Multiple testing with discrete data

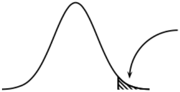
Alex Lewin and Elena Kulinskaya
alex.lewin@brunel.ac.uk

October 2016

Context

Hypothesis testing using p-values as measure of evidence against the null.

P-value (one-sided) $p_i = \mathbb{P}(T_i > t_i | H_{0i})$



Multiple testing: many different procedures available to decide which hypotheses to reject whilst controlling a required error rate at a given level α.

- Family Wise Error Rate (FWER),
- False Discovery Rate (FDR)

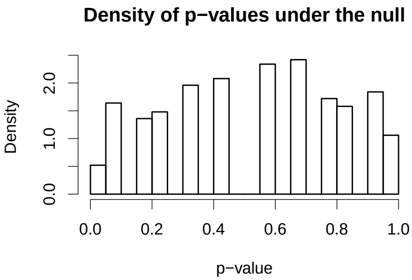
Procedures: Bonferroni (FWER), Hochberg (FWER), Benjamini and Hochberg (FDR), ...

A Problem with discrete data

Discrete data: p-values are not $U(0,1)$ under the null hypothesis

Example, conditional testing in 2x2 contingency tables

	A	$\bar{A}$	
B	x		100
$\bar{B}$			100
	80	120	200

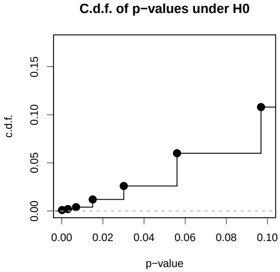


$\rightarrow$ using procedure for given nominal rate α, actual error rate can be much lower
$\rightarrow$ very conservative

A Solution

Old idea of randomised tests (Lehmann 1959):

1. Calculate p-value for observed data p^{obs},
2. Next smallest observable p-value p^{next} corresponds to the next most extreme possible test statistic
3. Draw RV $p^{rand} \sim U(p^{next}, p^{obs})$
   $\rightarrow$ under H_0, unconditionally $p^{rand} \sim U(0, 1)$,
4. Use p^{rand} to summarise evidence against H_0.



Standard (“crisp”) Decision Rule

Define a decision rule as

$\tau = 1 \quad \rightarrow \quad$ reject the null hypothesis,

$\tau = 0 \quad \rightarrow \quad$ don’t reject.

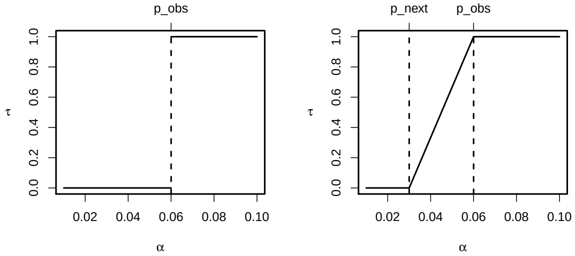
E.g. threshold p-value at the nominal value α:

$$\tau = \begin{cases} 0 & p^{obs} \geq \alpha \\ 1 & p^{obs} < \alpha, \end{cases}$$

Fuzzy Decision Rule

Define a decision rule as

$$\tau = P(p^{rand} < \alpha | X = x)$$
$$= \begin{cases} 0, & \alpha \leq p^{next}; \\ \frac{\alpha - p^{next}}{p^{obs} - p^{next}}, & p^{next} < \alpha \leq p^{obs}; \\ 1, & \alpha > p^{obs}. \end{cases}$$



Fuzzy Bonferroni Procedure

Now multiple testing (testing $i = 1, \cdots, m$ null hypotheses).

Bonferroni uses same threshold (α/m) for all tests:

$$\tau_i = P(p_i^{rand} < \alpha/m | X = x)$$
$$= \begin{cases} 0, & \alpha/m \leq p_i^{next}; \\ \frac{\alpha/m - p_i^{next}}{p_i^{obs} - p_i^{next}}, & p_i^{next} < \alpha/m \leq p_i^{obs}; \\ 1, & \alpha > p_i^{obs}. \end{cases}$$

Fuzzy Bonferroni Procedure

Toy Example $m = 7$ Binomial tests: $X_i \sim Bin(n_i, \pi_0), i = 1, \cdots, 7$.

$H_0 : \pi_0 = 0.5$ $\qquad$ $H_1 : \pi_0 < 0.5$

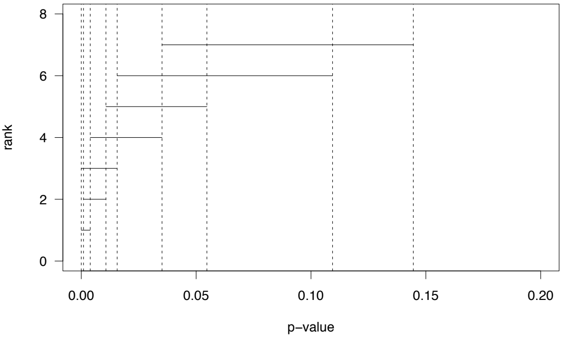
x	n	p^{next}	p^{obs}	$\tau_{crisp}(\alpha = 0.05)$	$\tau_{fuzzy}(\alpha = 0.05)$
0	8	0.000	0.004	1	1
1	10	0.001	0.011	0	0.63
0	6	0.000	0.016	0	0.46
1	8	0.004	0.035	0	0.10
2	10	0.011	0.055	0	0
1	6	0.016	0.109	0	0
2	8	0.035	0.145	0	0

Fuzzy False Discovery Rate (FDR)

Most procedures for controlling error rates (FWER or FDR) depend on ordering the p-values.

Different sets of randomised p-values will have different orderings

$\rightarrow$ need to integrate over the different orders.



Monte Carlo calculation

Generate K realisations of randomised p-values

$$p_{ik}^{rand} \sim Unif(p_i^{next}, p_i^{obs}), \quad i = 1, \cdots, m, \;\; k = 1, \cdots, K$$

For each realisation k, calculate adjusted randomised p-values, e.g. for BH

$$adj(p_{ik}^{rand}) = min_{j: r_{jk}^{rand} \geq r_{ik}^{rand}} (m p_{jk}^{rand} / r_{jk}^{rand})$$

Fuzzy decision rule simple in terms of adjusted p-values

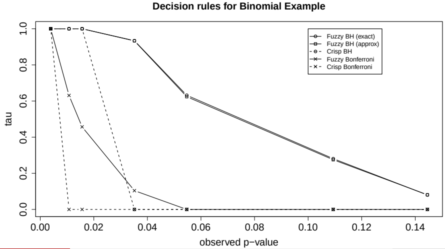
$$\hat{\tau}_i(\alpha) = \frac{1}{K} \sum_k I[adj(p_{ik}^{rand}) \leq \alpha]$$

We have shown this is valid for general error controlling procedure with adjusted p-values monotonic w.r.t. original p-values (Lewin and Kulinskaya 2016/2017, working paper).

Toy example again

x	n	p^{next}	p^{obs}	$\tau^{BH}_{crisp}(\alpha = 0.05)$	$\tau^{BH}_{fuzzy}(\alpha = 0.05)$
0	8	0.000	0.004	1	1
1	10	0.001	0.011	1	1
0	6	0.000	0.016	1	1
1	8	0.004	0.035	0	0.93
2	10	0.011	0.055	0	0.63
1	6	0.016	0.109	0	0.28
2	8	0.035	0.145	0	0.08

Comparison exact and monte carlo for BH



Comparing amino acid sequences

Two groups of patients with different subtypes of HIV. Aim is to compare the sequences of amino acids between the two types of virus.

Amino acids are compared at 118 positions $\rightarrow m = 118$ hypothesis tests (using Fisher exact test).

Data from Gilbert et al. JRSSC, 2005.

Comparing the standard BH procedure with fuzzy BH (**showing top results only**):

	p-values		$\tau(0.05)$	
	previous	observed	crisp	fuzzy
1	0	7×10^{-13}	1	1.00
2	8×10^{-13}	3×10^{-11}	1	1.00
3	3×10^{-10}	9×10^{-9}	1	1.00
4	2×10^{-6}	1×10^{-5}	1	1.00
5	2×10^{-5}	1×10^{-4}	1	1.00
6	0	1×10^{-4}	1	1.00
7	7×10^{-5}	0.001	1	1.00
8	0	0.001	1	1.00
9	0	0.001	1	1.00
10	3×10^{-4}	0.002	1	1.00
11	3×10^{-4}	0.005	1	1.00
12	0	0.006	0	0.98
13	0	0.006	0	0.98
14	0.002	0.007	0	0.90
15	0.005	0.015	0	0.17

Summary

- Fuzzy decision rules can be defined for hypothesis testing with discrete data, using the idea of randomised p-values
- Enables less conservative testing procedures
- We have extended the idea of fuzzy decision rules to the multiple testing situation
- The decision rules can be estimated for a wide class of multiple testing procedures

Acknowledgements

Thanks to my co-author, Elena Kulinskaya.

References:

- Geyer, C. and Meeden, G. (2005). Fuzzy and randomized Confidence intervals and P-values. *Statistical Science* **20**, 358–366.
- Kulinskaya, E. and Lewin, A. (2009). On fuzzy FWER and FDR procedures for discrete distributions. *Biometrika* **96**, 201 – 211.