

О самосогласованности статистической значимости

И. Б. Смирнов

Петербургский институт ядерной физики

10 января 2017 г.

Введение

- “Т”-величина (p -value): вероятность наблюдать данный или более несовместимый с гипотезой результат, если она верна.
- Значимость: $-F^{-1}(p\text{-value})$, F — интеграл ф. Гаусса.
- Статистика (statistic): любая функция набора независимых наблюдений (случайных величин). Но “статистика” (statisticS) используется в физике для обозначения набора или количества данных, как название этой науки и для других целей.
- Тестовая статистика (test statistic): некоторая функция наблюдения, предназначена для тестирования гипотез. Обычно большие или малые значения означают несогласие с тестируемой гипотезой H_0 .
- Доверительные интервалы:
 - Границы задаются данной малой вероятностью наблюдаемого или более несовместимого результата.
 - Корректность: “покрытие” (частота покрытия).
 - Неформальное понятие “покрытие из-за конструкции” (“coverage by construction”).
 - Нет мешающих параметров: есть “покрытие из-за конструкции”
 - Есть мешающие параметры: “покрытия из-за конструкции” может не быть, но покрытие можно проверить численными методами.
- p -величины:
 - Понятия “покрытие” нет!
 - Мешающие параметры и разные методы расчета дают **разные** p -величины.
 - Концептуальная проблема выбора “правильной” p -величины.

Выбор p -величины

- Максимизация p -величины подбором мешающих параметров:
 - 1) Дает консервативную оценку, но насколько консерватизм оправдан?
 - 2) Нетривиальные максимумы существуют не для всех тестовых статистик и выборок.
 - 3) Существование таких тестовых статистик и выборок не доказано для произвольной задачи.
 - 4) В некоторых случаях выбранные мешающие параметры несовместимы с измерениями.
 - 5) Установление ограничений при подборе приводит к зависимости от (произвольных) ограничений.
- Максимизация p -величины подбором мешающих параметров на доверительном множестве: [R. L. Berger, D. D. Boos, J.Am.Stat.Ass., 89 (1994) 1012] “Valid p -value”: $P(\rho \leq \alpha) \leq \alpha$ для любого $\alpha \in [0, 1]$. Пусть C_β есть доверительное множество соотв. $1 - \beta$ для θ , если H_0 — истинна. $\rho_\beta = \sup_{\theta \in C_\beta} \{\rho(\theta)\} + \beta$, ρ_β заменяется на $\min\{\rho_\beta, 1\}$, и этот минимум — “valid p -value”.
Устраняет (формально) проблему 4), но только её.

Реальный пример, условия, обозначения

Многоканальные Пуассоновские измерения с (возможно) неизвестным ожидаемым сигналом (аксептанс) \vec{a} и (возможно) неизвестным ожидаемым фоном \vec{b} : $\vec{f} = t_a \vec{a} s + t_b \vec{b}$.

Если \vec{a} и (или) \vec{b} неизвестны, они измеряются как \vec{n}_a и \vec{n}_b .

$$\prod_i P(n_i, n_{ai}, n_{bi} | s, a_i, b_i) = P(\vec{n}, \vec{n}_a, \vec{n}_b | s, \vec{a}, \vec{b}) = P(\vec{n} | t_a \vec{a} s + t_b \vec{b}) P(\vec{n}_a | \vec{a}) P(\vec{n}_b | \vec{b}).$$

Обозначения:

- $P(n|y)$: дискретная вероятность.
- $p(x|y)$: плотность вероятности.
(в англоязычной литературе обычно $f(x|y)$ или $f(x; y)$ но $p(x|y)$ тоже используется: [В. К. Гришин и др. 1988, Heinrich 2004, Kass and Wasserman 1996, §2.3.5 of James 2012].)
- ρ : p -величина (p -value) (чтобы отличить от плотности) (моё обозначение).
- z : статистическая значимость (significance), соответствующая ρ :

$$1 - \rho = F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$$

- \hat{s} : максимум $p(s)$.
- Индекс γ , например. \vec{n}_γ , \hat{s}_γ , и т.п.: данные полученные при “субгенерации”,
(в англоязычной литературе обычно “псевдоэксперименты” (pseudo-experiments) или “toy experiments”).

Фрекенвентистская обработка максимально правдоподобной оценки

Работы: M. Mandelkern J. Schultz, J. Math. Phys. 41 (2000) 5701.

S. Ciampolillo, Nuovo Cimento A 111 (1998) 1415.

Мой график, без мешающих параметров:

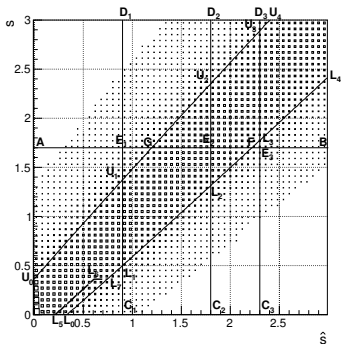


Figure 1: Распределение наблюдаемой наиболее вероятной величины s (обозначенной \hat{s}) как функции истинной s для 5 каналов с известным ожидаемым сигналом и фоном.

Постоянная вероятность оказаться внутри: $[U_0, U_4] - [L_5, L_6, L_7, L_4]$.

Постоянная односторонняя вероятность: $[U_0, U_4] - [L_0, L_4,]$ (лучше для физики).

Идея доказательства: (применима для любого метода дов. интервалов):

$$P(L_3 \text{ above } E_3) = P(E_3 \text{ to the right of } F) = \int_{\hat{s}_{\text{right}}}^{\infty} p(\hat{s}_\gamma | s) d\hat{s}_\gamma = \alpha$$

Но много деталей. Neuman, 1937:

6 страниц доказательства плюс 17 страниц введения.

$P(\vec{n}_\gamma | \hat{s}, \vec{a}, \vec{b})$ не может быть использована вместо \hat{s} .

Дискретная форма:

$$\sum_{\vec{n}_\gamma: \hat{s}_\gamma(\vec{n}_\gamma, \vec{a}, \vec{b}) \geq \hat{s}_{\text{obs}}} P(\vec{n}_\gamma | s, \vec{a}, \vec{b}) = \alpha$$

Чтобы представить результат в виде интервала, мешающие параметры должны быть удалены.

Фреквентистская обработка максимально правдоподобной оценки, варианты для мешающих параметров (из моего семинара от 2015 г.)

- Субгенерация:
 - Мешающие параметры:
 - Случайные согласно формуле Байеса с \vec{n}_a и \vec{n}_b с безопасными априорными (SSP: Subgeneration with Safe Priors).
 - Фиксированные наиболее вероятные для \vec{n}_a и \vec{n}_b (SM: Subgeneration with the Most probable nuisance parameters).
 - Переменные наиболее вероятные для \vec{n} , \vec{n}_a и \vec{n}_b при данном s (SA: Subgeneration with the Adjusted nuisance parameters).
 - Измерения мешающих параметров:
 - Фиксированные (обеспечивают хорошую модельную интерпретацию.)
 - Случайные (RN: Random Nuisance). (Интервалы короче, но нет модельной интерпретации .)
- Анализ:
 - \hat{s} с маргинализацией по мешающим параметрам (FMML: Frequency of Marginalized Maximum Likelihood)
 - \hat{s} с “глобальной” максимизацией по мешающим параметрам (FGML: Frequency of Global Maximum Likelihood)

Фреквентистская обработка максимально правдоподобной оценки,
варианты для мешающих параметров
основные комбинации (из моего семинара от 2015)

- SSP–FMML
- SSP–FGML
- SMRN–FGML
- SARN–FGML
- SSPRN–FMML
- SSPRN–FGML

Подробности в моем препринте: <https://arxiv.org/abs/1309.0531v5>
(последняя версия от 31.12.2016 г.)

"Analysis of multichannel measurements of rare processes with uncertain expected background and acceptance"

Фреквентистская обработка максимально правдоподобной оценки, “покрытие из-за конструкции”, случай с мешающими параметрами

Неформальное понятие “покрытие из-за конструкции” (“coverage by construction”).

- Подразумевает выборку и границы идентичные тем, что используются при субгенерации (at s_{true}).
- Ограничений может быть меньше. То же доказательство применимо, если
 - $p(\hat{s}_\gamma)$, сгенерированное для любого реального эксперимента (то-есть с $s = s_{\text{true}}$) идентично $p(\hat{s}|s_{\text{true}})$.
 - Линии U и L — возрастающие функции \hat{s} .

Тогда $P(L_3 \text{ above } E_3) = P(E_3 \text{ to the right of } F) = \int_{\hat{s}_{\text{right}}}^{\infty} p(\hat{s}_\gamma|s) d\hat{s}_\gamma = \alpha$

- Применимо для SSP-FMML (SSP-FGML):

$$\int p(\vec{a}_\gamma|\vec{n}_a) \int p(\vec{b}_\gamma|\vec{n}_b) \sum_{n_\gamma: \hat{s}_\gamma(\vec{n}_\gamma, \vec{n}_a, \vec{n}_b) \geq \hat{s}(\vec{n}, \vec{n}_a, \vec{n}_b)} P(\vec{n}_\gamma|s, \vec{a}_\gamma, \vec{b}_\gamma) d\vec{b}_\gamma d\vec{a}_\gamma = \alpha$$

$\hat{s}(\vec{n}, \vec{n}_a, \vec{n}_b)$ максимизирует $P(\vec{n}, \vec{n}_a, \vec{n}_b|s, \vec{a}, \vec{b})$

$\hat{s}_\gamma(\vec{n}_\gamma, \vec{n}_a, \vec{n}_b)$ максимизирует $P(\vec{n}_\gamma, \vec{n}_a, \vec{n}_b|s, \vec{a}, \vec{b})$

- Предположим, что мы повторяем только главный эксперимент и анализируем новые \vec{n} с изначально наблюдаемыми \vec{n}_a и \vec{n}_b .
- Предположим, что \vec{a} и \vec{b} — случайны и распределены по $p(\vec{a}|\vec{n}_a)$ и $p(\vec{b}|\vec{n}_b)$.

В каждом случае $[U_0, U_4]$ и $[L_0, L_4]$ — одинаковы.

Покрытие может быть доказано тем же способом. Это есть “модельная интерпретация” или “самосогласованность”.

Фреквентистская обработка максимально правдоподобной оценки, случай “RN”, слайд 1

SSPRN–FMML (SSPRN–FGML):

$$\int p(\vec{a}_\gamma | \vec{n}_a) \int p(\vec{b}_\gamma | \vec{n}_b) \sum_{\substack{\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma}: \\ \hat{s}_\gamma(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma}) \geq \hat{s}(\vec{n}, \vec{n}_a, \vec{n}_b)}} P(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma} | s, \vec{a}_\gamma, \vec{b}_\gamma) d\vec{b}_\gamma d\vec{a}_\gamma = \alpha$$

- Допустим, мы повторяем только главный эксперимент и анализируем новое \vec{n} с изначально наблюдаемыми \vec{n}_a и \vec{n}_b .

Распределение $\hat{s}_\gamma(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma})$ отличается от $\hat{s}_\gamma(\vec{n}_\gamma, \vec{n}_a, \vec{n}_b)$. U и L — отличаются. Покрытие НЕ гарантировано.

Фреквентистская обработка максимально правдоподобной оценки, случай “RN”, слайд 2

- Допустим, мы повторяем главный эксперимент и вспомогательные, и анализируем новые \vec{n} с новыми \vec{n}_a и \vec{n}_b .
- Допустим, \vec{a} и \vec{b} — случайны и распределены соответственно $p(\vec{a}|\vec{n}_a)$ и $p(\vec{b}|\vec{n}_b)$.

Для каждого \vec{n}^{new} , \vec{n}_a^{new} и \vec{n}_b^{new} решение нужно искать по

$$\int p(\vec{a}_\gamma|\vec{n}_a^{\text{new}}) \int p(\vec{b}_\gamma|\vec{n}_b^{\text{new}}) \sum_{\substack{\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma}: \\ \hat{s}_\gamma(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma}) \geq \hat{s}(\vec{n}^{\text{new}}, \vec{n}_a^{\text{new}}, \vec{n}_b^{\text{new}})}} P(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma}|s, \vec{a}_\gamma, \vec{b}_\gamma) d\vec{b}_\gamma d\vec{a}_\gamma = \alpha$$

Распределения $p(\vec{a}_\gamma|\vec{n}_a^{\text{new}})$ и $p(\vec{b}_\gamma|\vec{n}_b^{\text{new}})$ отличаются от тех, которые были использованы для анализа изначального эксперимента. Границы U и L будут разными для каждого следующего эксперимента. Покрытие НЕ гарантировано.

Невозможно найти последовательность экспериментов, для которых покрытие гарантировано для методов “Random-Nuisance”.

Нет “modeling interpretation” или “самосогласованности”.

Фреквентистская обработка отношения правдоподобий, тестовая статистика (из моего семинара от 2015)

$$Q(s) = \frac{P(\vec{n}|\hat{s})}{P(\vec{n}|s_{\text{ref}})}. \quad (1)$$

В начале было предложено $s_{\text{ref}} = 0$.

A. L. Read, 1st Workshop on Confidence Limits, CERN, Geneva Switzerland, 2000, CERN-2000-005, pp 81-101.

Моё обозначение: Background Related (BR).

Новый подход: s_{ref} максимизирует $P(\vec{n}|s_{\text{ref}})$, но ограничено в интервале $[0, s]$. Это может быть использовано для верхнего предела (верхней границы доверительного интервала).

The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*. August 2011, ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005.

Значимость, $s = 0$, следовательно s_{ref} просто максимизирует $P(\vec{n}|s_{\text{ref}})$ в интервале $[0, +\infty]$.

По аналогии (моё обобщение) для нижней границы s_{ref} может быть выбран так, что он максимизирует $P(\vec{n}|s_{\text{ref}})$ и ограничен условием $s_{\text{ref}} \geq s$.

Моё обозначение: the LHC-style method, constrained-maximum-related method (CMR).

Unified approach: Local notation: unconstrained-maximum-related (UMR).

Фреквентистская обработка отношения правдоподобий, получение верхнего предела

Вероятность, что Q для сигнала плюс фон совпадет или окажется меньше чем наблюдаемое значение

$$CL_{s+b} = P_{s+b}(Q \leq Q_{obs}) = \sum_{\vec{n}_\gamma: Q(\vec{n}_\gamma, \vec{a}, \vec{b}, s) \leq Q(\vec{n}, \vec{a}, \vec{b}, s)} P(\vec{n}_\gamma | s, \vec{a}, \vec{b}) = \alpha. \quad (2)$$

Моё обозначение: Frequency of Likelihood Ratio (FLR).

LHC-style CL_{s+b} : p -величина по той же формуле с $s = 0$ и $s_{ref} > 0$

Проблемы с CL_{s+b}

- Исключение гипотезы с истинной s с вероятностью равной α даже для экспериментов микроскопически зависящих от сигнала (Для “классического метода см. [Read et al, 2000])
- Бессмысленно малый верхний предел при флуктуациях вниз сигнала.

Устраняются путем “нормировки” CL_{s+b} на CL_b . Исследователь учитывает, насколько хорошо эксперимент описывается фоном.

Фреквентистская обработка отношения правдоподобий, Метод отношения к максимуму (Maximum Related Method), варианты для мешающих параметров (из моего семинара от 2015)

- Субгенерация:
 - Мешающие параметры:
 - Случайные согласно формуле Байеса с \vec{n}_a и \vec{n}_b с безопасными априорными (SSP: Subgeneration with Safe Priors).
 - Фиксированные наиболее вероятные для \vec{n}_a и \vec{n}_b (SM: Subgeneration with the Most probable nuisance parameters).
 - Переменные наиболее вероятные для \vec{n} , \vec{n}_a и \vec{n}_b при данном s
 - Измерения мешающих параметров:
 - Фиксированные (обеспечивают хорошую модельную интерпретацию.)
 - Случайные (RN: Random Nuisance). (Интервалы короче, но нет модельной интерпретации .)
- Анализ:
 - Отдельная маргинализация по мешающим параметрам (MM: Marginalized Maximum).
 - Отдельная максимизация по мешающим параметрам (GM: Global Maximum).

Фреквентистская обработка отношения правдоподобий, Модельная интерпретация (самосогласованность) интервалов

Для примера, LHC-style CL_s :

$$\sum_{\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma}:} P(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma} | s_u, \hat{\vec{a}}, \hat{\vec{b}}) = \alpha$$
$$Q(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma}, s) \leq Q(\vec{n}, \vec{n}_a, \vec{n}_b, s)$$

Субгенерация выполняется с сигналом s_u и с $\hat{\vec{a}}$ и $\hat{\vec{b}}$, которые максимизируют $P(\vec{n}, \vec{n}_a, \vec{n}_b | s_u, \hat{\vec{a}}, \hat{\vec{b}})$. $Q(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma}, s)$ вычисляется с разными $\hat{\vec{a}}_\gamma$ и $\hat{\vec{b}}_\gamma$ которые максимизируют $P(\vec{n}_\gamma, \vec{n}_{a\gamma}, \vec{n}_{b\gamma} | s, \hat{\vec{a}}_\gamma, \hat{\vec{b}}_\gamma)$, и с разными s_{ref} , \vec{a}_{ref} and \vec{b}_{ref} .
Если мы повторяем эксперимент вместе или без вспомогательных экспериментов, доверительная область всё-равно разная и интервалы тоже разные.

Нет “модельной интерпретации”.

Самосогласованность p -величин.

Существует много статей, требующих “однородность” p -величин:
 $\int_0^{\rho_t} p(\rho) d\rho \leq \rho_t$ for any threshold ρ_t .

- Если знак “<”: консервативный случай.
- Если знак “>”: антиконсервативный случай.
- Если знак “=”: однородность (с дискретными эффектами).

Интеграл — это p -величина для тестовой статистики ρ_t (или для соответствующей значимости, с учетом изменения направления для большего разногласия).

Нужна ли нам другая p -величина?

Если p -величина правильно рассчитана, почему она должна быть однородной?

[J. M. Robins, A. van der Vaart, V Ventura, J. Am. Stat. Ass. 95 (2000) 1143.]:

“ p -величина полезна, только если её распределение для нулевой модели известно Если нам говорят, что температура 30° , то существенно, чтобы назвали и шкалу, если надо знать, что планировать: заплыв или лыжную прогулку.”

Самосогласованность p -величин

Предположим, для любых данных \vec{n}_a и \vec{n}_b

существует одна и та же выборка для любого \vec{n} .

Для любых \vec{n}_1 и \vec{n}_2 , если $\hat{s}_{\text{obs}}(\vec{n}_1, \vec{n}_a, \vec{n}_b) > \hat{s}_{\text{obs}}(\vec{n}_2, \vec{n}_a, \vec{n}_b)$, то

- все \vec{n}_γ , которые учитываются для $\rho(\vec{n}_1, \vec{n}_a, \vec{n}_b)$, должны также учитываться для $\rho(\vec{n}_2, \vec{n}_a, \vec{n}_b)$,
- но как минимум одно $\vec{n}_\gamma = \vec{n}_2$ учтенное в $\rho(\vec{n}_2, \vec{n}_a, \vec{n}_b)$ не должно быть учтено в $\rho(\vec{n}_1, \vec{n}_a, \vec{n}_b)$.

Поэтому $\rho(\vec{n}_2, \vec{n}_a, \vec{n}_b) > \rho(\vec{n}_1, \vec{n}_a, \vec{n}_b)$.

Если использовать p -величину ρ как (альтернативную) тестовую статистику, то получим (альтернативную) p -величину, которая будет равна обычной p -величине, причем они обе будут распределены равномерно (с дискретными особенностями) на $[0, 1]$ для фиксированных \vec{n}_a и \vec{n}_b .

Я называю это “самосогласованностью” p -величин.

Это не “работает” для RN-методов,

поскольку $\vec{n}_{a\gamma}$ и $\vec{n}_{b\gamma}$ флуктуируют и выборка не идентична.

SSPRN-FMML, SSPRN-FGML, ... LHC- CL_s (not asymptotic) не обеспечивают самосогласованных p -величин.

Самосогласованность p -величин коррелирует с самосогласованностью интервалов: метод либо имеет и то, и другое, либо не имеет ни того, ни другого.

Простые примеры, слайд 1

Table 1: Сравнение значимостей и интервалов, полученных разными методами для простых примеров, интервалы по методу Байеса и методу отношения правдоподобий.

n	67	60		88		37, 51	standard	standard	7
t_a	1	1		1		1	cond.,	cond.,	1
a / n_a	1 / -	1 / -		1 / -		$a_1, a_2 / -, -$	5	30	- / 3
t_b	2	10		10		10	channels	channels	1
b / n_b	- / 15	- / 0		- / 3		$-, - / 3, 0$			2 / -
	z	z	z	limits	z	limits	z	z	z
Bayesian central	—	—	—	23.6–81.9	—	33.0–75.4	—	—	—
Bayesian central modified	—	—	—	23.6–81.9	—	33.1–75.4	—	—	—
Likelihood Ratio	—	—	—	28.0–79.9	—	49.1–74.0	—	—	—

Простые примеры, слайд 1

Table 2: Сравнение значимостей и интервалов, полученных разными методами для простых примеров, фреквентистские (частотные) методы.

n	67	60	88		37, 51		standard	standard	7
t_a	1	1	1		1		cond.,	cond.,	1
a / n_a	1 / -	1 / -	1 / -		$a_1, a_2 / -, -$		5	30	- / 3
t_b	2	10	10		10		channels	channels	1
b / n_b	- / 15	- / 0	- / 3		-, - / 3, 0				2 / -
	z	z	z	limits	z	limits	z	z	z
SSP-FGML	2.89	2.72	1.89	18.8-81.8	2.46	32.0-74.7	2.62	0.65	2.61
SSP-FMML	2.89	2.72	1.89	18.8-81.8	2.47	32.0-74.6	2.61	1.13	2.61
SSP-GM-CMR-NFLR	2.89	2.72	1.89	18.8-81.9	2.47	31.9-74.7	2.58	0.07	2.61
SSP-MM-CMR-NFLR	2.89	2.72	1.89	18.8-81.9	2.47	32.0-74.7	2.59	0.37	2.61
SSP-FGML-GM-CMR-FLR	2.89	2.72	1.89	18.8-81.8	2.40	30.8-74.7	2.52	0.07	2.61
SSP-FMML-MM-CMR-FLR	2.89	2.72	1.89	18.8-81.8	2.41	31.3-74.7	2.53	0.37	2.61
SEP-FBML	3.09	∞	2.35	32.1-74.8	≈ 4.6	52.2-67.3	≈ 3.4	> 5.0	2.61
SSPRN-FGML	3.63	4.02	2.80	31.7-84.4	3.69	45.3-74.2	3.67	1.96	1.73
SSPRN-FMML	3.70	3.93	2.85	33.1-83.2	3.61	44.2-73.9	3.70	2.51	2.08
SSPRN-GM-CMR-NFLR	3.01	4.30	2.01	23.7-79.6	3.90	47.6-74.2	3.07	1.24	2.70
SSPRN-MM-CMR-NFLR	3.03	4.29	2.20	27.5-79.8	3.90	45.8-73.8	3.35	1.65	2.71
SARN-GM-CMR-NFLR	3.01	3.07	2.09	25.4-80.0	2.85	45.6-74.2	2.76	2.42	2.65
SARN-GM-UMR-FLR	3.01	3.07	2.09	27.6-81.7	2.85	47.7-73.4	2.76	2.42	2.65
Asymptotic GM-MR-FLR	3.04	3.38	2.19	-	3.13	-	2.89	0.60	2.75
Used nuisance parameters	27.3	5.45	8.27	-	3.6, 4.6	-	-	-	-
Min/Max RN-GM-CMR-FLR	2.83	3.02	1.74	13.9-79.6	2.77	42.1-75.3	2.60	0.24	2.64
Min/Max RN-GM-UMR-FLR	[2.87]	[3.02]	[1.74]	21.6-81.7	[2.78]	45.8-75.5	[2.63]	[0.41]	[2.64]
Best nuisance parameters	4.5	6.1	2.8	-	29.2, 5.3	-	-	-	2.4
SARN-MM-CMR-NFLR	3.02	3.07	2.10	25.6-80.1	2.85	44.9-74.0	2.78	2.61	2.74
SARN-MM-UMR-FLR	3.02	3.07	2.10	27.1-86.5	2.85	40.4-74.8	2.78	2.61	2.74
Asymptotic MM-MR-FLR	3.02	3.11	2.16	-	2.85	-	2.80	0.53	2.62
Min/Max-RN-MM-CMR-FLR	3.01	3.01	2.01	25.2-83.9	2.77	37.9-74.9	2.77	0.52	2.61
Min/Max-RN-MM-UMR-FLR	[3.01]	[3.01]	[2.01]	20.1-87.3	[2.77]	28.3-84.5	[2.77]	[0.52]	[2.61]
Best nuisance parameters	$\gtrsim 15$	6.1	3.4	-	$\gtrsim 4.2, \gtrsim 5$	-	-	-	12.8
Min/Max-E-RN-MM-CMR-FLR	0.30	3.01	0.0	0.0 - 85.0	0.74	37.3-77.3	0.0	0.19	2.61

Простые примеры, пояснение к таблице

Методы Min/Max-RN-BM-CMR-FLR и Min/Max-RN-BM-UMR-FLR (“В” — либо “М”, либо “G”) дают идентичную по определению значимость, но разные пределы.

Чтобы показать, как влияют ограничения по мешающим параметрам, значимость в строчках Min-RN-BM-UMR-FLR вычислена со следующими ограничениями:

Каждая a_i or b_i ограничена согласно пуассоновским центральным доверительным интервалам, для измеренных n_{ai} or n_{bi} (которые используются как тестовые статистики), соответственно, и для односторонней α равной $(1 - (1 - \rho)^{1/k})/2 \approx \rho/(2k)$, где $\rho = F(-\max(z, 1.0))$, F определена на слайде 3, и z — значимость, которая находится в итерационной процедуре фитирования.

Подробности в моем препринте: <https://arxiv.org/abs/1309.0531v5>

Заключение

- Интервалы и значимости имеют свойство самосогласованности.
- Статистические методы либо имеют, либо не имеют оба этих свойства.
- Методы, которые имеют эти свойства, следует читать более надежными.