

АТРАМЕНТОВА Л. А.

Харьковский национальный университет имени В.Н. Каразина,

Украина, 61022, г. Харьков, пл. Свободы, 4, e-mail: lubov.atramentova@gmail.com, (095) 662-67-36

БАЙЕСОВСКАЯ СТАТИСТИКА В ГЕНЕТИКЕ ЧЕЛОВЕКА

Применение метода Байеса в генетическом консультировании позволяет уточнить вероятность заболевания у потомства, что в ряде случаев из-за рассчитанного крайне низкого риска может избавить от дополнительных лабораторно-диагностических исследований. Статья посвящена применению метода Байеса в генетике человека. Вводятся понятия априорной, условной, совместной и апостериорной вероятности. В качестве примера рассмотрена родословная с моногенным рецессивным X-сцепленным признаком. Приводится алгоритм расчета вероятности того, что консультирующийся является носителем патологического гена. Сравнивается результат прогноза, выполненный классическим методом и с помощью метода Байеса. Приводится пример, как на основе эпидемиологических и генетических данных рассчитать скорость мутирования гена. Рассмотренная тема может быть принята во внимание при изложении курсов генетики человека и медицинской генетики в классических и медицинских университетах, а приведенная задача служит примером для составления заданий для студентов с родословными различной конфигурации и другими признаками.

Ключевые слова: метод Байеса, априорная вероятность, апостериорная вероятность, преподавание генетики, медико-генетическое консультирование.

Томас Байес (1702–1761 гг.) – английский математик и священник. Он создал теорию, которую все активнее используют современные биологи для анализа данных [1, 2]. В основе теоремы Байеса – одной из основных теорем элементарной теории вероятностей – лежит идея об уточнении вероятности события с учетом дополнительной информации. По формуле Байеса можно более точно рассчитать вероятность, взяв в расчет как ранее известные факты, так и данные новых наблюдений. Идея, выраженная уравнением Байеса, формулируется так:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

выяснить вероятность события А, если оно связано с событием В, у которого есть собственная вероятность. Вероятность исходного события (А) в байесовской статистике называется *априорной*. Вероятность нового события (В), связанного с исходным событием (А), обозначается как *условная*. Используя исходную и дополнительную информацию, рассчитывают *апостериорную*, то есть окончательную, вероятность, которая в свою очередь может стать априорной для решения следующей задачи. Запись $P(A/B)$ означает вероятность события А при наступлении события В. $P(A/B)$ – это апостериорная вероятность, расчет которой составляет суть байесовского анализа: вычислить вероятность события А при условии наступления события В [1].

Рассмотрим, как работает метод Байеса в генетике [3].

Пример. В роду у Дианы, ожидающей ребенка, был случай гемофилии (рис.), и она хочет знать риск гемофилии для своего будущего сына.

Решение.

Исходная информация: гемофилия – сцепленный с X хромосомой рецессивный признак. Болезнь вызвана аллельным вариантом *h* гена *H*, локализованного в X хромосоме. У женщин две X хромосомы, и они имеют две дозы этого гена (генотип *HH*). Мужчины обладают одной X хромосомой, следовательно, имеют только одну дозу этого гена (генотип *H-*). Аллельный вариант *h* очень редок, поэтому больные гемофилией женщины (генотип *hh*) практически не встречаются, а гетерозиготные носительницы гена гемофилии *Hh* здоровы. Согласно закономерности наследования, вероятность рождения больного гемофилией сына (генотип *h-*) у таких женщин равна 0,5. Решение задачи сводится к выяснению вероятности того, что Диана является

ся носительницей гена гемофилии, то есть имеет генотип Hh .

Вот классическое решение задачи. Фенотип Бориса (рис.) однозначно указывает на его генотип h . Ген h Борис мог получить только с X хромосомой матери, так как от отца мальчики получают Y хромосому, в которой такого локуса нет. Из этого следует, что мать Бориса Анна является гетерозиготной носительницей гена гемофилии (Hh). Ее дочь Вера унаследует ген h с вероятностью $1/2$, внучка Галина – с вероятностью $1/4$, правнучка Диана – с вероятностью $1/8$. Отсюда риск гемофилии для будущего сына Дианы составляет $1/16$ (6,25 %).

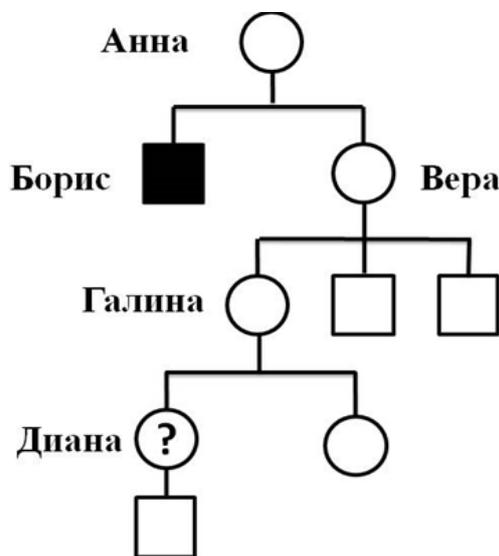


Рис. Родословная со случаем гемофилии.

С помощью байесовского метода задача решается иначе. Для этого используется дополнительная информация [4] (в качестве примера приведены ориентировочные числа):

- ✓ на 100 000 мужчин приходится 10 больных гемофилией;
- ✓ примерно в 20 % случаев мать мальчика, больного гемофилией, не является носительницей этого рецессивного гена.

Информация «10 больных на 100 000 мужчин» дает возможность рассчитать частоту аллеля гемофилии в популяции. У мужчин частота аллеля, сцепленного с X хромосомой, совпадает с частотой фенотипа, поэтому $p_h=0,0001$. Вероятность того, что потомок унаследует от матери любую из двух ее X хромосом, равняется $1/2$. Следовательно, только половина носительниц гена гемофилии передали потомкам хромосомы с аллелем h . Значит, среди 100 000 женщин, родивших 99990 здоровых и 10 боль-

ных сыновей, предположительно 20 были гетерозиготными носительницами гена h . Это соответствует частоте аллеля h , равной $p_h=0,0001$; она такая же, как и в мужской части населения, что свойственно равновесной панмиктической популяции.

Информация о том, что у 20 % больных мальчиков матери не имели в генотипе гена h , означает, что не десять, а только восемь из этих женщин были гетерозиготными. Всего же таких гетерозиготных женщин было в два раза больше, поскольку половина из них передали сыновьям хромосому с нормальным геном. Отсюда следует, что в женской части населения 0,016 % женщин являются носительницами гена гемофилии. Значит, вероятность иметь генотип Hh составляет 0,00016. Это является априорной вероятностью для Анны быть носительницей гена h , соответственно, вероятность не быть носительницей 0,99984.

Информация о рождении больных мальчиков (h -) от женщин не носительниц (HH) дает возможность рассчитать частоту мутирования $H \rightarrow h$. У 99984 женщин с генотипом HH родилось два больных гемофилией сына, получивших ген h как новую мутацию. Число 0,00002 ($2/99984$) является частотой мутирования и вероятностью для генетически здоровой женщины родить сына с гемофилией. Соответственно, с вероятностью 0,99998 ($1 - 0,00002$) у женщины с генотипом HH родится здоровый сын, два здоровых сына – с вероятностью $0,99998^2$ и т. д.

Все сказанное означает, что Анна априорно (до рождения у нее сыновей) с вероятностью 0,00016 является носительницей гена гемофилии (Hh) и с вероятностью 0,99984 не носительницей (HH) (табл. 1). Для нее имеется условие: один больной сын. Если генотип Анны Hh , то вероятность рождения одного больного сына равняется $1/2$, если ее генотип HH , то вероятность, что у нее произойдет мутация, которую унаследует сын, составляет 0,00002. В случае, если бы у нее было два больных сына, то вероятность такого события равнялась бы $0,00002^2$, если три – $0,00002^3$.

Расчеты, приведенные в табл. 1, показали, что Анна является гетерозиготной носительницей гена гемофилии с вероятностью 80 %. Вероятность для ее дочери Веры унаследовать этот ген составляет $1/2$, следовательно, априорно она носительница гена гемофилии с вероятностью 40 %. Используем эту вероятность как априорную для Веры и рассчитаем

окончательную вероятность, что она гетерозиготна. Учтем, что у нее двое здоровых сыновей. Вероятность рождения у гетерозиготной женщины двух здоровых сыновей равняется $1/4$ $(1/2)^2$. Если Вера нормальная гомозигота, то рождение у нее двух здоровых сыновей практически равняется единице (табл. 2).

Вера гетерозиготна с вероятностью 14 %, значит, априорная вероятность для ее дочери Галины иметь ген гемофилии составляет 7 %. У Галины нет сыновей, поэтому дополнительной информации для нее не имеется.

Вероятность, что дочь Галины унаследует ген гемофилии, составляет $1/2$, поэтому Диана априорно гетерозиготна с вероятностью 3,5 %. При вычислении окончательной вероятности, что Диана носительница, учтем, что у нее есть здоровый сын (табл. 3). Расчеты дают оценку в 1,8 %, а это означает, что риск гемофилии для следующего сына меньше 1 %, а это примерно в семь раз меньше, чем риск, рассчитанный классическим способом.

Таблица 1. Расчет вероятности носительства для Анны

Вероятность	Анна	
	<i>Hh</i>	<i>HH</i>
Априорная	0,00016	0,99984
Условная	0,5	0,00002
Совместная	0,00008	0,00002
Сумма	0,0001	
Апостериорная	0,8	0,2

Примечания: *Hh* и *HH* – возможные генотипы Анны. Априорная вероятность – частота генотипа в женском населении. Условная вероятность – рождение одного больного сына. Совместная вероятность – произведение априорной и условной вероятностей. Сумма – сложение альтернативных совместных вероятностей. Апостериорная вероятность – доля от суммы совместных вероятностей.

Таблица 2. Расчет вероятности носительства для Веры

Вероятность	Вера	
	<i>Hh</i>	<i>HH</i>
Априорная	0,4	0,6
Условная	$(1/2)^2 = 0,25$	$0,99998^2$
Совместная	0,1	0,6
Сумма	0,7	
Апостериорная	0,14	0,86

Примечания: *Hh* и *HH* – возможные генотипы Веры. Априорная вероятность носительства – половина вероятности для Анны. Условная вероятность – рождение двух здоровых сыновей. Совместная вероятность – произведение априорной и условной вероятностей. Сумма – сложение альтернативных совместных вероятностей. Апостериорная вероятность – доля от суммы совместных вероятностей.

Таблица 3. Расчет вероятности носительства для Дианы

Вероятность	Диана	
	<i>Hh</i>	<i>HH</i>
Априорная	0,035	0,965
Условная	0,5	0,99998
Совместная	0,018	0,965
Сумма	0,983	
Апостериорная	0,018	0,982

Примечания: *Hh* и *HH* – возможные генотипы Дианы. Априорная вероятность носительства – половина вероятности ее матери Галины. Условная вероятность – рождение одного здорового сына. Совместная вероятность – произведение априорной и условной вероятностей. Сумма – сложение альтернативных совместных вероятностей. Апостериорная вероятность – доля от суммы совместных вероятностей.

Как видим, применение метода Байеса в генетическом консультировании позволяет уточнить вероятность моногенного заболевания у потомства, что в ряде случаев из-за крайне низкого риска может избавить от дополнительных лабораторно-диагностических исследований.

Рассмотренная тема может быть принята во внимание при изложении курсов генетики человека и медицинской генетики в классических и медицинских университетах, а приведенная задача служит примером для составления задач с родословными различной конфигурации и другими X-сцепленными признаками.

References

1. Gmurman V.E. Teoriia veroiatnostey i matematicheskaia statistika: uchebnoe posobie. 12-e izdanie, pererabotannoe. Moskva: Vysshee obrazovanie, 2008. 479 s. [in Russian] / Гмурман, В.Е. Теория вероятностей и математическая статистика: учебное пособие. 12-е издание, переработанное. М.: Высшее образование, 2008. 479 с.
2. Trukhacheva N.V. Matematicheskaia statistika v mediko-biologicheskikh issledovaniiah s primeneniem paketa Statistica. Moskva: GEOTAR-Media, 2012. 384 s. [in Russian] / Трухачева Н.В. Математическая статистика в медико-биологических исследованиях с применением пакета Statistica. М.: ГЭОТАР-Медиа, 2012. 384 с.
3. Merfi E.A., Cheyz G.A. Osnovy mediko-geneticheskogo konsul'tirovaniia. Izdatel'stvo. Moskva: Meditsina, 1979. 398 s. [in Russian] / Мерфи Э.А., Чейз Г.А. Основы медико-генетического консультирования. М.: Медицина, 1979. 398 с.
4. Ginter E.K. Monogennoe nasledovanie priznakov u cheloveka. V sb. Nasledstvennye bolezni: natsional'noe rukovodstvo. Moskva: GEOTAR-Media, 2012. S. 126–146. [in Russian] / Гинтер Е.К. Моногенное наследование признаков у человека. В сб. Наследственные болезни: национальное руководство. М.: ГЭОТАР-Медиа, 2012. С. 126–146.

ATRAMENTOVA L.A.

*V.N. Karazin Kharkiv National University,
Ukraine, 61022, Kharkiv, Svobody sq., 4, e-mail: lubov.atramentova@gmail.com*

BAYESIAN STATISTICS IN HUMAN GENETICS

Using of the Bayes' method in genetic counseling allows clarifying of the probability of a disease in the offspring, which in some cases can save probands from additional laboratory diagnostic tests due to the extremely low risk which can be calculated. The article is devoted to the usage of the Bayes' method in human genetics. The concepts of a priori, conditional, total and posterior probabilities are introduced. As an example, a pedigree with a monogenic recessive X-linked trait is considered. An algorithm for calculating of the probability that a patient is a carrier of a pathological gene is presented. The prediction of results made with help of using the classical method of calculations and the Bayesian method are compared. An example for calculation of the gene mutation rate based on epidemiological and genetic data is given. The discussed topic can be taken into account for presenting of courses in human genetics and medical genetics at classical and medical universities, and the given problem can serve as an example for drawing up of tasks for students with pedigrees of various configurations and other attributes.

Keywords: Bayesian method, a priori probability, posterior probability, teaching of genetics, genetic counseling.

ATRAMENTOVA L. O.

*Харківський національний університет імені В. Н. Каразіна,
Україна, 61022, м. Харків, майд. Свободи, 4, e-mail: lubov.atramentova@gmail.com*

БАЙЄСІВСЬКА СТАТИСТИКА В ГЕНЕТИЦІ ЛЮДИНИ

Застосування методу Байєса в генетичному консультиванні дозволяє уточнити ймовірність захворювання у нащадків, що в багатьох випадках розрахунок вкрай низького ризику може позбавити пробанда від додаткових лабораторно-діагностичних досліджень. Стаття присвячена застосуванню методу Байєса в генетиці людини. Вводяться поняття апіорної, умовної, спільної та апостеріорної ймовірності. Як приклад розглянуто родовід з моногенною рецесивною X-зчепленою ознакою. Наводиться алгоритм розрахунку ймовірності того, що той, хто консультиється, є носієм патологічного гена. Порівнюється результат прогнозу, виконаної за допомогою класичного методу і методу Байєса. Наводиться приклад того, як на основі епідеміологічних і генетичних даних розрахувати швидкість мутування гена. Розглянута тема може бути взята до уваги під час викладання курсів генетики людини та медичної генетики в класичних і медичних університетах, а наведена задача може бути прикладом для складання завдань для студентів із родоводами різної конфігурації та іншими ознаками.

Ключові слова: метод Байєса, апіорна ймовірність, апостеріорна ймовірність, викладання генетики, медико-генетичне консультивання.