Мусин Артем Игоревич
Диакон, кандидат физико-математических наук, доцент кафедры физики и методики обучения физике Вятского государственного университета, преподаватель Центра дополнительного образования одаренных школьников, г. Киров



Как формулы Байеса помогают фильтровать спам и анализировать ВИЧ-инфекцию

В XVIII в. британский математик и священник Томас Байес (1702-1761) вывел формулы, которые позволяют оценить (и переоценить) вероятность тех или иных событий эмпирическим путем. В настоящее время формулы Байеса используются в разных областях математического знания. В статье рассматриваются примеры задач, посвященных применению формул Байеса. Часть из разобранных задач использовались автором при чтении лекций и проведении практических занятий в колледже Вятского государственного университета.

1. Условная вероятность. Гипотезы

Предполагается, что читатель знаком с основами теории вероятностей в размере учебника за 7-9 класс [1] или учебного пособия [2]. Автор указанного пособия даже приводит список основных положений теории вероятностей, которые должен знать каждый школьник [2, с. 61-61]. Среди них: классическое определение вероятности события A как отношение числа благоприятных исходов m к числу всех равновозможных исходов n; сложение и умножение событий, противоположное событие, теоремы суммы и произведения вероятностей.

Условной вероятностью $P_B(A)$ события A при условии наступления события B назовем следующее соотношение:

$$P_B(A) = \frac{P(AB)}{P(B)}.$$

Эту формулу можно получить и из классического определения вероятности, если под условной вероятностью понимать отношение числа исходов, благоприятных событию A и одновременно событию B, к числу всех исходов, при которых возможно событие-условие B.

Пример 1. В коробке лежит 3 синих и 2 красных ручки. Из нее наугад достают сначала одну ручку, потом еще одну. Вычислим вероятность того, что вторая ручка явля-

ется синей (событие A) при условии, что первая ручка также синяя (событие B).

На момент, когда событие B произошло, в коробке оставались 2 синих и 2 красных ручки. По определению, вероятность достать синюю ручку при таком раскладе равна $2/4 = 1/2$. Поэтому $P_B(A)=1/2$.

С другой стороны, чему равна вероятность события AB, то есть, что достали обе синих ручки? Из пяти ручек можно взять любые две $5 \cdot 4 = 20$ способами (это число всех исходов), а из трех синих любые две $3 \cdot 2 = 6$ способами (это число благоприятных исходов), поэтому $P(AB)=3/10$. Вероятность в первый раз достать синюю ручку $P(B) = 3/5$. По формуле условной вероятности получаем тот же ответ $P_B(A) = 1/2$.

Часто бывает так, что перед наступлением события A происходит одно из несовместных событий B_1, B_2, ..., B_n, причем заранее неизвестно, какое именно. Такую ситуацию можно представить в виде дерева (рис. 1). В исходной точке O кто-то (Господь Бог, природа, случай и т.д.) сделал свой ход, и реализовалась одна из возможных альтернатив B_i, то есть одна из веток дерева. Когда происходит событие A, мы можем лишь предполагать, какая из альтернатив реализовалась. Поэтому события B_1, B_2, ..., B_n называют *гипотезами*. Считается, что набор гипотез *исчерпывающий*, то есть у дерева возможностей не существует других веток, кроме уже нарисованных.

Событие A может происходить на каждой из веток с вероятностями $P_{B_1}(A)$, $P_{B_2}(A)$, ..., $P_{B_n}(A)$ соответственно. Тогда полная (безусловная) вероятность события A вычисляется по *формуле полной вероятности*

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) + \\ + \ldots + P(B_n)P_{B_n}(A).$$

Эта формула может быть выведена из определения условной вероятности. Действительно, каждое слагаемое здесь превращается в $P(AB_1)$, $P(AB_2)$, ..., $P(AB_n)$ соответственно. Вероятности этих событий можно изобразить в виде отдельных «ягод» на ветках, а полная вероятность события A — это «гроздь», содержащая все «ягоды», значит, равна сумме указанных вероятностей (рис. 1).

Пример 2. В коробке лежит 3 синих и 2 красных ручки. Из нее на-

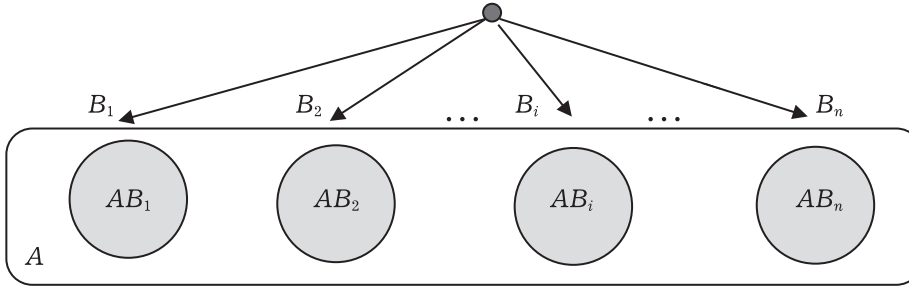

Рис. 1. Событие A и его реализация на ветках дерева гипотез B_1, B_2, …, B_n

угад достают сначала одну ручку и прячут, не показывая, после чего из коробки достают еще одну ручку. Вычислим вероятность того, что вторая ручка является синей (событие A).

У нас нет информации о том, какой была первая ручка, но есть две гипотезы. B_1: первая ручка — синяя. B_2: первая ручка — красная. Других вариантов нет.

Вероятность того, что первая ручка синяя $P(B_1) = 3/5$. Если так и случится, то вероятность вытащить синюю ручку во второй раз $P_{B_1}(A) = 1/2$.

Вероятность того, что первая ручка красная $P(B_2) = 2/5$. Если реализуется этот вариант, то вероятность вытащить во второй раз синюю ручку $P_{B_2}(A) = 3/4$.

По формуле полной вероятности $P(A) = 3/5$. Этот же результат можно получить и из определения, если правильно посчитать число благоприятных исходов (такими являются пары синяя+синяя в количестве $3 \cdot 2 = 6$ и красная+синяя в количестве $2 \cdot 3 = 6$).

2. Оценка и переоценка гипотез. Формулы Байеса

В только что разобранном примере вероятности гипотез вычислялись с использованием определения. Иногда вероятности гипотез напрямую вычислить не удается, и их приходится оценивать из других соображений.

Пример 3. В коробке лежит 2 шара неизвестного цвета. В нее кладут белый шар, а потом из нее наугад достают один шар. Пусть мы хотим вычислить вероятность того, что извлеченный шар — белый (событие A).

Занумеруем шары, которые изначально были в коробке (№1 и №2). Те шары, которые не белые, условно назовем «черными». Существует четыре гипотезы:

1. №1 и №2 — «черные» (B_1);
2. №1 — «черный», №2 — белый (B_2);
3. №1 — белый, №2 — «черный» (B_3);
4. №1 и №2 — белые (B_4).

Условные вероятности события A при наступлении одной из гипотез равны $P_{B_1}(A)=1/3$, $P_{B_2}(A)=2/3$, $P_{B_3}(A)=2/3$, $P_{B_4}(A)=1$ соответственно.

Однако есть проблема: вероятности самих гипотез неизвестна без дополнительных предположений! Если коробку наполняют двумя случайными шарами, которые выбирают из белых и черных (без кавычек) с равной вероятностью, то вероятности всех гипотез B_i одинаковы и равны $1/4$. В этом случае по формуле полной вероятности

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) +$$

$$+P(B_3)P_{B_3}(A) + P(B_4)P_{B_4}(A) =$$

$$= \frac{1}{4}\cdot\frac{1}{3} + \frac{1}{4}\cdot\frac{2}{3} + \frac{1}{4}\cdot\frac{2}{3} + \frac{1}{4}\cdot 1 = \frac{2}{3}.$$

Такое предположение довольно разумно. Однако все зависит от способа выбора шаров для наполнения коробки. Предположим, что два шара берутся из хранилища, где находятся белые, синие, красные, зеленые и т.д. шары в равном количестве. Тогда «черные» (т.е. все не белые) шары будут появляться чаще. В этом

случае гипотеза B_4 будет менее вероятна, и полная вероятность $P(A)$ снизится. Это же будет иметь место в случае подтасовки, когда человек, наполняющий коробку, предпочитает класть «черные» шары чаще белых.

Что же делать? Можно оценить вероятность гипотез, исходя из некоторого разумного предположения. Когда появляется новая информация, то теория вероятности помогает скорректировать вероятности гипотез. Эта операция носит название *переоценка гипотез*, и она возможна благодаря теореме Байеса.

Теорема Байеса. Пусть испытание проводится в предположении, что верна одна из гипотез B_1, B_2, ..., B_n, вероятности которых оцениваются как $P(B_1)$, $P(B_2)$, ..., $P(B_n)$. Если в результате испытания произошло событие A, то с учетом новой информации условная вероятность каждой из гипотез B_i равна

$$P_A(B_i) = \frac{P(B_i)P_{B_i}(A)}{P(A)}$$

(формулы Байеса).

Доказательство. Из определения условной вероятности

$$P_A(B_i) = \frac{P(AB_i)}{P(A)}, \quad P_{B_i}(A) = \frac{P(AB_i)}{P(B_i)}.$$

Выражая $P(AB_i)$ из второй формулы и подставляя в первую, получаем требуемый результат, ч.т.д.

Задача 1. В коробке лежит 2 шара, каждый из них либо черного, либо белого цвета. В коробку кладут белый шар, а потом из нее наугад достают один шар. Считая гипотезы о первоначальном составе коробки равновероятными, сделайте их переоценку, если извлеченный шар: а) белый; б) черный.

Решение. Четыре гипотезы о возможном первоначальном составе коробки разбирались в примере 3. По условию имеем

$$P(B_1) = P(B_2) = P(B_3) = P(B_4) = 1/4.$$

Пусть извлеченный шар — белый (событие A). По формуле полной вероятности $P(A) = 2/3$.

Формулы Байеса дают

$$P_A(B_1) = 1/8,\ P_A(B_2) = P_A(B_3) = 1/4,$$
$$P_A(B_4) = 3/8.$$

Пусть извлеченный шар — черный (событие $\overline{A}$). Вычислим условные вероятности этого события.

1. Если шар вытаскивали из набора «2 черных и 1 белый» (B_1), то $P_{B_1}(\overline{A}) = 2/3$.

2. Если шар вытаскивали из набора «1 черный и 2 белых» (B_2 или B_3), то $P_{B_2}(\overline{A}) = P_{B_3}(\overline{A}) = 1/3$.

3. Если шар вытаскивают из набора 3 белых (B_4), то событие $\overline{A}$ невозможно. $P_{B_4}(\overline{A}) = 0$.

По формуле полной вероятности

$$P(\overline{A}) = \frac{1}{4}\cdot\frac{2}{3} + \frac{1}{4}\cdot\frac{1}{3} + \frac{1}{4}\cdot\frac{1}{3} + \frac{1}{4}\cdot 0 = \frac{1}{3}$$

(и это логично, поскольку вытащить черный шар — значит не вытащить белый, как раз с вероятностью $1 - 2/3 = 1/3$). Формулы Байеса дают

$$P_{\overline{A}}(B_1) = 1/2,\ P_{\overline{A}}(B_2) = P_{\overline{A}}(B_3) = 1/4,$$
$$P_{\overline{A}}(B_4) = 0.$$

Интерпретация. Когда мы достаем шар определенного цвета, то мы получаем дополнительную информацию, которая изменяет вероятность гипотез. Так, если этот шар — белый, то, скорее всего, реализовалась не

гипотеза B_1 (2 черных), потому что в этом случае вероятность достать белый шар «всего» 1/3, а не 2/3 или 1. Поэтому информация о том, что шар — белый, повышает вероятность гипотезы B_4 и понижает вероятность гипотезы B_1. Вероятности гипотез B_2 и B_3 не изменяются.

Особенно видна важность переоценки в пункте б). Если достали черный шар, то очевидно, гипотеза B_4 неверна, ее вероятность понижается с 1/4 до 0. Скорее всего, реализовалась гипотеза B_1, потому что в таком случае вероятность достать черный шар выше всего — 2/3. Таким образом, вероятность гипотезы B_1 повышается с 1/4 до 1/2.

Задача 2 (Монти Холла). На столе лежит три одинаковых пустых коробки с номерами 1, 2, 3, в одну из которых ведущий помещает приз. Игрок выбирает одну коробку из трех, после чего ведущий открывает одну коробку из двух оставшихся и показывает, что там нет приза. После этого игрок может поменять свой выбор (если хочет), и вскрывает выбранную коробку. Как ему нужно действовать? С какой вероятностью он заберет приз?

Решение. Эта задача получила название по имени ведущего американского телешоу 1980-х «Let's Make a Deal», в котором вместо коробок были двери.

Введем гипотезы B_1, B_2, B_3, состоящие в том, что приз находится в коробке с соответствующим номером. В отсутствие дополнительной информации разумно считать, что $P(B_1) = P(B_2) = P(B_3) = 1/3$. Можно выбрать любую коробку, и если ее дадут открыть сразу, вероятность выигрыша равна 1/3.

Пусть, для определенности, игрок выбрал коробку №1, а ведущий открыл и показал коробку №3 (пустую). Назовем такой расклад игры событием A. По формуле полной вероятности

$$P(A) = P(B_1)P_{B_1}(A) + P(B_2)P_{B_2}(A) + \\ + P(B_3)P_{B_3}(A).$$

Рассмотрим каждый случай отдельно.

1. Приз в коробке №1 (B_1). Если игрок выбирает ее, то ведущий мог открыть коробку 2 или коробку 3 (они обе пустые). Поэтому разумно считать, что ведущий выберет коробку 3 в половине случаев и $P_{B_1}(A) = 1/2$.

2. Приз в коробке №2 (B_2). Игрок выбрал коробку 1, поэтому ведущий не может открыть ни первую, ни вторую коробку. Значит, он в любом случае выберет коробку 3 и $P_{B_2}(A) = 1$.

3. Приз в коробке №3 (B_3). Событие A невозможно. $P_{B_3}(A) = 0$.

Формула полной вероятности дает $P(A) = 1/2$.

Применим формулы Байеса для переоценки гипотез и получим: $P_A(B_1) = 1/3$, $P_A(B_2) = 2/3$, $P_A(B_3) = 0$. Последний результат очевиден и так: если коробка №3 открыта (и пуста), то вероятность нахождения в ней приза понижается с 1/3 до 0. Вместе с тем вероятность гипотезы B_2 повышается в 2 раза. Это означает, что игроку выгодно поменять свой выбор. В данном случае, если он вскроет коробку №2, он заберет приз с вероятностью не 1/2, а 2/3.

Задача 3 (элементарный спам-фильтр). Простейший фильтр спама для электронной почты проверяет наличие ключевого слова

«бесплатно» (или однокоренных) в каждом письме. В среднем ключевое слово встречается в 70% нежелательных писем и только в 5% обычных писем. Примерно 40% приходящих в почтовый ящик писем являются спамом. Фильтр удаляет все письма, которые являются спамом с вероятностью более 90%. Будет ли письмо удалено, если оно содержит слово «бесплатно»?

Решение. Для каждого письма верна одна из следующих гипотез: B_1 — письмо является спамом, B_2 — письмо не является спамом. Вероятности гипотез можно оценить, исходя из общей частоты появления спама в 40%, то есть $P(B_1) = 0{,}4$ и $P(B_2) = 0{,}6$.

Когда спам-фильтр находит в пришедшем письмо ключевое слово «бесплатно» (значит, случилось событие A), это дает дополнительную информацию, которая меняет вероятности гипотез. Рассмотрим два случая.

1. Письмо действительно является нежелательным (B_1). Вероятность наличия ключевого слова, по условию, $P_{B_1}(A) = 0{,}7$.

2. Письмо является полезным (B_2). Вероятность наличия ключевого слова, по условию, $P_{B_2}(A) = 0{,}05$.

По формуле полной вероятности $P(A) = 0{,}31$. То есть, в среднем, слово «бесплатно» встречается в 31% всех писем (желательных + нежелательных).

По формулам Байеса $P_A(B_1) \approx 0{,}903$ и $P_A(B_2) \approx 0{,}097$. Таким образом, с вероятностью 90,3% фильтр посчитает письмо нежелательным, и оно будет удалено.

Задача 4 (тест на ВИЧ). Вирусом иммунодефицита человека (ВИЧ) заражено примерно 0,02% населения. Пациент сдал анализ на ВИЧ, который дает верный результат в 97% случаев. Какова вероятность того, что пациент действительно заражен ВИЧ, если результат анализа: а) положительный; б) отрицательный?

Решение. О конкретном пациенте имеется две гипотезы: B_1 — он заражен ВИЧ, B_2 — не заражен. Исходные вероятности $P(B_1) = 0{,}0002$ и $P(B_2) = 0{,}9998$.

Пусть результат анализа положительный (событие A). Рассмотрим два случая.

1. Пациент заражен ВИЧ (B_1). Тогда вероятность положительного анализа $P_{B_1}(A) = 0{,}97$ согласно условию.

2. Пациент не заражен ВИЧ (B_2). В 97% случаев анализ даст верный результат (то есть отрицательный), но в 3% случаев будет ложноположительный результат, поэтому $P_{B_2}(A) = 0{,}03$.

По формуле полной вероятности $P(A) \approx 0{,}0302$.

Формулы Байеса дают $P_A(B_1) \approx 0{,}0064$ и $P_A(B_2) \approx 0{,}9932$. Обнадеживающе! То есть в случае положительного результата рано паниковать — вероятность быть здоровым очень высока, более 99%.

Пусть результат анализа отрицательный (противоположное событие $\overline{A}$). Опять два случая.

1. Пациент заражен ВИЧ (B_1). В 97% случаев анализ даст верный результат (то есть положительный), но в 3% случаев будет ложноотрицательный результат, поэтому $P_{B_1}(\overline{A}) = 0{,}03$.

2. Пациент не заражен ВИЧ (B_2). Вероятность отрицательного анализа $P_{B_2}(\overline{A}) = 0{,}97$.

По формуле полной вероятности $P(\overline{A}) \approx 0{,}9698$.

Формула Байеса дает $P_{\overline{A}}(B_1) \approx \approx 6 \cdot 10^{-6}$! Фактически, отрицательный результат практически полностью исключает возможность заражения.

3. Несколько переоценок гипотез

Формулы Байеса можно применять не один раз, а несколько. При этом гипотезы переоцениваются каждый раз при получении новой информации. Рассмотрим еще несколько задач.

Задача 5. Пациент, получивший положительный результат анализа на ВИЧ, сдал тот же анализ повторно. Какова вероятность, что он инфицирован ВИЧ, если второй результат: а) положительный; б) отрицательный?

Решение. При условии первого положительного результата (событие A_1) вероятности гипотез B_1 (инфицирован ВИЧ) и B_2 (здоров) были рассчитаны в задаче 4: $P_{A_1}(B_1) \approx \approx 0{,}0064$ и $P_{A_1}(B_2) \approx 0{,}9932$.

Пусть второй результат также положительный (событие A_2). Как и прежде, вероятность положительного результата при условии наличия ВИЧ 0,97, вероятность ложноположительного результата (при отсутствии ВИЧ) 0,03. Полная вероятность второго положительного результата при условии положительного первого

$$P_{A_1}(A_2) \approx 0{,}0064 \cdot 0{,}97 + 0{,}9932 \cdot 0{,}03 \approx \approx 0{,}0360.$$

По формулам Байеса

$$P_{A_1,A_2}(B_1) \approx \frac{0{,}0064 \cdot 0{,}97}{0{,}0360} \approx 0{,}1724,$$

$$P_{A_1,A_2}(B_2) \approx \frac{0{,}9932 \cdot 0{,}03}{0{,}0360} \approx 0{,}8277.$$

Таким образом, вероятность быть больным ВИЧ при получении второго положительного теста резко повышается с 0,64% до 17,24%.

Пусть второй результат отрицательный (событие $\overline{A}_2$). Как и прежде, вероятность ложноотрицательного результата (при наличии ВИЧ) 0,03, вероятность отрицательного результата при отсутствии ВИЧ 0,97. Полная вероятность второго отрицательного результата при условии положительного первого

$$P_{A_1}\left(\overline{A}_2\right) \approx 0{,}0064 \cdot 0{,}03 + 0{,}9932 \cdot 0{,}97 \approx \approx 0{,}9636.$$

По формуле Байеса

$$P_{A_1,\overline{A}_2}(B_1) \approx \frac{0{,}0064 \cdot 0{,}03}{0{,}9636} \approx 2 \cdot 10^{-4},$$

то есть можно быть относительно спокойным.

Задача 6. В спам-фильтр почтового ящика из задачи 3 добавили проверку ключевого слова «лотерея», которое встречается в среднем в 60% нежелательных писем и в 2%

обычных. Будут ли удалены письма, если в них обнаружили ключевые слова: №1 — только «бесплатно»; №2 — только «лотерея»; №3 — «бесплатная лотерея»?

Решение. Пусть A_1 — событие, состоящие в том, что в письме есть слово «бесплатно», A_2 — в письме есть слово «лотерея». Представим работу фильтра в два этапа: сначала проверяется, произошло или не произошло событие A_1, после этого переоцениваются вероятности гипотез, потом проверяется событие A_2, и вновь переоцениваются гипотезы.

В задаче 3 были введены гипотезы B_1 (спам) и B_2 (не спам) с вероятностями $P(B_1) = 0{,}4$ и $P(B_2) = 0{,}6$, и была сделана их переоценка в случае, если произошло событие A_1: $P_{A_1}(B_1) \approx 0{,}903$ и $P_{A_1}(B_2) \approx 0{,}097$. Переоценим гипотезы в случае, если событие A_1 не произошло, или, другими словами, произошло противоположное событие $\bar{A}_1$ (т.е. в письме нет слова «бесплатно»).

Полная вероятность $P(\bar{A}_1)$ проще всего вычислить как

$$1 - P(A_1) = 0{,}69.$$

Условная вероятность $P_{B_1}(\bar{A}_1) = 0{,}3$, поскольку из условия следует, что 30% нежелательных писем не содержат слова «бесплатно», и $P_{B_1}(\bar{A}_1) = 0{,}95$, поскольку 95% обычных писем не содержат этого же слова. Тогда по формулам Байеса $P_{\bar{A}_1}(B_1) \approx 0{,}174$ и $P_{\bar{A}_1}(B_2) \approx 0{,}826$.

Рассмотрим письмо №1. Так как в нем есть слово «бесплатно», то вероятность, что оно является спамом, равна 0,903, и не спамом — 0,097. В первом случае слово «лотерея» встречается в 60% писем, во втором — в 2%. Это значит, что полная вероятность найти в нем слово «лотерея» (событие A_2)

$$P_{A_1}(A_2) \approx 0{,}903 \cdot 0{,}6 + 0{,}097 \cdot 0{,}02 \approx \approx 0{,}544$$

и не найти в нем слово «лотерея» (событие $\bar{A}_2$)

$$P_{A_1}(\bar{A}_2) \approx 0{,}903 \cdot 0{,}4 + 0{,}097 \cdot 0{,}98 \approx \approx 0{,}456$$

В письме №1 слова «лотерея» нет, поэтому переоценка гипотез по формулам Байеса понижает вероятность спама:

$$P_{A_1,\bar{A}_2}(B_1) \approx \frac{0{,}903 \cdot 0{,}4}{0{,}456} \approx 0{,}792,$$

$$P_{A_1,\bar{A}_2}(B_2) \approx \frac{0{,}097 \cdot 0{,}98}{0{,}456} \approx 0{,}208.$$

Таким образом, письмо №1 не будет удалено.

Теперь рассмотрим письмо №3. В нем тоже есть слово «бесплатная», и вероятность спама 0,903. Переоценка гипотез по формулам Байеса повышает вероятность спама:

$$P_{A_1,A_2}(B_1) \approx \frac{0{,}903 \cdot 0{,}6}{0{,}544} \approx 0{,}996,$$

$$P_{A_1,A_2}(B_2) \approx \frac{0{,}097 \cdot 0{,}02}{0{,}544} \approx 0{,}004.$$

Итак, письмо №3 точно будет удалено.

Остается письмо №2. В нем нет слова «бесплатно», поэтому вероятность, что оно является спамом с учетом этого обстоятельства, 0,174, и не спамом — 0,826. Опять же, в первом случае слово «лотерея» встречается в 60% писем, во втором — в 2%. Это значит, что полная вероятность найти в нем слово «лотерея» (событие A_2)

$$P_{\bar{A}_1}(A_2) \approx 0{,}174 \cdot 0{,}6 + 0{,}826 \cdot 0{,}02 \approx$$

$$\approx 0{,}121.$$

Переоценка гипотез по формулам Байеса дает:

$$P_{\bar{A}_1, A_2}(B_1) \approx \frac{0{,}174 \cdot 0{,}6}{0{,}121} \approx 0{,}863,$$

$$P_{\bar{A}_1, A_2}(B_2) \approx \frac{0{,}826 \cdot 0{,}02}{0{,}121} \approx 0{,}137.$$

Письмо №2 не будет удалено.

4. Задачи для самостоятельного решения

Задача 7. В коробку кладут 1, 2, 3 или 4 белых шара, потом туда добавляют черные шары, чтобы общее количество шаров равнялось 5. Считая гипотезы о первоначальном составе коробки равновероятными, сделайте их переоценку, если извлеченный наугад из коробки шар — белый.

Задача 8. У пациентки болит горло — это симптом многих респираторных вирусных инфекций. Известно, что горло болит у 10% заболевших вирусом «К». Вероятность того, что этот симптом будет наблюдаться в иных случаях, равна 2%. Вычислите вероятность того, что пациентка действительно больна вирусом «К», если им переболело 3% населения.

Задача 9. Алиса и Боб играют в игру: из колоды 36 карт они случайным образом достают одну карту. Если карта красная, побеждает Алиса, если черная — то Боб. После каждого раунда карта возвращается в колоду. Алиса подозревает, что с вероятностью 5% Боб подменил 9 каких-то красных карт в колоде на черные. 1) Какова вероятность того, что колода действительно крапленая, если Алиса проиграла 4 раза подряд? 2) Как изменится эта оценка, если на пятый раз Алиса выиграет?

Задача 10. Один благородный статский советник, который никогда не обманывает, каким-то волшебным образом всегда выигрывает в азартные игры. Его противник изначально на 99% уверен в честности своего оппонента. На сколько процентов уменьшится эта оценка, когда сверхудачливый статский советник выиграет 10 раз подряд в игру, в которой обычные люди выигрывают с вероятностью 1/2?

Задача 11. В городе Эквилити на 10 тыс. населения приходится один шпион. Контрразведчики разработали анкету из 100 вопросов для определения шпионов. *Чувствительность* анкеты 90%, то есть если опрашиваемый — шпион, то он будет определен с вероятностью 90%. *Специфичность* анкеты 99%, то есть если опрашиваемый — честный гражданин, то он будет признан таковым по результатам

анкетирования с вероятностью 99%. Случайно взятого прохожего попросили пройти анкетирование. Какова вероятность получить при этом опросе: 1) ложноотрицательный; 2) ложноположительный; 3) верный результат? Какова вероятность, что этот прохожий — действительно шпион, если анкета показывает, что он шпион?

Ответы.

№7: «в коробке был 1 белый шар» — 10%, «в коробке было 2 белых шара» — 20%, «в коробке было 3 белых шара» — 30%, «в коробке было 4 белых шара» — 40%.

№8: 14%.

№9: 21,0%; 11,8%.

№10: на 90%.

№11: 0,001%; 0,9999%; 98,9991%; 0,89%.

Литература

1. Высоцкий И.Р., Макаров А.А., Тюрин Ю.Н., Ященко И.В. Теория вероятностей и статистика. 7–9 классы: учеб. пособие / под ред. И.В. Ященко. М.: Просвещение, 2020. 256 с.
2. Балакирев С.М. Теория вероятностей для школьников (с абсолютного нуля): учебное пособие. 2019. 73 с. URL: https://proproprogs.ru/files/balak_scool_tp.pdf (дата обращения: 06.02.2025).

НОВОСТИ

ИИ предсказал, каким будет абитуриент технического вуза в 2036 году

Подавляющее большинство абитуриентов к 2036 году будут чаще выбирать технические направления подготовки или дисциплины на стыке наук. Инженерные специальности постепенно станут менее мужскими, сам процесс поступления в вуз на 100% цифровизируется. Каким будет абитуриент спустя десятилетие?

Ученые МФТИ с помощью нейросетей построили прогноз, опираясь на данные Всероссийской переписи населения 2020 года, мониторинга качества приема в вузы, стратегию развития Физтеха до 2036 года, а также современные демографические и образовательные тренды. Директор по цифровизации образования МФТИ, и.о. директора Физтех-школы бизнеса высоких технологий МФТИ Дарья Гриц считает, что прогнозы с помощью ИИ могут помочь вузам выстраивать стратегии действий.

Генеративный искусственный интеллект не предсказывает будущее — он способен обратить внимание исследователя на неочевидные тренды. Это инструмент, который помогает увидеть, к чему мы придём, если будем действовать как сейчас. Если нас устраивает этот прогноз — замечательно. Если нет — значит, мы понимаем, куда стоит направить усилия. Так, например, если на образовательных программах по IT и инженерии сегодня обучается всего 15% девушек, а мы знаем, что разнообразие влияет на качество решений и устойчивость развития, — это сигнал, что нужно активнее работать с женским участием в STEM. Будущее не высчитывается ИИ — оно создается людьми. И в этом смысле ИИ — не пророк, а компас» — убеждена Дарья Гриц.

Продолжение статьи: https://mipt.ru/news