Математика для Data Science. Теория вероятностей.
Условия задач

Содержание

Замечание. Таким цветом отмечены ссылки на страницы внутри этого файла.

Вероятностное пространство

Задача 1

В этой задаче нет единственного правильного решения. Придумайте какое-нибудь :)

В каждой из этих ситуаций придумайте множество элементарных исходов и их вероятности. Придумайте какое-нибудь событие и найдите его вероятность.

1. У вас дома есть ёлка, украшенная двумя шариками, и кот. Вы выходите из дома и возвращаетесь через полчаса.
2. Вы играете в рулетку. Шарик будет катиться по колесу, пока не остановится в одном из 37-ми секторов: $0, 1, \ldots, 36$.
3. Вы играете в рулетку. Шарик будет катиться по колесу, пока не остановится в одном из 37-ми секторов: $0, 1, \ldots, 36$. За вашим столом также играет племянник владелицы казино, и он сделал ставку на сектор 25.

Задача 2

Определение. Пусть дано событие A. Тогда событием $\bar{A}$ называется событие, состоящие из всех элементарных исходов, которые не входят в A. Можно произносить $\bar{A}$ как "не A."

Докажите, что $P(\bar{A}) = 1 - P(A)$.

Равновероятные исходы

Задача 1

Пусть нам известно, что все исходы равновероятны. Докажите, что для любого события A выполнено $P(A) = \dfrac{|A|}{|\Omega|}$.

Можно переформулировать утверждение этой задачи так. Если все исходы равновероятны, то чтобы найти вероятность события A нужно

1. найти количество всех элементарных исходов,
2. найти количество элементарных исходов, принадлежащих A,
3. поделить результат второго пункта на результат первого пункта.

Задача 2

Мы бросаем честный шестигранный кубик два раза подряд и записываем результаты первого и второго броска. Тем самым, элементарным исходом будет пара цифр, где первая цифра – результат первого броска, а вторая – результат второго броска. Например, вот элементарный исход: $4, 3$ (на первом броске выпало 4, на втором 3). Можно считать, что вероятности всех элементарных исходов равны (мы докажем равновероятность исходов позже, когда будем обсуждать независимые события).

В этой задаче вероятностное пространство полезно представлять себе в виде таблицы. В таблице номер строки отвечает результату первого броска, номер столбца отвечает результату второго броска. Каждая клетка таблицы отвечает ровно одному исходу. Можно закрашивать исходы, которые относятся к вашему событию, и оставлять незакрашенными исходы, которые не относятся.

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

Опишите вероятностное пространство – множество Ω, алгебру событий F, и функцию P

Задача 3

Как и в задаче с предыдущего шага, мы бросаем честный шестигранный кубик два раза подряд и записываем результаты первого и второго броска. Можно считать, что вероятности всех элементарных исходов равны.

1. Какова вероятность, что сумма результатов бросков будет равна 5?

2. Какова вероятность, что сумма результатов бросков будет больше или равна 7?

3. Какова вероятность, что сумма результатов бросков будет чётной?

Задача 4

Пусть есть список из k объектов, и дано свойство, которым объект может обладать или не обладать. Например, объекты – батарейки, а свойство – наличие брака. Или объекты – клиенты, а свойство – совершил ли клиент хотя бы одну покупку в нашем приложении за последний месяц. Каждую из таких ситуаций можно описать строкой длины k из нулей и единиц. Место номер i в этой строке будет соответствовать i-ому объекту. На i-ом месте будет стоять единица, если i-ый объект обладает нужным свойством, и ноль в противном случае. Такая интерпретация очень пригодится нам, когда мы будем работать с повторяющимися испытаниями дальше в курсе.

Из-за этого (и по многим другим причинам) полезно рассматривать строки из нулей и единиц фиксированной длины. Более того, полезно рассматривать множество всех строк длины k и строить на нём вероятностное пространство. Это мы сделаем на следующем шаге. Для этого нам пригодится такая задача:

1. Докажите, что есть ровно 2^k разных строк длины k из нулей единиц.

2. Полезное следствие пункта 1. Докажите, что если в множестве X ровно k элементов, то множество всех подмножеств X состоит из 2^k элементов (это долг с конца шага предыдущего урока).

Задача 5

Давайте строить вероятностное пространство на множестве всех строк длины k, составленных из нулей и единиц.

Зафиксируем натуральное число k. Перед нами стоит шляпа с бумажками, на каждой бумажке написана строка длины k, составленная из нулей и единиц. Каждая возможная строка длины k встречается в шляпе ровно один раз. Например, если $k = 5$, то бумажка с надписью 10011 будет присутствовать в шляпе ровно в одном экземпляре.

1. Мы вслепую вытаскиваем из шляпы одну бумажку. Какова вероятность, что на вытянутой бумажке не будет ни одной единицы?

2. Какова вероятность, что на вытянутой бумажке будет ровно одна единица?

3. Какова вероятность, что на вытянутой бумажке будет хотя бы одна единица?

Условная вероятность

Задача 1

Это несложная задача. Мы добавили её затем, чтобы вы обсудили с преподавателем условную вероятность. И убедились, что вы поняли условную вероятность на нужном уровне строгости.

Мы подбросили монетку два раза. Тем самым, возможные исходы это: оо, ор, ро, рр (орёл это "о", решка это "р"). Будем считать все исходы равновероятными.

1. Известно, что выпал хотя бы один орёл. Какова вероятность того, что выпало два орла?

2. Известно, что выпал хотя бы один орёл. Какова вероятность того, что выпал сначала орёл, а затем решка?

3. Известно, что выпали два разных значения. Какова вероятность того, что выпал сначала орёл, а затем решка?

4. Известно, что выпало два орла. Какова вероятность того, что выпал хотя бы один орёл?

Задача 2

Мы сделали медицинский тест на наличие некоторой болезни тысяче человек. У 900 из них результат теста оказался отрицательным (тест говорит "здоров"), у 100 — положительным (тест говорит "болен").

Впоследствии выяснилось, что среди 900 людей с отрицательным результатом было 60 заболевших, а среди 100 людей с положительным результатом было 50 заболевших.

	Болен	**Здоров**
Отрицательный тест	60	$900 - 60 = 840$
Положительный тест	50	$100 - 50 = 50$

(здесь мы вписали в каждую клетку таблицы не вероятность, а соответствующее число людей)

1. Ещё один человек сдал тест и получил положительный результат. Какова вероятность того, что он болен?

2. Ещё один человек сдал тест и получил отрицательный результат. Какова вероятность того, что он болен?

3. Человек здоров, но пока не знает об этом. Он сдаёт тест. Какова вероятность, что тест по ошибке определит его как больного, то есть даст положительный результат? (такой результат называют ложноположительным)

4. Человек болен, но пока не знает об этом. Он сдаёт тест. Какова вероятность, что тест по ошибке определит его как здорового, то есть даст отрицательный результат? (такой результат называют ложноотрицательным)

Давайте считать, что все тестируемые люди живут в одном и том же городе примерно в одинаковых условиях, имеют схожее состояние здоровья, возраст и т.д (можете обсудить с преподавателем, почему это замечание стоило написать)

Независимые события

Задача 1

1. Докажите, что при $P(B) \neq 0$ эти утверждения эквивалентны: $P(A) = P(A|B)$ и $P(A \cap B) = P(A) \cdot P(B)$. Заметьте, что первая формула выглядит несимметричной по A и B, а во второй формуле роли A и B одинаковы.

2. Докажите, что при при $P(A) \neq 0, P(B) \neq 0$ эти утверждения эквивалентны: $P(A) = P(A|B)$ и $P(B) = P(B|A)$

Давайте словами опишем, что мы доказываем в каждом из пунктов.

В пункте 1 мы доказываем, что независимость A и B равносильна условию $P(A \cap B) = P(A) \cdot P(B)$.

В пункте 2 мы доказываем, что A независимо от B, если и только если B независимо от A. Поэтому в дальнейшем мы не будем писать, кто от кого независим, а просто будем говорить, что A и B независимы.

Обычно последнюю формулу из пункта 1 используют в определении независимости.

Определение. События A и B называются *независимыми*, если $P(A \cap B) = P(A) \cdot P(B)$.

Как видите, в таком определении уже не нужно требовать $P(A) \neq 0$ или $P(B) \neq 0$.

Задача 2

Докажите, что при $P(B) \notin \{0, 1\}$ эти утверждения эквивалентны: $P(A) = P(A|B)$ и $P(A) = P(A|\bar{B})$ (напоминание: символ $\bar{B}$ обозначает событие "не B")

Другими словами, мы доказываем, что независимость A и B равносильна независимости A и $\bar{B}$. На примере это можно понять так. Если событие "чёрная кошка перебежала вам дорогу" не меняет вероятность добраться до работы без происшествий, то и событие "чёрная кошка НЕ перебегала вам дорогу" не меняет вероятность добраться до работы без происшествий.

Задача 3

Мы бросаем честный кубик два раза подряд. Докажите, что вероятности всех возможных исходов равны. Это утверждение – наш долг с шага прошлого урока.

Задача 4

В колоде 52 карты. Есть 4 масти: пики, крести, бубны, черви. В каждой масти 13 различных номинаций: $2, 3, \ldots, 10$, валет, дама, король, туз.

1. Из колоды случайным образом вытягивается одна карта. Независимы ли события "вытянули валета" и "вытянули пику"?

2. Из колоды случайным образом вытягивается одна карта. Независимы ли события "вытянули валета" и "вытянули туз"?

Задача 5

В колоде 52 карты. Есть 4 масти: пики, крести, бубны, черви. В каждой масти 13 различных номинаций: $2, 3, \ldots, 10$, валет, дама, король, туз.

1. Из колоды случайным образом вытягивается одна карта, а затем ещё одна. Независимы ли события "первая карта это король" и "вторая карта это крести"?

2. Из колоды случайным образом вытягивается одна карта, а затем ещё одна. Независимы ли события "первая карта это король" и "вторая карта это десятка"?

Задача 6

Обозначение. Два события, которые не являются независимыми, называют зависимыми.

1. Вернёмся к задаче про арбузы. Являются ли зависимыми события "арбуз вкусный" и "у арбуза есть пятно"? Если вы покупаете арбуз и хотите, чтобы он был вкусным, стоит ли проверять наличие пятна?

2. Вернёмся к задаче про медицинский тест. Являются ли зависимыми события "положительный результат" и "человек болен"? Если эти события независимы, то стоит ли проводить такой тест?

Совместная независимость

Задача 1

В нашей картине мира разные броски монетки или кубика никак не связаны между собой.

Обоснуйте все ваши ответы строго – какие исходы в вероятностном пространстве, какие события, как считать условную вероятность, где и как использована совместная независимость.

1. Мы бросаем честную монетку два раза подряд. Какова вероятность исхода "два орла"?
2. Мы бросаем честную монетку четыре раза подряд. Какова вероятность исхода "решка-решка-орёл-решка"?
3. Мы бросаем честную монетку k раз подряд. Каждый исход – фиксированная последовательность орлов и решек. Докажите, что вероятность каждого исхода равна $\frac{1}{2^k}$

Задача 2

Приведите пример событий A_1, A_2, A_3, которые попарно независимы, но не совместно независимы. Другими словами, покажите, что для ваших событий первые три условия из конца предыдущего шага выполнены, а последнее условие – не выполнено.

Для описания вашего примера достаточно указать вероятности всех участков диаграммы Венна с прошлого шага (то есть мы не требуем интерпретации событий как каких-то конкретных событий из реального мира).

Задача 3

Задача о легкомысленном члене жюри.

Жюри состоит из трёх человек. Им нужно принять верное решение по какому-то вопросу. Первые два члена жюри независимо друг от друга принимают верное решение с вероятностью p. Третий для вынесения решения бросает монету, то есть его вероятность принять верное решение равна $\frac{1}{2}$. Окончательное решение жюри выносится большинством голосов.

С какой вероятностью это жюри выносит верное решение?

Задача 4

Это модификация задачи о легкомысленном члене жюри. В жюри заседают председатель, эксперт и Петя. Председатель принимает верное решение с вероятностью 0.7. Эксперт принимает верное решение с вероятностью 0.9. Председатель и эксперт принимают верные решения независимо. Петя не разбирается в обсуждаемом жюри вопросе.

Для каждой из перечисленных ниже стратегий Пети определите, с какой вероятностью комиссия примет верное решение.

1. Петя всегда копирует решение председателя
2. Петя всегда выдаёт решение, противоположное решению председателя
3. Петя всегда копирует решение эксперта
4. Петя подбрасывает честную монетку. Если выпал орёл – голосует "за", если выпала решка – "против".

Источник: журнал "Квантик", 2018 год

Задача 5

Точно такая же задача, как предыдущая, но с одним изменением – теперь председатель и эксперт голосуют <u>не независимо</u>.

- С вероятностью 0.7 эксперт и председатель одновременно приняли верное решение.

- С вероятностью 0.2 председатель принял неверное решение, а эксперт верное.
- С вероятностью 0.1 и председатель, и эксперт ошиблись.

Другими словами, невозможна ситуация, когда председатель принял верное решение, в то время как эксперт принял неверное решение. Заметьте, что как и в прошлой задаче, председатель принимает верное решение с вероятностью 0.7, а эксперт принимает верное решение с вероятностью $0.7 + 0.2 = 0.9$.

Для каждой из перечисленных ниже стратегий Пети определите, с какой вероятностью комиссия примет верное решение.

1. Петя всегда копирует решение председателя
2. Петя всегда выдаёт решение, противоположное решению председателя
3. Петя всегда копирует решение эксперта
4. Петя подбрасывает честную монетку. Если выпал орёл – голосует "за", если выпала решка – "против".

Задача 6

Ирине обещают приз, если она выиграет подряд по крайней мере две теннисные партии против своего тренера и клубного чемпиона. Всего играется три партии по одной из схем: тренер-чемпион-тренер или чемпион-тренер-чемпион. Чемпион играет лучше тренера. Какую схему лучше выбрать Ирине?

Задача 7

Нашим приложением пользуются 1000 человек. Будем считать, что каждый пользователь с вероятностью 0.25 уходит от вас в течение месяца. Люди прекращают пользоваться приложением независимо друг от друга.

1) Какова вероятность, что все 1000 прекратят им пользоваться в течение месяца?

2) В тех же условиях, что и в пункте 1. Какова вероятность, что никто не прекратит пользоваться в течение месяца?

Задача 8

Независимость ухода пользователей из условия предыдущей задачи выглядит странно. В реальной жизни, пусть вы знаете про всех пользователей, кроме пользователя по имени Саша, что большинство из них ушли. Интуиция подсказывает, что тогда и вероятность, что Саша уйдёт, велика. Аналогично, пусть вы знаете про всех пользователей, кроме Саши, что почти все из них не ушли. Тогда вероятность того, что Саша уйдет, низка.

Скорее всего, подсознательные цепочки рассуждений были следующими:

большинство пользователей уходит $\to$ приложение плохо работает $\to$ Саша тоже, вероятно, уйдёт;

большинство пользователей остаётся $\to$ приложение хорошо работает $\to$ Саша тоже, вероятно, останется.

Чтобы разобраться с этим затруднением, решите такую задачу.

К началу месяца вы закодили апдейт приложения, и первого числа месяца собираетесь выкатывать его на пользователей. Из прошлого опыта вы знаете, что половина ваших апдейтов удачные, а половина – неудачные. Известно, что если апдейт удачный, то пользователи будут уходить независимо друг от друга с вероятностью 0.1, а если неудачный – то с вероятностью 0.6. Являются ли уходы пользователей в текущем месяце независимыми событиями **до того**, как вы выкатили апдейт? Другими словами, если вы пока не знаете, удачен ли этот апдейт, то для вас уходы пользователей независимы?