2 Условная вероятность и теорема Байеса

Задача 1. Класс из 17 человек случайным образом делят на две группы по 9 и 8 человек соответственно. Найдите вероятность того, что ученицы этого класса - Лада и Варя - окажутся в одной группе.

Задача 2. Вы выучили 10 билетов из 25 и можете выбрать место в очереди на сдачу экзамена. Каждый экзаменуемый получает случайный билет из 25, билеты не возвращаются. Какое место в очереди вам занять выгоднее: первое или второе, чтобы с большей вероятностью получить известный вам билет?

Задача 3. У некоторого преподавателя есть 12 красных и 12 оранжевых ручек. Он раскладывает ручки в карманы по 12 в каждый таким образом, чтобы при случайном выборе кармана и вытаскивании из него ручки наугад, вероятность получить оранжевую ручку была максимально возможной. Какую вероятность вытащить оранжевую ручку преподавателю удастся получить?

Задача 4. Турнир по настольному теннису проводится по олимпийской системе: игроки случайным образом разбиваются на игровые пары; проигравший в каждой паре выбывает из турнира, а победитель выходит в следующий тур, где встречается со следующим противником, который определён жребием. Всего в турнире участвует 8 игроков, все они играют одинаково хорошо, поэтому в каждой встрече вероятность выигрыша и поражения у каждого игрока равна 0,5. Среди игроков два друга — Женя и Дима. Какова вероятность того, что этим двоим в каком-то туре придётся сыграть друг с другом?

Задача 5. У тёти Маши — двое детей разных возрастов. Вероятности рождения мальчика и девочки равны.

a) Какова вероятность того, что у тёти Маши оба ребёнка — мальчики, если известно, что у неё хотя бы один ребёнок — мальчик?

b) Какова вероятность того, что у тёти Маши оба ребёнка — мальчики, если известно, что у неё хотя бы один ребёнок — мальчик водолей по гороскопу?

Задача 6. В кофейне одинаковые с виду пирожки упакованы в пластиковые контейнеры по две штуки в каждом. В двух контейнерах по два пирожка с яблоком, в трех контейнерах по пирожку с яблоком и с малиной, и в одном контейнере оба пирожка с малиной. Взяли случайный контейнер, и один из пирожков в нем оказался с яблоком. Какова вероятность того, что второй пирожок в этом контейнере тоже с яблоком?

Задача 7. Энтомолог предполагает, что найденная им бабочка может относиться к редкому подвиду бабочек, так как узор на крыльях у неё несимметричен. Известно, что в редком подвиде у 99% бабочек узор на крыльях несимметричен, а среди обычных бабочек только 0.5% имеют несимметричный узор. Также известно, что бабочек редкого подвида насчитывается лишь 0.2% среди всей популяции. Какова вероятность того, что найденная энтомологом бабочка относится к редкому виду?

Задача 8 (Голосовой помощник)**.** Распознавание речи часто сталкивается с проблемой омофонии. Для её решения применяется система, основанная на формуле Байеса. Так, например, во французском языке слова *”mer”*, *”m‘ere”* 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 *"m‘ere"* 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:

a) нет дополнительной информации о контексте;

b) анализ контекста даёт следующие результаты: наиболее вероятным является слово *"mer"* (40%), а самая низкая вероятность встретиться у слова *"m‘ere"* (26%).

Задача 9. Известно, что в результате бросания десяти игральных костей выпала по крайней мере одна "шестерка". Какова вероятность того, что число выпавших "шестерок" больше единицы?

Решение. A - выпадение хотя бы одной шестерки, B - выпадение более одной шестерки, $B \subset A, p = 1/6$ - вероятность выпадения шестерки в отдельном бросании; искомая вероятность равна

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{1-\mathbb{P}(\bar{B})}{1-\mathbb{P}(\bar{A})} =$$

$$= \frac{1-(1-p)^{10} - C_{10}^{1} p(1-p)^9}{1-(1-p)^{10}} \approx 0.6148.$$

■

Задача 10 (Бунт ушастых)**.** Ровно половина населения острова Невезения – зайцы, а все остальные – кролики. Если житель острова Невезения что-нибудь утверждает, он всегда искренне верит в то, что говорит. При этом зайцы добросовестно заблуждаются в среднем в каждом четвертом случае, а кролики добросовестно заблуждаются в среднем в каждом третьем случае. Однажды в центр острова вышел зверь и закричал: "Я не заяц!". Подумал и грустно сказал: "Я не кролик". Какова вероятность того, что он все же заяц?

Задача 11 (Диплоидные организмы)**.** Ген карих глаз доминирует ген синих. Следовательно, у носителя пары bb глаза синие, а у носителя пар BB и Bb — карие. У диплоидных организмов (а мы такие :)) одна аллель наследуется от папы, а одна - от мамы. В семье у кареглазых родителей два сына — кареглазый и синеглазый. Кареглазый женился на синеглазой девушке.

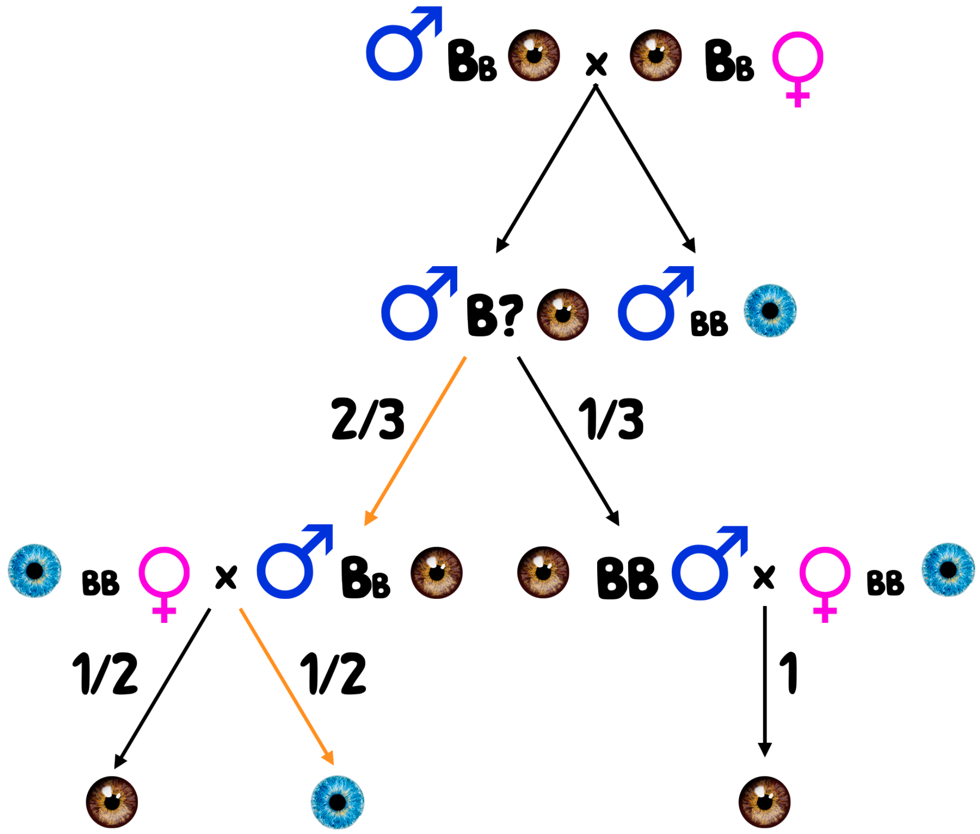
a) Какова вероятность рождения у них синеглазого ребенка?

b) Какова вероятность того, что ровно один из родителей синеглазой девушки - кареглазый?

Решение. По условию у кареглазых родителей $P_{♂}$ и $P_{♀}$ родился голубоглазый сын, а значит оба родителя — гетерозиготные и оба имеют как аллель карих глаз, так и аллель голубых. Имеем: $P_{♂}(Bb)$, $P_{♀}(Bb)$.

a) Девушка синеглазая, она гомозиготна с генотипом $F_{♀}(bb)$. В таком случае рассмотрим возможности для кареглазого сына: с $p_1 = \dfrac{1}{3}$ он гомозигота $F_{♂}(BB)$ и с $p_2 = \dfrac{2}{3}$ — гетерозигота $F_{♂}(Bb)$.

В первом случае с вероятностью 1 ребёнок от пары $F_{♀}(bb) \times F_{♂}(BB)$ будет кареглазым.



В противном же случае наблюдаем пару $F_{♀}(bb) \times F_{♂}(Bb)$, в которой с вероятностью $p_3 = 0.5$ рождается голубоглазый ребёнок. Отсюда

$$\mathbb{P}(\textit{голубоглазый ребёнок}) = 0.5 \cdot \frac{2}{3} \approx 0.33.$$

b) Раз девушка голубоглазая, то гомозиготная по этому признаку. Поэтому у каждого родителя есть аллель b. У каждого из родителей остается незаполненный аллель, ровно один кареглазый в двух случаях из 4, то есть $\mathbb{P}(\textit{ровно один кареглазый}) = 0.5$.

■

Задача 12 (Равновесие Харди-Вайнберга)**.** Предположим, что три возможных генотипа aa, Aa и AA изначально встречаются с частотами p_1, p_2 и p_3, где $p_1 + p_2 + p_3 = 1$. Ген не сцеплен с полом, поэтому частоты p_1, p_2 и p_3 одинаковы для мужчин и для женщин.

a) У семейных пар из этой популяции рождаются дети. Назовём этих детей первым поколением. Каковы частоты для трёх возможных генотипов в первом поколении?

b) У семейных пар первого поколения тоже рождаются дети. Назовём этих детей вторым поколением. Каковы частоты для трёх возможных генотипов во втором поколении?

c) Каковы частоты для трёх возможных генотипов в n-ном поколении?

d) Заметив явную особенность предыдущего ответа сформулируйте теорему о равновесии Харди-Вайнберга. Прокомментируйте утверждение: "Любой рецессивный ген со временем исчезнет".

Решение. Рассмотрим все варианты скрещивания и вероятности:

$$\begin{array}{ll} \mathrm{AA} * \mathrm{AA} & p_1^2 \\ \mathrm{AA} * \mathrm{Aa} & 2 \cdot p_1 \cdot p_2 \\ \mathrm{AA} * \ \mathrm{aa} & 2 \cdot p_1 \cdot p_3 \\ \mathrm{Aa} * \mathrm{Aa} & p_2^2 \\ \mathrm{Aa} * \ \mathrm{aa} & 2 \cdot p_2 \cdot p_3 \\ \mathrm{aa} \ * \ \mathrm{aa} & p_3^3 \end{array}$$

Зная, какие дети рождаются у всевозможных пар, составим аналогичную таблицу вероятностей для генотипов детей:

AA$p_1^2 + p_1 \cdot p_2 + 0.25 \cdot p_2^2 = p^2$

Aa $\quad p_1 \cdot p_2 + p_2 \cdot p_3 + 0.5 \cdot p_2^2 + 2 \cdot p_1 \cdot p_3 = 2pq$

aa $\quad p_3^2 + p_2 \cdot p_3 + 0.25 \cdot p_2^2 = q^2$

Видим, что всё это сворачивается в полные квадраты:

AA : $(p_1 + 0.5 \cdot p_2)^2 = p'^2$

Aa : $2\left(p_1 + 0.5 \cdot p_2\right)^2 \cdot \left(p_3 + 0.5 \cdot p_2\right)^2 = 2p'q'$

aa : $(p_3 + 0.5 \cdot p_2)^2 = q'^2$

Соотношение генов не изменилось и всё ещё удовлетворяет уравнению

$$p'^2 + 2p'q' + q'^2 = 1$$

Потому соотношение генов в любом поколении не меняется, а значит, рецессивный ген со временем не исчезает. ■

Задача 13. Для поиска пропавшего самолета выделено 10 вертолетов, каждый из которых может быть использован для поисков в одном из двух возможных районов, где самолет может находиться с вероятностями 0,8 и 0,2. Как следует распределить вертолеты по районам поисков, чтобы вероятность обнаружения самолета была наибольшей, если каждый вертолет обнаруживает находящийся в районе поиска самолет с вероятностью 0,2, а поиски осуществляются каждым вертолетом независимо от других? Найти вероятность обнаружения самолета при оптимальной процедуре поисков.

Решение. Решим задачу в общем виде. Пусть вертолётов $n = 10$; вероятности нахождения самолёта в одном и другом районе обозначим через $p = 0{,}8$ и $1 - p = 0{,}2$. Вероятность обнаружения самолёта каждым из вертолётов (в том районе, где самолёт находится) удобно обозначить через $1 - q$, где $q = 0{,}8$-вероятность необнаружения.

Пусть в первый район направили k вертолётов, а во второй $n - k$. Самолёт может быть в первом или втором районе; вероятности этих событий равны p и 1 - p соответственно. Вероятность того, что при нахождении самолёта в первом районе он никем не будет обнаружен, равна q^k. Вероятность обнаружения хотя бы одним вертолётом составит $1 - q^k$. Для второго района эта величина составит $1 - q^{n-k}$. Тогда по формуле полной вероятности, величина $p\left(1 - q^k\right) + (1 - p)\left(1 - q^{n-k}\right)$ будет вероятностью обнаружения самолёта. Она равна $1 - \left(pq^k + (1 - p)q^{n-k}\right)$, поэтому надо минимизировать величину, заключённую в скобки.

Положим $t = q^k$. Тогда минимизации подлежит величина $f(t) = pt + \frac{(1-p)q^n}{t}$, где $t \in (0; 1)$. Легко видеть, что минимум функции вида $At + \frac{B}{t}$ достигается в точке t, для которой слагаемые равны, то есть $t^2 = \frac{B}{A}$. Этот факт легко выводится из неравенства о среднем, либо при помощи производной. Последняя равна $A-\frac{B}{t^2}$, и она обращается в ноль именно в этой точке. При переходе через неё производная меняет знак с минуса на плюс, то есть это точка минимума. В нашем случае получается $q^{2k}=\frac{(1-p)q^n}{p}$, то есть $q^{2k-n}=\frac{1-p}{p}$. Из этого соотношения получается $2k-n=\log_q\frac{1-p}{p}=\frac{\ln\left(p^{-1}-1\right)}{\ln q}$. Эта формула позволяет найти k, но поскольку мы не можем послать дробное, и тем более иррациональное число вертолётов, результат надо округлить до целого, причём в обе стороны, и сравнить величины между собой. Несложные вычисления показывают, что $2k-10\approx 6{,}2$, откуда $k\approx 8{,}1$. Проверке и сравнению подлежат числа $k=8$ и $k=9$. Формула $1-\left(pq^k+(1-p)q^{n-k}\right)$ при этом даёт значения 0,73778 и 0,73262 (приближённо). Первое значение больше, поэтому оптимальным планом будет направить 8 вертолётов в первый район и 2 во второй. Вероятность обнаружения самолёта составит при этом около ■

Метод первого шага

Задача 14 (Амёба)**.** Популяция начинается с одной амебы. Для этого и для последующих поколений существует вероятность 0.75 того, что отдельная амеба разделится, чтобы создать двух амеб, и вероятность 0.25 того, что она вымрет, не произведя потомства. Пусть случайной величиной X будет число поколений до гибели всех амеб. Найдите вероятности:

a) $\mathbb{P}(X=2)$;

b) $\mathbb{P}(X=3)$;

c) $\mathbb{P}(X=\infty)$. *"Что мы говорим богу смерти? Не сегодня!"*

Решение. a) Для создания второго поколения первая амёба должна поделиться, а раз второе поколение - последнее, обе дочерние амёбы должны вымереть. Отсюда

$$\mathbb{P}(X=2)=\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}=\frac{5}{64}\approx 0.05.$$

b) Заметим, что событие $X=3$ раскладывается на события первая разделилась и из амёб второго поколения хотя бы одна не вымерла и все амебы третьего поколения вымерли. Тогда

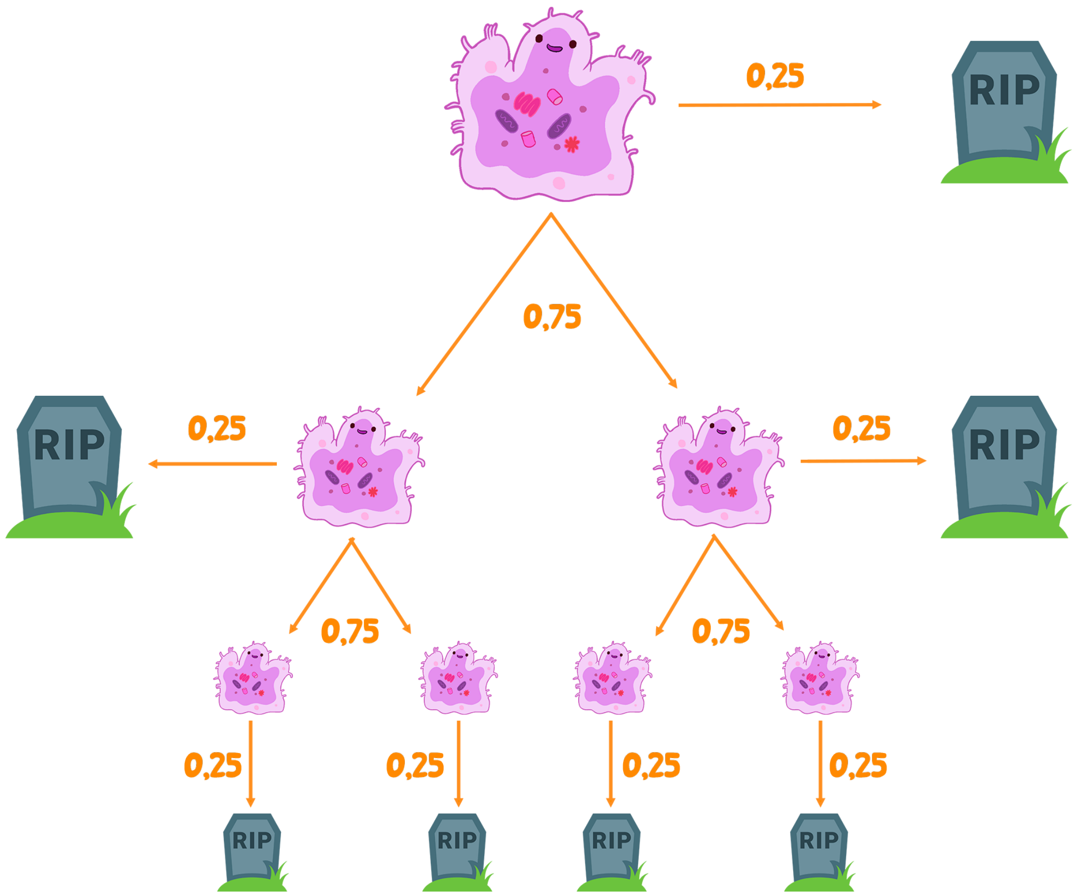
$$\mathbb{P}(X=3)=\frac{3}{4}\left(2\cdot\frac{1}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}+\left(\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}\right)^2\right)=\frac{315}{16384}\approx 0.02.$$

c) Решение 1. Пусть вероятность того, что генеалогическое древо конкретной амёбы будет бесконечным составляет p. Тогда заметим, что $p=\dfrac{3}{4}(1-(1-p)^2)$, ведь для бесконечного цикла необходимо, чтобы хотя бы одна из амёб-наследниц выжила. Тогда $p=\dfrac{3}{4}(2p-p^2)$, откуда, полагая $p\neq 0$, имеем $p=\dfrac{2}{3}\approx 0.67$.

Решение 2. Задача сводится к исследованию случайного блуждания. Находясь в единице за одну иттерацию можно попасть либо в ноль, либо в 2. Тогда нас интересует вероятность того, что при этом случайном блуждании мы никогда не попадём в ноль. Имеем: $p=1-\dfrac{1-\frac{3}{4}}{\frac{3}{4}}=\dfrac{2}{3}$.

Больше про случайные блуждания можно почитать здесь.

■



Задача 15. Илье Муромцу предстоит дорога к камню. От камня начинаются ещё три дороги. Каждая из тех дорог снова оканчивается камнем. И от каждого камня начинаются ещё три дороги. И каждые те три дороги оканчиваются камнем... И так далее до бесконечности. На каждой дороге живёт трёхголовый Змей Горыныч. Каждый Змей Горыныч бодрствует независимо от других с вероятностью (хм, Вы не поверите!) одна третья. У Василисы Премудрой существует Чудо-Карта, на которой видно, какие Змеи Горынычи бодрствуют, а какие — нет. Какова вероятность того, что Василиса Премудрая сможет найти на карте бесконечный жизненный путь Ильи Муромца, проходящий исключительно мимо спящих Змеев Горынычей?