



СОЦИОЛОГИЯ, СОЦИАЛЬНЫЕ СТРУКТУРЫ И ПРОЦЕССЫ, СОЦИАЛЬНЫЕ ТЕХНОЛОГИИ SOCIOLOGY, SOCIAL STRUCTURES AND PROCESSES, SOCIAL TECHNOLOGIES

УДК 519.226.3:303.447.3

DOI 10.52575/2712-746X-2024-49-1-26-38

Байесовская экспериментальная оценка социальных технологий: методологические и методические аспекты

Звонок А.А.

Луганский государственный педагогический университет,
Россия, 291011, Луганская Народная Республика, г. Луганск, ул. Оборонная, д. 2
al.zvonok@gmail.com

Аннотация. В современной отечественной социологии практически не представлена байесовская методология анализа социологических данных. Целью исследования является демонстрация методологических и методических аспектов и преимуществ применения байесовского подхода в области экспериментальной оценки социальных технологий. Автором были раскрыты основные принципы байесовского анализа экспериментальных данных на примере универсальной сравнительной модели BEST Дж. К. Крушке, построенной с учетом характерных особенностей экспериментов, проводимых в социальных и гуманитарных науках. Полученные результаты не только доказывают методологическое превосходство байесовского подхода в случае применения его для оценки социальных технологий, но и могут быть использованы для настройки существующих или построения новых сравнительных байесовских моделей, внося свой вклад в популяризацию байесовского подхода в социальных исследованиях.

Ключевые слова: байесовский анализ, байесовская модель, экспериментальный метод, размер эффекта, социальная технология, РуМС

Для цитирования: Звонок А.А. 2024. Байесовская экспериментальная оценка социальных технологий: методологические и методические аспекты. *NOMOTHETIKA: Философия. Социология. Право*, 49(1): 26–38. DOI: 10.52575/2712-746X-2024-49-1-26-38

Bayesian Experimental Assessment of Social Technologies: Methodological and Methodical Aspects

Aleksandr A. Zvonok

Lugansk State Pedagogical University,
2 Oboronnaya St, Lugansk 291011, LPR, Russia
al.zvonok@gmail.com

Abstract. In modern sociology the Bayesian methodology for analyzing sociological data is practically not represented. The basics of applying the Bayesian approach are not taught at Russian sociological faculties, and empirical sociological research is not conducted within the framework of this approach. At the same time, this methodology has significant prospects within the context of experimental sociological research, in particular, when solving the problem of assessing the effectiveness of new social technologies

and innovations. The main purpose of this study is demonstrating of the methodological and methodological aspects of the application of the Bayesian approach using modern statistical analysis software, as well as its advantages in comparison with the traditional frequentist approach in the field of experimental assessment of social technologies. The principles of operation of J.K. Kruschke's universal comparative Bayesian model BEST (Bayesian estimation supersedes the t test), reproduced with Python library PyMC, were revealed. The obtained results of the study not only prove the methodological superiority of the Bayesian approach when applied to the assessment of social technologies, but can also be used by other researchers to tune existing or build new comparative Bayesian models, contributing to the popularization of the Bayesian approach in social research.

Keywords: Bayesian analysis, Bayesian model, experimental method, effect size, effect size, social technology, PyMC

For citation: Zvonok A.A. 2024. Bayesian Experimental Assessment of Social Technologies: Methodological and Methodical Aspects. *NOMOTHETIKA: Philosophy. Sociology. Law*, 49(1): 26–38 (in Russian). DOI: 10.52575/2712-746X-2024-49-1-26-38

Введение

Термин «социальная технология» прочно вошел в обиход отечественной науки после публикации в 1977 году русского перевода зарубежной работы Н. Стефанова «Общественные науки и социальная технология», где понимался как разновидность технологии, т.е. специфической деятельности, ориентированной на достижение поставленной цели – изменения социального объекта или процесса [Щербина, Нечаева, 2015]. Хотя сам термин сегодня имеет огромное количество интерпретаций в социальных науках, подробное приведение которых было бы избыточным в рамках данной работы, метаанализ содержания термина Е.О. Нечаевой дает четкое понимание, что последний тесно связан с рационализацией деятельности, областью принятия решений. Среди характеристик, присваиваемых социальным технологиям различными авторами, особенно выделяются следующие: «являются алгоритмизированным средством практической деятельности», «рационализируют определенные направления управленческой деятельности», «гарантируют достижение заданного результата при соблюдении технологических предписаний», «отличаются надежностью, валидностью, простотой и возможностью их тиражирования» [Нечаева, 2014, с. 63].

В то же время в качестве одной из важнейших методологических проблем, препятствующих изучению, разработке и успешному применению социальных технологий, сегодня ряд исследователей называют низкий уровень воспроизводимости результатов их применения [Плотников, Смелцова, 2012]. Данный вопрос приводит нас к более широкой метанаучной проблеме, известной как «кризис воспроизводимости», обсуждение которой усилилось в последнее десятилетие. Суть данной проблемы заключается в том, что устоявшиеся практики проверки статистических гипотез в традиционной (частотной) статистике имеют ряд недостатков, которые ведут к неустойчивости результатов исследований при попытках повторить их другими лабораториями и исследовательскими группами. В качестве одного из решений данной проблемы рассматривается отход от частотной проверки гипотез, основанной на p -значениях, и внедрение в научную среду альтернативных подходов к анализу данных, таких как байесовская статистика [Cumming, 2008]. Таким образом, ключевым этапом разработки и внедрения новых социальных технологий и инноваций является количественная оценка их эффективности с помощью современных научных методов, в том числе использующих статистику и математическое моделирование.

Цель исследования заключается в доказательстве того, что для оценки социальных технологий методологически более подходящим является байесовский подход, основанный на построении сравнительных статистических моделей, использующих экспериментальные данные. Также нами будут продемонстрированы возможности современных байесовских моделей сравнения экспериментальных выборок и их интерпретации.



Оценка социальных технологий: базовые вопросы экспериментального дизайна

Исходя из объяснений В.А. Ядова, мы знаем, что достоверные познавательные результаты в социальных экспериментах достигаются двумя путями. В неконтролируемом эксперименте большое число повторений эксперимента погашает неконтролируемые факторы, оставляя только воздействие экспериментального фактора. Альтернативой является проведение эксперимента с контрольной и экспериментальной группой, при котором производится выравнивание условий в обеих группах [Ядов, 2009, с. 316–321]. Эксперимент с большим числом повторений рационально применять в условиях быстро протекающих социальных ситуаций, однако даже социальные технологии, применяемые на микроуровне, могут требовать месяцы и даже годы для их реализации. Таким образом, в контексте оценки социальных технологий рациональнее применять дизайн контролируемого эксперимента с контрольной и экспериментальной группой.

Другой важной стороной корректной экспериментальной оценки является выбор достоверного измерительного инструментария, позволяющего максимально точно отразить изменения (или их отсутствие) оцениваемых параметров. Ю.П. Лукашин и Л.И. Рахлина выделяют шкалы качественных признаков (номинальная и порядковая) и шкалы количественных признаков (интервальная, отношений, разностей, абсолютная) [Лукашин, Рахлина, 2012, с. 16–19]. Хотя в традиционной статистике существует возможность сравнения выборок порядковых значений с помощью непараметрических методов, количество потенциально возникающих фундаментальных проблем при этом столь велико, что даже применяющие частотный подход исследователи рекомендуют или вовсе этого избегать [Cowell, Flachaire, 2017; Bond, Lang, 2018] или преобразовывать порядковые шкалы в количественные [Knapp, 1990], при этом сравнительные исследования показывают, что чем ближе полученные шкалы к интервальным, тем лучше на них работают статистические методы, изначально создававшиеся для работы с количественными признаками, такие как факторный анализ [Зангиева, Ротмистров, 2018].

Таким образом, для измерений в процессе экспериментальной оценки социальных технологий нами рекомендуется изначально выбирать инструментарий, предоставляющий результаты измерений в шкале количественных признаков.

Проблемы традиционного подхода к экспериментальной оценке социальных технологий и основы байесовского статистического вывода

Ю.П. Сурмин и Н.В. Туленков, выделяя экспериментальный метод в качестве одного из главных способов оценки эффективности социальных технологий и инноваций, в качестве критерия «эффективности» социальной технологии понимают меру ее способности решать актуальные социальные проблемы, удовлетворять определенные социальные потребности людей.

В процессе проведения оценки эффективности социальной технологии требуется решить следующие задачи:

1. Определить размер позитивного эффекта социальной технологии.
2. Измерить размер негативного эффекта от социальной технологии.
3. Оценить возможные скрытые негативные последствия внедрения технологии, которые могут проявиться в будущем.
4. Определить интегральный эффект от социальной технологии как суммы позитивных и негативных эффектов [Сурмин, Туленков, 2004, с. 568–571].

Однако одной из самых больших проблем анализа экспериментальных данных в рамках частотной статистики является разделение концепции статистической значимости, основанной на р-значениях и выражающей ее количественной меры. Например, если мы рассчитали коэффициент корреляции между двумя переменными, то необходимо отдельно использовать статистические тесты, проверяющие, является ли взаимосвязь между

ними статистически значимой (в некоторых программах статистического анализа такие проверки выполняются автоматически). Если мы доказали, что существуют «статистически значимые» различия между выборками наблюдений в эксперименте и получили расчетный «сильный» размер экспериментального эффекта, мы все еще не можем с уверенностью утверждать, что применение экспериментального фактора приводит к «сильным» статистически значимым различиям в контрольной и экспериментальной выборках на определенном уровне значимости.

Преимуществом байесовского подхода в контексте экспериментальных исследований является точная численная оценка вероятности размеров экспериментального эффекта, при этом является возможным точный расчет вероятности как положительного, так и отрицательного воздействия в заданном интервале значений с учетом имеющихся данных наблюдений. В байесовском подходе (по крайней мере, в направлении «объективного» байесианства, которое является сегодня наиболее принимаемым в прикладной и фундаментальной науке) вероятности интерпретируются как степени уверенности (доверия), вычисленные на основе имеющихся данных («правдоподобия») [Williamson, 2010].

За подробными описаниями и сравнениями философских оснований байесовского подхода и его отличий от классического частотного подхода следует обратиться к современным трудам по философии и эпистемологии науки, например, к сравнительному исследованию Дж. Валлверду [Vallverdú, 2016].

Мы выделим лишь основные особенности байесовской методологии:

1. Итоговые результаты анализа получаются в результате перехода от априорных установок («предзнаний») к апостериорным результатам с учетом имеющихся данных наблюдений (для этого используется теорема Байеса). Следует отметить, что в настоящее время в байесовской статистике выделяют два основных философско-эпистемологических подхода. *Субъективное байесианство* тяготеет к использованию информативных априорных установок-предположений относительно исследуемых параметров, данное направление концентрируется вокруг рационализации принятия решений в условиях неопределенности (нехватки эмпирической информации). *Объективное байесианство* предпочитает использовать неинформативные априорные установки, не влияющие на итоговый результат анализа, т.е. апостериорное распределение параметра практически целиком формируется данными наблюдений. Это направление преимущественно используется в фундаментальных и прикладных научных исследованиях.

Субъективное байесианство сохраняет свои позиции в определенных областях, таких, как сфера управления и принятия решений (например, в социологии управления), так как позволяет принимать рациональные решения в условиях существенной нехватки объективной информации, однако в современных байесовских исследованиях общей направленности объективное байесианство преобладает, занимая промежуточную ступень между частотной статистикой и классической байесовской статистикой. В целом в отношении применения априорной информации в байесовском анализе выделяется три подхода:

– *неинформативный подход* (исследователь может делать априорные предположения о законе распределения, которому подчиняются данные, например, нормальному распределению, но в качестве априорного распределения будет использовано не влияющее на итоговый анализ, скажем, нормальное распределение со средним значением 0 и стандартным отклонением 10^6);

– *слабоинформативный (регуляризирующий) подход* (промежуточное направление, снижающее воздействие выбросов данных, устанавливая значения гиперпараметров в определенные границы, например, возраст человека может принимать только положительные значения, значения индекса человеческого развития всегда находятся между 0 и 1 и т.д.);

– *информативный подход* (делаются «сильные» предположения о значении параметра, например, о том, что на будущем референдуме 80 % участников проголосуют за принятие определенного закона, в дальнейшем поступающие данные будут подтверждать гипотезу или же делать ее все менее вероятной).



На сегодня байесовская методология слабо представлена в социологической сфере. Однако анализ байесовских исследований в других науках говорит о преобладании неинформативного подхода, например, результаты анализа 187 байесовских экологических исследований говорят о том, что в 56,7 % применялись неинформативные априорные установки [Lemoine, 2019]. Независимо от выбранного подхода к постановке априорных установок анализа (их еще называют «гиперпараметрами»), при количестве фактических данных, стремящемся к бесконечности, их влияние будет стремиться к нулю – это фундаментальное свойство байесовского статистического вывода [Хей, 1987, с. 134–135].

2. Байесовский анализ основан на построении вероятностных распределений (с учетом данных наблюдений и априорных предположений) и вычислении площадей их интервалов (а также других математических операций с данными распределениями), точечные оценки, такие как среднее значение или медиана не отвергаются, но используются в качестве вспомогательных инструментов. Например, альтернативой корреляции Пирсона для двух переменных в байесовском статистическом выводе будет являться двумерное нормальное распределение этих переменных. Это одна из главных причин, которые ограничивают распространение байесовского подхода в социальных и гуманитарных науках: популярные пользовательские программы для статистического анализа (например, SPSS Statistics) поддерживают лишь ограниченное количество простых байесовских методов. Применение продвинутых методов анализа требует от исследователей владеть навыками работы с наукоориентированными языками программирования R и Python, позволяющими строить и визуализировать сложные байесовские модели. До революции вычислительных вероятностных алгоритмов Марковских цепей Монте-Карло в 1990-х годах применение байесовской статистики в прикладной сфере было крайне ограничено. Сегодня МЦМК-алгоритмы позволяют строить и оценивать сложные многоуровневые иерархические байесовские модели за допустимое время, что позволило внедриться данной методологии в сферу эмпирических исследований значительного количества наук разного профиля: естественных, технических и т.д.

3. В байесовской статистике отсутствует жесткое принятие или отвержение статистических гипотез. Корректнее говорить о степени доверия к той или иной гипотезе при текущей совокупности данных наблюдений, например, результат байесовского анализа может сообщить нам, что «с вероятностью 95 % рассматриваемый параметр лежит в интервале между 16 и 35» или «с вероятностью 98 % размер экспериментального эффекта больше 0». Увеличение размеров выборки наблюдений сужает доверительные интервалы, давая более точные оценки, малые размеры выборки дают размытые оценки. Таким образом, в байесовской статистике проблемы «размера выборки» или «минимальной выборки» в исследованиях являются второстепенными: в отличие от частотной статистики, в которой многие методы имеют минимальные требования по количеству наблюдений, в байесовском подходе концепция «минимального объема выборки» отсутствует. Однако слишком малые выборки наблюдений в байесовском анализе дадут очень размытый нечеткий результат, на основе которого будет практически невозможно сделать обоснованные выводы об исследуемой проблеме. С другой стороны, байесовский подход позволяет проводить длительные исследования с постоянно расширяющейся выборкой наблюдений, предоставляя исследователю «срезы» состояний исследуемой проблемы. Это не только дает возможность предварительного оценивания, важного в прикладных исследованиях, но и позволяет определять интегральные эффекты и последствия от изучаемых феноменов, что особенно важно в социальных исследованиях.

Возможности сравнительных байесовских моделей

При использовании байесовского подхода одной из главных методологических проблем является проверка адекватности построенных математико-статистических моделей. В случае прогнозных и классификационных моделей одним из самых простых и в то же

время эффективных способов является разделением общей совокупности данных реальных наблюдений на две выборки – обучающую и проверочную. Первая выборка используется для построения и обучения модели, а вторая – для проверки того, насколько эффективно модель выполняет поставленные задачи (прогнозирования или классификации). Хотя сравнительные байесовские модели значительно проще по структуре, вышеозначенный способ обучения-проверки для них не подходит. В качестве альтернативы можно рассматривать проверку таких моделей с помощью наборов искусственно сгенерированных тестовых данных – выборок с заранее известными характеристиками, т. е. «истинными» значениями параметров. Следует заметить, что, согласно классическим соглашениям в области математической статистики, использование искусственно сгенерированных данных при разработке, тестировании и внедрении математико-статистических методов в конкретные научные области является адекватной практикой при соблюдении следующих условий: общие сведения о дизайне гипотетического эксперимента, раскрытие условий генерации (тип распределения, средние значения, стандартные отклонения и т. д.) и программного обеспечения, с помощью которого осуществлялась генерация [Hoaglin, Andrews, 1975]. Применение симулированных данных является достаточно устоявшейся практикой: согласно обзорному анализу 6 ведущих западных журналов, посвященных вопросам методологии статистических исследований в математике, психологии, биометрике и технометрии, на 2012 год 44,8 % всех научных работ в данных журналах применяли сгенерированные искусственные данные [Harwell et al., 2017]. В контексте нашего исследования, имеющего выраженный методологический характер, было бы недопустимо смешивать задачи тестирования адекватности сравнительной байесовской модели и задачи оценки реальной социальной технологии: в первом случае задача неосуществима без генерации данных с известными истинными параметрами, так как применение реального кейса байесовской оценки социальной технологии сделало бы невозможным доказательство работоспособности собственно метода оценки этой технологии (из-за возможного влияния неучтенных сторонних факторов в эксперименте).

В рамках текущей работы с помощью научной библиотеки Scipy языка программирования Python нами было сгенерировано 2 тестовых выборки по 35 наблюдений, которые представляют собой данные гипотетического эксперимента с контрольной и экспериментальной группой, тестирующей эффективность новой социальной технологии, призванной повысить социальную активность определенной категории индивидов. Настройки генерации предполагали истинный «средний» размер экспериментального эффекта, выраженный в величине d Коэна. Согласно Дж. Коэну классическая интерпретация d Коэна выглядит следующим образом: от 0,2 до 0,5 – «слабый» размер эффекта, от 0,5 до 0,8 – «средний» размер эффекта, выше 0,8 – «сильный» размер эффекта [Cohen, 1988]. При генерации данных предполагалось, что для измерения социальной активности использовался опросник Р.М. Шамионова и М.В. Григорьевой для диагностики компонентов социально-ориентированной активности личности, итоговая оценка которого выражается в количественной шкале от 15 до 105 баллов [Шамионов, Григорьева, 2019].

Сегодня в байесовской статистике одной из наиболее популярных сравнительных моделей (особенно в социальных и гуманитарных науках) является модель BEST (Bayesian Estimation Supersedes the t Test) Дж. К. Крушке для сравнения экспериментальных данных, выраженных в количественных (метрических) шкалах. Модель специально калибровалась под особенности социально-гуманитарных исследований: относительно малые выборки, высокая частота выбросов значений. Существует даже онлайн-реализация данной модели, которой может воспользоваться любой желающий без необходимости устанавливать и применять языки программирования R и Python¹. С диаграммой связей, полностью описывающей отношения в модели BEST, можно ознакомиться ниже (рис. 1).

¹ Bayesian Estimation Supersedes the t-test (BEST) – online. 2024. Publishable Stuff Rasmus Bååth's Blog. URL: https://www.sumsar.net/best_online (дата обращения 05.02.2024).

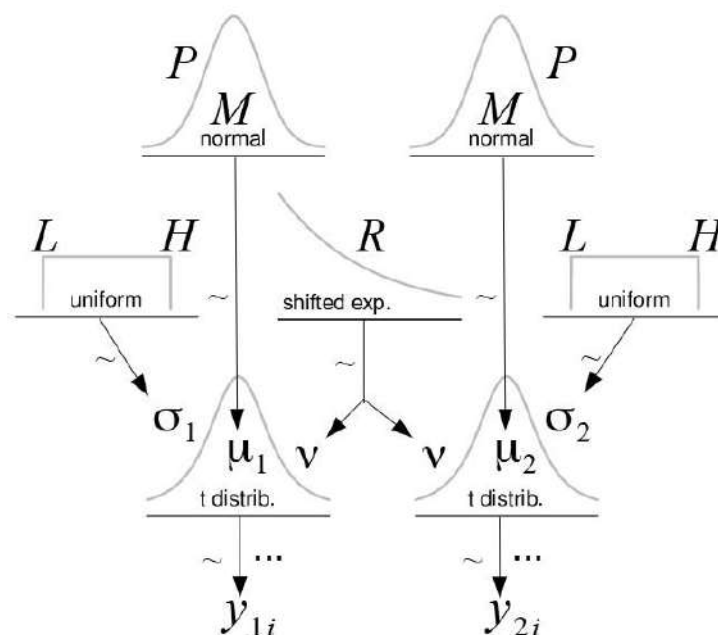


Рис. 1. Байесовская сравнительная модель BEST Дж. К. Крушке
 Fig. 1. Bayesian comparative model BEST J. K. Kruschke

Особенностью данной модели является использование не нормальных распределений в качестве априорных, а распределений Стьюдента (t-распределений), которые при большом значении степеней свободы (> 100) по свойствам приближаются к нормальному распределению (такое построение делает модель более устойчивой к выбросам данных). В самом низу иерархической модели находятся данные наблюдений y_1 и y_2 , под влиянием которых априорные распределения будут трансформироваться в апостериорные (например, в зависимости от «нормальности» распределения реальных данных наблюдений будет изменяться число степеней свободы апостериорного t-распределения). Все гиперпараметры модели, используя полностью неинформативный подход, вычисляются на основе реальных свойств наблюдаемых данных (среднего значения и стандартного отклонения объединенных выборок наблюдений, подлежащих сравнению). Например, гиперпараметры стандартных отклонений σ_1 и σ_2 выражены в виде равномерного распределения в интервале от одной тысячной до тысячи реальных значений объединенного стандартного отклонения данных наблюдений. Данный подход позволяет избежать калибровки модели под каждое новое исследование, предоставляя ей свойство универсальности. Гиперпараметры средних значений выборок μ_1 и μ_2 являются нормально распределенными величинами, где среднее значение M нормального распределения представляет собой реальное среднее значение объединенных сравниваемых выборок, а вместо стандартного отклонения используется параметр «точности» (precision) P – обратная величина дисперсии, которая часто применяется в байесовских моделях. Число степеней свободы ν выражено смещенным экспоненциальным распределением [Kruschke, 2013].

Хотя в первоначальной работе Дж. К. Крушке применялся инструментальный язык программирования R, для целей нашей работы мы воспроизвели модель BEST на языке программирования Python с использованием современных библиотек байесовского моделирования PyMC и Arviz, опустив часть второстепенных визуализаций модели, несущественных в рамках текущей работы. Несомненным преимуществом библиотеки PyMC является структура кода, повторяющая стиль вышеприведенной диаграммы связей, что значительно облегчает создание байесовских моделей. Кроме того, подобные высокоуровневые инструменты значительно снижают требования к уровню математической подготовки

исследователя, самостоятельно определяя оптимальные алгоритмы математических вычислений. С исходным кодом и тестовыми данными можно ознакомиться в репозитории автора данной работы ¹.

Полученные апостериорные распределения ранее сгенерированных выборок, проанализированных посредством модели BEST можно увидеть ниже (рис. 2).

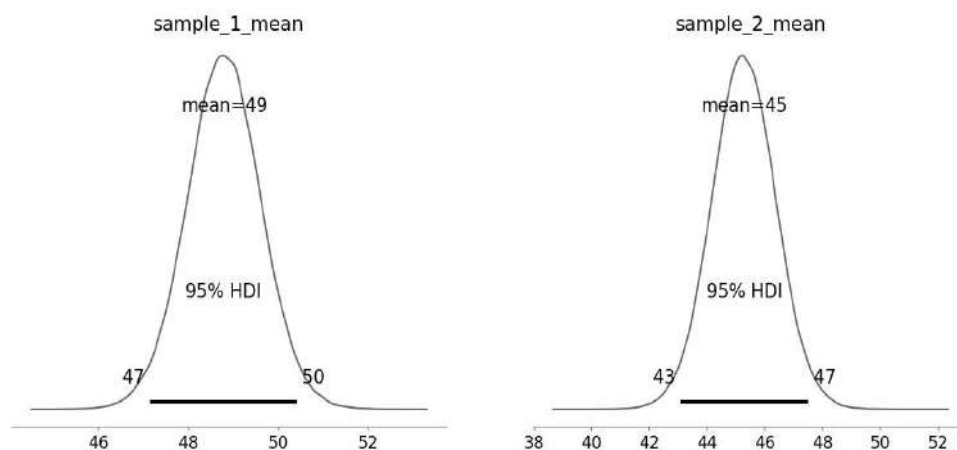


Рис. 2. Апостериорные распределения тестовых выборок при интервале плотности апостериорного распределения 95 %

Fig. 2. Posterior distributions of test samples with a highest posterior density interval of 95 %

Согласно полученным результатам, оцениваемые параметры в выборках с вероятностью 95 % находятся в интервалах [47,145, 50,424] и [43,103, 47,492]. На самом деле подобная точность является излишней для вероятностных алгоритмов, так как в рамках концепции приближенных вычислений («approximate computing») каждый запуск модели может давать различия итоговых результатов в пределах долей процента [Sunnåker et al., 2013]. Именно поэтому на вышеприведенных графиках значения округлены до целых. Очень важно различать частотные доверительные интервалы (confidence intervals) и байесовские доверительные интервалы (credible intervals), они же интервалы плотности распределения (highest density interval, HDI). Далее мы можем рассчитать размер эффекта d Коэна, в байесовской реализации представляющий собой разность апостериорных распределений средних значений выборок, разделенную на квадратный корень из половины суммы квадратов апостериорных распределений стандартных отклонений выборок. В данном случае арифметические операции производятся над вероятностными распределениями. Так, даже при сравнении зависимых связанных выборок (повторных измерений) в эксперименте, в байесовском анализе часто игнорируется равное число наблюдений в выборках и точная позиция единичных наблюдений (хотя существуют специальные байесовские модели, специально рассчитанные на применение для случаев повторных измерений, на практике они используются значительно реже, чем в частотной статистике).

На нижеприведенном графике (рис. 3) мы видим апостериорное распределение размера эффекта с установленным референсным значением 0,5 и областью практической эквивалентности 0,2 (region of practical equivalence или ROPE, другое название – null region, т. е. «нулевая область»), а также с вычисленными площадями соответствующих интервалов распределения. В реальных исследованиях референсные значения и область практической эквивалентности устанавливаются в зависимости от конкретных исследовательских задач.

¹ bayesian_BEST_comparative_model. 2024. GitFlic. URL: https://gitflic.ru/project/aazvonok/bayesian_best_comparative_model (дата обращения 05.02.2024).

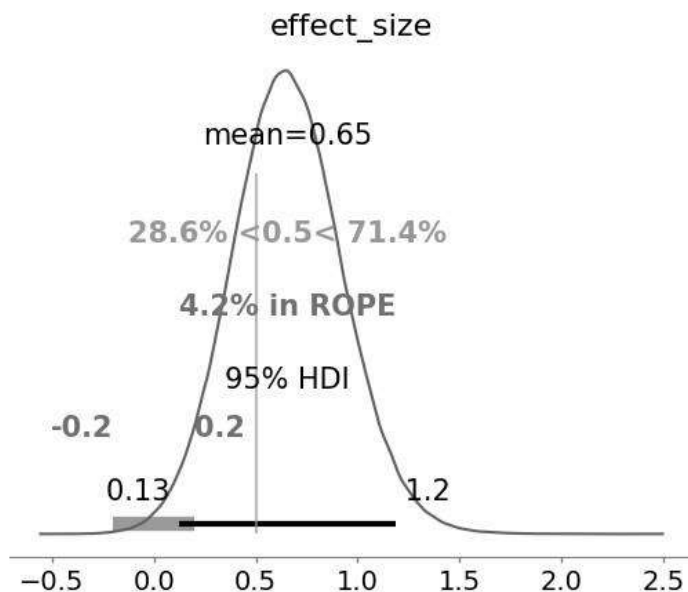


Рис. 3. Расчет размера эффекта для тестовых выборок при интервале плотности апостериорного распределения 95 %, референсном значении 0,5 и областью практической эквивалентности 0,2
 Fig. 3. Effect size calculation for test samples with a highest posterior density interval of 95 %, a reference value of 0.5 and a region of practical equivalence of 0.2

Исходя из анализа нашего тестового кейса, полученные результаты следует интерпретировать следующим образом. С вероятностью 95 % размер экспериментального эффекта находится в интервале между 0,13 («очень слабый» или «практически отсутствует») и 1,2 («сильный»). При этом существует 71,4 % вероятность того, что оцениваемая социальная технология оказывает не менее «среднего» экспериментального эффекта на тестовую выборку и 28,6 % вероятность, что размер эффекта менее установленного значения. В то же время существует всего лишь 4,2 % вероятность того, что между экспериментальной и контрольной выборкой различия практически отсутствуют (аналог «нулевой» гипотезы в частотной статистике).

Отдельно отметим, что не существует общепринятых значений области практической эквивалентности, как в случае уровней значимости в частотной статистике. Нами было выбрано значение ROPE от $-0,2$ до $0,2$, так как в классической интерпретации Дж. Коэна это размер эффекта ниже «слабого», т. е. попадание размера эффекта в данную область можно интерпретировать как «эффект практически отсутствует». Но в рамках иных научных трактовок размера эффекта в качестве ROPE можно было бы принять и интервалы $[-0,1, 0,1]$ или $[-0,01, 0,01]$ (некоторые авторы предлагают расширить традиционную градацию размера эффекта Дж. Коэна для ряда современных методов анализа вплоть до границы значимости в $0,01$ [Sawilowsky, 2009]). Таким образом, границы области практической эквивалентности должны устанавливаться в зависимости от применяемых методов и конкретных исследовательских задач. Отдельно отметим, что в современном байесовском анализе сравнительных экспериментальных исследований в последние десятилетия существует тенденция отхода от классической байесовской проверки и сравнения гипотез в сторону подхода «байесовский анализ мощности» («Bayesian power analysis») на основе ROPE, являющегося синтезом классического анализа размера эффекта Дж. Коэна и байесовского анализа. При данном подходе статистические гипотезы могут не выдвигаться вообще, а основная цель эксперимента – оценить размер экспериментального эффекта [Kruschke, 2010]. При частотном подходе к анализу экспериментов такая позиция была бы неприемлема: например, расчет размера эффекта d Коэна может показать «слабый» эф-

факт при том, что проверка гипотезы о наличии статистических различий в выборках вообще не находит статистически значимых различий. Байесовский анализ мощности лишен данных недостатков, так как получаемое апостериорное распределение размера эффекта по умолчанию является «статистически значимым». Рассматриваемая в работе модель BEST является образцовым примером модели, выполненной в рамках «байесовского анализа мощности» и не использующей выдвижение статистических гипотез. Так как задача экспериментальной оценки эффективности социальных технологий находится ближе к оценке величины эффекта на базе апостериорного распределения, нежели к выбору из двух конкурирующих гипотез – целесообразнее применять для ее решения именно данный подход.

В случае реализованного нами тестового кейса 95 % интервала размера эффекта является положительным, но, если бы распределение в значительной мере затрагивало отрицательную область шкалы, байесовский анализ позволил бы так же точно вычислить вероятность и размер отрицательного и положительного экспериментального эффекта (например, установив референсное значение 0). Таким образом, даже при использовании искусственно сгенерированного «социального» кейса с относительно малыми выборками по 35 наблюдений, байесовский анализ позволил нам с достаточно высокой точностью оценить истинный размер экспериментального эффекта, заложенный нами в тестовые данные, что доказывает адекватность рассмотренной модели. Дальнейшее увеличение объемов выборок позволило бы нам получать все более точные оценки, при этом в рамках байесовского подхода мы не связаны необходимостью заранее планировать уровень статистической значимости и объемы выборок эксперимента. Иначе говоря, при необходимости мы могли бы подключить к эксперименту дополнительные единицы отбора (будь это отдельные индивиды, группы или организации, на которых тестируется социальная технология) и использовать новые данные наблюдений в том же самом анализе, повысив его точность.

Подытоживая, мы можем заключить, что байесовский подход значительно лучше подходит для оценки эффективности социальных технологий, нежели классический подход, позволяя не просто заключить об отсутствии или присутствии «статистически значимых» различий, но дать численную оценку вероятности положительного и отрицательного эффекта в заданных интервалах. Кроме того, в отличие от традиционных статистических методов, байесовская экспериментальная оценка социальных технологий позволяет учесть экономическую целесообразность внедрения инновационной технологии в сравнении с ее отсутствием или уже внедренной старой технологией. Например, если для рентабельности технологии ее экспериментальный размер эффекта должен быть не менее значения X , то байесовский вывод дает числовую оценку вероятности эффекта не ниже такого порогового значения, а также предоставляет рассчитанные риски нулевого или отрицательного эффекта.

Однако главный недостаток байесовского подхода заключается в том, что итоговая интерпретация результатов байесовского анализа всегда остается за исследователем и не имеет жестких установленных пределов. Является ли оправданным внедрение социальной технологии, применение которой с 80 % вероятностью приводит к положительному эффекту, но с 20 % вероятностью приведет к негативным последствиям? Ответы на подобные вопросы лежат за рамками байесовской статистики. Очевидно, что для социальной технологии развития творческих способностей молодежи и для технологии социального сопровождения подростков после неудавшейся попытки самоубийства интерпретация оценки будет различаться. Выработка соглашений по интерпретации результатов байесовского анализа в рамках той или иной науки (как в случае «золотого стандарта» уровня значимости 0.05 в частотной методологии) является главным условием для успешного внедрения байесовского подхода в конкретные научные области и сферы исследований.



Заключение

В социологии сегодня существует слишком мало исследований, выполненных в рамках байесовской методологии, что значительно усложняет систематизацию подходов к применению байесовского подхода в социологических исследованиях. Например, с 1978 по 2017 годы было опубликовано всего лишь 142 научных статьи по социологии, использующих байесовские методы (согласно данным базы Web of Science). Этот показатель в десятки раз уступает соответствующим показателям даже в относительно близких областях научного знания (экономика, психология) [Lynch, Bartlett, 2019]. В сфере социологических исследований попросту недостаточно специалистов, владеющих методами байесовской статистики, и она, как правило, не преподается на социологических факультетах российских вузов. При этом байесовский подход уже активно применяется в таких близких к социологии прикладных областях, как цифровой маркетинг, то есть выпускник-социолог, желающий начать работать в данной сфере, оказывается попросту неподготовленным к требованиям современного рынка труда.

В данной работе нами были показаны богатые возможности байесовского подхода к экспериментальной оценке социальных технологий и аспекты его применения, а также доказаны преимущества этой методологии в решении научно-исследовательских задач такого рода. На наш взгляд, одной из ключевых проблем современной эмпирической социологии является необходимость внедрения в образовательные программы на социологических направлениях дисциплин, связанных с прикладным применением байесовской статистики в «социологических» кейсах. Кроме того, мы считаем приоритетной целью заимствование байесовских методов анализа и соглашений по интерпретации его результатов из близкородственных наук (экономических, психологических и т.д.), где байесовский подход является значительно более развитым, в сферу социологических исследований.

Список источников

- Сурмин, Ю.П., Туленков Н.В. 2004. Теория социальных технологий : учебное пособие. К., МАУП, 608 с.
- Ядов В.А. 2009. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности : учебное пособие. М., Омега-Л, 567 с.

Список литературы

- Зангиева И.К., Ротмистров А.Н. 2018. Сравнительный анализ способов проведения факторного анализа на порядковых переменных. *Мониторинг общественного мнения: экономические и социальные перемены*, 145(3): 29–46. DOI: 10.14515/monitoring.2018.3.02.
- Лукашин Ю.П., Рахлина Л.И. 2012. Современные направления статистического анализа взаимосвязей и зависимостей. М., ИМЭМО РАН, 54 с.
- Нечаева Е.О. 2014. Социальные технологии: история становления и версии понимания. *Вестник РГГУ. Серия «Философия. Социология. Искусствоведение»*, 126(4): 55–64.
- Плотников М.В., Смельцова С.В. 2012. Социальные технологии как объект социологического исследования. *В мире научных открытий*, 28(4–3): 169–187.
- Хей Д.Д. 1987. Введение в методы байесовского статистического вывода. М., Финансы и статистика, 335 с.
- Шамионов Р.М., Григорьева М.В. 2019. Методика диагностики компонентов социально-ориентированной активности. *Сибирский психологический журнал*, 74: 26–41. DOI: 10.17223/17267080/74/2
- Щербина В.В., Нечаева Е.О. 2015. Социальные технологии и их применение в сферах социальной практики. *Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки*, 38(2): 62–67.
- Bond T.N., Lang K. 2018. The Sad Truth About Happiness Scales. *Journal of Political Economy*, 127(4): 1629–1640. DOI: 10.1086/701679.



- Cohen J. 1988. *Statistical power analysis for the behavioral sciences*. New York, Lawrence Erlbaum Associates, 567 p.
- Cumming G. 2008. Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, 3(4): 286–300. DOI: 10.1111/j.1745-6924.2008.00079.x.
- Cowell F.A., Flachaire E. 2017. Inequality with Ordinal Data. *Economica*, 334(84): 290–321. DOI: 10.1111/ecca.12232.
- Knapp T.R. 1990. Treating Ordinal Scales as Interval Scales. *Nursing Research*, 39(2): 121–123. DOI: 10.1097/00006199-199003000-00019.
- Kruschke J.K. 2010. Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1: 658–676. DOI: 10.1002/wcs.72.
- Kruschke J.K. 2013. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology*, 142(2): 573–603. DOI: 10.1037/a0029146.
- Lemoine N.P. 2019. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128(7): 912–928. DOI: 10.1111/oik.05985.
- Lynch S.M., Bartlett B. 2019. Bayesian Statistics in Sociology: Past, Present, and Future. *Annual Review of Sociology*, 45(1): 47–68. DOI: 10.1146/annurev-soc-073018-022457.
- Sawilowsky S. 2009. New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2): 467–474. DOI: 10.22237/jmasm/1257035100.
- Sunnåker M., Busetto A.G., Numminen E., Corander J., Foll M., Dessimoz C. 2013. Approximate Bayesian Computation. *PLoS Computational Biology*, 9(1): 1–10. DOI: 10.1371/journal.pcbi.1002803.
- Vallverdú J. 2015. *Bayesians versus frequentists: a philosophical debate on statistical reasoning*. Heidelberg, Dordrecht, London, New York, Springer, 110 p.
- Williamson J. 2010. *In Defence of Objective Bayesianism*. New York, Oxford University Press, 191 p.

References

- Zangieva I.K., Rotmistrov A.N. 2018. Sravnitel'nyy analiz sposobov provedeniya faktornogo analiza na poryadkovykh peremennykh. [Factor analysis of ordinal variables: a comparative study]. *Monitoring obshchestvennogo mneniya: ekonomicheskije i sotsial'nye peremeny*, 145(3): 29–46 (in Russian). DOI: 10.14515/monitoring.2018.3.02.
- Lukashin YU.P., Rakhlina L.I. 2012. Sovremennyye napravleniya statisticheskogo analiza vzaimosvyazey i zavisimostey [Modern directions of statistical analysis of relationships and dependencies]. M., IMEMO RAN, 54 p.
- Nechayeva Ye.O. 2014. Sotsial'nyye tekhnologii: istoriya stanovleniya i versii ponimaniya [Social technologies: history of formation and versions of understanding]. *Vestnik RGGU. Seriya «Filosofiya. Sotsiologiya. Iskusstvovedeniye»*, 126(4): 55–64 (in Russian).
- Plotnikov M.V., Smeltsova S.V. 2012. Sotsial'nyye tekhnologii kak ob"yekt sotsiologicheskogo issledovaniya [Institutional dead ends of the Russian teacher training system]. *V mire nauchnykh otkrytiy*, 28(4–3): 169–187 (in Russian).
- Khey D.D. 1987. Vvedeniye v metody bayesovskogo statisticheskogo vyvoda [Introduction to Bayesian statistical inference methods]. M., Finansy i statistika, 335 p.
- Shamionov R.M., Grigor'yeva M.V. 2019. Metodika diagnostiki komponentov sotsial'no-oriyentirovannoy aktivnosti [Methodology for diagnosing the components of socially oriented activity]. *Sibirskiy psikhologicheskij zhurnal*, 74: 26–41 (in Russian). DOI: 10.17223/17267080/74/2.
- Shcherbina V.V., Nechayeva Ye.O. 2015. Sotsial'nyye tekhnologii i ikh primeneniye v sferakh sotsial'noy praktiki [Social technologies and their application in areas of social practice]. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo. Seriya: Sotsial'nyye nauki*, 38(2): 62–67 (in Russian).
- Bond T.N., Lang K. 2018. The Sad Truth About Happiness Scales. *Journal of Political Economy*, 127(4): 1629–1640. DOI: 10.1086/701679.
- Cohen J. 1988. *Statistical power analysis for the behavioral sciences*. New York, Lawrence Erlbaum Associates, 567 p.
- Cumming G. 2008. Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, 3(4): 286–300. DOI: 10.1111/j.1745-6924.2008.00079.x.



- Cowell F.A., Flachaire E. 2017. Inequality with Ordinal Data. *Economica*, 334(84): 290–321. DOI: 10.1111/ecca.12232.
- Khey D.D. 1987. Vvedenie v metody bayesovskogo statisticheskogo vyvoda [An introduction to bayesian statistical inference]. M., Publ. Finansy i statistika, 335 p. (In Russian)
- Knapp T.R. 1990. Treating Ordinal Scales as Interval Scales. *Nursing Research*, 39(2): 121–123. DOI: 10.1097/00006199-199003000-00019.
- Kruschke J.K. 2010. Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1: 658–676. DOI: 10.1002/wcs.72.
- Kruschke J.K. 2013. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology*, 142(2): 573–603. DOI: 10.1037/a0029146.
- Lemoine N.P. 2019. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128(7): 912–928. DOI: 10.1111/oik.05985.
- Lukashin Yu.P., Rakhlina L.I. 2012. Sovremennye napravleniya statisticheskogo analiza vzaimosvyazey i zavisimostey [Modern directions of statistical analysis of relationships and dependencies]. M., Publ. IMEMO RAN, 54 p. (In Russian).
- Lynch S.M., Bartlett B. 2019. Bayesian Statistics in Sociology: Past, Present, and Future. *Annual Review of Sociology*, 45(1): 47–68. DOI: 10.1146/annurev-soc-073018-022457.
- Nechaeva E.O. 2014. Sotsial'nye tekhnologii: istoriya stanovleniya i versii ponimaniya [Social technologies: history of formation and versions of understanding]. *Vestnik RGGU. Seriya Filosofiya. Sotsiologiya. Iskusstvovedenie*, 126(4): 55–64. (In Russian)
- Plotnikov M.V., Smel'tsova S.V. 2012. Sotsial'nye tekhnologii kak ob'ekt sotsiologicheskogo issledovaniya [Social technologies as an object of sociological study]. *V mire nauchnykh otkrytiy*, 28(4–3): 169–187. (In Russian)
- Sawilowsky S. 2009. New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2): 467–474. DOI: 10.22237/jmasm/1257035100.
- Shamionov R.M., Grigor'eva M.V. 2019. Metodika diagnostiki komponentov sotsial'no-orientirovannoy aktivnosti [Method for diagnosing components of socially oriented activity]. *Sibirskiy psikhologicheskii zhurnal*, 74: 26–41. DOI: 10.17223/17267080/74/2 (In Russian)
- Shcherbina V.V., Nechaeva E.O. 2015. Sotsial'nye tekhnologii i ikh primenenie v sferakh sotsial'noy praktiki [Social technologies and their application in areas of social practice]. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo. Seriya: Sotsial'nye nauki*, 38(2): 62–67. (In Russian)
- Sunnåker M., Busetto A.G., Numminen E., Corander J., Foll M., Dessimoz C. 2013. Approximate Bayesian Computation. *PLoS Computational Biology*, 9(1): 1–10. DOI: 10.1371/journal.pcbi.1002803.
- Vallverdú J. 2015. Bayesians versus frequentists: a philosophical debate on statistical reasoning. Heidelberg, Dordrecht, London, New York, Springer, 110 p.
- Williamson J. 2010. In Defence of Objective Bayesianism. New York, Oxford University Press, 191 p.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.

Conflict of interest: no potential conflict of interest has been reported.

Поступила в редакцию 05.02.2024

Поступила после рецензирования 21.02.2024

Принята к публикации 27.02.2024

Received February 5, 2024

Revised February 21, 2024

Accepted February 27, 2024

ИНФОРМАЦИЯ ОБ АВТОРЕ

Звонко Александр Анатольевич, кандидат философских наук, доцент кафедры социальной педагогики и организации работы с молодежью, Луганский государственный педагогический университет, г. Луганск, Россия.

INFORMATION ABOUT THE AUTHOR

Aleksandr A. Zvonok, Candidate of Philosophy, Associate professor of the Department of Social Pedagogy and Organization of Work with Youth, Lugansk State Pedagogical University, Lugansk, Russia.