

Байесовское моделирование биномиальных экспериментов в социологии: проблемный анализ

УДК 303.447.3:303.714

DOI 10.26425/2658-347X-2024-7-1-14-25

Получено 31.11.2023

Доработано после рецензирования 24.12.2023

Принято 10.01.2024

Звонко Александр Анатольевич

Канд. филос. наук, доц. каф. социальной педагогики и организации работы с молодежью

ORCID: 0009-0007-7332-1330

E-mail: al.zvonok@gmail.com

Луганский государственный педагогический университет, г. Луганск, Россия

АННОТАЦИЯ

Статья посвящена байесовскому моделированию простых сравнительных биномиальных экспериментов с двоичными наборами данных (формата «успех» и «неудача») в социологии и других социальных науках. Кратко рассмотрены основные методологические основы применения байесовского подхода в статистике: применение априорных установок в анализе, особенности байесовского статистического вывода, различия частотных и байесовских доверительных интервалов, особенности проверки гипотез в байесовской статистике. Построена байесовская модель сравнительного биномиального эксперимента, поддерживающая сравнение независимых и зависимых выборок биномиальных переменных, а также допускающая различия в размерах сравниваемых выборок. Возможности модели, а также принципы байесовской проверки гипотез были продемонстрированы

на тестовых данных с применением современных свободных пакетов байесовского моделирования и анализа – PyMC и ArviZ. Использование указанных инструментов позволяет реализовывать прямые тензорные операции с получаемыми апостериорными распределениями, предоставляя исследователю результативный способ расчета размера эффекта при сравнении двух биномиальных выборок без необходимости прибегать к усложненным формам вычисления данного параметра. Показаны возможности и ограничения байесовского подхода в контексте сравнительного анализа результатов биномиальных экспериментов в социальных науках путем расчета вероятности гипотез посредством вычисления и сравнения площади интервалов апостериорных распределений.

Ключевые слова

Байесовский анализ, байесовская модель, экспериментальный метод, биномиальный эксперимент, биномиальные переменные, размер эффекта, PyMC, ArviZ

Для цитирования

Звонко А.А. Байесовское моделирование биномиальных экспериментов в социологии: проблемный анализ // Цифровая социология. 2024. Т. 7. № 1. С. 14–25.

© Звонко А.А., 2024.

Статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0. всемирная (<http://creativecommons.org/licenses/by/4.0/>).



Bayesian modeling of binomial experiments in sociology: problem analysis

Received 31.11.2023

Revised 24.12.2023

Accepted 10.01.2024

Aleksandr A. Zvonok

Cand. Sci. (Philos.), Assoc. Prof. at the Social Pedagogy and Organisation of Work with Youth Department

ORCID: 0009-0007-7332-1330

E-mail: al.zvonok@gmail.com

Lugansk State Pedagogical University, Lugansk, Russia

ABSTRACT

The article is devoted to Bayesian modeling of simple comparative binomial experiments with binary data sets (of “hit” and “miss” format) in sociology and other social sciences. The main methodological foundations of application of Bayesian approach in statistics are briefly reviewed: the use of priors in analysis, features of Bayesian statistical inference, differences in frequency and Bayesian confidence intervals, features of hypothesis testing in Bayesian statistics. A Bayesian model of a comparative binomial experiment has been constructed. It supports comparison of independent and dependent samples of binomial variables, and also allows for differences in sizes of the compared samples. The capabilities of the model, as well as the principles of the

Bayesian hypothesis testing, were demonstrated on test data using PyMC and ArviZ, contemporary free packages of the Bayesian modeling and analysis. The use of these tools allows implementing direct tensor operations with the obtained posterior distributions and provides the researcher with an effective way to calculate the effect size when comparing two binomial samples without having to resort to complicated forms of calculating this parameter. The possibilities and limitations of the Bayesian approach are shown in the context of comparative analysis of the results of binomial experiments in social sciences by estimating the probability of hypotheses via finding and comparing the area of intervals of posterior distributions.

Keywords

Bayesian analysis, Bayesian model, experimental method, binomial experiment, binomial variables, effect size, PyMC, ArviZ

For citation

Zvonok A.A. (2024) Bayesian modeling of binomial experiments in sociology: problem analysis. *Digital sociology*. Vol. 7, no 1, pp. 14–25. DOI: 10.26425/2658-347X-2024-7-1-14-25

© Zvonok A.A., 2024.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).



ВВЕДЕНИЕ / INTRODUCTION

Хотя экспериментальные исследования являются рядовым явлением для социальных наук, биномиальные эксперименты, основанные на биномиальных (биномиальных) переменных, все еще наблюдаются крайне редко в социологии, несмотря на обширный спектр задач, оптимальным решением для которых выступает именно данный тип исследований. Например, Центру содействия трудоустройству студентов и выпускников вуза может потребоваться выяснить, влияет ли на успешность поиска работы новый специальный курс по подготовке к трудоустройству для выпускников. В этом случае биномиальным значением 1 может выступить успешное трудоустройство выпускника в течение года после окончания учебного заведения, а под значением 0 будет подразумеваться невыполнение данных условий, при этом экспериментальной группой станут студенты, успешно прошедшие специальный курс, а контрольной группой – учащиеся, не включенные в прохождение курса. Другим типичным примером биномиального эксперимента является так называемое А/В-тестирование, ставшее де-факто стандартом в цифровом маркетинге при тестировании конверсии сайтов: пользователи интернет-ресурса при первом посещении случайно получают А-вариант или В-вариант сайта, после чего сравниваются пропорции «успехов» обеих групп в отношении неких целевых действий пользователей (регистрация, совершение покупки и т.д.).

В традиционной (частотной) статистике основной проблемой прикладного применения биномиальных экспериментов в течение многих десятилетий является нехватка информативных доказательных инструментов анализа их результатов. Существующий одновыборочный биномиальный критерий достаточно неудобен в контексте прикладного применения в сравнительных экспериментальных исследованиях и, кроме того, не лишен проблем, которые присущи самой системе проверки нулевой и альтернативной гипотез, основанной на р-значениях. Так, для обнаружения статистически значимых различий при малых уровнях значимости требуются выборки больших объемов, а «золотой стандарт» социологических исследований 0,05 во многих научных отраслях, таких как медицина, уже стал синонимом некачественных исследований из-за нарастающего метанаучного кризиса воспроизводимости. То есть при практических попытках воспроизведения подобных исследований в действительности количество экспериментов,

в которых оцениваемые значения выходят за рамки доверительных интервалов, оказывается гораздо больше, чем прогнозируемое 1 из 20 при уровне значимости 0,05. На самом деле в социологическом контексте данная проблема была показана еще в начале 2000-х гг. А.О. Крыштановским, проводившим эксперименты с данными нескольких исследований Всероссийского центра изучения общественного мнения: даже в рамках одного вопроса с 5 вариантами ответов из 100 случайных выборок в 26 из них был обнаружен выход за границы доверительных интервалов значений хотя бы одного из вариантов ответов (что, по сути, означает ошибочную оценку вопроса в целом) [Добренков, Кравченко, 2004]. В дальнейшем мы объясним, почему в этом эксперименте проблема заключается не в организации исследований, но в самой концепции частотных доверительных интервалов.

В числе прочих причин снижения достоверности исследований можно обозначить сознательные факторы, например очевидные фальсификации эмпирических данных наблюдений или манипуляции с выборками реальных данных и методами их анализа, призванные повысить статистическую значимость исследования (так называемый взлом р-значений, или р-хакинг). Другой важный фактор – низкий уровень квалификации исследователей в области анализа данных: непонимание ими эпистемологических и методологических оснований повсеместно применяемой частотной статистики, пренебрежение строгими принципами экспериментальных научных исследований. Так, Ф.А. Шродт одним из существенных недостатков современного количественного анализа определяет интерпретацию результатов в частотной статистике, как если бы она была байесовской [Schrodtd, 2013]. В числе распространенных нарушений, которые можно сюда отнести: свободные изменения уровней статистической значимости и объемов планируемых выборок после начала исследования; применение параметрических методов анализа в отношении данных, нормальность которых не была явно проверена соответствующими тестами; неправильное понимание доверительных интервалов в частотной статистике как байесовских интервалов; неправильное понимание концепции нулевых и альтернативных гипотез (вплоть до принятия нулевых гипотез) и т.д.

Методологические ограничения традиционного подхода заставляют специалистов по прикладным исследованиям обращаться к альтернативным направлениям анализа эмпирических данных, таким как байесовская статистика, но если

представители естественнонаучных дисциплин (например вычислительной физики или молекулярной биологии) имеют многолетний опыт применения подобных инструментов в своей прикладной деятельности, то распространенность байесовских методов в социальных и гуманитарных науках остается невысокой. Особенно выделяется в данном контексте социология: западные исследования публикационной активности свидетельствуют, что с 1978 г. по 2017 г. социологических исследований, содержащих байесовские методы, было опубликовано в 20 раз меньше, чем в психологии, и в 30 раз меньше, чем в экономических науках [Lynch, Bartlett, 2019]. При этом все социальные науки по данному показателю заметно уступают физико-математическим, химическим, биологическим наукам, в которых байесовская статистика давно применяется наравне с частотной. Не способствует решению проблемы и практически полное отсутствие преподавания байесовских методов в учебных программах социологов: даже среди зарубежных англоязычных изданий большинство университетских учебников предназначено для технических специальностей, а единственное найденное нами переводное издание, достаточно подробно затрагивающее социальные кейсы, датируется 1987 г. [Хей, 1987].

Существующая ситуация привела к тому, что, скажем, в сети интернет присутствует множество пособий в стиле «how-to» (англ. как сделать), содержащих пошаговую байесовскую реализацию А/В-тестирования сайтов, но не объясняется методологическая составляющая производимых действий. Более того, многие из этих реализаций чрезмерно усложнены и фактически используют философию частотной парадигмы анализа данных вместо байесовской. Многие специалисты по прикладным исследованиям обоснованно высказывают опасения, что при сохранении подобных тенденций неправильное использование частотных статистических методов просто будет заменено на некорректное применение байесовской статистики, только усугубив нарастающий кризис воспроизводимости в эмпирических науках.

Целью данной работы является не только построение типовой байесовской модели сравнительных биномиальных экспериментальных исследований (их частной реализацией выступает ранее упомянутое А/В-тестирование), но и объяснение базовых эпистемологических принципов, на которых основываются познавательные возможности байесовских статистических моделей. Важно отметить, что, несмотря на кажущуюся

специфичность биномиальных экспериментов, они имеют особое значение именно в контексте сравнительных исследований в области цифровой социологии: например, один из часто исследуемых параметров в данной сфере – количество лайков у поста (или статьи) в социальной сети (или блоге) – изначально, без всяких преобразований, можно представить в виде биномиальной переменной. В этом случае количество отметок «Нравится» соответствует количеству «успехов», а число «неудач» представляет собой разность между общим количеством просмотров данного контента и количеством отметок «Нравится». Следовательно, сравниваемые различные единицы онлайн-контента можно представить как аналогии выборок в классическом эксперименте. Подобные преобразования переменных относительно редко применяются в частотной статистике, но являются рядовой практикой при использовании байесовского подхода. В целом вопрос дихотомизации анализируемых переменных, когда они приводятся в бинарный вид, представляет собой узкую, но важную область data science (англ. наука о данных), хотя более глубокое раскрытие этой темы выходит за рамки данной работы.

МЕТОДОЛОГИЧЕСКИЕ ОСНОВАНИЯ / METHODOLOGICAL GROUNDS

Байесовский подход возник в XVIII в. и тесно связан с исследователями Т. Байесом и Дж. Р. Прайсом, чью концепцию субъективной вероятности как меры человеческого незнания о природе стохастических процессов доработал П.-С. де Лаплас в виде итоговой формулы под названием «Теорема Байеса». Подход основан на получении апостериорных (извлеченных из опыта) знаний с учетом наблюдаемых данных и априорных субъективных предзнаний об изучаемом явлении. Само слово «субъективный» в контексте данной концепции на протяжении столетий вызывало отторжение у многих ученых, вплоть до того, что современные байесианцы предпочитают использовать термин «эпистемная вероятность». Кроме того, чрезмерная сложность практических вычислений в байесовской статистике сделала невозможным ее массовое применение до распространения компьютеров.

Разработка основных принципов частотной статистики в начале XX в. Р. Фишером, Е. Нейманом и Э.Ш. Пирсоном сделала частотный подход доминирующим в прикладных научных исследованиях на протяжении целого века в немалой степени из-за простоты расчетов, реализуемых даже вручную. С распространением компьютерных

технологий и из-за невозможности решить многие научные проблемы в рамках частотной статистики интерес к байесовским методам возрос с новой силой. Дальнейшее развитие группы вычислительных алгоритмов марковских цепей Монте-Карло для моделирования вероятностных распределений и появление первых пользовательских программ для байесовского анализа в конце 1980-х гг. – начале 1990-х гг. обусловили внедрение данного подхода во все передовые научные отрасли. Тем не менее, ведущие программные пакеты для статистического анализа, такие как SAS и SPSS Statistics, стали поддерживать применение байесовских методов на пользовательском уровне только в последние годы (например, SPSS Statistics – с 2017 г.). Они не содержат наиболее трудные техники анализа, поэтому основным инструментарием сложных байесовских исследований сегодня выступают наукоориентированные языки программирования R и Python с подключением соответствующих научных модулей и библиотек для байесовских вычислений.

Для начала следует обозначить основные методологические отличия байесовского подхода от общепринятого частотного подхода в статистике. Полному объяснению философских и эпистемологических основ байесовского подхода посвящено множество фундаментальных трудов по философии науки, философии математики и эпистемологии, рассмотрение которых выходит за рамки данной работы, выделим лишь главные постулаты.

1. Априорные установки. Основой байесовской статистики является использование априорных знаний об изучаемом объекте. В качестве таких знаний может выступать вероятность протекания хорошо изученной химической реакции или результаты эксперимента, проведенного несколькими годами ранее. В продолжительном исследовании с использованием байесовского подхода апостериорные результаты каждой итерации байесовского анализа выступают как априорные знания в следующей итерации, придавая исследованию форму непрерывного процесса обновления знаний об исследуемом феномене с учетом получения новых данных. Следует заметить, что любые субъективные предположения ученого о характере изучаемого явления, например его предположение о том, что возраст респондентов в выборочном исследовании подчиняется нормальному распределению, по сути являются неосознанным применением байесовского подхода. Так, при строгом следовании частотной парадигме в отношении любых данных следует проводить явные проверки их на нормальность,

чтобы применять к ним параметрические методы анализа.

Наиболее спорным вопросом в байесовском анализе выступает использование первоначальных априорных установок в начале исследования, и к этому вопросу даже среди ученых-байесовцев существует несколько подходов. В социальных науках сегодня сложилась ситуация, при которой все ведущие ученые-приверженцы байесовского анализа (Д.Б. Каплан, Э.Э. Гельман, К.Р. Шализи и др.) используют так называемый неинформативный подход, то есть применяют априорные установки, оказывающие минимальное влияние на результаты анализа. Неинформативный подход не отрицает предположения «исследуемый параметр, скорее всего, подчиняется нормальному распределению», но противоречит таким априорным суждениям, как «скорее всего, на референдуме 70 % граждан проголосуют за принятие поправок в Конституцию». Однако главными в байесовском анализе остаются наблюдаемые данные, и при достаточно больших объемах данных наблюдений влияние даже информативных априорных данных будет стремиться к нулю.

2. Байесовский статистический вывод базируется на моделировании вероятностных распределений, а не на точечных оценках. На основе имеющихся данных и априорных предзнаний строится апостериорное вероятностное распределение параметра. Дальнейшие аналитические процедуры проводятся с полученным распределением, например, вычисляется площадь распределения, соответствующая определенному доверительному интервалу. В простых вероятностных задачах существует возможность численных решений, например площадь интервала в нормальном распределении можно вычислить даже вручную (при условии, что исследователь хорошо знаком с интегральным исчислением). Однако в случае сложных прикладных задач с сотнями переменных подобные вычисления возможно реализовать за приемлемое время только с помощью вероятностных алгоритмов класса «марковские цепи Монте-Карло», и именно их развитие в 1990-х гг., впоследствии получившее название «МЦМК-революции», привело к массовому внедрению байесовской статистики в прикладные научные сферы. Однако следует понимать, что результат вычислений МЦМК-алгоритмов – приближенные значения, точность которых зависит от количества генерируемых случайных выборок. В реальных исследованиях аналитик всегда вынужден искать компромисс между точностью результатов и временем расчетов, так как даже

на современных компьютерах байесовские вычисления занимают значительное время в сравнении с частотными методами.

3. Байесовские доверительные интервалы отличаются от частотных доверительных интервалов. Частотный доверительный интервал 95 % означает, что при проведении 20 независимых экспериментов в одном из них доверительный интервал не будет содержать настоящее значение параметра (что ранее нами было упомянуто в контексте эксперимента А.О. Крыштановского). Байесовский доверительный интервал 95 % представляет собой интервал плотности апостериорного распределения, то есть с 95 %-ной вероятностью значение параметра будет находиться в интервале между 2,5- и 97,5-м процентилем графика плотности апостериорного распределения параметра. Следует заметить, что, хотя байесовские оценки дают более информативные сведения, чем частотные, в условиях малых объемов выборок наблюдений они не способны решить такие проблемы эмпирических социальных исследований, как смещение выборки и другие подобные проблемы, возникающие на полевом этапе сбора данных.

4. Байесовские статистические гипотезы не имеют жестко установленных границ, как нулевая и альтернативная гипотезы в частотной статистике. Байесовский анализ может показать, что одна гипотеза заслуживает большего доверия, чем другая, во сколько именно раз, но он не дает однозначных указаний: отклонить или принять данную конкретную гипотезу.

Таким образом, проведенный анализ подсказывает, какое предположение рациональнее принять с учетом текущих данных, но не создает иллюзий относительно объективности полученных результатов. Современная байесовская статистика имеет множество высокоуровневых инструментов оценки построенных сложных статистических моделей, но на базовом уровне все основывается на простом сравнении вероятностей. Именно поэтому байесовская статистика сегодня является неотъемлемой частью областей, связанных с теорией принятия решений, особенно в условиях неопределенности.

ПОСТРОЕНИЕ МОДЕЛИ / MODEL CONSTRUCTION

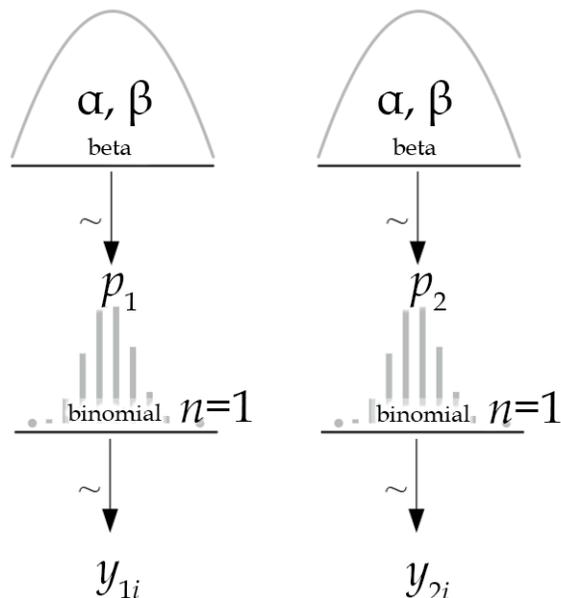
Обозначив основные постулаты байесовского анализа, мы можем приступить к непосредственному построению байесовской модели сравнительного биномиального эксперимента. Существует множество экспериментальных дизайнов,

но для биномиальных экспериментов, как правило, характерен формат с фиксацией «успехов» и «неудач» только после эксперимента (в отличие от ставшего классическим дизайна «до-после» с контрольной группой). В этом случае, например, подсчитывается количество студентов, успешно сдавших тест, или число пациентов в тяжелом состоянии, переживших сложную операцию. Однако даже в дизайне биномиального эксперимента с повторными измерениями байесовские принципы сравнительного анализа выборок не будут различаться. Так как байесовская статистика работает с вероятностными распределениями, построенными на основе данных, а не с единичными значениями, порядок отдельных наблюдений в выборке не имеет значения, в отличие от частотной статистики. Кроме того, необязательно иметь равное количество испытуемых в обеих выборках.

Также отметим, что, какой бы подход ни применялся для анализа данных биномиальных экспериментов, базовые принципы экспериментальных исследований сохраняются. Например, для формирования контрольной и экспериментальной групп должны применяться либо методы попарного или случайного отбора, либо метод частотного выравнивания или выравнивания по квоте (в зависимости от исследовательского контекста). Точно также, в случае если в качестве сравниваемых выборок выступают различные единицы цифрового контента, отличающиеся содержанием (например, при тестировании стратегий продвижения контента), все прочие параметры (оформление, время размещения и т.д.) должны быть приведены к максимальному единообразию при соблюдении основоположных принципов экспериментального метода для повышения достоверности результатов исследования.

Наша байесовская биномиальная параллельная модель, основанная на числе «успехов» в последовательности независимых испытаний в каждой выборке, будет иметь следующую структуру. Данные наблюдений в каждой выборке y_1 и y_2 определяются биномиальным распределением с параметрами $n = 1$ (которое в данном случае является распределением Бернулли) и переменными p_1 и p_2 . Последние в свою очередь являются вероятностями успеха в каждой выборке, которые характеризуются непрерывным вероятностным бета-распределением, содержащим переменные α (количество «успехов», 1) и β (количество «неудач», 0). В качестве априорных гиперпараметров модели мы использовали неинформативное бета-распределение Дж. Кермана ($\alpha = 1/3$, $\beta = 1/3$), оказывающее минимальное влияние на итоговые

результаты даже при очень малых выборках в несколько испытуемых. Модель можно визуализировать в виде популярной среди байесианцев диаграммы Дж. К. Крушке¹, как показано на рис. 1.



Составлено автором по материалам исследования /
Compiled by the author on the materials of the study

Рис. 1. Простая байесовская модель сравнительного биномиального эксперимента с контрольной и экспериментальной группами с измерением только после эксперимента

Fig. 1. Simple Bayesian model of a comparative binomial experiment with control and experimental groups with measurement only after the experiment

Как мы можем видеть, построение подобных моделей осуществляется из исходных данных наблюдений (правдоподобия, или likelihood в байесовской парадигме). Исследователь определяет, каким законам распределения подчиняются данные (нормальному, экспоненциальному и т.д.) и последовательно, снизу вверх, исходя от данных, устанавливает каждый параметр распределения (например, в нормальном распределении такими параметрами чаще всего будут среднее значение и стандартное отклонение). При необходимости указываются гиперпараметры – априорные установки, которые при окончательном построении модели изменятся под воздействием поступающих в модель данных.

Для более подробного рассмотрения нашей модели мы сгенерировали тестовый набор данных, содержащий две выборки по 120 наблюдений.

При этом в одной выборке (назовем ее экспериментальной) содержится 65 % «успехов», а в другой выборке (контрольной) – 55 % «успехов». Для построения модели нами были использованы библиотеки PyMC и ArviZ языка программирования Python, представляющие сегодня один из передовых инструментов серьезного байесовского анализа. Другим их преимуществом является очень простая структура программного кода, практически изолирующая исследователя от огромного пласта сложных математических вычислений, таких как дифференциальное исчисление и тензорная алгебра. С исходным кодом и тестовыми данными можно ознакомиться в репозитории автора данной работы².

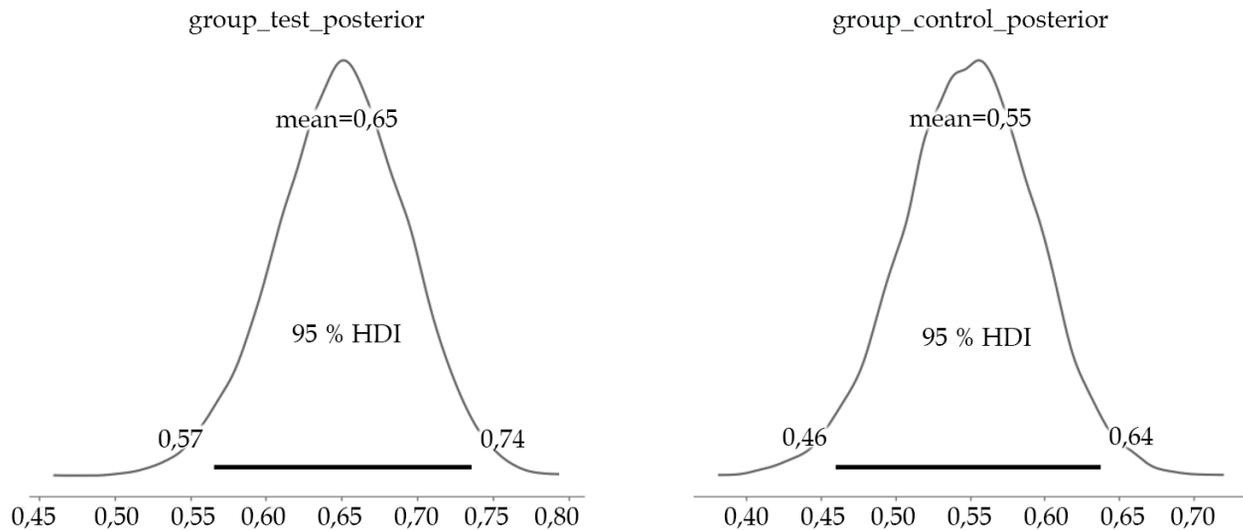
Важным принципом в байесианстве выступает пропаганда раскрытия и публикации исходных моделей и необработанных эмпирических данных исследований. Во-первых, эти данные другие исследователи могут использовать в качестве априорных в своей работе. Во-вторых, другие ученые могут заметить ошибки и недочеты в опубликованных моделях и помочь их улучшить (на практике специалисты по байесовскому анализу достаточно свободно критикуют и модифицируют модели, предложенные даже широко известными учеными-байесианцами, такими как Дж.К. Крушке и Э. Гельман, что, несомненно, приветствуется последними).

Полученные итоговые апостериорные распределения в построенной модели можно увидеть на рис. 2.

Как мы видим, при частотных оценках средних значений 0,65 и 0,55 в выборках истинные значения параметров, согласно байесовской интерпретации, с 95 %-ной вероятностью должны находиться в интервалах [0,57, 0,74] и [0,46, 0,64]. Следует обратить особое внимание на то, что результаты получены с помощью вероятностных алгоритмов с не очень высокими установками точности, и повторное моделирование может дать различия в итоговых интервалах вплоть до сотых долей (что является приемлемым для социальных исследований и простых байесовских моделей). Можно было бы значительно повысить установки точности ценой кардинального увеличения времени вычислений, но текущая погрешность приемлема в контексте решаемой задачи. Для сложных моделей потребовались бы гораздо более высокие требования к точности. В рамках библиотеки PyMC мы можем устанавливать точность путем указания размеров сэмпла – количества

¹ Bääth R. DIY Kruschke style diagrams. Режим доступа: <https://www.sumsar.net/blog/2013/10/diy-kruschke-style-diagrams/> (дата обращения: 25.11.2023).

² Zvonok A.A. Bayesian_binomial_simple_comparison_model. Режим доступа: https://gitflic.ru/project/aazvonok/bayesian_binomial_simple_comparison_model (дата обращения: 28.11.2023).



Примечание: $n = 120$, интервал плотности апостериорного распределения равен 95 %; HDI – плотность апостериорного распределения (англ. highest density interval)

Составлено автором по материалам исследования / Compiled by the author on the materials of the study

Рис. 2. Апостериорные бета-распределения контрольной и экспериментальной тестовых выборок

Fig. 2. Posterior beta distributions of the control and experimental test samples

выборок, извлекаемых из апостериорных распределений МЦМК-алгоритмами. Для тестового примера используется значение 4 000, но можно было бы свободно выставить и значение 100 000, однако с увеличением размеров сэмпла рост точности вычислений растет значительно медленнее, чем рост времени вычислений.

Полученные интервалы дают не очень много информации, тем не менее они являются более честными оценками истинного положения вещей, чем частотные средние: показанные интервалы всего лишь иллюстрируют общую неопределенность при небольшом объеме наблюдений ($n = 120$). Мы могли бы провести эксперимент, взяв выборки по 10 испытуемых, и результаты стали бы еще более неопределенными, или же мы могли бы взять 1 000 человек, и доверительные интервалы сузились бы почти до средних значений. Таким образом, в байесовской статистике не существует минимальных объемов выборок для проведения анализа, но объем выборки влияет на точность оценок. Для сравнения, применение частотного t-критерия для уровня значимости 0,05 при прочих идеальных условиях требует хотя бы по 6 наблюдений в каждой выборке [Kenny, 1987].

Дальнейшая интерпретация модели невозможна без расчета размеров экспериментального эффекта. Можно было бы просто вычислить разницу между двумя полученными апостериорными распределениями. Однако для большей наглядности влияния экспериментального фактора

мы вычтем из апостериорного бета-распределения нашей экспериментальной выборки апостериорное бета-распределение контрольной выборки и получившийся результат разделим на апостериорное бета-распределение последней, как показано в следующей формуле:

$$P_{\text{э.э.}} = \frac{\Theta_{\text{эксп.}} - \Theta_{\text{конт.}}}{\Theta_{\text{конт.}}}, \quad (1)$$

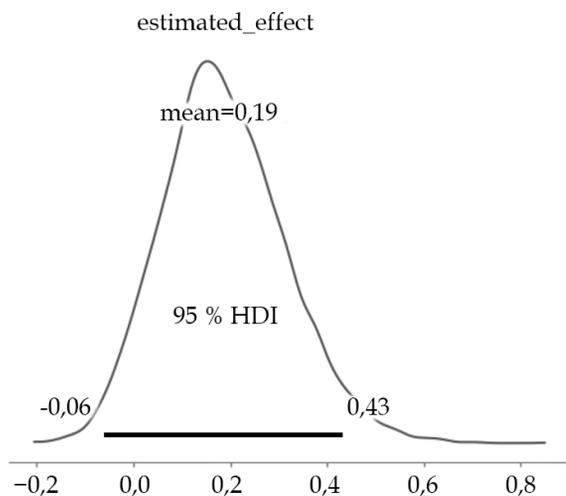
где $P_{\text{э.э.}}$ – размер экспериментального эффекта, $\Theta_{\text{эксп.}}$ – апостериорное бета-распределение экспериментальной выборки, $\Theta_{\text{конт.}}$ – апостериорное бета-распределение контрольной выборки.

Отметим, что сегодня существует несколько десятков способов расчета размера эффекта в зависимости от типа данных и прочих условий. Мы лишь выбрали наиболее простой метод, относящийся к биномиальным переменным. Например, для нормально распределенных переменных при расчете размера эффекта разницы в обеих группах чаще всего применяется метод d Коэна – разница средних в обеих группах, разделенная на объединенное стандартное отклонение групп. Некоторые аналитики пытаются применять его же и для биномиальных переменных, но, на наш взгляд, это излишнее усложнение для биномиальной модели, к тому же в байесовской реализации.

Также следует заметить, что простой вышеприведенная формула выглядит для пользователя библиотеки РумС, однако на программном

уровне для подобных арифметических операций с распределениями используется достаточно сложная математика тензоров. Далеко не все программы, включающие инструменты байесовского анализа, поддерживают подобные арифметические операции над вероятностными распределениями, и тогда для анализа приходится прибегать к другим способам с расчетом средних, стандартных отклонений и т.д. Эти параметры не отвергаются в байесовской статистике и используются для построения вероятностных распределений различных типов, но когда апостериорные распределения получены – лучше всего работать с ними напрямую, если это позволяет доступный технический инструментарий.

Важно обратить внимание на то, что в частотной статистике расчет таких параметров, как, например, коэффициент корреляции, отделен от доказательства их статистической значимости, то есть полученная корреляция должна проверяться на статистическую значимость отдельным тестом (в некоторых программах, например SPSS Statistics, подобные проверки осуществляются автоматически). Так как байесовская статистика работает с вероятностными распределениями, она не нуждается в подобных доказательствах, что может показаться крайне непривычным исследователям, всю жизнь работавшим в частотной парадигме. Полученный результат расчета величины экспериментального эффекта можно видеть на рис. 3.

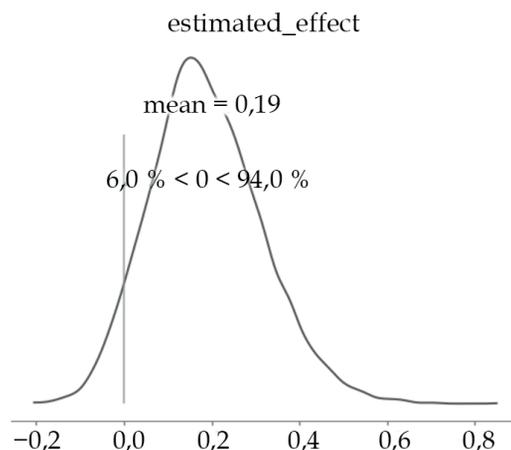


Составлено автором по материалам исследования /
Compiled by the author on the materials of the study

Рис. 3. Апостериорное бета-распределение размера эффекта влияния экспериментального фактора на экспериментальную выборку

Fig. 3. Posterior beta distribution of the size of the influence effect of the experimental factor on the experimental sample

Однако вышеприведенный результат анализа не является информативным с точки зрения исследователя, стремящегося решить прикладную исследовательскую задачу: расчетный эффект с 95 % лежит в интервале $[-0,06, 0,43]$, то есть от -6% до 43% . Иначе говоря, существует вероятность даже отрицательного влияния экспериментального фактора. Можно увеличивать объемы выборок наблюдений, пока доверительный интервал не сузится до однозначно положительных значений, однако в прикладном байесовском анализе гораздо чаще применяется выдвижение гипотез, основанных на расчете площади апостериорных распределений. Например, для полученного распределения можно выдвинуть гипотезы H_0 (истинное значение параметра меньше 0) и H_1 (истинное значение параметра больше 0). Расчет площади распределения (рис. 4), показывает, что существует приблизительно 6 %-ная вероятность H_0 и 94 %-ная вероятность H_1 . Иначе говоря, вторая гипотеза примерно в 15,7 раз более достойна доверия, чем первая. Это дает нам возможность обоснованно назвать H_1 сильной гипотезой, отвергнуть нулевую гипотезу (заметим, что в байесовской статистике, в отличие от частотной, нулевую гипотезу можно и принять) и на основе этого сделать какие-либо выводы об исследуемом параметре.



Составлено автором по материалам исследования /
Compiled by the author on the materials of the study

Рис. 4. Вычисление площади апостериорного распределения для байесовского сравнения гипотез

Fig. 4. Computing the area of the posterior distribution for Bayesian comparison of the hypotheses

ВОЗМОЖНОСТИ МОДЕЛИ / CAPABILITIES OF THE MODEL

Таким же образом мы могли бы выдвинуть в исследовании целый ряд гипотез, например «истинное значение параметра больше 0,1» или

«истинное значение параметра меньше 0,05», и провести их попарное сравнение. Именно данный факт делает такой привлекательной байесовскую статистику для прикладных областей, связанных с принятием экономических и управленческих решений. Итоги байесовского анализа не просто показывают, что результаты статистически значимы, но могут ответить на вопрос «С какой вероятностью ожидаемый эффект от внедрения данного решения обеспечит рост эффективности хотя бы на 10 %?» Тем не менее, точность подобных ответов все еще будет зависеть от наблюдаемых данных, объем и качество которых определяются действиями, совершенными на предварительном и полевом этапах социологического исследования.

Концептуально байесовская проверка гипотез является чрезвычайно простой, однако практическое ее применение стало возможным только в последние десятилетия благодаря развитию компьютерных технологий. Даже построение и расчет показанной относительно простой байесовской модели с учетом тестовых данных на современном компьютере занимают считанные минуты, более сложные модели могут требовать на порядок более серьезных вычислительных мощностей. Однако сам принцип байесовского моделирования и проверки гипотез остается тем же самым. Например, по тем же принципам строится популярная байесовская реализация t-критерия Стьюдента BEST (англ. Bayesian estimation supersedes the t-test) и другие подобные ей модели [Kruschke, 2013]. Понимая общие принципы байесовского моделирования, можно строить огромные иерархические модели, заметно превосходящие сложность текущей (хотя на прикладном уровне для достаточно сложных задач потребуются выделение гораздо более значительных вычислительных мощностей, чем те, что доступны рядовому пользователю).

Описываемая нами байесовская модель имеет широкие перспективы применения в цифровой социологии при условии изначального построения дизайна исследований, предусматривающего сравнительный анализ биномиальных переменных. В качестве альтернативы можно рассматривать дихотомизацию уже имеющихся данных из порядковой или интервальной шкалы в двоичную форму. Например, результаты онлайн-тестирования в баллах могут преобразовываться в бинарное значение «сдал-не сдал», используя определенное пороговое значение, устанавливаемое исследователем. Вот лишь некоторые потенциальные кейсы применения биномиальных экспериментов в байесовской реализации:

- традиционное А/В-тестирование (следует отметить, что в рамках цифровой социологии данный метод позволяет выполнять значительно больше задач, нежели классическая оптимизация сайтов в цифровом маркетинге: тестируются стиль и оформление сообществ в социальных сетях, отдельных постов в данных сообществах и т.д. в целях достижения наилучших показателей «социальной эффективности» [Кибакин, 2020, с. 15]);

- определение контент-стратегии в SMM (англ. social media marketing – маркетинг в социальных сетях). Для решения данной задачи отсутствует универсальная модель, поэтому только сравнительное тестирование определенных моделей и практика могут найти оптимальное решение [Аржанова, Еремеева, 2023] (например, в данном случае сравниваемой биномиальной переменной может выступать наличие или отсутствие отметки «Нравится», поставленной пользователем под контентом определенной направленности);

- математически обоснованная численная оценка влияния или эффективности социальных технологий в цифровой среде в случаях, когда подобные оценки можно дихотомизировать. Преимуществом байесовского подхода в этом случае является четкий ответ на вопрос «Какова вероятность того, что конкретная технология лучше другой на некий определенный минимальный порог?» Это особенно важно в прикладных коммерческих исследованиях (и может стать одним из шагов в решении актуальной проблемы социологии – замещения экспертов-социологов аналитиками общего профиля в процессе перемещения социальных исследований из академических институтов в корпоративные лаборатории [Мещерякова, 2020]).

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ / STUDY RESULTS

Таким образом, нами были рассмотрены методологические основы применения байесовского анализа и выполнено построение простой модели биномиального эксперимента, применимой для сравнительных экспериментальных социологических исследований в цифровой социологии. Поняв принципы, по которым проводятся байесовское моделирование и байесовская проверка статистических гипотез, исследователь сможет начать работать и с более сложными построениями, включая байесовские прогнозные регрессионные модели. Существующие пособия по байесовской статистике, как правило, неудобны читателям из-за достаточно сложной математики,

но автор данной работы начинал свое ознакомление с байесовской методологией по философским, а не математическим трудам. С точки зрения философии науки, байесовский подход является чрезвычайно простым и интуитивным и более согласуется с концепциями Т.С. Куна, П.К. Фейерабенда, Б. Латура и других постпозитивистов, в сравнении с традиционной частотной статистикой.

На теоретическом уровне классический байесовский анализ представляет собой простой интуитивный непрерывный процесс установки и перераспределения уровней доверия к гипотезам с учетом первоначальных субъективных убеждений исследователя и поступления новых данных. Мы умышленно минимизировали и упростили математическое наполнение данной работы, так как байесовская статистика, как правило, не преподается на социологических факультетах в России. Более того, последние эмпирические исследования свидетельствуют о нарастающем кризисе даже в области базового статистического образования российских социологов [Зырянов, 2022]. К счастью, современные достижения в области компьютерных технологий сделали байесовские методы доступными даже для специалистов без математического образования (а при наличии навыков программирования хотя бы на уровне средней школы они могут пользоваться целым набором высокоуровневых продвинутых инструментов для сложного анализа). Однако для прикладного применения байесовской методологии нужно четко понимать философию данной исследовательской парадигмы.

Существуют и методические проблемы байесовских вычислений. Например, в рамках данного направления очень удобно работать с биномиальными или нормально распределенными переменными, но анализ данных, выраженных в порядковых шкалах, потребует или дополнительных преобразований подобных данных или создания таких сложных моделей, что будет проще обратиться за решением задачи к традиционной частотной статистике. Кроме того, применяя байесовские методы, исследователь должен обладать более специфичными познаниями в теории вероятности и математической статистике,

например более глубокими знаниями типологии и характеристик вероятностных распределений, которые применяются в байесовских моделях.

Наконец, многим исследователям в области социальных и гуманитарных наук, в отличие от ученых-специалистов естественных наук, как ни парадоксально, часто не нравится концепция приближенных вычислений (англ. approximate computing), характерная для байесовской статистики. Тот факт, что одна и та же программа при одинаковых настройках и одинаковых введенных данных может выдавать результаты 6 %, 5,5 % или 5,7 % при повторных запусках, заставляет некоторых специалистов сомневаться в объективности результатов анализа. В прикладном байесовском анализе проблема нахождения компромисса между точностью и вычислительными затратами всегда незримо присутствует в любой научно-исследовательской лаборатории.

ЗАКЛЮЧЕНИЕ / CONCLUSION

Таким образом, не следует заблуждаться, что байесовская статистика способна исправить серьезные недочеты, допущенные на этапе планирования исследования, или некачественно проведенный полевой этап получения первичных социологических данных. Данная парадигма допускает некоторую свободу действий: позволяет расширить объем выборки в рамках одного и того же исследования или же начать анализировать данные по мере их поступления, обновляя результаты с появлением новых наблюдений, что является недопустимым с точки зрения частотной статистики. Тем не менее, она не освобождает исследователя от необходимости следовать правилам формирования выборочных совокупностей и не помогает добиться более искренних ответов от респондентов. С другой стороны, байесовский подход дает более честную оценку в отношении исследуемых проблем, не называя обнаруженные закономерности «статистически значимыми» только потому, что р-значения оказались ниже уровня значимости 0,05, но лишь констатируя, какая из научных гипотез более достойна доверия на текущий момент с учетом имеющейся информации и в какой именно степени.

СПИСОК ЛИТЕРАТУРЫ

Аржанова К.А., Еремеева А.И. Ситуативный и коммуникационный контент в рамках SMM-продвижения бренда в социальных сетях. *Цифровая социология*. 2023;2(6):4–11. <https://doi.org/10.26425/2658-347X-2023-6-2-4-11>

Добреньков В.И., Кравченко А.И. *Фундаментальная социология: в 15 томах. Том 2. Эмпирическая и прикладная социология*. М.: Инфра-М; 2004. 986 с.

Зырянов В.В. Социальная статистика в социологическом образовании. Социологические исследования. 2022;2:129–141. <https://doi.org/10.31857/S013216250017138-4>

Кибакин М.В. Вебометрика как диагностический инструментальный цифровой социологии: содержание, предназначение, опыт использования. Цифровая социология. 2020;1(3):12–18. <https://doi.org/10.26425/2658-347X-2020-1-12-18>

Меццьякова Н.Н. Методология познания цифрового общества. Цифровая социология. 2020;2(3):17–26. <https://doi.org/10.26425/2658-347X-2020-2-17-26>

Хей Дж. Введение в методы байесовского статистического вывода. Пер. с англ. М.: Финансы и статистика; 1987. 336 с.

Kruschke J.K. Bayesian estimation supersedes the t-test. Journal of Experimental Psychology: General. 2013;2(142):573–603. <https://doi.org/10.1037/a0029146>

Lynch S.M., Bartlett B. Bayesian statistics in sociology: past, present, and future. Annual Review of Sociology. 2019;45:47–68. <http://dx.doi.org/10.1146/annurev-soc-073018-022457>

Schrodt Ph.A. Seven deadly sins of contemporary quantitative political analysis. Journal of Peace Research. 2013;2(51):287–300. <https://www.doi.org/10.1177/0022343313499597>

REFERENCES

Arzhanova K.A., Ereemeeva A.I. Situational and communication content in the framework of SMM brand promotion in social networks. Digital Sociology. 2023;2(6):4–11. (In Russian). <https://doi.org/10.26425/2658-347X-2023-6-2-4-11>

Dobrenkov V.I., Kravchenko A.I. Fundamental sociology: in 15 volumes. Volume 2. Empirical and applied sociology. Moscow: Infra-M; 2004. 986 p. (In Russian).

Hey J.D. An introduction to methods of Bayesian statistical inference. Trans. from Eng. Moscow: Finansy i statistika; 1987. 336 p. (In Russian).

Kibakin M.V. Webometric as diagnostic tools of digital sociology: contents, purpose, usage experience. Digital Sociology. 2020;1(3):12–18. (In Russian). <https://doi.org/10.26425/2658-347X-2020-1-12-18>

Kruschke J.K. Bayesian estimation supersedes the t-test. Journal of Experimental Psychology: General. 2013;2(142):573–603. <https://doi.org/10.1037/a0029146>

Lynch S.M., Bartlett B. Bayesian statistics in sociology: past, present, and future. Annual Review of Sociology. 2019;45:47–68. <http://dx.doi.org/10.1146/annurev-soc-073018-022457>

Meshcheryakova N.N. Methodology for cognition of digital society. Digital Sociology. 2020;2(3):17–26. (In Russian). <https://doi.org/10.26425/2658-347X-2020-2-17-26>

Schrodt Ph.A. Seven deadly sins of contemporary quantitative political analysis. Journal of Peace Research. 2013;2(51):287–300. <https://www.doi.org/10.1177/0022343313499597>

Zyryanov V.V. Social statistics in sociological education. Sociological Studies. 2022;2:129–141. (In Russian). <https://doi.org/10.31857/S013216250017138-4>