Исследование непараметрических методов проверки гипотез с помощью статистического моделирования

Сальников Дмитрий Игоревич, гр. 622

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Статистическое моделирование

Научный руководитель: д.ф.-м.н., профессор Мелас В.Б.
Рецензент: доктор технических наук, профессор Григорьев Ю. Д.

Санкт-Петербург
2019г.

- в последнее время при проверке гипотез приобрели популярность перестановочные критерии, обладающие высокой мощностью, гибкостью и универсальностью;
- теоретическое сравнение мощности подобных критериев является сложной задачей, часто неразрешимой;
- в работе введен ряд новых перестановочных критериев для проверки гипотезы равенства двух распределений;
- с помощью статистического моделирования установлено, что для многих типичных распределений один из введенных критериев является наиболее мощными.

Постановка задачи

$X = \left(X_1, \ldots, X_{n_1}\right)^{\mathrm{T}} \sim F_1$, $Y = \left(Y_1, \ldots, Y_{n_2}\right)^{\mathrm{T}} \sim F_2$ — независимые выборки, где F_1 и F_2 — непрерывные функции распределения.

Гипотеза о равенстве двух распределений

$$H_0 \; : \; F_1 = F_2, \tag{1}$$

$$H_1 \; : \; F_1 \neq F_2. \tag{2}$$

Без потери общности положим $n_1 = n_2 = n$.

Перестановочные критерии. Обозначения

Определим вектор

$$Z = Z(\pi_0) = \Big(X_1, \ldots, X_n, Y_1, \ldots, Y_n\Big)^{\mathrm{T}}. \qquad (3)$$

Введем множество перестановок элементов вектора Z:

$$\Big\{Z(\pi_k) = \Big(\underbrace{X_1(\pi_k), \ldots, X_n(\pi_k)}_{X(\pi_k)}, \underbrace{Y_1(\pi_k), \ldots, Y_n(\pi_k)}_{Y(\pi_k)}\Big)^{\mathrm{T}}\Big\}_{k=1}^{(2n)!}. \qquad (4)$$

Степенные критерии

Определим статистики на векторах (3), (4) при $\gamma > 0$:

$$K_\gamma\left(Z(\pi_k)\right) = K_\gamma\left(X(\pi_k), Y(\pi_k)\right) = \sum_{i,j=1}^{n} |X_i(\pi_k) - Y_j(\pi_k)|^\gamma.$$

Алгоритм проверки H_0 (1) при альтернативе H_1 (2):

- пусть d — общее число перестановок, r — число перестановок π_k, для которых $K_\gamma(Z(\pi_k)) \geq K_\gamma(Z(\pi_0))$;
- если $\frac{r}{d} < \alpha$, то H_0 отвергается в пользу H_1 с уровнем значимости α.

K_1 и K_2: известные результаты

Согласно Corain et al. 2013 [1] критерий K_2 эквивалентен

$$\widetilde{K}_2(Z(\pi_k)) = \left(\overline{X}(\pi_k) - \overline{Y}(\pi_k)\right)^2,$$

где $\overline{X}(\pi_k)$, $\overline{Y}(\pi_k)$ — выборочные средние.

Из работ [2], [3] и [4] получены следующие выводы:

- критерий K_1 обладает высокой мощностью для широкого набора распределений;
- K_1 особенно эффективен в случае симметричных распределений с совпадающими центрами;
- мощность критерия K_2 близка к мощность классического критерия Стьюдента.

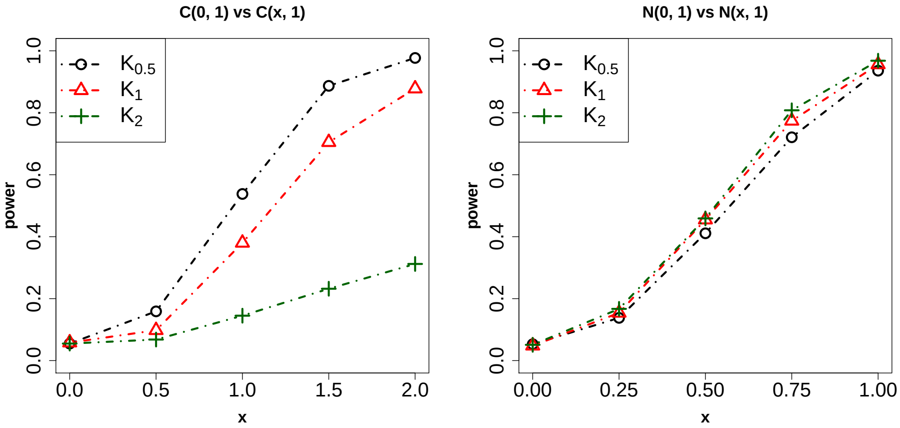
Степенные критерии. Постановка экспериментов

- уровень значимости $\alpha = 0.05$;
- критерии: K_γ, $\gamma = 0.5, 1, 1.5, 2$, критерий Стьюдента;
- распределения: нормальные, Коши, Вейбулла;
- объем выборок: $n_1 = n_2 = n = 30$;
- $m = 1000$ независимых испытаний в каждом случае;
- $d = 800$ случайных перестановок в каждом испытании (Keller-McNulty, Higgins 1987 [5], Marozzi 2004 [6]).

Степенные критерии. Выводы

- критерий $K_{0.5}$ является наиболее мощным в случаях
  - распределений Коши;
  - нормальных распределений с совпадающими центрами;
  - распределений Вейбулла, различающихся параметром формы;
- критерий $K_{0.5}$ является наименее мощным в случаях
  - нормальных гомоскедастичных распределений;
  - распределений Вейбулла, различающихся параметром масштаба;
- среди остальных рассмотренных критериев мощность критерия K_1 является наибольшей или статистически равна мощности лидирующего критерия.

Мощность критериев, объем выборок $n = 30$



Логарифмические критерии

Введем семейство статистик на векторах (3), (4):

$$
\begin{aligned}
L_\gamma(Z) &= \sum_{i,j=1}^{n} \ln(1 + |X_i - Y_j|^\gamma), \quad \gamma > 0, \\
L_\infty(Z) &= \sum_{i,j=1}^{n} \ln(|X_i - Y_j|).
\end{aligned}
$$

Алгоритм проверки гипотезы H_0 (1), H_1 (2) для критериев L_γ такой же, как и для критериев K_γ.

Логарифмические критерии. Асимптотика

$$\frac{L_\gamma(Z(\pi_k))}{K_\gamma(Z(\pi_k))} \xrightarrow[\max\limits_{1\le i<j\le 2n}|Z_i - Z_j| \to 0]{} 1, \quad \gamma > 0,$$

$$\frac{L_\gamma(Z(\pi_k))}{\gamma L_\infty(Z(\pi_k))} \xrightarrow[\min\limits_{1\le i<j\le 2n}|Z_i - Z_j| \to \infty]{} 1, \quad \gamma > 0.$$

Введем нормирующий множитель

$$C(Z) = \frac{\sum\limits_{1 \le i < j \le 2n} |Z_i - Z_j|}{n(2n-1)}.$$

Расширим семейство L_γ критериями со статистиками

$$L_\gamma^C(Z(\pi_k)) = \sum_{i,j=1}^{n} \ln\left(1 + \left(\frac{|X_i(\pi_k) - Y_j(\pi_k)|}{C(Z)}\right)^\gamma\right), \quad \gamma > 0.$$

Логарифмические критерии. Постановка экспериментов

- уровень значимости $\alpha = 0.05$;
- критерии: L_γ и L_γ^C при $\gamma = 0.5, 1, 2,$ L_∞;
- распределения: нормальные, Коши, Вейбулла;
- объем выборок: $n_1 = n_2 = n = 30$;
- $m = 1000$ независимых испытаний в каждом случае;
- $d = 800$ случайных перестановок в каждом испытании.

Логарифмические критерии. Табличные результаты

Таблица: Мощность критериев L_γ, $\gamma = 0.5, 1, 2, \infty$, объем выборок $n = 30$

F_1	F_2	$L_{0.5}$	L_1	L_2	L_∞
$C(0,1)$	$C(1.5,1)$	**0.9**	**0.902**	**0.901**	**0.89**
$C(0,1)$	$C(0,4)$	**0.899**	**0.902**	**0.902**	**0.889**
$N(0,1)$	$N(1,1)$	0.916	0.936	**0.953**	0.869
$N(0,1)$	$N(0,2.5)$	**0.888**	**0.898**	**0.901**	0.86
$W(5,1)$	$W(2,1)$	**0.852**	0.814	0.378	0.817
$W(1,3)$	$W(1,1)$	0.934	**0.943**	**0.954**	0.887

Жирным шрифтом выделены критерии, различие мощностей которых с лидирующим критерием статистически незначимо с вероятностью 0.95.

Логарифмические критерии. Выводы

- в большинстве рассмотренных случаев:
  - критерий L_∞ является наименее мощным среди рассмотренных логарифмических критериев;
  - мощность критерия L_1 заключена между значениями мощностей критериев $L_{0.5}$ и L_2;
- критерий L_1 является наиболее универсальным среди рассмотренных логарифмических критериев в случае отсутствия априорной информации о распределениях;

Сравнение перестановочных и классических критериев. Постановка экспериментов

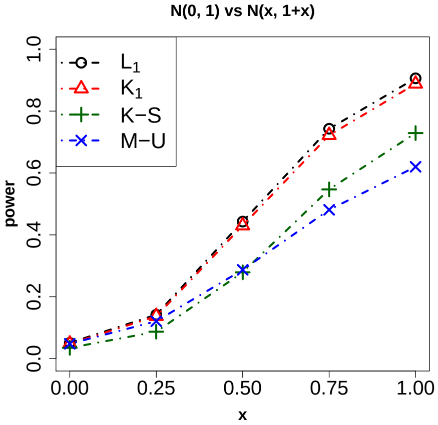
- уровень значимости $\alpha = 0.05$;
- критерии: $K_{0.5}$, K_1, L_1, L_1^C, Стьюдента, Манна-Уитни, Колмогорова-Смирнова;
- распределения: нормальные, Коши, смеси 90% нормального и 10% Коши, Вейбулла, Фишера, Парето;
- объем выборок: $n_1 = n_2 = n = 5, 30, 50$;
- $m = 1000$ независимых испытаний в каждом случае;
- $d = 800$ случайных перестановок в каждом испытании при $n = 30, 50$.

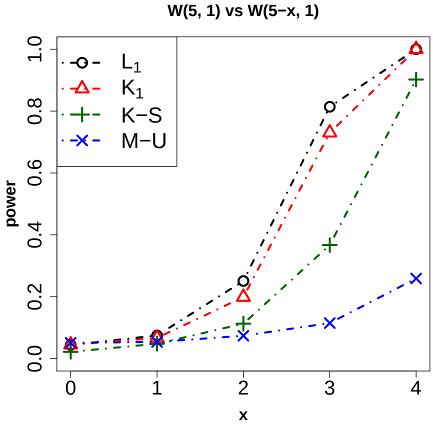
Таблица: Мощность критериев, объем выборок $n = 5$

F_1	F_2	$K_{0.5}$	K_1	L_1	L_1^C	t.test	M-U	K-S
$C(0,1)$	$C(5,1)$	**0.80**	0.73	**0.82**	**0.79**	0.45	0.51	0.41
$C(0,1)$	$C(0,20)$	**0.7**	0.22	**0.73**	0.66	0.03	0.06	0.05
$N(0,1)$	$N(2,1)$	0.72	**0.78**	0.73	0.74	**0.77**	0.67	0.38
$N(0,1)$	$N(0,9)$	**0.77**	0.25	**0.77**	0.73	0.05	0.06	0.04
$NC(0,1)$	$NC(2,1)$	**0.68**	**0.70**	**0.68**	**0.69**	0.66	0.61	0.34

Сравнение перестановочных и классических критериев. Графические результаты

Мощность критериев, объем выборок $n = 30$





Выводы

- мощность перестановочных критериев выше мощности неперестановочных в большинстве рассмотренных случаев;
- преимущество перестановочных критериев особенно велико в случае выборок малого объема $(n = 5)$ (аналогичные выводы смотри, например, в Ludbrook, Dudley 1998 [7]);
- среди рассмотренных перестановочных критериев:
  - критерий L_1 является наиболее мощным для большинства рассмотренных распределений;
  - преимущество критерия L_1 особенно велико в случаях распределений Коши и распределений Парето с параметром формы, равным единице.

New insights on permutation approach for hypothesis testing on functional data / L. Corain, V. Melas, A. Pepelyshev, L. Salmaso // Advances in Data Analysis and Classification. — 2014. — Vol. 8. — P. 339–356.

Sirsky M. On the Statistical Analysis of Functional Data Arasing from Designed Experiments : Ph. D. thesis / M. Sirsky ; University of Manitoba. — 2012.

Sturino J., Zorych I., Mallick B. et al. Statistical methods for comparative phenomics using high-throughput phenotype microarrays // The International Journal of Biostatistics. — 2010. — Vol. 6(1). — P. 29.

Мелас В.Б., Сальников Д.И., Гудулина А.О. Численное сравнение перестановочных и классических методов проверки статистических гипотез // Вестник СПбГУ. — 2016. — Т. 3(61). — С. 415–423.

Keller-McNulty S., Higgins J. Effect of tail weight and outliers on power and type-I error of robust permutation tests for location // Communications in Statistics — Simulation and Computation. — 1987. — Vol. 16. — P. 17–35.

Marozzi M. Some remarks about the number of permutations one should consider to perform a permutation test // Statistica. — 2004. — Vol. 64, no. 1. — P. 9.

Ludbrook J., Dudley H. Why Permutation Tests Are Superior to t and F Tests in Biomedical Research // American Statistician. — 1998. — Vol. 52. — P. 127–132.