

Школа лингвистики, 2021-22 уч. год**Теория вероятностей****Формула полной вероятности и формула Байеса (18.01.2022)***И. В. Щуров, Д. А. Филимонов, Р. Я. Будылин*

Задача 1. У меня в кармане три монетки. Две нормальные, а одна с двумя орлами. Я достал одну монетку наугад и подбросил её.

- Какова вероятность, что выпадет орёл?
- Допустим, что выпал орёл. Какова вероятность, что я достал «нормальную» монетку?
- Я подкинул ту же монетку ещё раз, и снова выпал орёл. Какова теперь вероятность, что я достал нормальную монетку?
- Я подкинул ту же монетку ещё раз, и выпала решка. Какова теперь вероятность, что я достал нормальную монетку?

Задача 2. Тест на ВИЧ (вирус иммунодефицита человека, причина СПИДа) выдаёт верный результат в 97% случаев. Известно, что ВИЧ заражено 1/5000 часть всего населения.

- С какой вероятностью случайно взятый человек заражён ВИЧ? Не заражен ВИЧ?
- С какой вероятностью случайно взятый здоровый человек получит положительный тест на ВИЧ?
- С какой вероятностью случайно взятый заражённый ВИЧ человек получит положительный тест?
- Какую долю из всех людей составляют здоровые люди, которым бы тест показал, что они заражены ВИЧ?
- Какую долю из всех людей составляют заражённые ВИЧ люди, которым бы тест показал, что они заражены ВИЧ?
- С какой вероятностью у случайно взятого человека тест на ВИЧ будет положительным?
- Человек получил положительный результат теста на ВИЧ. С какой вероятностью он действительно заражён ВИЧ?

Задача 3. В космосе летают метеориты двух сортов — тяжёлые и лёгкие, причём тяжёлые встречаются втрое чаще лёгких. Если в спутник GPS попадёт лёгкий метеорит, он с вероятностью 70% останется на орбите, и с вероятностью 30% упадёт в океан, а если тяжёлый — то он с вероятностью 90% упадёт в океан, а с вероятностью 10% останется на орбите.

- В спутник попал метеорит. С какой вероятностью он окажется на дне морском?
- Допустим, мы знаем из новостей, что спутник упал в океан из-за столкновения с каким-то метеоритом. С какой вероятностью это был лёгкий метеорит?

Задача 4. Для фильтрации спама используются так называемые *байесовские фильтры*. Простейший механизм работы такого фильтра выглядит следующим образом. Допустим, есть некоторое слово-маркер, характерное для спамерских писем. Если это слово встречается в сообщении, это увеличивает вероятность того, что письмо является спамерским. Для расчёта вероятности используется формула Байеса.

Предположим, что всего 10% от всех писем являются спамерскими, причём слово-маркер встречается в 95% спамерских писем и в 10% неспамерских. Если это слово встретилось в письме, какова вероятность того, что оно спамерское?

Задача 5. Один из 100 тысяч людей — террорист. Система Обнаружения Террористов, разработанная компания SecurityForever по заказу ФБР за 2 миллиарда долларов, автоматически распознает лица всех людей, проходящих паспортный контроль в аэропортах, и сверяет их с Базой Данных Всех Террористов. Система очень хорошая, и если человек действительно входит в Базу Данных Всех Террористов, она его идентифицирует как террориста с вероятностью 95%, а если человек не входит в Базу Данных, то она его идентифицирует как террориста с вероятностью 0.5%. Пассажир был идентифицирован системой как террорист — какова вероятность, что он действительно террорист?

Задача 6. На улице расположены подряд 2 светофора. Каждый устроен так, что зелёный свет горит в нём две трети времени. Автомобилист заметил, что когда он проезжает первый светофор на зелёный, то в 3 из 4 случаях и второй светофор его не задерживает. Допустим, что автомобилист проскочил на красный. Какова вероятность того, что и на втором светофоре будет красный?

Задача 7. В одном маленьком городке полиция разыскивает бродягу. Можно считать, что есть 4 шанса из 5, что он находится в одном из 8 баров городка, безразлично в каком - он не отдаёт предпочтения ни одному из них. Двое полицейский посетили семь баров, но бродягу не обнаружили. Каковы шансы найти его в восьмом баре?

Задача 8. Распознавание речи часто сталкивается с проблемой омофонии. Для её решения применяется система, основанная на формуле Байеса¹. Так, например, во французском языке слова «mer», «mère» и «maître» (обозначим эти слова W_1 , W_2 и W_3 соответственно) разными носителями произносятся чуть по-разному, но в целом очень похоже. Для обучения системы записывают различные варианты произношения, и по полученным данным оценивают вероятности для разных слов получить конкретное произношение A . Дополнительно, при распознавании не одного слова, а хотя бы фразы, из контекста можно оценить вероятности появления каждого из слов (или же считать их равновероятными, если оценка по контексту невозможна). Пусть для слов выше вариант произношения A встречается у 40% носителей для слова «mer», у 60% носителей для слова «mère» и у 50% носителей для слова «maître». Найти вероятности, что было произнесено каждое из слов при условии получения варианта произношения A и выбрать из них предпочтительное, если

- нет дополнительной информации о контексте;
- анализ контекста даёт следующие результаты: чуть выше вероятность встретить слово «mer» (40%), а два оставшихся варианта равновероятны.

Задача 9. Имеются две урны с чёрными и белыми шарами, из второй урны в первую перекладывают определённое количество шаров (неизвестно каких именно), вопрос: какова вероятность вытащить из первой урны белый шар, если

- В первой урне было 6 белых и 3 чёрных шара, во второй 2 белых и 8 чёрных шаров, а переложили ровно один шар?
- В первой урне было 6 белых и 3 чёрных шара, во второй 2 белых и 8 чёрных шаров, а переложили три шара?
- В первой урне было 600 белых и 300 чёрных шаров, во второй 300 белых и 700 чёрных шаров, а переложили 200 шаров?

¹Подробнее про эту систему можно почитать в книге books.google.ru/books?id=-7Z-GA73MMAC&pg=PA840

Дополнительные задачи

Задача 10 (Парадокс Монти Холла). Представьте, что вы стали участником игры, в которой вы находитесь перед тремя дверями. Ведущий, о котором известно, что он честен, поместил за одной из дверей автомобиль, а за двумя другими дверями — по козе. У вас нет никакой информации о том, что за какой дверью находится. Ведущий говорит вам:

Сначала вы должны выбрать одну из дверей. После этого я открою одну из оставшихся дверей, за которой находится коза. (Если у меня будет возможность выбрать, какую из двух дверей открывать, я выберу её случайным образом, с равными вероятностями.) Затем я предложу вам изменить свой первоначальный выбор и выбрать оставшуюся закрытую дверь вместо той, которую вы выбрали вначале. Вы можете последовать моему совету и выбрать другую дверь, либо подтвердить свой первоначальный выбор. После этого я открою дверь, которую вы выбрали, и вы выиграете то, что находится за этой дверью.

Вы выбираете дверь номер 1.

- (a) С какой вероятностью автомобиль находится за дверью номер 1? за дверью номер 2? за дверью номер 3?
- (b) С какой вероятностью ведущий откроет третью дверь, при условии, что машина за второй дверью? За третьей дверью? За первой дверью?
- (c) С какой вероятностью машина за первой дверью, при условии, что ведущий открыл третью дверь? С какой вероятностью машина за второй дверью, при условии, что ведущий открыл третью дверь? Следует игроку ли менять свой выбор?
- (d) (почему этому решению так сопротивляется интуиция) Ведущий, о котором известно, что он честен, поместил за одной из дверей автомобиль, а за двумя другими дверями — по козе. У вас нет никакой информации о том, что за какой дверью находится. Чему равна вероятность, что машина за второй дверью, при условии, что за третьей дверью — коза?
- (e) Чем отличается ситуация пункта 10d от ситуации в задаче Монти Холла?