

---

## ПЕРВАЯ СТЕПЕНЬ

---

**А.О. ДОМАНОВ\***

### **ОСНОВЫ БАЙЕСОВСКОГО ПОДХОДА К КОЛИЧЕСТВЕННОМУ АНАЛИЗУ (НА ПРИМЕРЕ ЕВРОСКЕПТИЦИЗМА)**

*Аннотация.* В данной статье предпринята попытка выявить основные допущения, предпосылки и приемы методик, использующих теорему Т. Байеса для оценки взаимосвязи признаков социальных акторов и явлений. Выделены некоторые преимущества предложенного подхода по отношению к более традиционным количественным методам, а также ключевые направления исследований, которые могут быть усовершенствованы при помощи байесовских оценок коэффициентов. Речь идет о совместимости изложенного подхода, прежде всего, с теориями игр и принятия решений, event-анализом, скрытыми цепями Маркова, прогнозированием при помощи нейронных сетей и других предиктивных алгоритмов искусственного интеллекта.

Байесовский подход значительно отличается от традиционных статистических методов (прежде всего, он ориентирован на поиск наиболее вероятного, а не единственно верного значения коэффициента связи признаков). В связи с этим предложена графическая интерпретация таких базовых понятий и приемов, как вероятностный вывод, показатель с максимальным правдоподобием и байесовская сеть доверия.

---

\* **Доманов Алексей Олегович**, младший научный сотрудник Центра документации ЕС, Институт Европы РАН; преподаватель кафедры интеграционных процессов, научный сотрудник Центра пространственного анализа международных отношений Института международных исследований, МГИМО МИД России (Москва, Россия), e-mail: domanov.aleksey@gmail.com

Описанный инструментарий использован для проверки гипотезы о влиянии ухудшения качества жизни на рост евроскептицизма граждан всех стран ЕС. Дисперсионный и корреляционный анализ ответов 27 тыс. человек, опрошенных в рамках проекта «Евробарометр» в ноябре-декабре 2019 г., показал высокую степень правдоподобия данного предположения. Более того, байесовский подход позволил сделать вероятностный вывод о том, что более вероятно подтверждение именно этой гипотезы, а не связи евроскептицизма с нынешним финансовым положением респондентов (т.е. относительно большую объясняющую силу имеет сопоставление этой ситуации с прошлым).

*Ключевые слова:* байесовское оценивание; евроскептицизм; количественные методы; байесовская сеть доверия; качество жизни; ЕС.

*Для цитирования:* Доманов А.О. Основы байесовского подхода к количественному анализу (на примере евроскептицизма) // Политическая наука. – 2021. – № 1. – С. 301–321. – DOI: <http://www.doi.org/10.31249/poln/2021.01.13>

Рост интереса к методам, основанным на идеях английского математика XVIII в. Т. Байеса, вызван увеличением объема научного знания и компьютерных мощностей, позволяющих совершать множество итераций для поиска оптимальных коэффициентов. По мере количественной обработки все более обширного материала, накопленного различными дисциплинами, исследователи приступают к изучению неосвоенных тем, эмпирические данные о которых не соответствуют требованиям традиционного статистического анализа. В частности, затруднена обработка информации по некоторым политическим темам, касающимся редких явлений или событий: исторические наблюдения не позволяют построить достаточно большую выборку.

Поскольку с помощью байесовского подхода указанные проблемы можно решить или обойти, желательнее рассмотреть его принципы, заложенные основателями этой методики [Lynch, 2007; Kass, Raftery, 1995; McNeish, 2016; Raftery, 1995; van Ravenzwaaij, Cassey, Brown, 2018; Western, 1999]. При этом наиболее уместным представляется нетехническое изложение процедуры анализа взаимосвязей с использованием минимального математического аппарата. Примеры из политической сферы, приведенные по ходу этого описания, помогли бы приложить эти методы к изучению политики. В частности, наглядным примером могло бы стать использование байесовского инструментария для проверки гипотезы о влиянии восприятия качества жизни на уровень евроскептицизма граждан ЕС.

## Роль условной вероятности в байесовских вычислениях

Цель большинства количественных исследований – охарактеризовать генеральную совокупность объектов, оценив какой-либо ее параметр. Дальнейшие иллюстрации будут приведены на примере оценки показателя, который одновременно и достаточно прост, и широко используется в относительно сложных методах анализа в качестве компонент вычисляемых коэффициентов – частот распределения признака. Например, на измерении частот основаны многие типы регрессионного анализа: то, на сколько градусов нагревается вода за секунду при увеличении мощности нагревателей на 1 ватт, определяется пропорцией различных итоговых температур воды на различных конфорках к количеству этих электроплит (какой прирост температуры происходил чаще всего).

Часто у исследователя нет возможности напрямую узнать характеристики каждого объекта генеральной совокупности. Например, в случае опроса общественного мнения или анализа корпуса высказываний в социальных сетях проводить исчерпывающее исследование затратно, при анализе голосования с низкой явкой получить бюллетени от граждан, проигнорировавших выборы, и вовсе невозможно. По этой причине специалисты часто вынуждены полагаться на информацию о некой выборке из элементов генеральной совокупности, чтобы экстраполировать полученные выводы на объекты, не вошедшие в выборку.

Одна из предпосылок байесовского подхода – признание того, что любой способ конструирования выборки может быть несовершенным. Любая выборка потенциально нерепрезентативна, т.е. статистические коэффициенты и распределение измеряемого признака в генеральной совокупности могут отличаться от полученных эмпирически. В случае параметра, взятого для примера оценки (соотношение количества объектов с неким признаком и без него), допускается возможность того, что пропорция этих типов элементов генеральной совокупности расходится с обнаруженной долей выборки, обладающей этим свойством.

Предлагаемый подход учитывает эту неполноту нашей информации о генеральной совокупности, не полагаясь на закон больших чисел (как более традиционные статистические методы), а используя понятие *условной вероятности*. Эта пропорция частот удобна тем, что присутствует в любой выборке, сколь бы малой

она ни была (следовательно, независимо от применимости к ней закона больших чисел).

Вычислим условные вероятности для величин из табл. 1, позволяющие в дальнейшем оценить искомый параметр – найти пропорцию количества элементов в двух верхних ее ячейках к общему числу объектов, вероятность случайно попасть в эту зону генеральной совокупности:  $P$  (наличия признака) или, для краткости,  $P(П)$ . Стоит отметить, что для наглядности признаков, распределение которого требуется проанализировать, будем считать дихотомическим и фиксировать только наличие или отсутствие признака (без градаций).

Из табл. 1 можно выразить вероятности, относящиеся к компонентам верхнего ряда, следующим образом:

– вероятность обнаружить признак у объектов в выборке (или *при условии* принадлежности к выборке) называется условной  $P(П|B)$ <sup>1</sup>, поскольку вертикальная черта показывает дополнительное условие (которому соответствуют указанные элементы), и составляет  $40/(40+20)=2/3$ ;

– доля объектов вне выборки с признаком, т.е.  $P(П|-B)$ , равна  $60/60+80=3/7$ . Стоит отметить, что знаменатель в формуле расчета этого коэффициента изменился, поскольку новый коэффициент означает пропорцию признака среди других элементов: вне выборки, а не вошедших в нее.

Английский математик Т. Байес в XVIII в. заметил, что совместную вероятность нахождения объекта в левой верхней ячейке таблицы (включающей элементы с признаком и в выборке) можно выразить одновременно двумя такими парами сомножителей, в которых вероятности условные попарно совмещались бы с вероятностями их условий:

– и как долю  $P(П|B)$  выборки, которая в своей полноте занимает часть  $P(B)$  генеральной совокупности;

---

<sup>1</sup> Допустив нерепрезентативность выборки, исследователь предполагает, в указанных терминах, что следующее утверждение ложно: доля элементов с признаком во всей генеральной совокупности не равна доле таких объектов, для которых выполнено условие нахождения в выборке;  $P(П) \neq P(П|B)$ . К примеру, результаты опроса на выходе с избирательных участков (exit poll) обычно расходятся с итогами выборов из-за трудностей формирования репрезентативной выборки при сборе эмпирического материала таким способом.

– и как выборочную долю  $P(B | \Pi)$  от количества всех элементов с признаком (включая таковые за пределами выборки), которых насчитывается  $P(\Pi)$  размера генеральной совокупности.

Поскольку эта двояко выраженная доля элементов из левой верхней ячейки эквивалентна самой себе, то справедливо следующее равенство между попарными произведениями, обозначающими оба этих утверждения:

$$P(\Pi | B) \cdot P(B) = P(B | \Pi) \cdot P(\Pi)$$

Стоит отметить, что последний сомножитель является искомым значением параметра. Выразим его из приведенной *теоремы Байеса*, поделив обе части на  $P(B | \Pi)$ :

$$P(\Pi) = P(\Pi | B) \cdot P(B) / P(B | \Pi)$$

В этом произведении нам известен только первый множитель – из эмпирических данных, ведь это пропорция в одной из ячеек таблицы сопряженности. Доля выборки среди всех объектов и доля наших наблюдений в общем количестве элементов с признаком во всей генеральной совокупности неизвестна.

При оценке коэффициентов байесовским путем исходное уравнение также часто сокращают на  $P(B)$ :

$$P(\Pi | B) = P(B | \Pi) \cdot P(\Pi) / P(B).$$

### **Байесовский способ оценки статистических параметров**

Уравнения такого вида могут увязывать и другие характеристики выборки с остальными свойствами генеральной совокупности. Например, некоторые множители могут отражать различие между долей элементов с признаком в выборке (пропорцией, известной из эмпирического материала) и в генеральной совокупности (неизвестным соотношением): часто исследователь стремится узнать вторую, располагая информацией о первой. В терминах условной вероятности поставленная задача состоит в том, чтобы найти соотношение во всей генеральной совокупности при условии, что уже получены данные для расчета этого показателя в рамках выборки. В дальнейших обозначениях величина параметра в выборке приведена с маленькой буквы, по всей генеральной сово-

купности – с большой: требуется вычислить некий коэффициент с характеристиками вида [ *Ratio|ratio* ] или, для краткости, [ *R|r* ].

Компоненты этой записи можно поменять местами: например, вероятность получить в случайной выборке пропорцию  $r$ , если бы мы знали значение  $R$ , записывается как  $P(r|R)$ . Смысл обозначений с подобными аргументами в скобках раскрывается на примере приведения выборки в репрезентативное состояние: в ходе этого процесса аналитик стремится сделать соотношение элементов с признаком и без него в рамках выборки (например, социологического опроса;  $r$ ) максимально близким к значению этого показателя по результатам переписи населения (почти<sup>1</sup> равному  $R$ ). Многие традиционные количественные методы недооценивают воздействие разницы  $r$  и  $R$  на итоги исследования: закон больших чисел позволяет нивелировать это различие, взяв бесконечное количество выборок из одинаковой генеральной совокупности (тогда значение пропорции  $r$ , усредненное по бесконечному количеству выборок, было бы примерно равно  $R$ ).

Однако байесовский подход фокусируется на этой разнице, поскольку фактическое количество выборок, доступных для анализа, не бесконечно, а конечно; часто – единично (если речь не идет, например, о метаисследованиях). При анализе политических феноменов указанное обстоятельство тем более важно: стремясь сделать выводы о событиях какого-либо типа (например, революциях; не произошедших в конкретные годы, а как о классе явлений), мы располагаем информацией о них лишь на ограниченном временном промежутке. Более того, если этот период времени ограничен не хронологическими рамками и источниковой базой, определенными исследователем, а более фундаментально – исследовательским вопросом – то необоснованность допущения о бесконечном количестве выборок становится еще более явной. Например, задача изучить революции XIX в. предполагает, что сама изучаемая генеральная совокупность ограничена числом этих преобразований, поэтому и количество возможных выборок из нее конечно.

Приведенная запись  $P(r|R)$  в более общем виде представляет собой вероятность получить эмпирические данные (по которым

---

<sup>1</sup> Если количество людей, не охваченных переписью, считать пренебрежимо малым.

уже сделаны расчеты о распределении признака в выборке и рассчитаны различные характеристики выборки, в том числе пропорция  $r$ ) при следующем условии: что признак распределен в генеральной совокупности определенным образом (например, что он встречается с частотой  $R$ ), а именно как  $P$  (данные|параметр распределения).

В теореме Байеса указанная условная вероятность занимает ту же позицию, которую занимали бы  $P(B|H)$  или  $P(r|R)$  в своих соответствующих уравнениях; а именно первого множителя в числителе формулы, помогающей найти ответ на упомянутый исследовательский вопрос – значение параметра в том случае, если эмпирический материал уже собран.

$$P(\text{параметр}|\text{данные}) = \frac{P(\text{данные}|\text{параметр}) \cdot P(\text{параметр})}{P(\text{данные})}$$

Допуская существование коэффициента  $P(\text{параметр} | \text{данные})$ , мы признаем вероятностную природу параметров, а не только полученных данных. С точки зрения байесовского подхода имеющиеся данные позволяют полагать некоторые значения обобщенных параметров маловероятными (например, трудно ожидать, что монета является обычной неперекосенной, если из 10 бросков она девять раз выпала «орлом» и лишь один – «решкой»). Напротив, отдельные их значения крайне вероятны, поскольку «поддерживаются» проведенными наблюдениями (например, монета вполне может оказаться обыкновенной, т.е. параметр частоты «орлов» у нее равен  $\frac{1}{2}$ , а не  $\frac{9}{10}$ , если из 10 бросков она выпала «орлом» лишь шесть раз).

Допуская, что размер искомого параметра может быть не единственно верным, а многовариантным с некоторой вероятностью для каждого потенциально значения, байесовский подход расходится с традиционными количественными методами. Дело в том, что рассчитываемый с их помощью показатель значимости дает основания усомниться лишь в одном значении коэффициента, а не оценивать обоснованность нескольких сразу.

Эта степень «достоверности» каждого значения параметра, обозначенная ранее как  $P(\text{параметр}|\text{данные})$ , называется *апостериорной* вероятностью, поскольку рассчитывается после получения данных. Напротив, *априорная* вероятность неформально опре-

деляется как вероятность зафиксировать некоторые значения искомого коэффициента (для каждого – своя), которую наблюдатель предполагает перед проведением наблюдений или эксперимента – указанная  $P$  (*параметр*). В свою очередь, эти предположения могут строиться на основе литературы о более ранних исследованиях интересующего объекта: например, что вероятность увидеть черного лебедя близка к нулю, поскольку якобы до того момента встречались только белые лебеди (в этом случае встреча с птицей этого цвета при расширении выборки на австралийских животных приведет к тому, что апостериорная вероятность увидеть черного лебедя будет пересчитана в сторону повышения).

Для сравнения обоснованности различных значений параметров их апостериорные вероятности вычисляются по последней из указанных формул. Поскольку целью является сравнение правдоподобий, а не расчет их абсолютных значений, опускается величина  $P$  (*данные*): она одинакова для всех вариантов, так как их апостериорные вероятности рассчитываются по одной и той же выборке наблюдений с тождественными характеристиками. Если перед проведением исследования все допустимые значения предполагаются равновероятными, то множителем  $P$  (*параметр*) тоже пренебрегают: при дальнейшем сравнении сократятся величины  $P$  (*значение<sub>1</sub> параметра*) =  $P$  (*значение<sub>2</sub> параметра*) = ...

Таким образом, в этом случае (когда априорные вероятности равномерно распределены в допустимом диапазоне)<sup>1</sup> на результат проводимого сравнения (множества апостериорных вероятностей различных значений оцениваемого параметра) воздействует лишь оставшийся сомножитель формулы –  $P$  (*данные|параметр*). Поскольку эта величина сама зависит от искомой характеристики, анализируется именно эта функциональная зависимость, а конкретно – сопоставляются результирующие показатели при разных значениях параметра. При этом его уровни лишь предполагаются исследователем: он «настраивает» искомый коэффициент таким образом, чтобы конфигурацию элементов разных типов в выборке можно было считать не маловероятным статистическим выбросом,

---

<sup>1</sup> Обычно полностью избавиться от предубеждений (добиться полностью равномерного распределения – не исключать ни одно значение параметра и полагать равновероятными все в диапазоне от  $-\infty$  до  $+\infty$ ) было бы некорректно, поскольку многие искомые коэффициенты опираются на частоты признаков, а они не принимают значения меньше нуля и больше единицы.



а одним из рутинных проявлений описываемой закономерности (которое можно повторить с тем же результатом – получив ту же пропорцию элементов с признаком и без, как и в эмпирическом материале). Определившись с диапазоном потенциальных значений параметра (в нашем примере эта пропорция может составлять от 0 до 1), исследователь сопоставляет свои предположения в ходе симуляции случайных распределений признака методом Монте-Карло (усовершенствованного генератора случайных чисел и объектов выборки).

Показатели  $P$  (*данные | параметр<sub>n</sub>*), которые вычисляются и сравниваются для каждого аргумента этой функции (т.е. для любого потенциального значения искомого параметра), представляют собой классическую оценку так называемого *правдоподобия*<sup>1</sup> – совместной вероятности появления объектов с определенными характеристиками в случайной выборке. Поскольку в описанном примере с проверкой простого наличия или отсутствия признака вероятность обнаружить элемент с признаком –  $P(\Pi)$  – не достигает единицы, шансы на это (произведение равных вероятностей для каждого из них, т.е.  $P(\Pi)$  многократно) также меньше единицы. Примечательно, что эта совместная вероятность – так называемая – уменьшается по мере того, как исследователь предполагает все меньшее значение параметра (ведь перемножаются все меньшие числа), и наоборот.

Чем больше элементов с признаком было фактически найдено в ходе эмпирического исследования, тем больше множителей меньше единицы в этом произведении. Следовательно, уменьшая предполагаемое значение параметра, исследователь снижает совместную вероятность обнаружить большое количество элементов с признаком. Тем не менее эти объекты были найдены, а значит, вероятность такого события была не настолько мала, как предполагалось.

Следуя этой логике, находится баланс между количеством объектов с признаком в эмпирическом материале (т.е. количеством

---

<sup>1</sup> С точки зрения байесовского подхода, исходящего из субъективности оценок, характерно стремление англоязычных авторов избежать ассоциаций с аналогами русских слов «правда» или «истина». Соотечественники Т. Байеса предпочитают обозначать правдоподобие более релятивистским термином *likelihood* (который тоже отличается от перевода русской «вероятности» – *probability*).

этих множителей величиной меньше 1) и желанием проверить вариант крайне низкого искомого коэффициента. Дисбаланс возникает при попытке занизить значение параметра, несмотря на свидетельства того, что признак широко распространен во взятой выборке: произведение множества околонулевых вероятностей является крайне малым числом. Напротив, если эти расчеты повторить с большим значением параметра (т.е., предположить относительно широкое распространение признака в генеральной совокупности, а не только в выборке), то результат будет выше, поскольку пройдет перемножение чисел, находящихся ближе к единице<sup>1</sup>.

В итоге произведение достигает наивысшего уровня (т.е. проявляется *максимальное правдоподобие*) в тех случаях, когда предполагаемое значение параметра не противоречит имеющимся данным. Проверив все возможные уровни искомого параметра, исследователь выявляет те раунды, которые не привели к серьезным отклонениям от эмпирики, и после этого наиболее вероятными значениями коэффициента считаются обладатели наибольшего правдоподобия.

Описанная процедура оценки того, насколько «необычно» существование полученной выборки наблюдений для каждого нашего предположения о значении параметра, выражается в следующих расчетах для монеты, которая выпала «орлом» четыре раза из пяти. Выдвинув гипотезу о какой-либо склонности монеты выпадать «решкой» лишь в редких случаях (допустив, что общая вероятность «орла» равна, к примеру, 0,7), можно подставить степень этой склонности в формулу правдоподобия: перемножить вероятность исходов одного типа столько раз, сколько они произошли; а также проделать ту же операцию с наблюдениями другого вида.

$P(4 \text{ «орла» и } 1 \text{ «решка»} \mid \text{степень «фальшивости» монеты и вероятность выпадения «орла» равны } 0,7) = 0,7^4 \cdot (1-0,7)^1 = 0,07203.$

Стоит подчеркнуть, что при этом значения параметра с меньшим уровнем правдоподобия не отбрасываются полностью,

---

<sup>1</sup> Однако попытка рассчитать правдоподобие для слишком большого значения параметра приведет к примерно такой же ситуации, что и при его экстремально низком гипотетическом уровне: так как вероятность появления элемента без признака дополняет упомянутое значение до единицы, на этот раз близкими к нулю будут множители, соответствующие этим объектам.

как поступили бы с промежуточными результатами многие традиционные статистические процедуры. Допускается, что величина искомого коэффициента может быть далека от максимально правдоподобного: что распределение признака в генеральной совокупности абсолютно не похоже на пропорции, зафиксированные в ходе наблюдения, а произошедшее искажение выборки – случайность (настолько маловероятная, насколько низок показатель правдоподобия для этого варианта). К примеру, гипотеза об обычном характере монеты (что вероятность «орла» равна  $\frac{1}{2}$ ) считается менее «правдоподобной» чем предположение о фальшивом, если «орел» наблюдался четыре раза, а «решка» – один:  $P(\text{данные} \mid \text{вероятность «орла» } \frac{1}{2}) = (\frac{1}{2})^4 \cdot (1-\frac{1}{2})^1 = 0,03125$ , т.е. меньше рассчитанного ранее значения  $P(\text{данные} \mid 0,7)$ .

Ключевое свойство байесовского подхода состоит в том, что после этого промежуточного вывода коэффициент  $\frac{1}{2}$  не игнорируется, а лишь признается менее обоснованным; хотя и тоже возможным (в отличие от традиционных количественных методов, результатом применения которых является лишь одно значение параметра). Из-за того, что в результате расчета часто выделяют множество маловероятных, ряд возможных и один или несколько крайне вероятных коэффициентов связи, такой вывод называется *вероятностным*.

Описанная методика расчета позволяет вычислять правдоподобие более сложных параметров: например, коэффициентов взаимосвязи двух и более признаков, т.е. проверять гипотезы более широкого спектра. Это возможно из-за того, что теорема Байеса применима не только к анализу соотношения содержимого выборки с распределением одной переменной, но и к таблицам сопряженности (выражающим совместное распределение двух переменных).

Как и при обработке данных о единственной переменной, задача оценить взаимосвязь формулируется иначе, чем при традиционных статистических расчетах. Оцениваются шансы обнаружить в генеральной совокупности то распределение значений признаков (в том числе, совместное распределение), которое соответствует условным вероятностям, выявленным в выборке.

Тем не менее предлагаемая операция аналогична вопросу традиционного статистического подхода о допустимости отклонить нулевую гипотезу. Исходя из полученных наблюдений, оценивается правдоподобие различных значений коэффициента связи;

затем вероятность того, что этот параметр равен нулю, сравнивается с правдоподобием различных ненулевых значений. Оценка правдоподобия предполагает формулировку промежуточных вопросов в следующем виде: если на данной выборке значения одной переменной позволяют вычислить значения другой в  $p\%$  случаев, то насколько правдоподобно, что в генеральной совокупности первый признак детерминирует второй в той же мере?

Процедура расчета правдоподобия для сравнения гипотез о взаимосвязи (либо нулевой и альтернативной, либо нескольких альтернативных) аналогична описанному способу оценки коэффициентов для единственной переменной. Критерием наличия связи переменных считается высокий уровень правдоподобия для ненулевых предположительных коэффициентов связи.

Поскольку байесовский подход предполагает сравнение правдоподобия разных значений параметров связи, этим способом можно сравнивать на одном и том же эмпирическом материале несколько гипотез разного типа:

– как о различных видах связи двух переменных (не только о разных значениях коэффициентов, но и о нормальной, биномиальной или другой форме распределения одного признака по значениям другого);

– так и о связи нескольких переменных с зависимой – с целью выявить признак, связанный с ней наиболее тесно или сильно.

Сравнение моделей причинно-следственных связей (т.е. выбор между ними, называемый в англоязычной литературе Bayesian model selection) проводится по индикатору, называемому байесовским фактором. Он позволяет утверждать, насколько правдоподобие одного значения коэффициента превосходит правдоподобие другого (хотя методика его расчета включает не только условную вероятность тестируемой модели, из-за чего разница между значениями фактора обычно контрастнее, чем между двумя индикаторами правдоподобия. Таким образом, выявляется максимально достоверная («правдоподобная») гипотеза, наиболее подходящая к полученным данным.

Поскольку можно рассчитать вероятность обнаружить признак не только при одном условии, но и при двух, трех и более  $P(\Pi | a_1, b_1, \dots)$ , то степень правдоподобия определима и для моделей со сложной конфигурацией нескольких взаимосвязей. Теорема Байеса позволяет вычислить коэффициенты взаимодействия

независимых и промежуточных переменных, а по ним – соотношение этого подмножества признаков с зависимыми величинами общей гипотезы исследования.

Вычисления проводятся на основе расширенного описания соотношений между вероятностями значений переменных – *байесовских сетей доверия*, называемых также *графовыми вероятностными моделями* (см. рис.). На таких схемах условные вероятности представлены направленными ребрами, соединяющими вершины-предки с переменными, разбивающими совокупность элементов с «родительским» значением какого-либо признака. Такая точка зрения на условное распределение проявлений этого свойства и направленный характер этого графа позволяет проверять модели причинно-следственных связей.

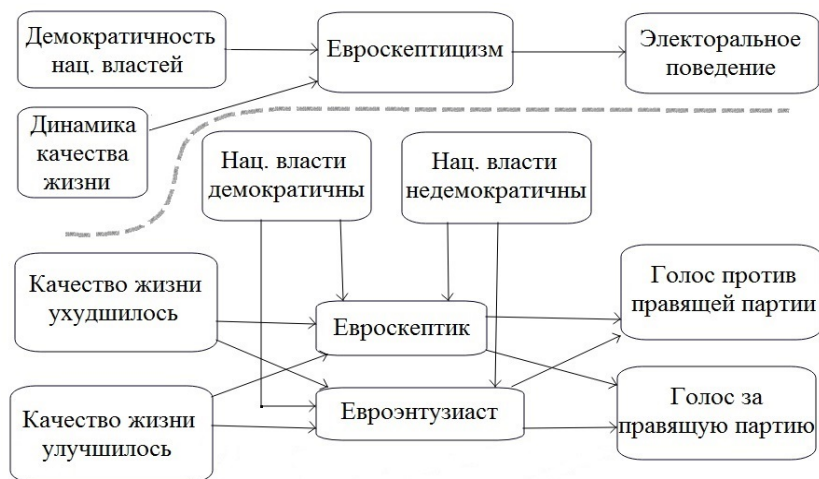


Рис.  
**Представление байесовской сети доверия  
через условные вероятности**

Распространенный способ проверки связей одного из типов – взаимовлияния разных переменных при их совместном воздействии на одну и ту же зависимую – многоуровневый регрессионный анализ (другие названия – регрессия иерархическая, вложенная или со смешанными эффектами). Данная процедура позволяет установить размер эффекта от группировки объектов с различными значениями

признака. Обычно сначала распределение признака по выборке анализируется такими привычными методами, как линейная регрессия, и затем в уравнение включается информация о принадлежности объектов каким-либо группам и обобщенных характеристиках этих категорий (например, общей доле проголосовавших за праворадикальные партии в каждой из стран – обычно эти агрегированные значения заменяют отсутствующие данные об индивидуальных предпочтениях каждого избирателя). В результате прослеживается взаимодействие переменных на разных уровнях анализа – индивидуальном и групповом.

В демонстрационной модели факторов евроскептицизма на рис. также присутствует несколько уровней разбиения электората. В частности, на одном уровне выделяются две категории людей: считающих национальные власти демократичными и недемократичными. Кроме того, внутри этих групп выделяются категории второго уровня: те, чье качество жизни ухудшилось и улучшилось за последнее время. При необходимости исследователи меняют уровни местами для того, чтобы оценить влияние принадлежности единиц анализа к наиболее общим группам, выделенным по другому признаку. Так, часто выявляется влияние гражданства людей на уровень их евроскептицизма, а не наоборот (в этом случае более абстрактное разделение – на первом уровне – проводится по проживанию в какой-либо из стран, более детальное – по степени евроскептичности жителей каждого из выделенных государств).

Такое описание выглядит особенно подходящим для моделирования цепи событий. В этом случае направление разбиения выборки вершинами-потомками демонстрирует последовательность явлений, а также вероятность наступления нового события после определения исхода предыдущего (например, склонность голосовавших за правящую партию выходить на демонстрации после обнаружения фальсификаций на выборах).

Алгоритмы вероятностного вывода широко используются для обработки дерева игры в теории игр (переменными считаются последовательные ходы игроков), в теории принятия решений (события, разбивающие выборку в узлах ветвления, вводятся по инициативе человека) и многих областях event-анализа.

Благодаря акценту на расчете условной вероятности обнаружить какой-либо признак у объекта по его другим зафиксированным характеристикам байесовские сети получили распростра-

нение в прогнозных аналитических исследованиях и разработках предиктивного искусственного интеллекта. Специалисты по машинному обучению широко используют описанные методы, поскольку определить количество будущих событий невозможно, а значит, предполагать репрезентативность доступной выборки заведомо бесперспективно (тем более если она мала, т.е. события редки или начались недавно – как при прогнозе поведения систем, только что вошедших в фазу кризиса). Например, нейронные сети для временных рядов оценивают наиболее вероятный тип следующего элемента, сопоставляя условные вероятности согласно предложенной схеме взаимосвязей множества независимых переменных.

При этом последовательные соединения переменных могут быть проанализированы и в обратном направлении – противоположном ребрам графа. В частности, скрытые цепи Маркова с помощью байесовских методов помогают обнаружить наиболее вероятные события, предшествовавшие зафиксированным явлениям.

### **Байесовский способ оценки взаимосвязи евроскептицизма граждан и некоторых его факторов**

Продемонстрировать применение байесовских методов можно на примере отдельной закономерности с рис., предполагавшейся в одной из работ о факторах евроскептицизма граждан ЕС [Доманов, 2020]. На основании прошлых исследований политологов и когнитивных психологов в ней выдвинута гипотеза следующего вида: люди, доход которых снизился в последнее время (не только малообеспеченные, но и предприниматели, испытывающие затруднения), склонны сопротивляться интернационализации в целом и европейской интеграции в частности.

Не вдаваясь в обзор литературы, обобщенной в упомянутой статье, можно свести механизм действия этого фактора к следующему: граждане, потерпевшие какие-либо убытки от трансграничной торговли, миграции и перетока инвестиций, склонны не замечать выгоды интернационализации и делать акцент на ее издержках. Вследствие этих когнитивных особенностей такие пострадавшие стремятся защититься от продолжения указанных неблагоприятных

(для них) процессов, и поддерживают политиков, предлагающих принять евроскептические и протекционистские меры.

Как показано далее, эта закономерность выявлена на материале опроса общественного мнения «Евробарометр». Ко времени проведения данной работы в открытом доступе находилась база данных за 14 ноября – 13 декабря 2019 г.<sup>1</sup> Анкетирование проводилось среди жителей 33 государств, однако далее будут рассмотрены ответы 27 464 опрошенных, проживавших только в государствах – членах ЕС.

Проанализированы ответы респондентов на вопрос о независимой переменной в следующей формулировке: «По каждому из следующих утверждений скажите, пожалуйста, вы с ним полностью согласны, склонны согласиться, склонны не согласиться или абсолютно не согласны: в целом, что касается качества Вашей жизни, раньше было лучше». О зависимой – «Скажите, пожалуйста, в какой степени Вы согласны или не согласны со следующими утверждениями: [нашу страну] могло бы ожидать лучшее будущее вне ЕС». Стоит отметить, что значения этих переменных взяты не из дополнительных источников – например, из пресс-релиза Европейской комиссии или отчетов центральных избирательных комиссий, – а из того же опроса общественного мнения.

Множители, подставляемые в уравнение Байеса (условные вероятности наблюдений; доли респондентов с одним признаком среди опрошенных с другим признаком), выведены из таблицы сопряженности вариантов ответа респондентов. Обработка базы данных проводилась в компьютерной программе Jasp, специализирующейся на байесовских вычислениях.

В табл. 2 приведены частоты респондентов с различными значениями обоих анализируемых признаков. При оценке коэффициентов связи байесовскими способами использованы такие условные вероятности, как доля крайних евроэнтузиастов среди тех, кто полностью не согласен с утверждением о снизившемся качестве жизни, –  $P$  (не лучше вне ЕС | качество жизни улучшилось), которая составляет 7,6% (или 2099/27464).

---

<sup>1</sup> База данных с индивидуальными ответами респондентов и характеристиками опроса доступна в Eurobarometer: Data Access // GESIS. Leibniz-Institut für Sozialwissenschaften. 2019. – Mode of access: <http://www.gesis.org/eurobarometer-data-service/search-data-access/data-access/> (accessed: 15.08.2020).



Таблица 2

**Распределение граждан ЕС с разным уровнем  
евроскептицизма по ретроспективным оценкам  
качества жизни (кол-во и некоторые доли)**

	Качество жизни было лучше	Полностью согласны	Скорее согласны	Скорее не согласны	Полностью не согласны	<i>Не знаю</i>	<b>Всего в строке</b>
Стране лучше вне ЕС							
Полностью согласны		1177 (4,3%)	749	406	193	101	<b>2626</b>
Скорее согласны		1295	2505 (9,1%)	1220	298	166	<b>5484</b>
Скорее не согласны		1100	2729	3326	653	429	<b>8237</b>
Полностью не согласны		1188	1807	3225 (11,7%)	2099 (7,6%)	448	<b>8767</b>
<i>Не знаю</i>		515	687	540	168	440	<b>2350</b>
<b>Всего в столбце</b>		<b>5275</b>	<b>8477</b>	<b>8717</b>	<b>3411</b>	<b>1584</b>	<b>27 464</b>

Наличие взаимозависимости признаков подтверждено дисперсионным анализом. Цель этой процедуры в байесовском исполнении та же, что и при оценке традиционными статистическими приемами: показать, что получить приведенные результаты опроса в отсутствие зависимости между ними было бы крайне маловероятно или, наоборот, вполне возможно. Проверяемое выражение можно обозначить как условную вероятность: *P (имеющиеся результаты | независимость переменных)*.

В ходе симуляции для оценки правдоподобия различных значений этого коэффициента выяснилось, что влияние воспринимаемой динамики качества жизни на уровень евроскептицизма «более правдоподобно», чем независимость этих переменных. Суммарная вероятность получить указанные итоги опроса в первом случае составила 0,932, во втором – лишь  $7,9 \cdot 10^{-75}$ .

Байесовская оценка коэффициентов дисперсионного анализа позволила сравнить не только нулевую и альтернативную гипотезы, но и предположения о влиянии различных переменных на зависимую. Степень правдоподобия упомянутой закономерности была сопоставлена с правдоподобием другой модели – связи евроскептицизма респондентов с их материальным положением (ответами на вопрос «Как бы Вы оценили ситуацию в каждой из следующих областей: финансовое положение Вашего домохозяйства?»).

В сравнении с моделью, предполагающей связь нежелания находиться в ЕС с финансовым положением респондентов, «более

правдоподобной» признана модель взаимодействия евроскептицизма и ухудшения качества жизни. Этот вывод сделан на основании того, что возможные потенциальные частоты переменных и взаимное распределение этих симулированных значений больше напоминает эмпирически полученные частотные таблицы наблюдаемых индикаторов в том случае, если это разнообразие возможных значений признаков порождено симуляцией на основе второй, а не первой схемы взаимозависимости переменных. В итоге вторая гипотеза считается более вероятной, поскольку соответствующие байесовские факторы для каждой из моделей равны 0,000343 и 0,857.

Как следует из методического раздела данной работы, байесовские методы позволяют не только оценить наличие связи признаков, но и охарактеризовать выявленную зависимость. В частности, с помощью этого подхода рассчитывается такая мера тесноты связи, как коэффициент корреляции. Наиболее правдоподобным признано значение 0,32 (что, впрочем, не исключает вероятности получить на других выборках корреляцию в 0,5 или 0,3). Примечательно, что коэффициент корреляции, рассчитанный традиционным способом, близок к байесовскому – 0,321. Его значимость меньше 0,001, поэтому эта процедура также не дает оснований полностью отвергнуть итоговое значение. Однако основания такого вывода – не величина правдоподобия (как после байесовского анализа), а классический показатель значимости.

Действительно, судя по частотам в табл. 2, предложение вывести свою страну из состава ЕС встречает все меньше поддержки респондентов по мере нарастания их несогласия с тем, что качество их жизни ухудшилось. И наоборот, больший евроскептицизм зафиксирован среди тех, кто чувствует себя все менее комфортно.

### **Заключение**

В данной работе приведены некоторые базовые принципы подхода к оценке взаимосвязей признаков социальных акторов и явлений, который основан на теореме Т. Байеса. В последние годы идеи этого английского математика XVIII в. позволили современным специалистам в области статистики усовершенствовать приемы расчета традиционных количественных показателей за счет

интерпретации вероятности как характеристики, воспринимаемой человеком, а не только присутствующей в окружающем мире.

По этой причине байесовские методы исходят из того, что наши знания о вероятностях явлений основаны на некоторой выборке наблюдений, которая никогда в точности не отражает все особенности генеральной совокупности. Учесть сопутствующие погрешности вычислений предлагается с помощью процедур, основанных на условной вероятности одного или нескольких явлений. А именно: не искать единственную величину требуемого коэффициента связи переменных, но предположить разброс его значений и стремиться выбрать из них то, в котором можно быть наиболее уверенным на основании полученной выборки, – имеющее максимальную степень правдоподобия.

Наиболее вероятные показатели связи признаков, обнаруженные при помощи метода Монте-Карло, могут быть рассчитаны не только для двух, но и для большего количества переменных, взаимодействующих внутри байесовской сети доверия. Вероятностные выводы о вершинах и ребрах этого графа используются для обработки дерева игры, дерева решений, в event-анализе и сферах, подлежащих исследованию с помощью нейронных сетей и других предиктивных алгоритмов искусственного интеллекта. В ходе таких исследований байесовские методы дают возможность не только прогнозировать редкие явления по малым выборкам, но и, наоборот, выявлять предшествующие события путем расчета скрытых цепей Маркова.

Описанный инструментарий использован для проверки гипотезы, входящей в предложенную графовую вероятностную модель и обоснованную политологическими и психологическими исследованиями, проведенными ранее. Дисперсионный и корреляционный анализ ответов 27 тыс. человек, опрошенных в рамках проекта «Евробарометр» в ноябре-декабре 2019 г., показал следующее. Предположение о повышении уровня евроскептицизма из-за того, что респонденты констатировали снижение качества своей жизни, имеет высокую степень правдоподобия.

## Список литературы

- Доманов А.О.* Когнитивные факторы усиления антиглобалистских настроений праворадикальных избирателей // Правый популизм: глобальный тренд и региональные особенности : монография / Н.К. Капитонова, И.Э. Магадеев, В.О. Печатнов [и др.] ; под ред. Л.С. Окуновой, А.И. Тэвдой-Бурмули. – М. : МГИМО-Университет, 2020. – С. 269–281.
- Kass R., Raftery A.* Bayes factors // Journal of the American Statistical Association. – 1995. – Vol. 90, N 430. – P. 773–795. – DOI: <https://doi.org/10.1080/01621459.1995.10476572>
- Lynch S.M.* Introduction to applied Bayesian statistics and estimation for social scientists. – New York : Springer, 2007. – 359 p. – DOI: <https://doi.org/10.1007/978-0-387-71265-9>
- McNeish D.* On using Bayesian methods to address small sample problems // Structural equation modelling : A multidisciplinary journal. – 2016. – Vol. 23, N 5. – P. 750–773. – DOI: <https://doi.org/10.1080/10705511.2016.1186549>
- Raftery A.* Bayesian model selection in social research // Sociological methodology. – 1995. – Vol. 25. – P. 111–163. – DOI: <https://doi.org/10.2307/271063>
- Van Ravenzwaaij D., Cassey P., Brown S.* A simple introduction to Markov Chain Monte-Carlo Sampling // Psychonomic bulletin and review. – 2018. – Vol. 25, N 1. – P. 143–154. – DOI: <https://doi.org/10.3758/s13423-016-1015-8>
- Western B.* Bayesian analysis for sociologists: an introduction // Sociological methods and research. – 1999. – Vol. 2, N 1. – P. 7–34. – DOI: <https://doi.org/10.1177/0049124199028001002>

**A.O. Domanov\***

### **The basics of Bayesian approach to quantitative analysis (at the example of Euroscepticism)**

*Abstract.* This article attempts to identify the main assumptions, prerequisites and techniques of the methods developed by some modern statisticians on the basis of T. Bayes' theorem for the purposes of social variables interactions assessment. The author underlined several advantages of the given approach as compared to more traditional quantitative methods and highlighted key research areas subject to evaluation by Bayesian estimates. First of all, this approach is compatible with game and decision theory, event analysis, hidden Markov chains, prediction using neural networks and other predictive algorithms of artificial intelligence.

The Bayesian approach differs significantly from traditional statistical methods (first of all, it is focused on finding the most probable, rather than the only true value of the feature coupling coefficient), hence a graphical interpretation was provided for such

---

\* **Domanov Aleksey**, Institute of Europe RAS; Institute for international studies, MGIMO University (Moscow, Russia), e-mail: [domanov.aleksey@gmail.com](mailto:domanov.aleksey@gmail.com)

basic concepts and techniques as probabilistic inference, maximum likelihood estimation and Bayesian confidence network.

The described tools were used to test the hypothesis about the impact of life quality decrease on rise in Euroscepticism of EU citizens. ANOVA and correlation analysis of 27 thousand people's responses to Eurobarometer questions addressed in November-December 2019 attributed strong likelihood to this assumption. Moreover, Bayesian approach allowed for a probabilistic conclusion that this hypothesis is more plausible than the link between Euroscepticism and respondents' current financial situation (explanatory power of comparison to the past is relatively greater).

*Keywords:* Bayesian estimation; Euroscepticism; quantitative methods; Bayesian belief network; quality of life; the EU.

*For citation:* Domanov A.O. The basics of Bayesian approach to quantitative analysis (at the example of Euroscepticism). *Political Science (RU)*. 2021, N 1, P. 301–321. DOI: <http://www.doi.org/10.31249/poln/2021.01.13>

## References

- Domanov A. Cognitive factors of anti-globalist attitudes of right-wing voters. In: Okuneva L., Tevdoy-Burmuli A. (eds). *Right-Wing Populism: Global Trend and Regional Features*. Moscow : MGIMO University, 2020, P. 269–281. (In Russ.)
- Kass R., Raftery A. Bayes factors. *Journal of the American Statistical Association*. 1995, Vol. 90, N 430, P. 773–795. DOI: <https://doi.org/10.1080/01621459.1995.10476572>
- Lynch S.M. *Introduction to applied Bayesian statistics and estimation for social scientists*. New York : Springer, 2007, 359 p. DOI: <https://doi.org/10.1007/978-0-387-71265-9>
- McNeish D. On using Bayesian methods to address small sample problems. *Structural equation modelling: a multidisciplinary journal*. 2016, Vol. 23, N 5, P. 750–773. DOI: <https://doi.org/10.1080/10705511.2016.1186549>
- Raftery A. Bayesian model selection in social research. *Sociological methodology*. 1995, Vol. 25, P. 111–163. DOI: <https://doi.org/10.2307/271063>
- Van Ravenzwaaij D., Cassey P., Brown S. A simple introduction to Markov Chain Monte-Carlo Sampling. *Psychonomic bulletin and review*. 2018, Vol. 25, N 1, P. 143–154. DOI: <https://doi.org/10.3758/s13423-016-1015-8>
- Western B. Bayesian analysis for sociologists: an introduction. *Sociological methods and research*. 1999, Vol. 28, N 1, P. 7–34. DOI: <https://doi.org/10.1177/0049124199028001002>