой.

Ротамстед (другое) — Место работы Рональда Фишера, где развивалось планирование экспериментов.

NumPy (инструмент) — Библиотека для симуляций: генерация случайных чисел и вычисления на массивах.

matplotlib (инструмент) — Библиотека для построения графиков гистограмм, линий и тепловых карт на семинаре.

Google Colab (инструмент) — Среда, в которой выполняются семинарские примеры с кодом.

WinRAR, zip (инструмент) — Архиваторы как пример сжатия информации через поиск избыточности, по аналогии с человеческим восприятием паттернов.

Хабр (ссылка) — Ресурс, где лектор обещал популярные статьи о генераторах случайных чисел.

Карточки для повторения

#	ВОПРОС	ОТВЕТ
1	Для чего в начале лекции нужна теория вероятностей, если говорить о принятии решений?	Чтобы принимать решения в условиях неопределённости и понимать, какие шансы у возможных исходов.
2	Как предлагали разделить банк прерванной игры, если игрок А выиграл два раунда, а игрок Б один?	По шансам на победу в гипотетическом продолжении игры: в примере оставшиеся варианты дают 75 для А и 25 для Б. Так учитываются будущие возможности, а не только текущий счёт.
3	Почему при делёже по шансам на победу учитывают не только текущий счёт, но и возможное продолжение игры?	Потому что честный дележ должен отражать, кто с какой вероятностью выиграл бы, если бы игру доиграли. Текущий счёт этого не показывает, а раздел 50 на 50 признаёт игру несостоявшейся.
4	Какие два вопроса возникали у учёных, когда данные накапливались?	Как эти данные получились и почему они такие, то есть какой механизм их порождает. Из этих вопросов выросли анализ данных и теория их порождения.
5	Что сделал Граунт в XVII веке с записями о смертности?	Опубликовал исследование, в котором объединял разрозненные записи в таблицы и сравнивал их между собой, чтобы делать выводы.
6	Кто в начале XIX века разрабатывал методы объединения неточных наблюдений?	Лежандр и Гаусс. Такие задачи возникали, например, в астрономии.
7	Что даёт усреднение повторных измерений и чего оно не исправляет?	Усреднение уменьшает случайные колебания. Систематическую ошибку, например дефект самой линейки, одно усреднение не убирает.
8	Зачем в полевом опыте случайно распределять условия по участкам?	Случайное назначение не даёт расположению участков и свойствам почвы систематически совпасть с одним условием. Так эффект условия честнее отделяется от других факторов. Этот подход предложил Фишер, работавший в Ротамстеде.
9	Что такое аксиоматика Колмогорова в том порядке, в каком её формулируют?	Сначала задаются все возможные исходы, затем указывается событие, которое интересует, и задаются вероятности.

#	ВОПРОС	ОТВЕТ
10	Чем вероятность отличается от статистики по направлению вывода?	Вероятность идёт от модели мира к возможным данным: модель задана, и мы вычисляем, какие результаты вероятны. Статистика идёт обратно: по наблюдаемым данным пытаемся понять, какая модель их породила.
11	Почему 14 орлов из 20 бросков не доказывает, что шанс орла равен 0,7?	Частота 0,7 в одной серии может быть результатом случайного разброса. Из наблюдаемой частоты нельзя напрямую вычислить вероятность.
12	В чём идея частотного подхода к вероятности?	Опыт повторяют много раз, смотрят доли исходов и связывают вероятность с этими долями. Вероятность опирается на повторяемость эксперимента.
13	Что выражает байесовский подход к вероятности?	Степень уверенности в исходе с учётом всей доступной информации. Новые данные могут изменить эту оценку.
14	Почему в А/В-тестировании чаще используют частотный подход, а не байесовский?	Частотный подход опирается на повторяемость экспериментов и хорошо работает при большом объёме данных. Байесовский требует учесть связи между признаками и физику процесса, что на практике сложно.
15	Что означает слово «случайно» в вероятностных задачах?	Либо механизм опыта, при котором никто не влиял на результат, либо то, что наблюдатель не знает результата. Полезно уточнять, какой смысл имеется в виду.
16	Почему редкость серии из 10 орлов подряд не доказывает, что монета нечестная?	Такая серия маловероятна, но возможна: шанс около $1/1024$. Чтобы делать вывод о монете, нужно провести гораздо больше бросков, а доверительный интервал по текущим данным остаётся широким.
17	Какова вероятность ровно одного орла при трёх честных бросках и как её получить?	$3/8$. Из восьми равновозможных исходов три содержат ровно один орёл.
18	Как вычислить вероятность хотя бы одного орла при трёх бросках через дополнение?	Сначала найти вероятность противоположного события «ни одного орла»: $P = 1/8$. Тогда $P(\text{хотя бы один орёл}) = 1 - 1/8 = 7/8$.

#	ВОПРОС	ОТВЕТ
19	Что такое дерево исходов?	Схема, в которой каждый бросок добавляет ветвь с орлом или решкой. Путь от корня до конца дерева задаёт один исход, и все такие пути равновероятны при честных бросках.
20	Почему строки с одинаковыми шансами кажутся нам разными?	Мы замечаем паттерны и воспринимаем длинные серии или регулярные чередования как неслучайные, хотя каждая конкретная строка имеет одинаковую вероятность.
21	Почему нельзя складывать шансы пересекающихся окон серии, например окон 1-5 и 2-6?	Окна используют общие броски, и один и тот же исход был бы засчитан несколько раз. Поэтому вероятность хотя бы одной серии нужно считать по всей последовательности целиком.
22	Что такое случайность с точки зрения сжатия и непредсказуемости?	Строка случайна, если её нельзя описать значительно более короткой программой или правилом. Чем предсказуемее строка, тем лучше её можно сжать и запомнить.
23	Какова вероятность выпадения одной конкретной грани честного шестигранного кубика?	1/6, потому что шесть граней равновозможны.
24	Как считается вероятность дополнения события?	$P(\text{не } A) = 1 - P(A)$. Дополнение содержит все исходы, которые не входят в A.
25	Как считается вероятность объединения двух событий с пересечением?	$P(A \text{ или } B) = P(A) + P(B) - P(A \text{ и } B)$. Пересечение вычитается, чтобы не посчитать общие исходы дважды.
26	Что происходит с пространством исходов, если второй кубик всегда выставляют равным первому?	Броски перестают быть независимыми. Сумма всегда чётная, поэтому сумма 7 исчезает, а случайная величина меняется.
27	Почему выбор сначала альбома, затем песни, даёт другие шансы, чем выбор из общей библиотеки?	Вероятности выбора перемножаются, поэтому песни короткого альбома получают больший вес. Из общей библиотеки каждая песня получает одинаковый шанс.
28	Какие шансы имеют песни в примере с альбомами из двух и восьми песен при выборе альбома с вероятностью 1/2?	Каждая песня короткого альбома получает $1/2 \times 1/2 = 1/4$, каждая песня длинного получает $1/2 \times 1/8 = 1/16$. При выборе из общей библиотеки каждая из десяти получает 1/10.

#	ВОПРОС	ОТВЕТ
29	Что такое псевдослучайный генератор?	Детерминированный алгоритм, который выдаёт последовательность чисел, похожую на случайную. По виду последовательности качество генератора не определить, его свойства изучают аналитически.
30	Почему одинаковой доли орлов недостаточно, чтобы описать последовательность бросков?	Порядок результатов несёт информацию. Строго чередующаяся строка и нерегулярная строка с тем же числом орлов и решек различаются по структуре.
31	Зачем фиксировать seed в случайных симуляциях?	Seed задаёт одну и ту же последовательность случайных чисел, поэтому результаты воспроизводятся у вас и у другого человека.
32	Почему миллион опытов не исправит ошибочную модель?	Повторения уточняют оценку внутри заданной модели, но не проверяют её предположения. Например, связанные кубики дадут только чётные суммы при любом числе бросков, поэтому модель нужно проверять отдельно.
33	Какие цели у семинара по курсу?	Сформировать базовую интуицию и умение проводить симуляции, которые нужны в курсах математической статистики и A/B-тестированию.
34	Что показывает признак «число переключений» в строке монеты?	Насколько строка типична для честной монеты. Число переключений сравнивают с распределением для нескольких тысяч случайно сгенерированных строк той же длины.
35	Как задаётся случайное блуждание в семинаре?	Орёл означает шаг вправо на +1, решка означает шаг влево на -1. Баланс накапливается за все броски, и его значение показывает перевес орлов над решками.
36	Почему небольшое преимущество шанса даёт устойчивый перевес на длинной дистанции?	При шансе 0,52 на коротких сессиях выигрыш незаметен, но при многократном повторении сессий перевес становится значительным и накапливается в пользу того, у кого шанс больше.
37	Что происходит с долей орлов и перевесом при шансе орла 0,55?	Доля орлов сходится к 0,55, а перевес растёт почти линейно с числом бросков.
38	Как проверяется вероятность ровно 10 орлов в 20 бросках?	Точно $C(20,10)/2^{20} \approx 0,176$. В повторённых экспериментах доля получилась около 0,178, поэтому расчёт и симуляция совпадают.

#	ВОПРОС	ОТВЕТ
39	Почему пять орлов где-нибудь среди 100 бросков вероятнее, чем пять орлов в первых пяти?	Для первых пяти бросков шанс $1/32$, а серия может появиться в любом из многих окон. В симуляции шанс серии где-нибудь среди 100 бросков около 0,81.
40	Что вероятнее при трёх бросках монеты: хотя бы один орёл или ровно три орла?	Хотя бы один орёл, с вероятностью $7/8$, против $1/8$ для трёх орлов.
41	Почему при коротких сериях бросков частоты граней кубика заметно отличаются от равных?	Случайный разброс при малом числе бросков велик. При 60 бросках частоты заметно различаются, а при 6000 бросках они близки к $1/6$.
42	Почему при сумме двух кубиков чаще всего выпадает 7?	Этой сумме соответствует больше пар исходов: шесть из 36 равновозможных, тогда как сумма 2 или 12 получается только одной парой.
43	Какой кубик выгоднее при пороге 7 и при пороге 10?	При пороге 7 выгоднее $2d6$: $21/36$ против $6/12$. При пороге 10 выгоднее $d12$: $3/12$ против $6/36$.
44	Какова вероятность хотя бы одного повтора соседней песни в плейлисте из 20 песен при случайном выборе из 10?	Около 86 процентов. Вероятность отсутствия повтора на одном шаге равна $9/10$, и для 19 переходов получается $1 - (9/10)^{19} \approx 0,865$.
45	Почему случайная перестановка исключает повтор песен, а выбор с возвращением допускает его?	Перестановка использует каждую песню не больше одного раза. При выборе с возвращением каждый шаг заново выбирает из всего каталога, поэтому песня может повториться.
46	Что означает запрет соседнего повтора в плейлисте?	Одна и та же песня не может стоять сразу после себя, но может вернуться позже в плейлисте.