A new omnibus test for the global null hypothesis

Sonja Zehetmayer, Andreas Futschik

Medical University of Vienna, Johannes Kepler University Linz

Outline

- Testing the global null hypothesis, popular approaches
- Omnibus method
- Simulation study

Introduction - The global null hypothesis

Setting

- m null hypotheses $H_{01}, \ldots, H_{0m}$
- **Independent** p-values $p_1, \ldots, p_m$
- Test the global null hypothesis, that none of the null hypotheses is false:

$$H_0 = \bigcap_{i=1}^{m} H_{0i}$$

Examples

- Meta analyses
- Independent subgroups in a clinical trial
- Experimental Evolution

Methods for testing the global null hypothesis I

Based on individual test statistics

- Maximum test statistic, minimum p-value
- E.g., Bonferroni, Simes (1986), Higher critisism - HC (Tukey, 1976, Donoho and Jin, 2004)
- E.g., the Bonferroni test rejects the global null hypothesis, if the minimum p-value falls below α/m.

Remark

Efficient, if only one or a few null hypotheses are false and effect sizes are large.

Methods for testing the global null hypothesis II

Based on combined test statistics

- Data from several endpoints (e.g., test statistics, p-values,...) are combined to a single test statistic
- E.g., Fisher (1932), Stouffer (1949)
- Fisher combination function: A combined test statistic is given by $T = -\sum_{i=1}^{m} 2 \log p_i$.

Remark

Efficient, if the alternative holds for most hypotheses even if effect sizes are moderate.

Omnibus tests

To maximize power

- Usually there is no a priori knowledge on the estimated number of false individual null hypotheses.
- The decision for a combination test or a test based on individual test statistics is not straigthforward.
- We propose non-parametric omnibus tests based on cumulative sums of the sorted p-values.

Omnibus tests

Simple assumptions:

- Independent p-values $p_1, \ldots, p_m$ obtained from testing $H_{01}, \ldots, H_{0m}$.
- We assume that the p-values are uniformly distributed under the global null hypothesis (or stochastically larger).

Omnibus tests

Algorithm

- The p-values are sorted, $p_{(1)} \leq \cdots \leq p_{(m)}$ and transformed with a monotonously **decreasing** function $h(.)$.

Omnibus tests

Algorithm

- The p-values are sorted, $p_{(1)} \leq \cdots \leq p_{(m)}$ and transformed with a monotonously **decreasing** function $h(.)$.
- Cumulative sums are calculated

$$S_i = \sum_{j=1}^{i} h(p_{(j)}), i = 1, \ldots, m$$

Omnibus tests

Algorithm

- The p-values are sorted, $p_{(1)} \leq \cdots \leq p_{(m)}$ and transformed with a monotonously **decreasing** function $h(.)$.
- Cumulative sums are calculated

$$S_i = \sum_{j=1}^{i} h(p_{(j)}), i = 1, \ldots, m$$

- The sums are transformed using the distribution function G_i of S_i under the global null hypothesis

Omnibus tests

Algorithm

- The p-values are sorted, $p_{(1)} \leq \cdots \leq p_{(m)}$ and transformed with a monotonously **decreasing** function $h(.)$.
- Cumulative sums are calculated

$$S_i = \sum_{j=1}^{i} h(p_{(j)}), i = 1, \ldots, m$$

- The sums are transformed using the distribution function G_i of S_i under the global null hypothesis
- The final test statistic $T^* = \max_{1 \leq i \leq m} G_i(S_i)$.

Omnibus tests

Algorithm

- The p-values are sorted, $p_{(1)} \leq \cdots \leq p_{(m)}$ and transformed with a monotonously **decreasing** function $h(.)$.
- Cumulative sums are calculated

$$S_i = \sum_{j=1}^{i} h(p_{(j)}), i = 1, \ldots, m$$

- The sums are transformed using the distribution function G_i of S_i under the global null hypothesis
- The final test statistic $T^* = \max_{1 \leq i \leq m} G_i(S_i)$.

The distribution function G_i is estimated by Monte Carlo method via simulation of uniformly distributed p-values.

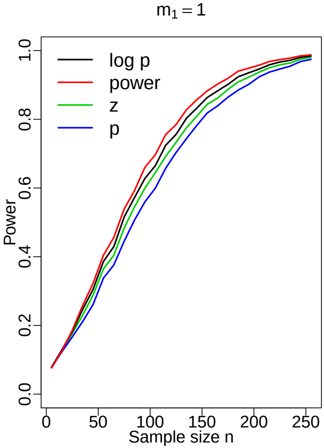
Omnibus tests

The following transformations $h(p)$ of the p-values were considered:

- $h(p) = 1 - p$ (omnibus p)
- $h(p) = -\log p$ **(omnibus log p)**
- $h(p) = \Phi^{-1}(1 - p)$ (omnibus z)
- $h(p) = p^{-\alpha}$ with $\alpha = 0.5$ (omnibus power).

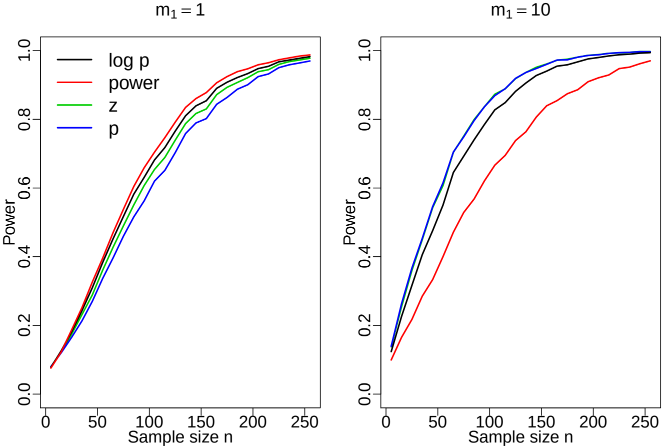
Simulation study for normally distributed data: Comparing transformations

- $m = 10$ hypotheses with m_1 alternatives, effect size $\Delta = 0.3/\sqrt{m}_1$
- $\alpha = 0.05$



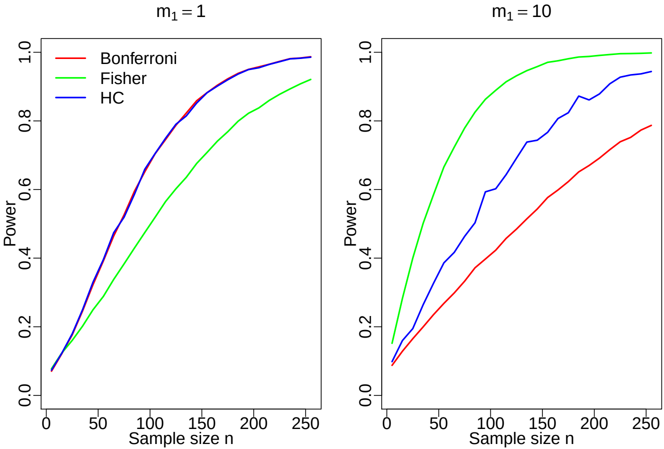
Simulation study for normally distributed data: Comparing transformations

- $m = 10$ hypotheses with m_1 alternatives, effect size $\Delta = 0.3/\sqrt{m_1}$
- $\alpha = 0.05$



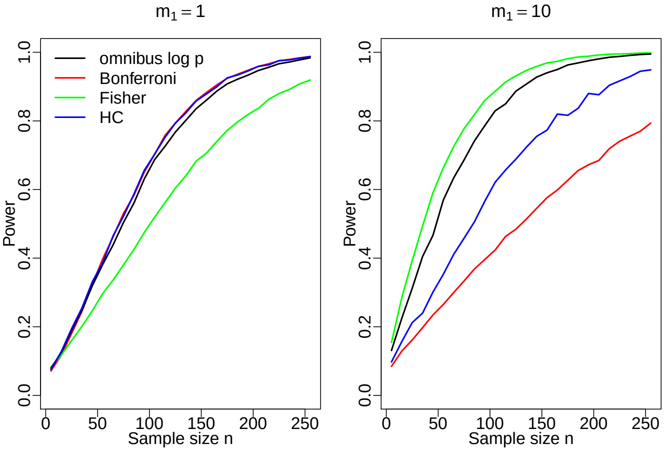
Simulation study for normally distributed data: Comparison with other methods

- $m = 10$ hypotheses with m_1 alternatives, effect size $\Delta = 0.3/\sqrt{m_1}$
- $\alpha = 0.05$



Simulation study for normally distributed data: Comparison with other methods

- $m = 10$ hypotheses with m_1 alternatives, effect size $\Delta = 0.3/\sqrt{m_1}$
- $\alpha = 0.05$



Maximin power

To assess overall performance: Worst case power values for m_1 **from 1 to** m for $\Delta = 0.3/\sqrt{m_1}$.

	$m = 10$	
	$n = 100$	$n = 200$
log p	0.63	0.92
Bonferroni	0.41	0.68
Fisher	0.49	0.83
HC	0.53	0.86

Maximin power

To assess overall performance: Worst case power values for m_1 **from 1 to** m for $\Delta = 0.3/\sqrt{m_1}$.

	$m = 10$		$m = 20$		$m = 1000$	
	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
log p	0.63	0.92	0.50	0.84	0.23	0.50
Bonferroni	0.41	0.68	0.28	0.47	0.11	0.18
Fisher	0.49	0.83	0.35	0.67	0.15	0.28
HC	0.53	0.86	0.40	0.73	0.14	0.30

Meta analysis

- 17 studies comparing survivors of post operative radiation therapy with or without adjuvant chemotherapy in patients with malignant gliomas (R-package metafor, Fine et al. 1993)
- Several follow-ups.
- Resulting **p-values**:

	6 months	18 months
log p	0.12	
Bonferroni	0.12	
Fisher	0.51	
HC	0.50	

Meta analysis

- 17 studies comparing survivors of post operative radiation therapy with or without adjuvant chemotherapy in patients with malignant gliomas (R-package metafor, Fine et al. 1993)
- Several follow-ups.
- Resulting **p-values**:

	6 months	18 months
log p	0.12	0.12
Bonferroni	0.12	0.28
Fisher	0.51	0.09
HC	0.50	0.15

Conclusion

- Omnibus test is a non-parametric method applicable for all kind of independent p-values.
- Omnibus test is not the overall optimal test for all scenarios, but we did not find scenarios, where the omnibus test is really bad compared to alternative methods.
- Best minimax criterium in the considered scenarios.

Futschik A, Taus T, Zehetmayer S (2018) An omnibus test for the global null hypothesis. Statistical Methods in Medical Research. To Appear.

R-package: omnibus (github)

Methods for testing the global null hypothesis III

Higher critisism (HC)

- Donoho and Jin (2004, 2015), original idea Tukey (1976)
- It is defined by

$$HC_m^* = \max_{1 \leq i \leq \alpha_0 m} \left\{ \sqrt{m} \frac{i/m - p_{(i)}}{\sqrt{p_{(i)}(1 - p_{(i)})}} \right\}$$

α_0 is a tuning parameter often set to $1/2$.