

A COMPARISON OF BOOTSTRAP BASED MULTIPLE TESTING PROCEDURES

Wiebke Werft and Axel Benner

German Cancer Research Center, Heidelberg, Germany

ISCB Prague 2009

24. August 2009

Translational Clinical Trials

- Clinical trials with microarray data to identify genes which may influence clinical outcome, e.g. response to treatment.
- Each gene is analysed as covariate in a univariable regression model and its effect on the outcome variable is tested.
- Especially logistic regression models for binary clinical outcome variables.
- To account for multiple testing the resulting p -values have to be adjusted adequately.
- Incorporate complex and unknown dependence structures among genes.

Translational Clinical Trials

- Clinical trials with microarray data to identify genes which may influence clinical outcome, e.g. response to treatment.
- Each gene is analysed as covariate in a univariable regression model and its effect on the outcome variable is tested.
- Especially logistic regression models for binary clinical outcome variables.
- To account for multiple testing the resulting p -values have to be adjusted adequately.
- Incorporate complex and unknown dependence structures among genes.

Translational Clinical Trials

- Clinical trials with microarray data to identify genes which may influence clinical outcome, e.g. response to treatment.
- Each gene is analysed as covariate in a univariable regression model and its effect on the outcome variable is tested.
- Especially logistic regression models for binary clinical outcome variables.
- To account for multiple testing the resulting p -values have to be adjusted adequately.
- Incorporate complex and unknown dependence structures among genes.

Translational Clinical Trials

- Clinical trials with microarray data to identify genes which may influence clinical outcome, e.g. response to treatment.
- Each gene is analysed as covariate in a univariable regression model and its effect on the outcome variable is tested.
- Especially logistic regression models for binary clinical outcome variables.
- To account for multiple testing the resulting p -values have to be adjusted adequately.
- Incorporate complex and unknown dependence structures among genes.

Translational Clinical Trials

- Clinical trials with microarray data to identify genes which may influence clinical outcome, e.g. response to treatment.
- Each gene is analysed as covariate in a univariable regression model and its effect on the outcome variable is tested.
- Especially logistic regression models for binary clinical outcome variables.
- To account for multiple testing the resulting p -values have to be adjusted adequately.
- Incorporate complex and unknown dependence structures among genes.

Multiple Hypotheses Testing - Notations

- testing M null hypotheses simultaneously
- precisely: testing slope coefficients of M generalized linear models $E(Y) = g^{-1}(\beta_{0,m} + \beta_{1,m}X_m)$ for $m = 1, \dots, M$, g a link function, e.g. logit for a logistic model

$$H_{0,m} : \beta_{1,m} = 0 \quad \text{and} \quad H_{A,m} : \beta_{1,m} \neq 0 \quad m = 1, \dots, M$$

- $M \times n$ matrix of realizations; m th row with entries x_{m1}, \dots, x_{mn} corresponds to covariate X_m
- p -values $p_m = P(T_{n,m} \geq T_{n,m}(x_{m1}, \dots, x_{mn}))$ derived from a test statistics $T_{n,m}$, e.g. Likelihood Ratio test statistics
- Q_{n0} joint null distribution of the M vector of test statistics, $Q_{n,m}$ marginal distribution of $T_{n,m}$

Multiple Hypotheses Testing - Notations

- testing M null hypotheses simultaneously
- precisely: testing slope coefficients of M generalized linear models $E(Y) = g^{-1}(\beta_{0,m} + \beta_{1,m}X_m)$ for $m = 1, \dots, M$, g a link function, e.g. logit for a logistic model

$$H_{0,m} : \beta_{1,m} = 0 \quad \text{and} \quad H_{A,m} : \beta_{1,m} \neq 0 \quad m = 1, \dots, M$$

- $M \times n$ matrix of realizations; m th row with entries x_{m1}, \dots, x_{mn} corresponds to covariate X_m
- p -values $p_m = P(T_{n,m} \geq T_{n,m}(x_{m1}, \dots, x_{mn}))$ derived from a test statistics $T_{n,m}$, e.g. Likelihood Ratio test statistics
- Q_{n0} joint null distribution of the M vector of test statistics, $Q_{n,m}$ marginal distribution of $T_{n,m}$

Multiple Hypotheses Testing - Notations

- testing M null hypotheses simultaneously
- precisely: testing slope coefficients of M generalized linear models $E(Y) = g^{-1}(\beta_{0,m} + \beta_{1,m}X_m)$ for $m = 1, \dots, M$, g a link function, e.g. logit for a logistic model

$$H_{0,m} : \beta_{1,m} = 0 \quad \text{and} \quad H_{A,m} : \beta_{1,m} \neq 0 \quad m = 1, \dots, M$$

- $M \times n$ matrix of realizations; m th row with entries x_{m1}, \dots, x_{mn} corresponds to covariate X_m
- p -values $p_m = P(T_{n,m} \geq T_{n,m}(x_{m1}, \dots, x_{mn}))$ derived from a test statistics $T_{n,m}$, e.g. Likelihood Ratio test statistics
- Q_{n0} joint null distribution of the M vector of test statistics, $Q_{n,m}$ marginal distribution of $T_{n,m}$

Multiple Hypotheses Testing - Notations

- testing M null hypotheses simultaneously
- precisely: testing slope coefficients of M generalized linear models $E(Y) = g^{-1}(\beta_{0,m} + \beta_{1,m}X_m)$ for $m = 1, \dots, M$, g a link function, e.g. logit for a logistic model

$$H_{0,m} : \beta_{1,m} = 0 \quad \text{and} \quad H_{A,m} : \beta_{1,m} \neq 0 \quad m = 1, \dots, M$$

- $M \times n$ matrix of realizations; m th row with entries x_{m1}, \dots, x_{mn} corresponds to covariate X_m
- p -values $p_m = P(T_{n,m} \geq T_{n,m}(x_{m1}, \dots, x_{mn}))$ derived from a test statistics $T_{n,m}$, e.g. Likelihood Ratio test statistics
- Q_{n0} joint null distribution of the M vector of test statistics, $Q_{n,m}$ marginal distribution of $T_{n,m}$

Multiple Hypotheses Testing - Notations

- testing M null hypotheses simultaneously
- precisely: testing slope coefficients of M generalized linear models $E(Y) = g^{-1}(\beta_{0,m} + \beta_{1,m}X_m)$ for $m = 1, \dots, M$, g a link function, e.g. logit for a logistic model

$$H_{0,m} : \beta_{1,m} = 0 \quad \text{and} \quad H_{A,m} : \beta_{1,m} \neq 0 \quad m = 1, \dots, M$$

- $M \times n$ matrix of realizations; m th row with entries x_{m1}, \dots, x_{mn} corresponds to covariate X_m
- p -values $p_m = P(T_{n,m} \geq T_{n,m}(x_{m1}, \dots, x_{mn}))$ derived from a test statistics $T_{n,m}$, e.g. Likelihood Ratio test statistics
- Q_{n0} joint null distribution of the M vector of test statistics, $Q_{n,m}$ marginal distribution of $T_{n,m}$

Type I Error Rates

Number of H_0	not rejected	rejected	
true	U	V	h_0
false	W	S	h_1
	$M - R$	R	M

False discovery rate:

$$FDR = E \left[\frac{V}{R} \mid V > 0 \right] P(V > 0)$$

- Benjamini-Hochberg (1995) for independent data
- Benjamini-Yekutieli (2001) for any dependency, but too conservative
- Westfall & Young (1993) resampling-based multiple testing, but rely on subset pivotality condition
- Resampling methods by Dudoit et al. (2004) do not rely on subset pivotality condition (R package `multtest`)

Type I Error Rates

Number of H_0	not rejected	rejected	
true	U	V	h_0
false	W	S	h_1
	$M - R$	R	M

False discovery rate:

$$FDR = E \left[\frac{V}{R} \mid V > 0 \right] P(V > 0)$$

- Benjamini-Hochberg (1995) for independent data
- Benjamini-Yekutieli (2001) for any dependency, but too conservative
- Westfall & Young (1993) resampling-based multiple testing, but rely on subset pivotality condition
- Resampling methods by Dudoit et al. (2004) do not rely on subset pivotality condition (R package `multtest`)

Type I Error Rates

Number of H_0	not rejected	rejected	
true	U	V	h_0
false	W	S	h_1
	$M - R$	R	M

False discovery rate:

$$FDR = E \left[\frac{V}{R} \mid V > 0 \right] P(V > 0)$$

- Benjamini-Hochberg (1995) for independent data
- Benjamini-Yekutieli (2001) for any dependency, but too conservative
- Westfall & Young (1993) resampling-based multiple testing, but rely on subset pivotality condition
- Resampling methods by Dudoit et al. (2004) do not rely on subset pivotality condition (R package `multtest`)

Type I Error Rates

Number of H_0	not rejected	rejected	
true	U	V	h_0
false	W	S	h_1
	$M - R$	R	M

False discovery rate:

$$FDR = E \left[\frac{V}{R} \mid V > 0 \right] P(V > 0)$$

- Benjamini-Hochberg (1995) for independent data
- Benjamini-Yekutieli (2001) for any dependency, but too conservative
- Westfall & Young (1993) resampling-based multiple testing, but rely on subset pivotality condition
- Resampling methods by Dudoit et al. (2004) do not rely on subset pivotality condition (R package `multtest`)

Type I Error Rates

Number of H_0	not rejected	rejected	
true	U	V	h_0
false	W	S	h_1
	$M - R$	R	M

False discovery rate:

$$FDR = E \left[\frac{V}{R} \mid V > 0 \right] P(V > 0)$$

- Benjamini-Hochberg (1995) for independent data
- Benjamini-Yekutieli (2001) for any dependency, but too conservative
- Westfall & Young (1993) resampling-based multiple testing, but rely on subset pivotality condition
- Resampling methods by Dudoit et al. (2004) do not rely on subset pivotality condition (R package `multtest`)

Roadmap of Resampling Method (Dudoit et al.)

Define $H_{0,m}$ und $H_{A,m}$ and specify $T_{n,m}$

Estimate test statistics joint null distribution Q_{n0} ←

Select Type I error rate, e.g. FDR, FWER

Apply MTP:

FWER:

- single-step maxT
- single-step minP
- step-down maxT
- step-down minP

FDR:

- augmentation of FWER-controlling procedures ←
- Empirical Bayes

Summarize results via adjusted p values

Roadmap of Resampling Method (Dudoit et al.)

Define $H_{0,m}$ und $H_{A,m}$ and specify $T_{n,m}$

Estimate test statistics joint null distribution Q_{n0} ←

Select Type I error rate, e.g. FDR, FWER

Apply MTP:

FWER:

- single-step maxT
- single-step minP
- step-down maxT
- step-down minP

FDR:

- augmentation of FWER-controlling procedures ←
- Empirical Bayes

Summarize results via adjusted p values

Roadmap of Resampling Method (Dudoit et al.)

Define $H_{0,m}$ und $H_{A,m}$ and specify $T_{n,m}$

Estimate test statistics joint null distribution Q_{n0} ←

Select Type I error rate, e.g. FDR, FWER

Apply MTP:

FWER:

- single-step maxT
- single-step minP
- step-down maxT
- step-down minP

FDR:

- augmentation of FWER-controlling procedures ←
- Empirical Bayes

Summarize results via adjusted p values

Roadmap of Resampling Method (Dudoit et al.)

Define $H_{0,m}$ und $H_{A,m}$ and specify $T_{n,m}$

Estimate test statistics joint null distribution Q_{n0} ←

Select Type I error rate, e.g. FDR, FWER

Apply MTP:

FWER:

- single-step maxT
- single-step minP
- step-down maxT
- step-down minP

FDR:

- augmentation of FWER-controlling procedures ←
- Empirical Bayes

Summarize results via adjusted p values

Roadmap of Resampling Method (Dudoit et al.)

Define $H_{0,m}$ und $H_{A,m}$ and specify $T_{n,m}$

Estimate test statistics joint null distribution Q_{n0} ←

Select Type I error rate, e.g. FDR, FWER

Apply MTP:

FWER:

- single-step maxT
- single-step minP
- step-down maxT
- step-down minP

FDR:

- augmentation of FWER-controlling procedures ←
- Empirical Bayes

Summarize results via adjusted p values

Roadmap of Resampling Method (Dudoit et al.)

Define $H_{0,m}$ und $H_{A,m}$ and specify $T_{n,m}$

Estimate test statistics joint null distribution Q_{n0} ←

Select Type I error rate, e.g. FDR, FWER

Apply MTP:

FWER:

- single-step maxT
- single-step minP
- step-down maxT
- step-down minP

FDR:

- augmentation of FWER-controlling procedures ←
- Empirical Bayes

Summarize results via adjusted p values

Roadmap of Resampling Method (Dudoit et al.)

Define $H_{0,m}$ und $H_{A,m}$ and specify $T_{n,m}$

Estimate test statistics joint null distribution Q_{n0} ←

Select Type I error rate, e.g. FDR, FWER

Apply MTP:

FWER:

- single-step maxT
- single-step minP
- step-down maxT
- step-down minP

FDR:

- augmentation of FWER-controlling procedures ←
- Empirical Bayes

Summarize results via adjusted p values

Roadmap of Resampling Method (Dudoit et al.)

Define $H_{0,m}$ und $H_{A,m}$ and specify $T_{n,m}$

Estimate test statistics joint null distribution Q_{n0} ←

Select Type I error rate, e.g. FDR, FWER

Apply MTP:

FWER:

- single-step maxT
- single-step minP
- step-down maxT
- step-down minP

FDR:

- augmentation of FWER-controlling procedures ←
- Empirical Bayes

Summarize results via adjusted p values

Test Statistics Null Distribution

Shift and Scale Transformed Test Statistics (SST):

$$Z_{n,m} = \sqrt{\min\left(1, \frac{\tau_0}{\text{Var}[T_{n,m}]}\right)} (T_{n,m} - E[T_{n,m}]) + \lambda_0$$

λ_0 and τ_0 user-supplied upper bounds for mean and variance of test statistics under $H_{0,m}$

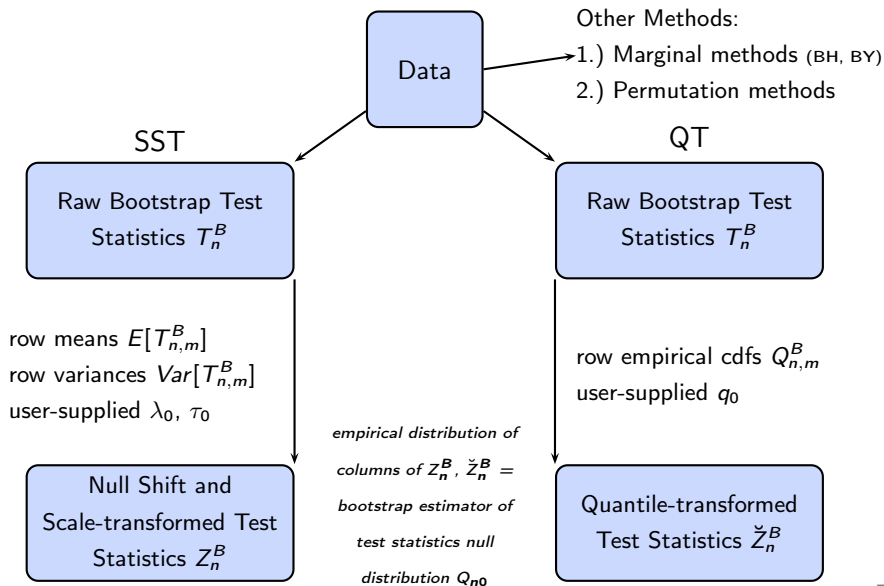
Quantile Transformed Test Statistics (QT):

$$\check{Z}_{n,m} = q_0^{-1} Q_{n,m}(T_{n,m})$$

q_0 user-supplied marginal test statistics null distributions.

Main feature: joint null distributions based on SST and QT test statistics provide asymptotic control of the Type I error rate with arbitrary dependence structures, data distributions and test statistics

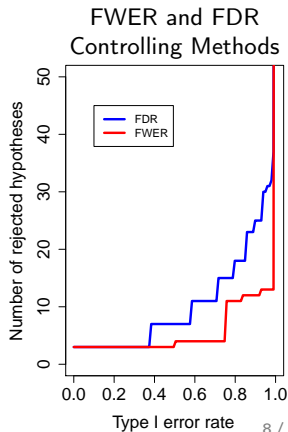
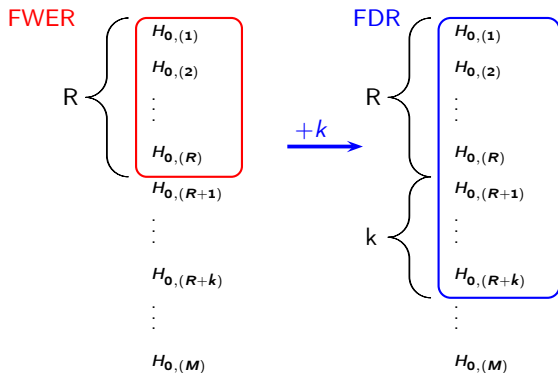
Bootstrap estimation of SST or QT null distribution



Augmentation Procedures

Basic concept: add suitably chosen null hypotheses to the set of hypotheses already rejected by an initial FWER-controlling MTP

R package `multtest`: `fwcr2fdr`



Simulation Study I - Resampling Methods

Hypotheses and Test Statistics:

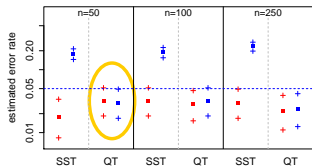
- linear Model: $Y = \beta_{0,m} + \beta_{1,m}X_m + \epsilon$, with $\epsilon \sim N(0, 1)$, $m = 1, \dots, M$
- Hypotheses $H_{0,m} : \beta_{1,m} = 0$ vs. $H_{a,m} : \beta_{1,m} \neq 0$, $m = 1, \dots, M$
- Likelihood Ratio (LR) and Wald (W) Test Statistics

Model Parameters:

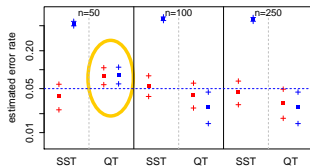
- $M = 50, 500, 1000$
- $n = 50, 100, 250$
- proportion of true null hypotheses: $h_0/M = 1, 0.95$
- *SST* or *QT*
 - parameters for *SST*:
for Wald $\lambda_0 = 0$, $\tau_0 = 1$ and for LR $\lambda_0 = 1$, $\tau_0 = 2$
 - marginal distribution for *QT*:
for Wald $q_0 \sim N(0, 1)$ and for LR $q_0 \sim \chi_1^2$

Results Simulation Study I

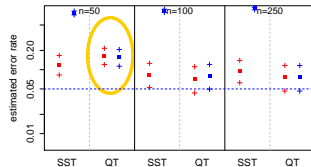
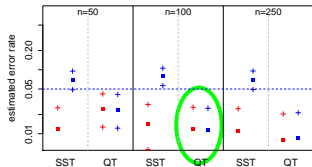
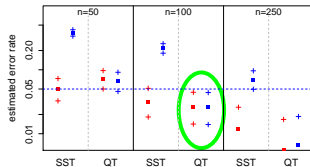
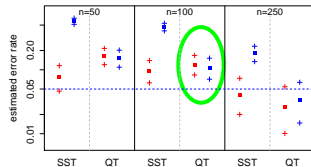
M=50

 $h_0/M = 1$ 

M=500

 $h_0/M = 1$ 

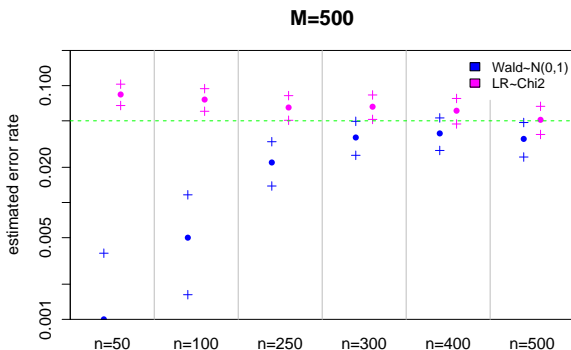
M=1000

 $h_0/M = 1$  $h_0/M = 0.95$  $h_0/M = 0.95$  $h_0/M = 0.95$ 

- Wald Test Statistics
- Likelihood Ratio Test Statistics

Simulation Study II - Marginal Methods

- Logistic Model: $\text{logit}(Y = 1|X_m) = \beta_{0,m} + \beta_{1,m}X_m$, $m = 1, \dots, 500$
- $H_{0,m} : \beta_{1,m} = 0$ vs. $H_{A,m} : \beta_{1,m} \neq 0$, $m = 1, \dots, 500$
- test statistics: Wald, LR
- marginal method: Benjamini-Hochberg



Summary

Findings:

- quantile transformation works better for Likelihood Ratio test statistics
- resampling-methods do not control FDR when based on asymptotic distributions for small sample sizes and large numbers of tests
- marginal methods are unreliable for small sample sizes when based on asymptotic p-values

Recommendations:

- rather use QT instead of SST (`nulldist=boot.qt`)
- within marginal methods use permutation tests to obtain raw p-values when dealing with small sample sizes

Summary

Findings:

- quantile transformation works better for Likelihood Ratio test statistics
- resampling-methods do not control FDR when based on asymptotic distributions for small sample sizes and large numbers of tests
- marginal methods are unreliable for small sample sizes when based on asymptotic p-values

Recommendations:

- rather use QT instead of SST (`nulldist=boot.qt`)
- within marginal methods use permutation tests to obtain raw p-values when dealing with small sample sizes