FWER Controlling Multiple Testing Procedures for Discrete Data

Yalin Zhu, Wenge Guo

Department of Mathematical Sciences, New Jersey Institute of Technology

Abstract

In applications such as flagging adverse events (AEs) in clinical safety studies or identifying differentially expressed genes in microarray experiments, the data of the experiments usually consists of frequency counts. In the analysis of such data, researchers often face multiple hypotheses testing based on discrete test statistics. Incorporating this discrete property of the data we propose several stepwise procedures controlling FWER, which allow to use the CDF of p-values to determine the testing threshold. We show that the proposed procedures strongly control the FWER and are more powerful than the existing ones for discrete data. Through real data analysis and simulation studies, the proposed procedures are shown to outperform the existing procedures in terms of the FWER control and power.

Background

In statistical analysis of biomedical experiments, there are often a number of hypotheses to test. Examples include:

- Evaluating **efficacy** of a new drug for multiple endpoints.
- Detecting **adverse events** across body systems in a clinical trial of drug or medical treatment.
- Identifying **differentially expressed genes** in microarray experiments.

Discrete data sets are often seen in these experiments. Thus, some discrete statistics (for instance, *Fisher's Exact Test*, *Binomial Test*, etc.) are commonly used in practice.

A Motivating Example

Suppose there are two groups of individuals, group 1 is study group with the treatment, group 2 is control group with the placebo. The clinical trial is conducted to monitor the counts of m types of adverse events (AEs) in patients' body systems. Our goal is to detect "reasonable" or "correct" AEs (flagging) from many AEs. Thus, the null and alternative hypotheses of interest are:

$$H_{ij} : p_{1j} = p_{2j} \quad \text{against} \quad H'_{ij} : p_{1j} \neq p_{2j}.$$

Here, p_{ij} is proportion of individuals from i-th group experiencing the j-th AE, where $i = 1, 2$; $j = 1, \ldots, m$.
If $p_{1j} = p_{2j}$ then the j-th AE is not flagging (or is not caused by this treatment). Otherwise this AE is caused by the treatment.

⋆ Such a clinical safety problem can be formulated as a multiple hypotheses testing for discrete data.

Related Concepts

- **Family-wise Error Rate:** $FWER = Pr\{V > 0\}$, the probability of making at least one false rejection.
- **Minimal Power:** $Pr\{S > 0\}$, the probability of rejecting at least one false null.
- **Assumption of p-value for discrete test:** For the discrete test statistic T_i, let P_i denote the corresponding p-value for testing H_i and $\mathbb{P}_i$ denote the set of all attainable p-values for H_i such that $P_i \in \mathbb{P}_i$. Suppose F_i denote the cumulative distribution function (CDF) of P_i under H_i is true, that is, $F_i(u) = Pr\{P_i \leq u | H_i \text{ is true}\}$. For any $u \in \mathbb{P}_i$, $F_i(u) = u$; otherwise, $F_i(u) < u$, where $i = 1, \ldots, m$.

Existing Methods

Traditional stepwise procedures:

- Bonferroni procedure
- Holm procedure (Holm, 1979)
- Hochberg procedure (Hochberg, 1988)

Existing Methods (cont.)

Procedures for discrete data:

- Tarone procedure (Tarone, 1990)
- Tarone-Holm (Hommel and Krummenauer, 1998)
- Roth-Hochberg (Roth, 1999)

Proposed Procedures

Let P_i be the p-value of hypothesis H_i, where $i = 1, \ldots, m$ and $P_{(1)} \leq \ldots \leq P_{(m)}$ be the ordered p-values of the corresponding hypotheses $H_{(1)}, \ldots, H_{(m)}$.

Modified Bonferroni Procedure: Let $t = \max\left\{p \in \bigcup_{i=1}^{m} \mathbb{P}_i : \sum_{i=1}^{m} F_i(p) \leq \alpha\right\}$ and set $t = \frac{\alpha}{m}$ if the maximum does not exist. Then reject H_i if its corresponding p-value $P_i \leq t$.

Modified Holm Procedure: Let $\alpha_i = \max\left\{p \in \bigcup_{j=i}^{m} \mathbb{P}_{(j)} : \sum_{j=i}^{m} F_{(j)}(p) \leq \alpha\right\}$ with $\alpha_0 = 0$. Set $\alpha_i = \max\left\{\alpha_{i-1}, \frac{\alpha}{m-i+1}\right\}$ if the maximum does not exist for some i. Then reject no null hypotheses if $P_{(1)} > \alpha_1$; otherwise, reject $H_{(1)}, \ldots, H_{(r)}$ and retain $H_{(r+1)}, \ldots, H_{(m)}$, where r is the largest index satisfying $P_{(1)} \leq \alpha_1, \ldots, P_{(r)} \leq \alpha_r$.

Modified Hochberg Procedure: Using the same α_i as in the Modified Holm procedure, reject all hypotheses $H_{(1)}, \ldots, H_{(m)}$ if $P_{(m)} \leq \alpha_m$; otherwise, reject $H_{(1)}, \ldots, H_{(r)}$ and retain $H_{(r+1)}, \ldots, H_{(m)}$, where r is the largest index satisfying $P_{(r)} \leq \alpha_r$.

Theoretical Results

1. **Modified Bonferroni Procedure** strongly controls the FWER at level α under arbitrary dependence.
2. **Modified Holm Procedure** strongly controls the FWER at level α under arbitrary dependence.

Properties of Proposed Procedures

- **Modified Bonferroni Procedure** is universally more powerful than Tarone Procedure and Modified Tarone Procedure.
- **Adjusted p-value:**
  1. The adjusted p-value of H_i for **Modified Bonferroni Procedure** is $\tilde{P}_i^{\text{MBonf}} = \min\left\{1, \sum_{j=1}^{m} F_j(P_i)\right\}$, where $i = 1, \ldots, m$.
  2. The adjusted p-value of $H_{(i)}$ for **Modified Holm Procedure** is
  $$\tilde{P}_{(i)}^{\text{MHolm}} = \begin{cases} \min\left\{1, \sum_{j=1}^{m} F_{(j)}(P_{(1)})\right\}, & i = 1 \\ \max\left\{\tilde{P}_{(i-1)}^{\text{MHolm}}, \min\left\{1, \sum_{j=i}^{m} F_{(j)}(P_{(i)})\right\}\right\}. & i = 2, \ldots, m \end{cases}$$
  3. The adjusted p-value of $H_{(i)}$ for **Modified Hochberg Procedure** is
  $$\tilde{P}_{(i)}^{\text{MHoch}} = \begin{cases} F_{(m)}(P_{(m)}), & i = m \\ \min\left\{\tilde{P}_{(i+1)}^{\text{MHoch}}, \sum_{j=i}^{m} F_{(j)}(P_{(i)})\right\}. & i = m-1, \ldots, 1 \end{cases}$$
- All proposed procedures satisfy **α-consistency** and **p-value monotonicity**.

Real Data Analysis

Conduct a safety trial of a candidate vaccine against measles, mumps, rubella and varicella (MMRV). (see Mehrotra and Heyse, 2004)

- Monitor Nine **Tier 2 AE types** in the skin body system.
- $N_1 = 148 \longrightarrow$ MMR $\longrightarrow X_{1i}$
  $N_2 = 132 \longrightarrow$ MMRV $\longrightarrow X_{2i}$
- Two-sided **Fisher's Exact Test**

Real Data Analysis (cont.)

Table 1: A comparison of adjusted p-values for single step procedures

AE_i	X_1	X_2	P_i	$\tilde{P}_i^{\text{Bonf}}$	$\tilde{P}_i^{\text{Tarone}}$	$\tilde{\boldsymbol{P}}_{\boldsymbol{i}}^{\mathbf{MBonf}}$
1	13	3	0.0209	0.1880	0.0836	0.0534
2	8	1	0.0388	0.3490	0.1551	0.1343
3	4	0	0.1248	1.0000	0.8734	0.7134
4	0	2	0.2214	1.0000	1.0000	1.0000
5	6	2	0.2885	1.0000	1.0000	1.0000
6	2	0	0.4998	1.0000	1.0000	1.0000
7	1	2	0.6033	1.0000	1.0000	1.0000
8	4	2	0.6872	1.0000	1.0000	1.0000
9	2	1	1.0000	1.0000	1.0000	1.0000

Table 2: A comparison of adjusted p-values for step-down procedures

$AE_{(i)}$	X_1	X_2	$P_{(i)}$	$\tilde{P}_{(i)}^{\text{Holm}}$	$\tilde{P}_{(i)}^{\text{TH}}$	$\tilde{\boldsymbol{P}}_{\boldsymbol{(i)}}^{\mathbf{MHolm}}$
(1)	13	3	0.0209	0.1880	0.0836	0.0534
(2)	8	1	0.0388	0.3103	0.1163	0.0982
(3)	4	0	0.1248	0.8734	0.6238	0.5050
(4)	0	2	0.2214	1.0000	1.0000	1.0000
(5)	6	2	0.2885	1.0000	1.0000	1.0000
(6)	2	0	0.4998	1.0000	1.0000	1.0000
(7)	1	2	0.6033	1.0000	1.0000	1.0000
(8)	4	2	0.6872	1.0000	1.0000	1.0000
(9)	2	1	1.0000	1.0000	1.0000	1.0000

Table 3: A comparison of adjusted p-values for step-up procedures

$AE_{(i)}$	X_1	X_2	$P_{(i)}$	$\tilde{P}_{(i)}^{\text{Hoch}}$	$\tilde{P}_{(i)}^{\text{Roth}}$	$\tilde{\boldsymbol{P}}_{\boldsymbol{(i)}}^{\mathbf{MHoch}}$
(1)	13	3	0.0209	0.1880	0.0836	0.0534
(2)	8	1	0.0388	0.3103	0.1552	0.0982
(3)	4	0	0.1248	0.8734	0.7246	0.5050
(4)	0	2	0.2214	1.0000	1.0000	1.0000
(5)	6	2	0.2885	1.0000	1.0000	1.0000
(6)	2	0	0.4998	1.0000	1.0000	1.0000
(7)	1	2	0.6033	1.0000	1.0000	1.0000
(8)	4	2	0.6872	1.0000	1.0000	1.0000
(9)	2	1	1.0000	1.0000	1.0000	1.0000

Simulation Studies

Basic settings: number of hypotheses $m = 5$, true null proportion $\pi_0 = \{0.2, 0.4, 0.6, 0.8\}$, iteration $B = 2000$, sample size $N = 25$.

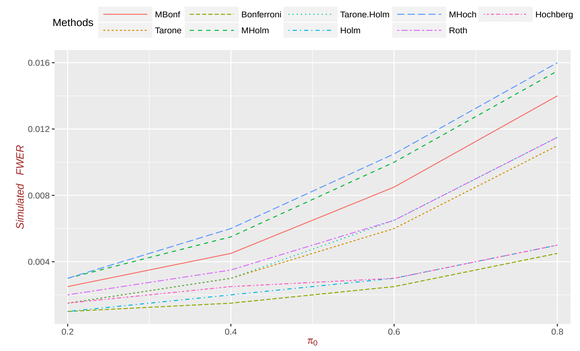

Figure 1: The simulated FWER comparisons using Fisher's Exact Test.

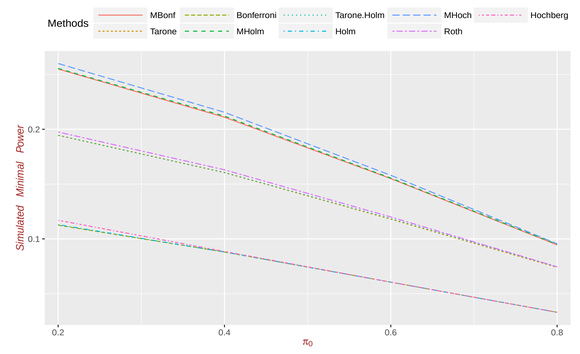

Figure 2: The simulated minimal power comparisons using Fisher's Exact Test.

Conclusions

- We have developed several FWER controlling procedures, which exploit the information of discreteness for test statistics. Some desired properties of proposed procedures are discussed as well.
- Real data analysis and simulation studies illustrate that the proposed procedures can perform better than existing procedures in some cases, especially when the proportion of true null hypotheses (π_0) is large, or when the sample size (N) is small.

Computing Tools

⋆ R package: **MHTdiscrete**
cran.r-project.org/web/packages/MHTdiscrete

⋆ Web Application: **MTPs**
allen.shinyapps.io/MTPs

Email: yalin.zhu@outlook.com **Tel:** +1 (973)-991-9996