Multiple Testing in Group Sequential Clinical Trials

Tian Zhao

Supervisor: Michael Baron

Department of Mathematical Sciences
University of Texas at Dallas

txz102020@utdallas.edu

7/20/2013
1 Introduction
   • Sequential statistics
   • Problem

2 Results
   • Bonferroni approach

3 Improvement
   • Holm-type stepwise methods
     • Truncated Sequential Probability Ratio Test
   • Fallback approach

4 Future plans
Multiple testing in clinical trials

- Multiple endpoints (efficacy and safety)
- Multiple treatment arms or doses of a drug
- Overall, get an answer to several questions
Sequential statistics

- Multiple testing in clinical trials
  - Multiple endpoints (efficacy and safety)
  - Multiple treatment arms or doses of a drug
  - Overall, get an answer to several questions

- Sequential / group sequential clinical trials
  - Interim analysis (Flector trial: stop for futility)
  - Sequential clinical trials always have a restriction on the maximum number of sampled groups (K)
Test multiple hypotheses based on sequentially observed data, achieving a decision on each individual test

- Group sequential sampling
Our problem

Test multiple hypotheses based on sequentially observed data, achieving a decision on each individual test

- Group sequential sampling
- Formulation
Our problem

Test multiple hypotheses based on sequentially observed data, achieving a decision on each individual test

- Group sequential sampling
- Formulation
- Goal
A number of group sequential procedures have been proposed for testing ONE hypothesis.

Pocock (1977) applies repeated significance tests to group sequential trials with equal-size groups and derives a constant critical value on the standardized normal Z scale across all stages that control Type I error. O’Brien and Fleming (1979) propose a sequential procedure that has boundary values decrease over the stages on the standardized normal Z scale.
Formulation

Observe a sequence of iid random vectors \( X_n, n = 1, 2, ... \)
\( X_n = (X_1^{(1)}, ... X_n^{(d)}) \in R^d \) is a set of measurements on unit \( n \).
\( X_i^{(j)} \sim f_j(x|\theta(j)) \)

**Multiple hypotheses**

Test \( d \) hypotheses:

\[
H_0^{(j)} : \theta^{(j)} = \theta_0^{(j)} \text{ vs } H_A^{(j)} : \theta^{(j)} = \theta_1^{(j)}, \text{ for } j = 1, 2, ..., d
\]

Let \( m = \text{constant} \) be the group size and \( T \) be a stopping time

Tests are based on a group sequential experiment. Data are collected in groups of \( m \) patients.
Goal

- Control
  - **Familywise Type I error rate** = FWER-I = \( P( \text{At least one Type I error}) \leq \alpha \)
  - **Familywise Type II error rate** = FWER-II = \( P( \text{At least one Type II error}) \leq \beta \)

- \( P(T \leq K) = 1 \), \( T \) is stopping time, \( K = \) given integer = max allowed number of groups

- Do it at **low expected cost**
Bonferroni approach:

Bonferroni approach: control of both error rates.
Choose

\[ \alpha_j \in (0, 1), \quad \sum \alpha_j = \alpha; \quad \beta_j \in (0, 1), \quad \sum \beta_j = \beta \]

Test \( H_0^{(j)} \) at level \( \alpha_j \) and power \( 1 - \beta_j \) by any known sequential procedure

\[
P \left( \bigcup_{j=1}^{d} \text{Type I error on } H_0^{(j)} \right) \leq \sum P \left( \text{Type I error on } H_0^{(j)} \right) = \sum \alpha_j = \alpha
\]

\[
P \left( \bigcup_{j=1}^{d} \text{Type II error on } H_A^{(j)} \right) \leq \sum P \left( \text{Type II error on } H_A^{(j)} \right) = \sum \beta_j = \beta
\]
Improvement of Bonferroni Methods

Bonferroni inequality is very crude for moderate to large $d$, extend to Holm’s stepwise approach.

- Order the log-likelihood ratio (LLR) statistics in decreasing (step-down)/increasing (step-up)
- Choose stopping boundaries for the ordered LLR
- Define the stopping rule and decision rules
- This requires analysis of the group sequential test
- Boundaries of Pocock, O’Brien-Fleming, Wang-Tsiatis tests are computed numerically, not tractable to analyze
- Derive equations for Whitehead triangular test and Truncated SPRT
Group sequential tests

- **Pocock test**
- **Whitehead test**
- **O’Brien-Fleming test**
- **Truncated SPRT**
Derive Truncated SPRT

**SPRT**

- Choose stopping boundaries $a, b$ to control $\alpha$ and $\beta$
- where,
  
  \[
  a = \frac{1 - \beta}{\alpha}
  \]
  
  \[
  b = \frac{\beta}{1 - \alpha}
  \]
- SPRT control both $\alpha, \beta$ among all sequential tests that have
  
  \[
P(\text{Type I error}) \leq \alpha
  \]
  
  \[
P(\text{Type II error}) \leq \beta
  \]
Wald’s SPRT

\[ \Lambda_k \]

\( R e j e c t \ H_0 \)

\( A c c e p t \ H_0 \)

Tian Zhao  Supervisor: Michael Baron
Multiple Testing in Group Sequential Clinical Trials
Truncated SPRT

Stopping time and Decision rule

Following SPRT approach for single hypothesis testing, truncated SPRT is based on log-likelihood ratios

\[ \Lambda_n = \ln \frac{f(x|\theta_1)}{f(x|\theta_0)} \]

Stopping time

\[ T = \min(K, \min \{ n : \Lambda_n \notin (b, a) \}) \leq K \]

Decision rule,

\[
\delta = \begin{cases} 
\text{reject } H_0 & \text{if } \Lambda_T \geq a \text{ or } \Lambda_T \geq c \cap T = K \\
\text{accept } H_0 & \text{if } \Lambda_T \leq b \text{ or } \Lambda_T < c \cap T = K 
\end{cases}
\]
Truncated SPRT (cont’d)

\[ \Lambda_k \]

\[ \text{Reject } H_0 \]

\[ a \]

\[ b \]

\[ c \]

\[ K \]

\[ \text{Accept } H_0 \]

\[ k \]
Plan

- Evaluate performance of TSPRT, probabilities of Type I and Type II errors.
- Find equations of the stopping boundaries and the required group size to control $P\{\text{Type I error}\}$ and $P\{\text{Type II error}\}$.
- Then develop Holm-type stepwise sequential procedures to test $d$ hypotheses controlling familywise error rates I and II.
Evaluate probabilities of Type I and Type II errors by comparison between TSPRT and SPPRT.

Both tests reject $H_0$
Truncated SPRT. Performance evaluation.

\[ \Lambda_k \]

- Reject \( H_0 \)
- Accept \( H_0 \)

TSPRT rejects \( H_0 \) but SPRT accepts it

\[ \begin{align*}
\Lambda_k & \quad R e j e c t \quad H_0 \\
0 & \quad A c c e p t \quad H_0 \\
\end{align*} \]
Truncated SPRT. Performance evaluation.

Both tests accept $H_0$
Truncated SPRT. Performance evaluation.

\[ \Lambda_k \]

Reject \( H_0 \)

Accept \( H_0 \)

TSPRT accepts \( H_0 \) but SPRT rejects it

\( A c c e p t H_0 \)

\( R e j e c t H_0 \)

\( a \)

\( b \)

\( c \)

\( k \)

\( K \)
Choose $c$ to attain $P(\text{Type I error}|\text{TSPRT}) \leq \alpha$.

Let $\Lambda_K = \sum_{j=1}^{Km} \ln \frac{f(x_j|\theta_1)}{f(x_j|\theta_0)}$. For a given $\alpha_1$ = level for SPRT,

\[
P(\text{Type I error}|\text{TSPRT}) = P_{H_0}(\text{Type I error}|\text{SPRT}) + P_{H_0}(\{\Lambda_1, \ldots \Lambda_{K-1} \in (b, a)\} \cap \{\Lambda_K \in [c, a]\} \cap \{\Lambda_n \leq b \text{ before } \Lambda_n \geq a\}) - P_{H_0}(\{\Lambda_1, \ldots \Lambda_{K-1} \in (b, a)\} \cap \{\Lambda_K \in (b, c)\} \cap \{\Lambda_n \geq a \text{ before } \Lambda_n \leq b\}) \leq \alpha_1 + P(A) - P(B)
\]

We need to make $P(A) \leq \alpha - \alpha_1$
Type I error

\[
P(A) \leq P_{H_0}(\Lambda_K \in [c, a))P_{H_0}(\Lambda_n \leq b \text{ before } \Lambda_n \geq a | \Lambda_K \geq c) \\
\leq P_{H_0}(\Lambda_K \geq c)P_{H_0}(\Lambda_n \leq b \text{ before } \Lambda_n \geq a | \Lambda_K \geq c) \\
\leq P_{H_0}(\Lambda_K \geq c)P_{H_0}(\Lambda_n \leq b \text{ before } \Lambda_n \geq a | \Lambda_K = c)
\]
Truncated SPRT (cont’d)

Derivations

Assumption: (General) Exponential family

Test

\[ H_0 : \theta = \theta_0 \text{ vs } H_A : \theta = \theta_1. \]

Exponential family is written in the canonical form with

\[ \eta = \text{canonical parameter} \]

as follows,

\[ f(x|\eta) = f(x|0)e^{\eta x - \psi(\eta)} \]

where, \( \psi(0) = 0 \) and \( \mu(\eta) = EX = \psi'(\eta) \), \( \text{Var}X = \psi''(\eta) \)
Truncated SPRT (cont’d)

Chernoff Inequality

In general case, we have

\[ P\left(\sum_{i=1}^{n} Z_j \geq ny\right) \leq e^{-n\sup_{t \geq 0}\{ty - \psi_z(t)\}} \]

here,

\[ Z = \ln \frac{f(x|\theta_1)}{f(x|\theta_0)}, \quad c = ny \]
Normal as example

Observe $X_1, X_2, ..., X_n \sim N(\theta, \sigma^2)$, we can write it as

$$f(x|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} e^{\frac{\theta x}{\sigma^2} - \frac{\theta^2}{2\sigma^2}}$$

here,

$$\eta = \frac{\theta}{\sigma^2}, \quad f(x|0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, \quad \psi_x(\eta) = \frac{\sigma^2\eta^2}{2}$$
Normal as example

By Chernoff Inequality,

\[
\sup_{t \geq 0} \{ ty - \psi_z(t) \} \text{ attained at } \hat{t} = [\psi'_z]^{-1}(y)
\]

Therefore,

\[
P(\sum_{i=1}^{n} Z_j \geq ny) \leq e^{-n\left\{ \frac{\sigma^2 y^2}{2(\theta_1 - \theta_0)^2} + \frac{1}{2} y + \frac{(\theta_1 - \theta_0)^2}{8\sigma^2} \right\}}
\]
Type I error

\[
P(A) \leq P_{H_0}(\Lambda_K \in [c, a])P_{H_0}(\Lambda_n \leq b \text{ before } \Lambda_n \geq a|\Lambda_K \geq c)
\]
\[
\leq P_{H_0}(\Lambda_K \geq c)P_{H_0}(\Lambda_n \leq b \text{ before } \Lambda_n \geq a|\Lambda_K \geq c)
\]
\[
\leq P_{H_0}(\Lambda_K \geq c)P_{H_0}(\Lambda_n \leq b \text{ before } \Lambda_n \geq a|\Lambda_K = c)
\]

Hence, we have

\[
P(A) \leq \frac{1 - \alpha_1 e^c}{1 - \alpha_1 \beta} e^{-n\left\{\frac{\sigma^2 c^2}{2n^2(\theta_1 - \theta_0)^2} + \frac{c}{2n} + \frac{(\theta_1 - \theta_0)^2}{8\sigma^2}\right\}} \leq \alpha - \alpha_1
\]
Type II error

For $\alpha_1 = \text{level for SPRT} \in (0, \alpha)$, $\beta_1 = \text{power for SPRT} \in (0, \beta)$

Let $a = -\ln \alpha_1$, and $b = -\ln \beta_1$

- Choose $c_1$ to control Type I error:

$$\frac{1 - \alpha_1 e^{c_1}}{1 - \alpha_1 \beta_1} e^{-n\left\{\frac{\sigma^2 c_1^2}{2n^2(\theta_1 - \theta_0)^2} + \frac{c_1}{2n} + \frac{(\theta_1 - \theta_0)^2}{8\sigma^2}\right\}} \leq \alpha - \alpha_1$$

- Choose $c_2$ to control Type II error:

$$\frac{1 - \beta_1 e^{c_2}}{1 - \alpha_1 \beta_1} e^{-n\left\{\frac{\sigma^2 c_2^2}{2n^2(\theta_1 - \theta_0)^2} - \frac{c_2}{2n} + \frac{(\theta_1 - \theta_0)^2}{8\sigma^2}\right\}} \leq \beta - \beta_1$$
Type II error

Therefore, if \( \frac{\theta_1 - \theta_0}{\sigma} > 0 \), we have,

\[
c_1 \geq \left( \sqrt{-\frac{2}{n} \ln (\alpha - \alpha_1)(1 - \alpha_1 \beta_1) - \left(\frac{\theta_1 - \theta_0}{2\sigma}\right)} \right) \frac{n(\theta_1 - \theta_0)}{\sigma}
\]

\[
c_2 \leq \left( -\sqrt{-\frac{2}{n} \ln (\beta - \beta_1)(1 - \alpha_1 \beta_1) + \left(\frac{\theta_1 - \theta_0}{2\sigma}\right)} \right) \frac{n(\theta_1 - \theta_0)}{\sigma}
\]

Minimize \( n \) to control both Type I error and Type II error

\[
n^* = \frac{2\sigma^2 \left( \sqrt{-\ln (\alpha - \alpha_1)(1 - \alpha_1 \beta_1)} + \sqrt{-\ln (\beta - \beta_1)(1 - \alpha_1 \beta_1)} \right)^2}{(\theta_1 - \theta_0)^2}
\]

Tian Zhao  Supervisor: Michael Baron  Multiple Testing in Group Sequential Clinical Trials
Stepwise method for Truncated SPRT

- Test d hypotheses:
  \[ H_0^{(j)} : \theta^{(j)} = \theta_0^{(j)} \text{ vs } H_A^{(j)} : \theta^{(j)} = \theta_1^{(j)} \quad j = 1, 2, \ldots, d \]

- Test statistics \( \Lambda_k^{(j)} = \sum \ln \frac{f_j(X_i^{(j)}|\theta_1^{(j)})}{f_j(X_i^{(j)}|\theta_0^{(j)})} \)

- Order test statistics \( \Lambda_k^{[1]} \geq \ldots \geq \Lambda_k^{[d]} \), let \( H_0^{(j)} \) be the corresponding tested null hypotheses arranged in the same order.

- \( a_j = -\ln \frac{\alpha - \alpha_1}{d-j+1}, \quad b_j = \ln \frac{\beta - \beta_1}{j} \)
Stepwise method for Truncated SPRT (cont’d)

Decision rule (reject one by one, accept all)

\[
\text{If } \Lambda_k^{[1]} \leq b_1 \Rightarrow \text{stop, accept } H_0^{[1]}, ..., H_0^{[d]} \\
\text{If } \Lambda_k^{[1]} \geq a_1 \Rightarrow \text{reject } H_0^{[1]}, \text{go to } H_0^{[2]} \\
\text{If } \Lambda_k^{[1]} \in (b_1, a_1), \Rightarrow \text{continue sampling}
\]

......
Stepwise method for Truncated SPRT (cont'd)

At $k = K$,

- If $\Lambda_{[j]}^K \notin (b_j, a_j)$, $\Rightarrow$ stop
- If $\Lambda_{[j]}^K \in (b_j, a_j)$,
  - $\Lambda_{[j]}^K > c_j$ $\Rightarrow$ stop, reject
  - $\Lambda_{[j]}^K < c_j$ $\Rightarrow$ stop, accept

This stepwise method can control FWER-I and FWER-II
Pocock Test

Pocock derives the constant boundary \([-C_p, C_p]\) on the standardized Z scale, if \(Z \notin [-C_p, C_p]\), stop and reject \(H_0\), otherwise, keep sampling, after K groups, if \(Z \notin [-C_p, C_p]\), stop and reject \(H_0\), otherwise, stop and accept \(H_0\).

Fallback method for Pocock: recycle \(\beta_j\) on all \(H_0^{(j)}\) that were rejected, then at \(k = K\), test remaining \(H_0^{(j)}\) at higher \(\beta_j\)

\[
\beta_j(K) = \frac{\beta_j(1)}{1 - \sum_{i: H_0^{(i)} \text{ was rejected}} \beta_i}
\]

Testing with higher \(\beta\) means that we can accept more often, expand the acceptance region. \(P(\text{reject} | H_0) \text{ reduces}\), we can shrink boundary \([-C_p, C_p]\) to meet the \(\alpha\) level, and then reduce \(m\).
Use stepwise approach to multiple hypotheses starting from various standard group sequential methods

Thank you!
Future plans

- Use stepwise approach to multiple hypotheses starting from various standard group sequential methods
- Evaluate performance of fallback procedures, compare fallback and stepwise testing approaches

Thank you!
Future plans

- Use stepwise approach to multiple hypotheses starting from various standard group sequential methods
- Evaluate performance of fallback procedures, compare fallback and stepwise testing approaches
- Composite hypotheses and nuisance parameters. Sequential t-test.

Thank you!
Future plans

- Use stepwise approach to multiple hypotheses starting from various standard group sequential methods
- Evaluate performance of fallback procedures, compare fallback and stepwise testing approaches
- Composite hypotheses and nuisance parameters. Sequential $t$-test.
- Examples of clinical data

Thank you!