

Школа лингвистики, 2019-20 уч. год**Теория вероятностей****Формула полной вероятности и формула Байеса (20.01.2020/21.01.2020)***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.

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

Задача 2. 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.

- (а) В спутник попал метеорит. С какой вероятностью он окажется на дне морском?
- (б) 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 3. Для фильтрации спама используются так называемые *байесовские фильтры*. Простейший механизм работы такого фильтра выглядит следующим образом. Допустим, есть некоторое слово-маркер, характерное для спаммерских писем. Если это слово встречается в сообщении, это увеличивает вероятность того, что письмо является спаммерским. Для расчёта вероятности используется формула Байеса.

Предположим, что всего 10% от всех писем являются спаммерскими, причём слово-маркер встречается в 95% спаммерских писем и в 10% неспаммерских. Если это слово встретилось в письме, какова вероятность того, что оно спаммерское?

Задача 4. 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 5. 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 6. 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 7. 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 «maire» (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 «maire». Найти вероятности, что было произнесено каждое из слов при условии получения варианта произношения A и выбрать из них предпочтительное, если

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

Дополнительные задачи

Задача 8. 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

- (а) В первой урне было 6 белых и 3 чёрных шара, во второй 2 белых и 8 чёрных шаров, а переложили ровно один шар?
- (б) В первой урне было 6 белых и 3 чёрных шара, во второй 2 белых и 8 чёрных шаров, а переложили три шара?
- (с) 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