

19.00.00 Psychological sciences

19.00.00 Психологические науки

UDC 519/316

Bayesian Information Criterion as an Alternative way of Statistical Inference¹Nadejda Yu. Gubanova²Simon Zh. Simavoryan¹Sochi State University, Russia

PhD (Psychology)

E-mail: ngubanova@hotmail.com

²Sochi State University, Russia

PhD (Technical)

E-mail: simsim58@mail.ru

Abstract. The article treats Bayesian information criterion as an alternative to traditional methods of statistical inference, based on NHST. The comparison of ANOVA and BIC results for psychological experiment is discussed.

Keywords: statistical inference; hypothesis testing; Kullback–Leibler information; Bayesian information criterion.

В последние десятилетия в экспериментальные исследования в области гуманитарных и общественных наук активно внедряются методы статистической обработки и анализа данных. Между тем, существующая методология статистического вывода (*NHST*), далеко не свободна от внутренних противоречий. Одной из её главных проблем является оценка $P(D/H_0)$ -вероятности получения экспериментальных данных при гипотезе H_0 , вместо необходимой с точки зрения логики исследования апостериорной байесовской вероятности $P(H_0/D)$. Поскольку эти вероятности связаны между собой при помощи формулы Байеса, рассматриваемая проблема получила название «дилеммы Байеса». Невозможность прямой вероятностной трактовки результатов, полученных в эксперименте, влечет за собой целый ряд проблем прикладного и практического характера.

Наиболее ярко эти проблемы проявляются в задачах сравнения выборок. Применение метода однофакторного дисперсионного анализа ANOVA с поставленной гипотезой H_0 о равенстве средних в k группах испытуемых ($a_1 = a_2 = \dots = a_k$) в случае принятия H_1 дает лишь результат о различном среднем уровне показателя, ничего не отвечая на главный в этом случае для исследователя вопрос: какие именно группы и какие именно значения качественного признака являются наиболее вероятными причинами возникновения в совокупности статистической неоднородности. Решение этих вопросов при помощи попарных сравнений групп, влечет за собой проблему множественных сравнений Шеффе, поскольку ситуация, в которой результаты методов ANOVA и t-критерия Стьюдента противоречат друг другу, как теоретически, так и практически не исключена.

Несмотря на целый ряд попыток улучшить традиционную методологию, к которым можно отнести: использование различных мер тесноты связи (effect size measures) между психологическими переменными (типа ϵ^2 , ω^2 , η^2); выявление величины β -мощности критерия для экспериментальных исследований, использование множества статистических моделей для одних и тех же данных, использование доверительных интервалов и т.п., существенных улучшений в сторону повышения корректности статистических процедур *NHST* пока не наблюдается. К сожалению, то, что методология статистического вывода должна развиваться в русле общей теории принятия решений и, по-существу, обосновывать выбор исследователя между предлагаемыми различными моделями реальности, пока еще не стало нормой для практических экспериментальных исследований.

В то же время, одной из альтернатив методологии статистического вывода, основанной на проверке нулевых гипотез (*NHST*), являются методы теории информации. Эти методы позволяют осуществлять естественный выбор между конкурирующими моделями (гипотезами) на основе расчетов потери информации, происходящей при замене реальной функции распределения функцией распределения каждой из рассматриваемых моделей.

Базовым понятием этих методов является информация Кульбака-Лейблера $I(g, f)$, являющаяся статистическим аналогом энтропии Больцмана. Под $I(g, f)$ имеется в виду информация, потерянная при замене реальности f моделью g с параметрами θ :

$$I(g, f) = E_f \left(\log \frac{f(x)}{g(x/\theta)} \right).$$

В качестве $I(g, f)$ в статистических задачах принято использовать функции вида $-2L + V(k, n)$, где L -функция максимума правдоподобия, k -число степеней свободы модели, n -число независимых испытаний. Наилучшей моделью из рассматриваемого множества считается модель, минимизирующая значение информационного критерия.

Наиболее часто на практике применяются критерии Акайка и Ханнана-Куина и Байеса, вычисляемые по формулам: $AIC = -2 \ln(L) + 2k$, $HQC = -2 \ln(L) + 2k \ln(\ln(n))$, $BIC = -2 \ln(L) + k \ln(n)$.

Байесовский информационный критерий дает возможность оценить апостериорные вероятности нулевой и альтернативной гипотез как $p_{BIC}(H_0/D) = \frac{BF}{BF+1}$ и

$p_{BIC}(H_1/D) = 1 - p_{BIC}(H_0/D)$, где байесовский фактор BF определяется как $BF = \exp(\Delta BIC/2)$, $\Delta BIC = BIC(H_1) - BIC(H_0)$.

Согласно работе [2], в случае регрессионных моделей и в задачах сравнения выборок для расчета информационного критерия Байеса может быть использовано соотношение

$$BIC = n \ln(1 - R^2) + k \ln(n), \text{ где } 1 - R^2 = \frac{SSE}{SS_{total}},$$

квадратов.

В соответствии с предложенной в работе [1] классификацией, принято рассматривать вероятности P_{BIC} в интервале (0,50;0,75) как слабое свидетельство в пользу альтернативной гипотезы H_1 , вероятности в интервале (0,75;0,95) как положительное, (0,95;0,99) как сильное и свыше 0,99 как очень сильное.

Сравним результаты использования методов *ANOVA* и *BIC* для исследования зависимости готовности к риску от типа темперамента (ЕJ, EP, IJ, IP) на выборке 62 человек (студенты СГУТиКД); использовались опросники ЛФР-25 Т.В. Корниловой и типа темперамента Д. Кейрси. Рассматривались следующие модели: однородность всей выборки испытуемых, различие 4-х групп, отличие экстравертов от интровертов (Е от I), отличие решающего типа от перцептивного (J от P), отличие каждого из четырех выделенных типов от всей остальной выборки.

Полученные результаты приведены в таблице 1. В последней графе приведена вероятностная оценка шансов каждой из 7-ми гипотез о статистических различиях против гипотезы однородности выборки, позволяющая оценить риск принятия каждой из альтернативных гипотез.

Согласно критерию Байеса, из предложенных моделей лучшей оказалась модель отличия типа EP от остальной выборки (ранг 1). Обращает на себя внимание факт соответствия рангов гипотез по *BIC* точным значениям *F*- дисперсионного отношения Фишера, но не их *p*-уровням значимости по *NHST*.

Таблица 1.

**Результаты статистического тестирования
гипотез по методам ANOVA и BIC**

N	Вид модели	SSE	F-Фишера	p-уровень	BIC	Δ BIC	BF	Ранг по BIC	$P_{BIC}(H_1)$ шансы (%) против H_0
1	Однородность	794,97	-	-	416,06			-	-
2	Отличие 4-х групп	607,97	5,95	0,0013	405,42	-4,48	0,11	4	90,37
3	Отличие E от I	674,49	10,72	0,0018	407,83	-6,30	0,04	2	95,88
4	Отличие J от P	739,97	4,54	0,0373	413,61	-0,55	0,76	5	56,85
5	Отличие EJ от других	792,73	0,17	0,6823	417,88	3,72	60,42	7	13,47
6	Отличие EP от других	666,56	11,56	0,0012	407,13	-7,03	0,03	1	97,11
7	Отличие IJ от других	693,69	8,79	0,0044	409,61	-4,56	0,10	3	90,70
8	Отличие IP от других	791,55	0,26	0,6103	417,79	3,63	60,13	6	14,03

Предлагаемый подход обладает по сравнению с *NHST* рядом преимуществ: рассмотрение одномерных и многомерных моделей в рамках единой методологической концепции, применимость к различным задачам статистики с проверкой множественных альтернатив, при этом модели не обязательно должны быть вложенными. Таким образом, на основе внедрения методов теории информации в статистическую науку становится возможной прямая вероятностная оценка выдвигаемых исследователями гипотез и их сравнение между собой с использованием общих принципов и критериев оптимальности.

Примечания:

1. Raftery, A. E. (1999). Bayes factors and BIC: Comment on "A critique of the Bayesian Information Criterion for model selection". *Sociological Methods & Research*, 27, 411-427.
2. Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, 14, 779-804.

УДК 519/316

**Использование байесовского информационного критерия в задачах
статистического вывода**

¹ Надежда Юрьевна Губанова

² Симон Жоржевич Симаворян

¹⁻² Сочинский государственный университет, Россия
35400, г. Сочи, ул. Советская, 26а

¹ кандидат психологических наук, доцент
E-mail: ngubanova@hotmail.com

² кандидат технических наук, доцент
E-mail: simsim58@mail.ru

Аннотация. В статье обсуждается возможность применения информационного критерия Байеса как альтернативы традиционным методам статистической проверки гипотез. Приведены сравнительные результаты методов *ANOVA* и *BIC* на примере психологического исследования.

Ключевые слова: статистический вывод; тестирование гипотез; информация Кульбака-Лейблера; байесовский информационный критерий.