Multiple hypothesis testing - recent developments and future challenges

Ingelin Steinsland

`ingelins@math.ntnu.no`

Norwegian University of Science and Technology

Outline

- Single hypothesis testing
- Multiple hypothesis testing
    - Quantities and issues
    - False discovery rates
- Future challenges
    - Within false discovery rates.
    - Multiple hypothesis tests, the right tool?

Single hypothesis testing, example

Typical question: Does treatment A give the wished effect?

Hypothesis:

$H = 0$: Non or negative effect.

$H = 1$: Positive effect

Single hypothesis testing, example

Typical question: Does treatment A give the wished effect?

Hypothesis:

$H = 0$: Non or negative effect.

$H = 1$: Positive effect

- Collect data.
- IF the collected data is very unlikely given $H = 0$;
  - $H = 0$ rejected and $H = 1$ accepted.
  - Treatment A has positive effect.
- ELSE
  - $H = 0$ accepted.
  - Treatment A does not have significant positive effect.

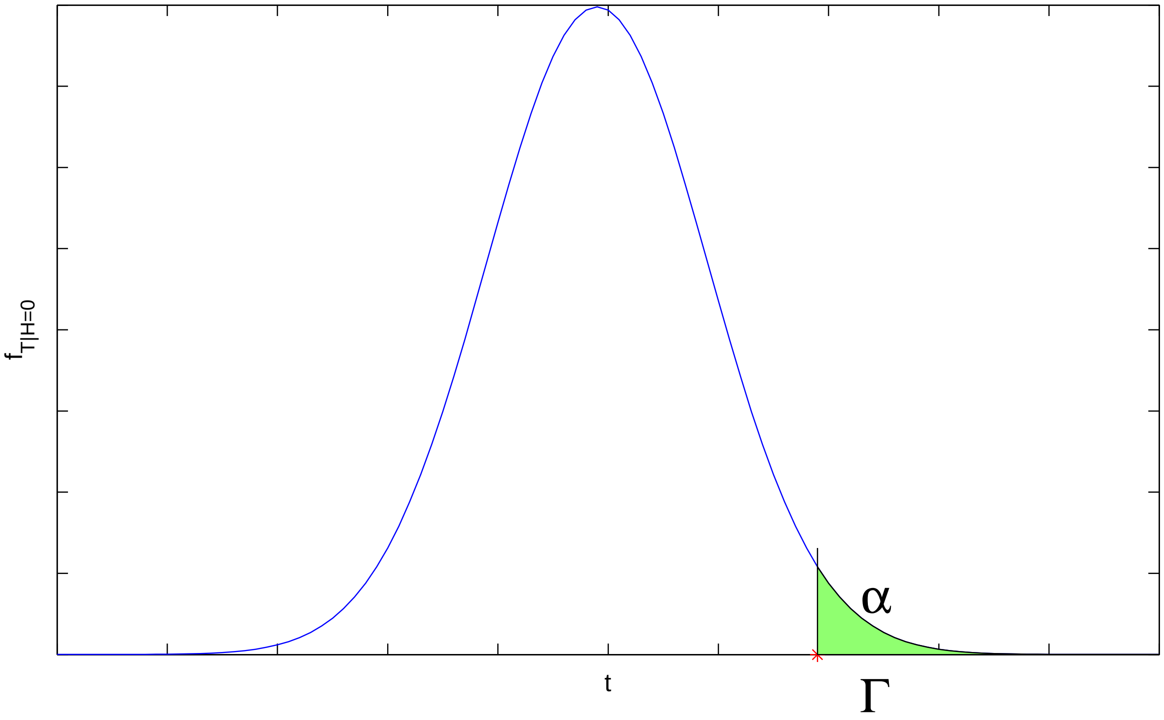
Single hypothesis testing

- Hypothesis test:
  - $H = 0 : \theta \in \Theta_0$ versus
  - $H = 1 : \theta \in \Theta_1$ ($\Theta_0 \cap \Theta_1 = \emptyset$).
- Test statistics: $T(X)$, observed t.

Single hypothesis testing

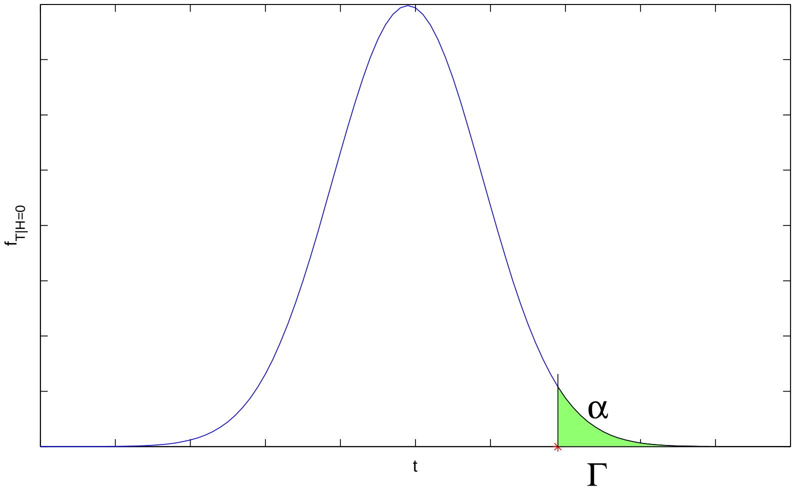
- Hypothesis test:
  - $H = 0 : \theta \in \Theta_0$ versus
  - $H = 1 : \theta \in \Theta_1$ ($\Theta_0 \cap \Theta_1 = \emptyset$).
- Test statistics: $T(X)$, observed t.
- Rejection region: Γ
  - If $t \in \Gamma$ reject $H = 0$.
  - If $t \notin \Gamma$ accept $H = 0$.

Single hypothesis testing



- Rejection region: Γ
  - If $t \in \Gamma$ reject $H = 0$.
  - If $t \notin \Gamma$ accept $H = 0$.

Single hypothesis testing

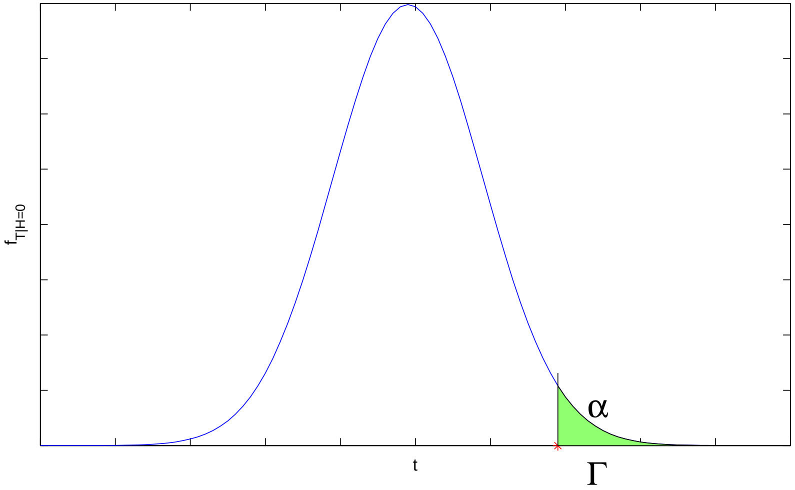


- Two types of errors:

	accept H_0	reject H_0
H_0		type-I error
H_1	type-II error	

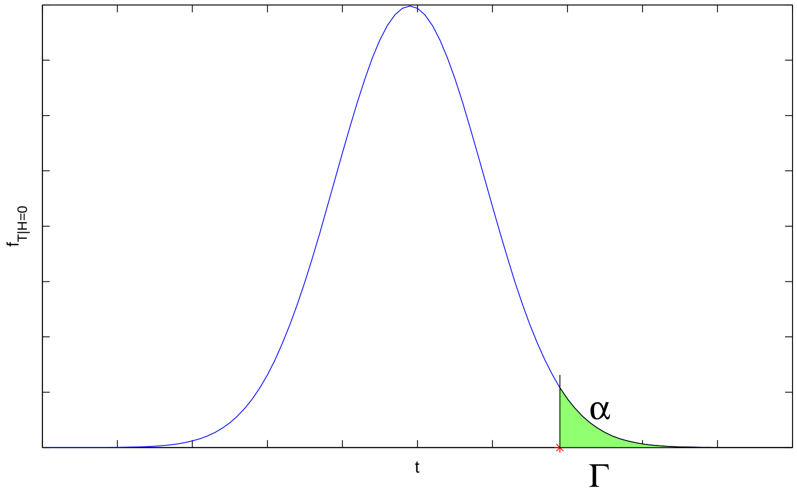
  - Type I error (false positive), $\theta \in \Theta_0$ yet $t \in \Gamma$.
  - Type II error (false negative), $\theta \in \Theta_1$ yet $t \notin \Gamma$

Single hypothesis testing



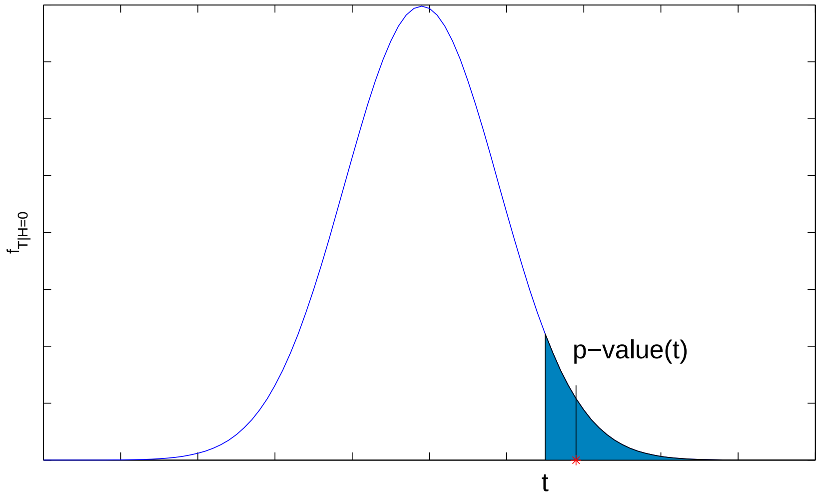
- Want to control type I error rate; $Pr(t \in \Gamma | H = 0)$,
- and minimise type II error rate; $Pr(t \notin \Gamma | H = 1)$.
- Power $= 1 - Pr(t \notin \Gamma | H = 1)$.

Single hypothesis testing



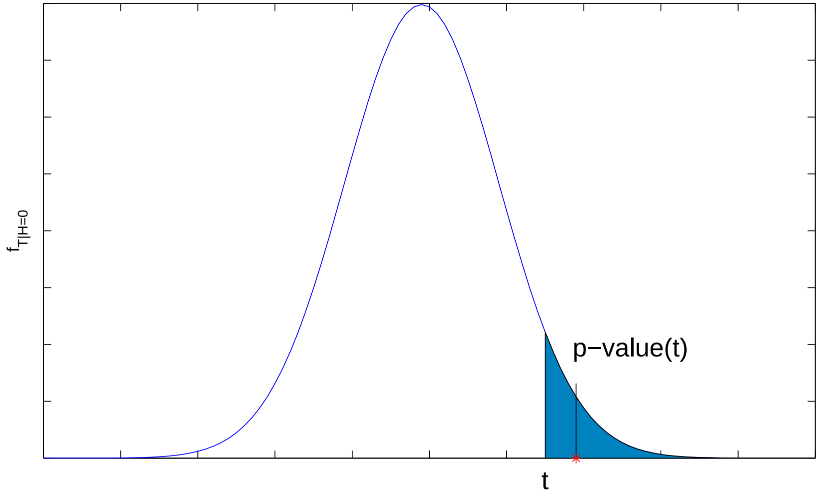
- Significant level $\alpha = Pr(t \in \Gamma | H = 0)$.

Single hypothesis testing



- Significant level $\alpha = Pr(t \in \Gamma | H = 0)$.
- $$\text{p-value} = \inf_{\Gamma : t \in \Gamma} Pr(t \in \Gamma | H = 0)$$
- Can use p-values as tests statistics.

Single hypothesis testing



- Significant level $\alpha = Pr(t \in \Gamma | H = 0)$.
-
$$\text{p-value} = \inf_{\Gamma : t \in \Gamma} Pr(t \in \Gamma | H = 0)$$
- Can use p-values as tests statistics.

Book: *Testing Statistical Hypotheses* E.L. Lehmann (1986)

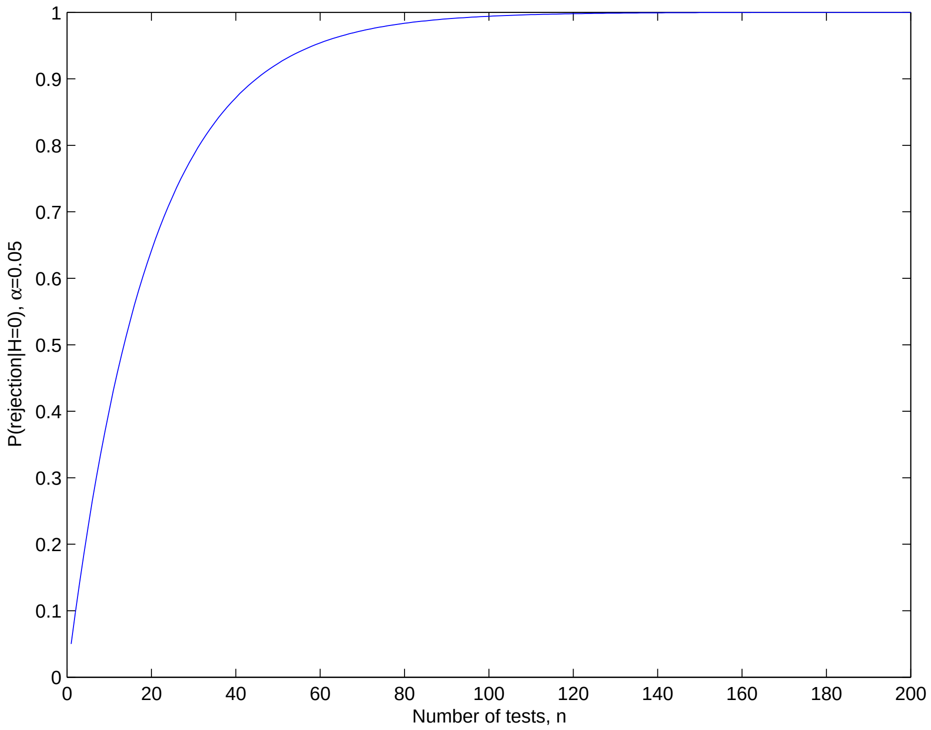
Multiple hypothesis testing

- m hypothesis tests
  - $H_1 = 0$ versus $H_1 = 1$
  - $H_2 = 0$ versus $H_2 = 1$
  -
  -
  - $H_m = 0$ versus $H_m = 1$
- Want to make simultaneous inference.
- Rejection regions?

Multiple hypothesis testing

- m hypothesis tests $(H_1, H_2, \ldots, H_m)$
- Want to make simultaneous inference.
- Rejection regions?

Same as in single hypothesis testing?



Multiple hyp. testing quantities

	accept null	reject null	total
$H = 0$	U	V	m_0
$H = 1$	T	S	m_1
total	W	R	m

- Total number of misclassifications: $V + T$.

Multiple hyp. testing quantities

	accept null	reject null	total
$H = 0$	U	V	m_0
$H = 1$	T	S	m_1
total	W	R	m

- Compound error rates:
  - Family wise error rate: $FWER = P(V \geq 1)$
  - Per comparison error rate: $PCER = E(V)/m$
  - False discovery rate: $FDR = E(V/R|R > 0)P(R > 0)$
  - Positive false discovery rate: $pFDR = E(V/R|R > 0)$

Multiple hyp. testing quantities

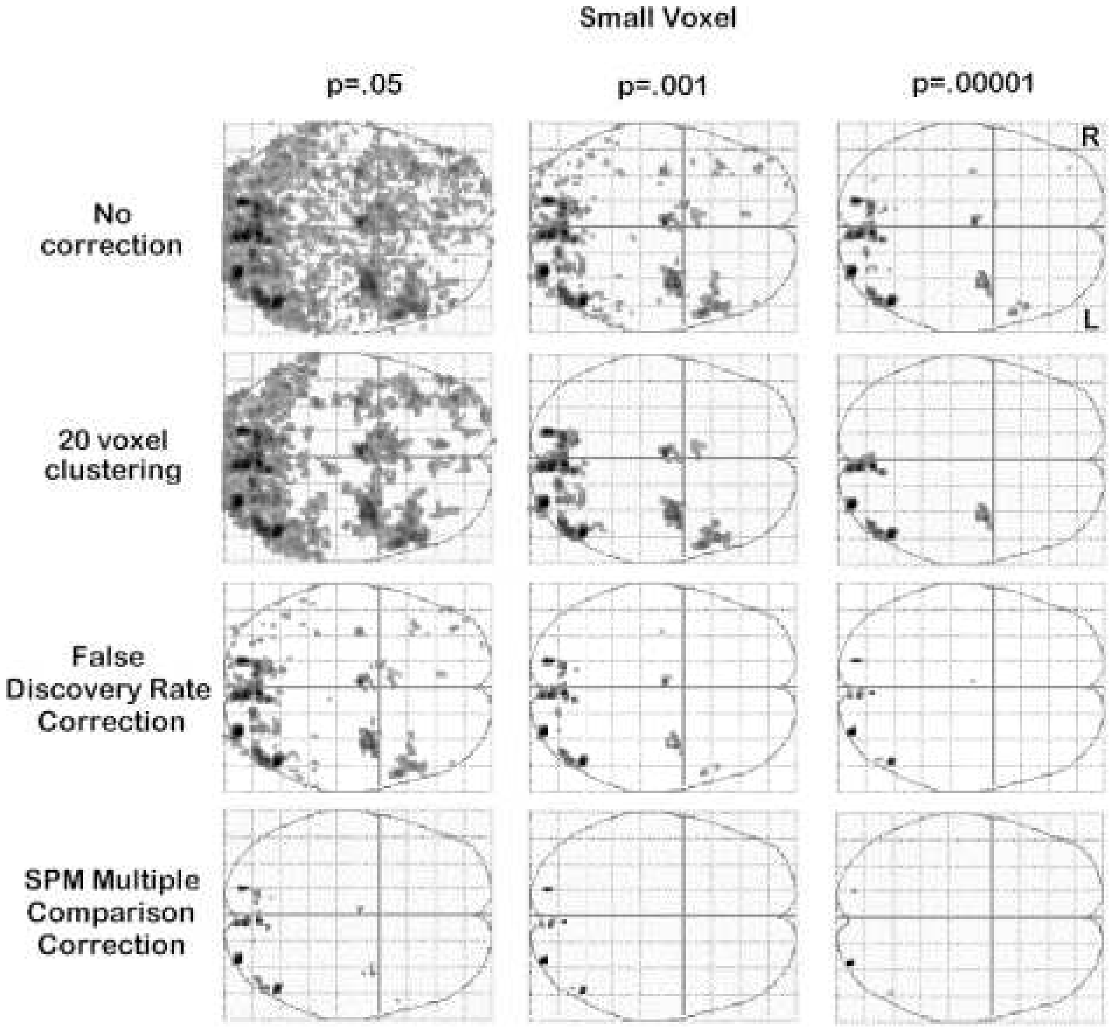
	accept null	reject null	total
$H = 0$	U	V	m_0
$H = 1$	T	S	m_1
total	W	R	m

- Weak control: Only when $m_0 = m$
- Strong control: Holds for all m_0 simultaneously.

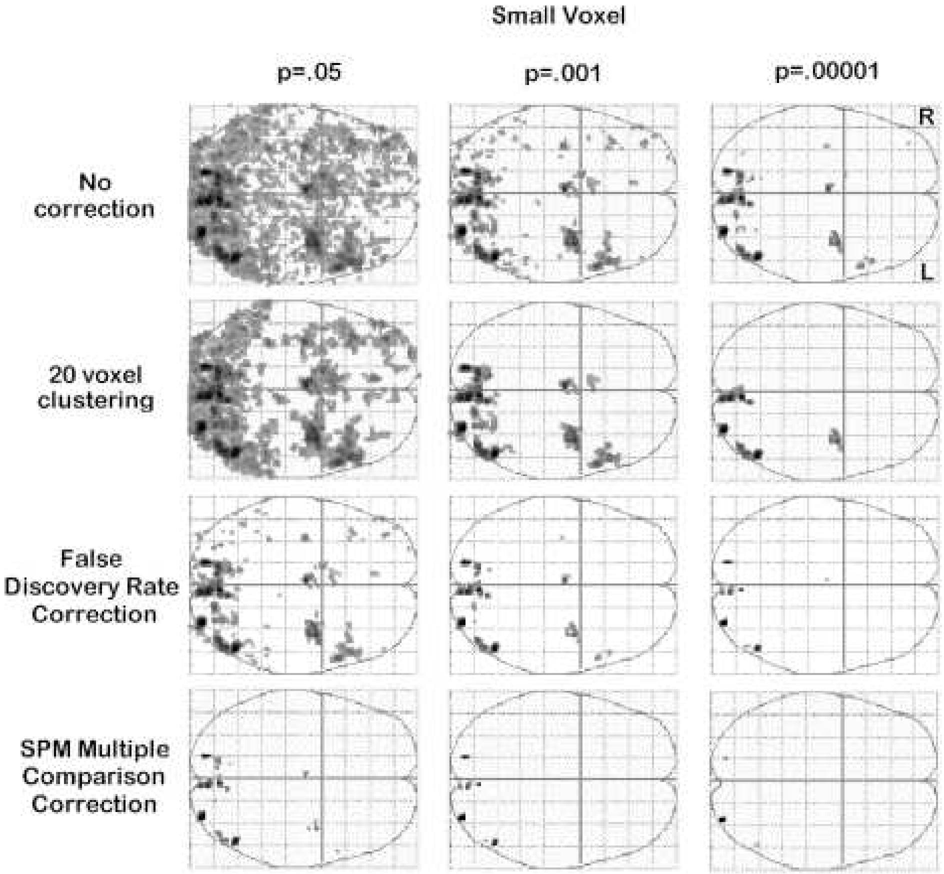
Example, fMRI

- *Now you see it, now you don't: statistical and methodological considerations in fMRI.*
  D.W. Loring et al., Epilepsy & Behavior 3 (2002)
- Each voxel is tested if activation causes difference.
- Pure exploratory study of method and significance level.

Example, fMRI



Example, fMRI



"(...), apparent random activation decreased as more conservative statistical approaches were employed, but activation in areas considered to be functionally significant was also reduced"

Multiple hyp. testing and microarray experiments

- DNA microarrays; method for measuring expression levels for thousands of genes simultaneous.
- Purpose: Identify different expressed genes.
- These can be further investigated using more expensive methods.
- Review article: *Multiple Hypothesis Testing in Microarray Experiments* S. Dudoit, J.P. Shaffer & J.C. Boldrick. Statistical Science 18 (2003).

False discovery rate

- Can accept some false rejections if they are relatively few.
- *Controlling the False Discovery rate: A Practical and Powerful Approach to Multiple Testing* by Y. Benjamini and Y. Hochberg, JRSS-B Vol 57 (1995).

  $$FDR = E(V/R|R > 0)P(R > 0)$$

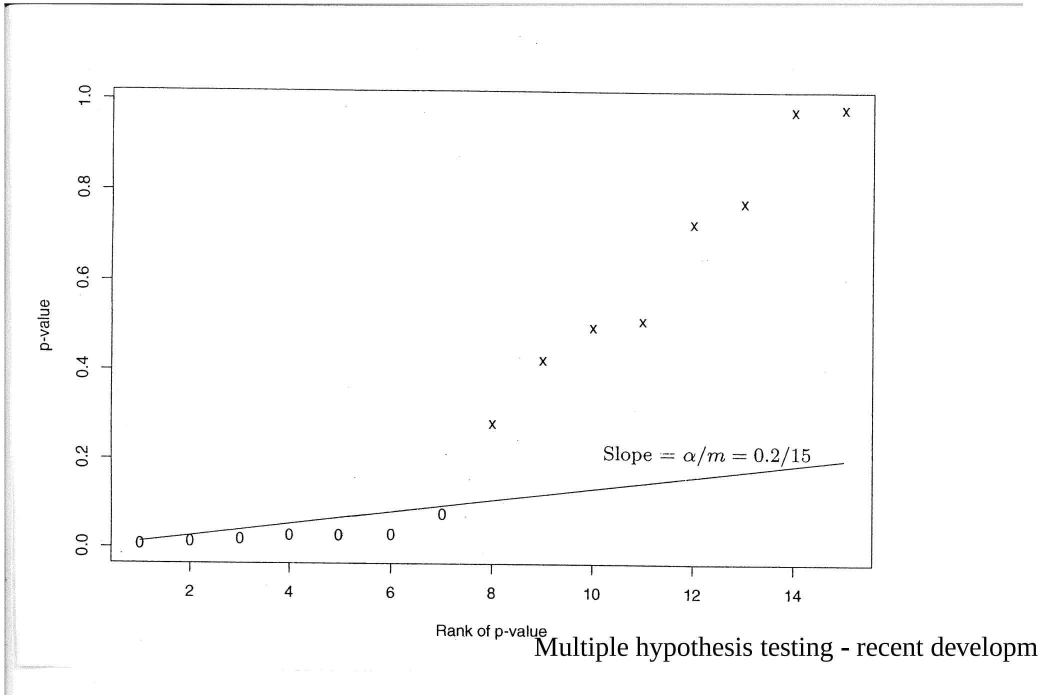
  V: Number of false rejections.
  R: Number of rejections.
- $FDR = E(V/R)$ with $V/R \equiv 0$ when $R = 0$.

FDR, BH-procedure

- Algorithm:
  - Find ordered observed p-values: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
  - Calculate $\hat{k} = \max\{k : p_{(k)} \leq \alpha \cdot k/m\}$
  - Reject null hyp. corresponding to $p_{(1)} \ldots p_{(\hat{k})}$

FDR, BH-procedure

- Algorithm:
  - Find ordered observed p-values: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
  - Calculate $\hat{k} = \max\{k : p_{(k)} \leq \alpha \cdot k/m\}$
  - Reject null hyp. corresponding to $p_{(1)} \ldots p_{(\hat{k})}$



FDR, BH-procedure

- Algorithm:
  - Find ordered observed p-values: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
  - Calculate $\hat{k} = \max\{k : p_{(k)} \leq \alpha \cdot k/m\}$
  - Reject null hyp. corresponding to $p_{(1)} \ldots p_{(\hat{k})}$
- Weakly controls FWER.

  An improved Bonferroni procedure for multiple tests of significance by R.J. Simes, Biometrica 73 (1986).
- Strongly controls FDR, Benjamini & Hochberg (1995)
- Also valid under some kind of dependences.

  Benjamini & Yekutieli, Annals of Statistics Vol 29 (2001)

Recent developments FDRs, outline and key references

- *A direct approach to false discovery rate* by J.D. Storey, JRSS-B vol 64 (2002)
  - Fixed rejection region procedure
  - The q-value

Recent developments FDRs, outline and key references

- Storey (2002)
  - Fixed rejection region procedure
  - The q-value
- *Strong Control, Conservative Point Estimation, and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach* by J.D. Storey, J.E. Taylor & D. Siegmund, in press JRSS-B
  - Improved fixed significance level procedure
  - Some theoretical results

Recent developments FDRs, outline and key references

- Storey (2002)
  - Fixed rejection region procedure
  - The q-value
- Storey et al. (2003)
  - Improved fixed significance level procedure
  - Some theoretical results
- *The positive false discovery rate: A Bayesian interpretation and the q-value* by John D. Storey, accepted in Annals of Statistics.
  - A Bayesian interpretation.
  - Classification theory.

Recent developments FDRs, outline and key references

- Storey (2002)
  - Fixed rejection region procedure
  - The q-value
- Storey et al. (2003)
  - Improved fixed significance level procedure
  - Some theoretical results
- Storey (2003)
  - A Bayesian interpretation.
  - Classification theory.
- *Operating characteristics and extensions of the false discovery rate procedure* by C. Genovese & L. Wasserman, JRSS-B (2002).
- Benjamini & Yekutieli (2001)

Positive false discovery rate

- $pFDR = E(V/R|R > 0)$

Positive false discovery rate

- $pFDR = E(V/R|R > 0)$
- Algorithm
  - Fix rejection region Γ
  - Calculate $pFDR$

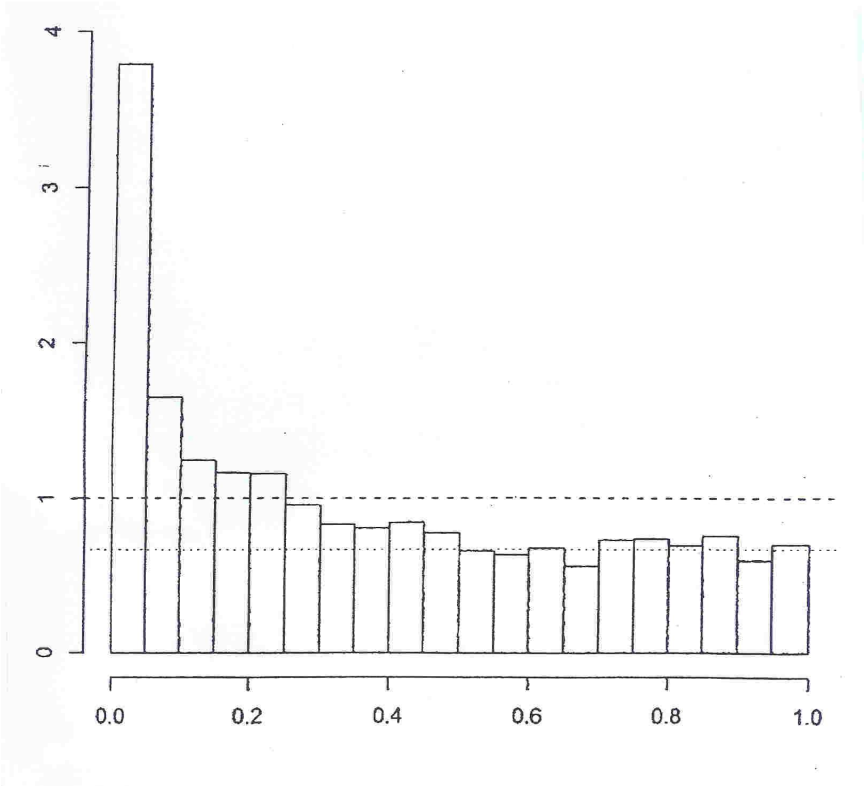
Positive false discovery rate

- $pFDR = E(V/R|R > 0)$
- Algorithm
  - Fix rejection region Γ
  - Calculate $pFDR$
- Useful approach?
  - Set Γ from experience from similar experiments.
  - Better power than FDR-procedure.

Positive false discovery rate

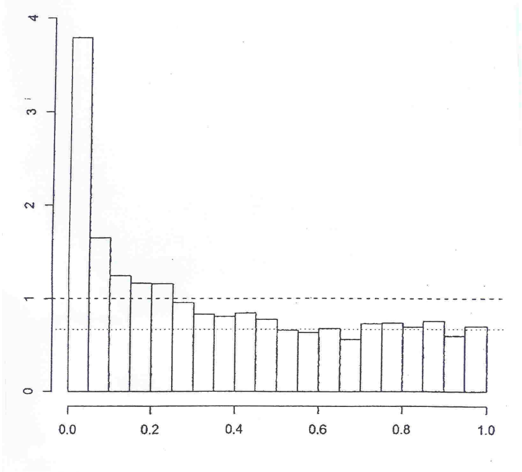
- $pFDR = E(V/R|R > 0)$
- Algorithm
  - Fix rejection region Γ
  - Calculate $pFDR$
- Useful approach?
  - Set Γ from experience from similar experiments.
  - Better power than FDR-procedure.
- Estimates $\pi_0 = \frac{m_0}{m}$
  - m: Number of tests
  - m_0: Number of true alternative hypothesis

Storeys estimaton of π_0



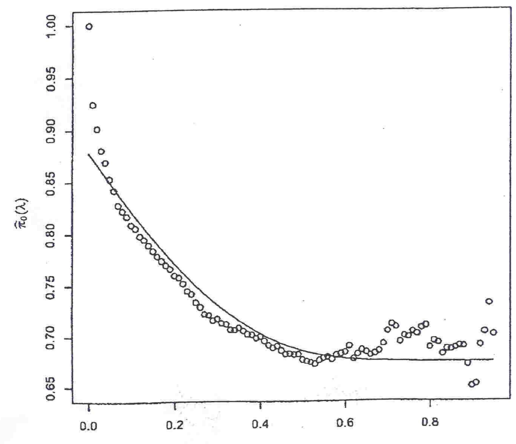
- Under the null-hyp p_i-s ars uniformly distributed.

Storeys estimaton of π_0



- Procedure
  - Choose a $0 < \lambda < 1$.
  - Assume $p_i > \lambda$ from uniform distribution.
  - Use $\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{(1-\lambda)m}$, where $W(\lambda) = \#\{p_i > \lambda\}$.

Storeys estimaton of π_0



- Procedure
  - Choose a $0 < \lambda < 1$.
  - Assume $p_i > \lambda$ from uniform distribution.
  - Use $\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{(1-\lambda)m}$, where $W(\lambda) = \#\{p_i > \lambda\}$.
- Can choose λ from minimising MSE obtained from bootstrapping.

Storeys estimaton of π_0

- Procedure
  - Choose a $0 < \lambda < 1$.
  - Assume $p_i > \lambda$ from uniform distribution.
  - Use $\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{(1-\lambda)m}$, where $W(\lambda) = \#\{p_i > \lambda\}$.
- Can choose λ from minimising MSE obtained from bootstrapping.
- Much research currently done.
- Has interest on its own.

Fixed rejection region procedure

- Calculate p-values $p_1, p_2, \ldots, p_m$.

Fixed rejection region procedure

- Calculate p-values $p_1, p_2, \ldots, p_m$.
- Estimate $\hat{\pi}_0(\lambda)$ and $\hat{P}r(P \leq t)$ by
  - $\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{(1-\lambda)m}$
  - $\hat{P}r(P \leq t) = \frac{R(t) \vee 1}{m}$

with $R(t) = \#\{p_i \leq t\}$ and $W(\lambda) = \#\{p_i > \lambda\}$

Fixed rejection region procedure

- Calculate p-values $p_1, p_2, \ldots, p_m$.
- Estimate $\hat{\pi}_0(\lambda)$ and $\hat{Pr}(P \leq t)$
- For rejection region of interest $[0, t]$, estimate $pFDR(t)$

$$\widehat{pFDR}_{\lambda}(t) = \frac{\hat{\pi}_0(\lambda) \cdot t}{\hat{Pr}(P \leq t) \cdot (1 - (1 - t)^m)}$$

Fixed rejection region procedure

- Calculate p-values $p_1, p_2, \ldots, p_m$.
- Estimate $\hat{\pi}_0(\lambda)$ and $\hat{P}r(P \leq t)$
- For rejection region of interest $[0, t]$, estimate $pFDR(t)$
- For B bootstrap samples of $p_1, p_2, \ldots, p_m$ find $\widehat{pFDR}_\lambda^{*b}(t)$.
- Use $(1 - \alpha)$ quantile of $\widehat{pFDR}_\lambda^{*b}(t)$ as the $(1 - \alpha)$ upper confidence bound for $pFDR(t)$.

Fixed rejection region procedure

- Calculate p-values $p_1, p_2, \ldots, p_m$.
- Estimate $\hat{\pi}_0(\lambda)$ and $\hat{Pr}(P \leq t)$
- For rejection region of interest $[0, t]$, estimate $pFDR(t)$
- For B bootstrap samples of $p_1, p_2, \ldots, p_m$ find $\widehat{pFDR}_\lambda^{*b}(t)$.
- Use $(1-\alpha)$ quantile of $\widehat{pFDR}_\lambda^{*b}(t)$ as the $(1-\alpha)$ upper confidence bound for $pFDR(t)$.
- If FDR of interest use $\widehat{FDR}_\lambda(t) = \frac{\hat{\pi}_0(\lambda) \cdot t}{\hat{Pr}(P \leq t)}$

The q-value

- A pFDR parallel to p-values.
  - $$\text{p-value} = \min_{\Gamma : t \in \Gamma} \{Pr(T \in \Gamma | H = 0)\}$$
  - $$\text{q-value} = \inf_{\Gamma : t \in \Gamma} (pFDR(\Gamma))$$
  - The minimum pFDR that can occur when rejecting a statistic with value t.
- For test with independent p-values, for observed p-value p

$$q(p) = \inf_{\gamma \geq p} \left\{ \frac{\pi_0 \gamma}{Pr(P \leq \gamma)} \right\}$$

The q-value

- For test with independent p-values, for observed p-value p

$$q(p) = \inf_{\gamma \geq p} \left\{ \frac{\pi_0 \gamma}{Pr(P \leq \gamma)} \right\}$$

- Estimation algorithm:
  - Calculate p-values $p_1, \ldots, p_m$.
  - Order the p-values: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
  - Set $\hat{q}(p_{(m)}) = \widehat{pFDR}(p_{(m)})$
  - for i=(m-1):1
    - Set $\hat{q}(p_{(i)}) = \min(\widehat{pFDR}(p_{(i)}), \hat{q}(p_{(i+1)}))$

BH vs. Storeys procedure

- BH-procedure:
  - Find ordered observed p-values: $p_{(1)} \leq p_{(2)} \leq \dots p_{(m)}$
  - Calculate $\hat{k} = \max\{k : p_{(k)} \leq \alpha \cdot k/m\}$
  - Reject null hyp. corresponding to $p_{(1)} \dots p_{(\hat{k})}$
- Threshold t found such that $\left(\frac{t \cdot m}{R(t)}\right) \leq \alpha$.

BH vs. Storeys procedure

- BH-procedure:
  - Find ordered observed p-values:
    $p_{(1)} \leq p_{(2)} \leq \dots p_{(m)}$
  - Calculate $\hat{k} = \max\{k : p_{(k)} \leq \alpha \cdot k/m\}$
  - Reject null hyp. corresponding to $p_{(1)} \dots p_{(\hat{k})}$
- Threshold t found such that $\boxed{\left(\frac{t \cdot m}{R(t)}\right)} \leq \alpha$.
- The natural empirical estimator for FDR.
- Corresponds $\widehat{FDR}_{\lambda=0}(t)$ (and $\pi_0 = 1$).

BH vs. Storeys procedure

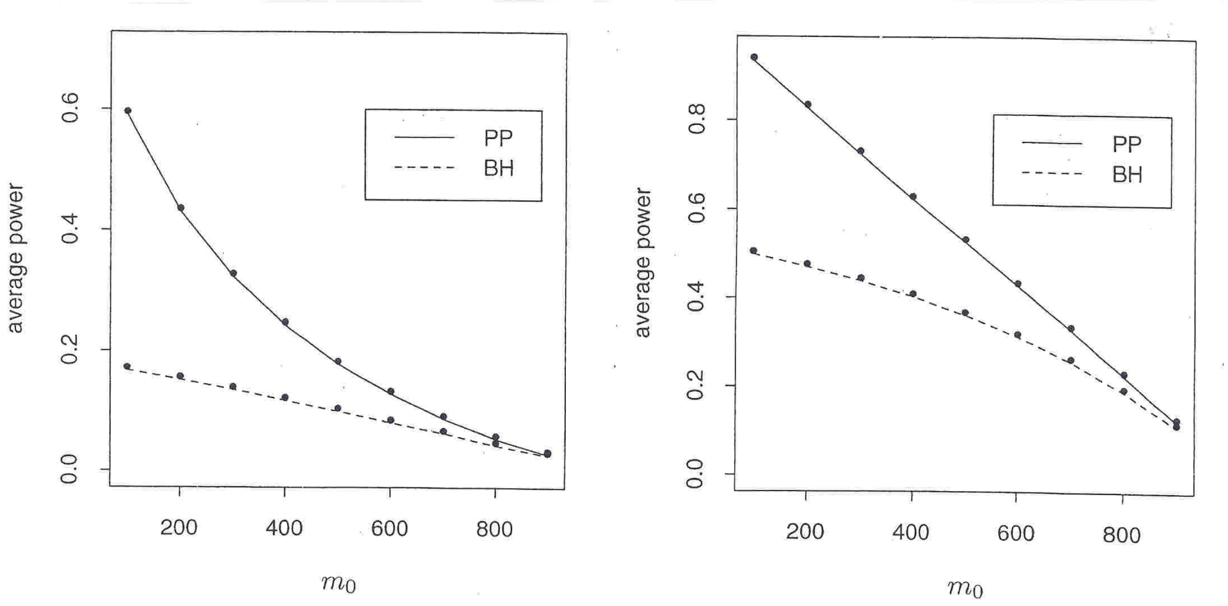
- New procedure:
  - Estimate $\hat{\pi}_0(\lambda)$, $(t \leq \lambda)$
  - Find ordered observed p-values: $p_{(1)} \leq p_{(2)} \leq \dots p_{(m)}$
  - Calculate $\hat{k} = \max\{k : p_{(k)} \leq \alpha \cdot k/(m \cdot \hat{\pi}_0(\lambda))\}$
  - Reject null hyp. corresponding to $p_{(1)} \dots p_{(\hat{k})}$
- Use estimated $\hat{\pi}_0$?
- A less conservative test.

BH vs. Storeys procedure

- New procedure:
  - Find ordered observed p-values: $p_{(1)} \leq p_{(2)} \leq \dots p_{(m)}$
  - Calculate $\hat{k} = \max\{k : p_{(k)} \leq \alpha \cdot k/(m \cdot \hat{\pi}_0(\lambda))\}$
  - Reject null hyp. corresponding to $p_{(1)} \dots p_{(\hat{k})}$
- Use estimated $\hat{\pi}_0$?
- If the p-values corresponding to the true null hypothesis are independent the procedure strongly controls the FDR at level α for any λ.
  Some technical adjustments needed.
- Asymptotically also valid under weakly dependence.

Example power, Storey et al. (2003)

- $m = 1000$ one-sided hypothesis tests.
- Null distribution $N(0,1)$, alternative $N(2,1)$
- $m_0 = 100, 200, \ldots, 900$
- 1000 sets of 1000 variables for each m_0
- Levels $\alpha = 0.05$ and $\alpha = 0.01$ and $\lambda = 0.5$



Bayesian interpretation

- Prior:
  - Let $Pr(H_i = 0) = \pi_0$ and $Pr(H_i = 1) = \pi_1$,
  - and assume H_i i.i.d. Bernoulli.

Bayesian interpretation

- Prior:
  - Let $Pr(H_i = 0) = \pi_0$ and $Pr(H_i = 1) = \pi_1$,
  - and assume H_i i.i.d. Bernoulli.

for $m = 1$

- $Pr(H = 0|T \in \Gamma) =$ Probability of false rejection given stat. is significant.
- $\frac{V(\Gamma)}{R(\Gamma)}|R > 0 = 0 \vee 1$
- $pFDR(\Gamma) = Pr(H = 0|T \in \Gamma)$, posterior probability that the rejection is false.

Bayesian interpretation

- Prior:
  - Let $Pr(H_i = 0) = \pi_0$ and $Pr(H_i = 1) = \pi_1$,
  - and assume H_i i.i.d. Bernoulli.

For general m

Theorem 1

Let T_i be test stat. corresponding to H_i. If

- (T_i, H_i) i.i.d., and
- $T_i | H_i \sim (1 - H_i)F_0 + H_i F_1$ then

$$pFDR(\Gamma) = Pr(H = 0 | T \in \Gamma)$$

Bayesian interpretation

- Prior:
  - Let $Pr(H_i = 0) = \pi_0$ and $Pr(H_i = 1) = \pi_1$,
  - and assume H_i i.i.d. Bernoulli.

For general m

Theorem 1

Let T_i be test stat. corresponding to H_i. If

- (T_i, H_i) i.i.d., and
- $T_i | H_i \sim (1 - H_i)F_0 + H_i F_1$ then

$$pFDR(\Gamma) = Pr(H = 0 | T \in \Gamma)$$

Posterior Bayesian type I error.

Does not depend on m

Properties pFDR

$$pFDR(\Gamma) = Pr(H = 0 | T \in \Gamma)$$

Properties pFDR

$$\begin{aligned} pFDR(\Gamma) &= Pr(H=0|T\in\Gamma) \\ &= \frac{\pi_0 \cdot Pr(T\in\Gamma|H=0)}{\pi_0 \cdot Pr(T\in\Gamma|H=0) + \pi_1 \cdot Pr(T\in\Gamma|H=1)} \end{aligned}$$

Properties pFDR

$$
\begin{aligned}
pFDR(\Gamma) &= Pr(H=0|T\in\Gamma) \\
&= \frac{\pi_0 \cdot Pr(T\in\Gamma|H=0)}{\pi_0 \cdot Pr(T\in\Gamma|H=0) + \pi_1 \cdot Pr(T\in\Gamma|H=1)} \\
&= \frac{\pi_0 \cdot (\text{Type-I-error of } \Gamma)}{\pi_0 \cdot (\text{Type-I-error of } \Gamma) + \pi_1 \cdot (\text{Power of } \Gamma)}
\end{aligned}
$$

- Increases with increasing type-I-errors.
- Decreases with increasing power.

Properties pFDR

$$
\begin{aligned}
pFDR(\Gamma) &= Pr(H=0|T\in\Gamma) \\
&= \frac{\pi_0 \cdot Pr(T\in\Gamma|H=0)}{\pi_0 \cdot Pr(T\in\Gamma|H=0) + \pi_1 \cdot Pr(T\in\Gamma|H=1)} \\
&= \frac{\pi_0 \cdot (\text{Type-I-error of } \Gamma)}{\pi_0 \cdot (\text{Type-I-error of } \Gamma) + \pi_1 \cdot (\text{Power of } \Gamma)}
\end{aligned}
$$

- $E[V(\Gamma)] = m \cdot \pi_0 \cdot Pr(T\in\Gamma|H=0)$
- $E[R(\Gamma)] = m \cdot Pr(T\in\Gamma)$

Properties pFDR

$$\begin{aligned} pFDR(\Gamma) &= Pr(H=0|T\in\Gamma) \\ &= \frac{\pi_0 \cdot Pr(T\in\Gamma|H=0)}{\pi_0 \cdot Pr(T\in\Gamma|H=0) + \pi_1 \cdot Pr(T\in\Gamma|H=1)} \\ &= \frac{\pi_0 \cdot (\text{Type-I-error of } \Gamma)}{\pi_0 \cdot (\text{Type-I-error of } \Gamma) + \pi_1 \cdot (\text{Power of } \Gamma)} \end{aligned}$$

Corollary

Under the assumptions of Theorem 1:

$$pFDR = E[\frac{V(\Gamma)}{R(\Gamma)}|R(\Gamma)>0] = \frac{E[V(\Gamma)]}{E[R(\Gamma)]}$$

Interpretation of the q-value

- Def:

$$\text{q-value} = \inf_{\Gamma_\alpha : t \in \alpha} pFDR(\Gamma_\alpha)$$

- The pFDR of the smallest possible rejection region s.t. $t \in \Gamma_\alpha$.

Corollary
Under the assumptions of Theorem 1:

$$\text{q-value} = \inf_{\Gamma_\alpha : t \in \Gamma_\alpha} Pr(H = 0 | T \in \Gamma_\alpha)$$

Connection to classification theory

- Misclassification penalties:

	Classify H_i as 0	Classify H_i as 1
$H_i = 0$	0	$1-\lambda$
$H_i = 1$	λ	0

- Bayes error:
  - $$\begin{aligned} BE(\Gamma) &= (1-\lambda)\cdot Pr(T_i \in \Gamma, H_i = 0) \\ &\quad +\lambda \cdot Pr(T_i \notin \Gamma, H_i = 1) \end{aligned}$$
  - Expected loss under misclassification penalties.

The positive non-discovery rate

- $$pFNR = E[\frac{T}{W}|W > 0]$$
  - W: Number of non-rejected hypothesis.
  - T: Number of non-rejected alternative hypothesis.

The positive non-discovery rate

- $$pFNR = E[\frac{T}{W}|W > 0]$$

Theorem 2

Under the assumptions of theorem 1 is

$$pNDR(\Gamma) = Pr(H = 1|T \notin \Gamma)$$

with $\pi_1 = 1 - \pi_0$ as prior; $Pr(H = 1) = \pi_1$.

- Posterior Bayesian type-II error

The positive non-discovery rate

- $$pFNR = E[\frac{T}{W}|W > 0]$$
- Posterior Bayesian type-II error

Corollary
Under the assumptions of theorem 1;

$$\begin{aligned} BE(\Gamma) &= (1-\lambda)\cdot Pr(T\in\Gamma)\cdot pFDR(\Gamma) \\ &+\lambda\cdot Pr(T\notin\Gamma)\cdot pNDR(\Gamma) \end{aligned}$$

Choosing rejection region

Two ways of fixing the rejection region beforehand:

- Rejection region Γ that minimise the Bayes error (based on relative cost λ)

$$\begin{aligned} BE(\Gamma) &= (1-\lambda)\cdot Pr(T\in\Gamma)\cdot pFDR(\Gamma) \\ &\quad +\lambda\cdot Pr(T\notin\Gamma)\cdot pFNR(\Gamma) \end{aligned}$$

- Rejection region Γ that minimise the weighted average

$$(1-\omega)\cdot pFDR(\Gamma)+\omega\cdot pFNR(\Gamma)$$

Choosing rejection region

Two ways of fixing the rejection region beforehand:

- Rejection region Γ that minimise the Bayes error (based on relative cost λ)

$$\begin{aligned} BE(\Gamma) &= (1-\lambda) \cdot Pr(T \in \Gamma) \cdot pFDR(\Gamma) \\ &\quad + \lambda \cdot Pr(T \notin \Gamma) \cdot pFNR(\Gamma) \end{aligned}$$

- Rejection region Γ that minimise the weighted average

$$(1-\omega) \cdot pFDR(\Gamma) + \omega \cdot pFNR(\Gamma)$$

- PS: Can not find Γ and estimate $pFDR$ from the same data.

Future challenges, false discovery rates

- Estimator properties:
    - Optimal conservative estimates for $\widehat{FRD}_\lambda$ and $p\widehat{FDR}_\lambda$?
    - Convergence properties.
    - Operational properties of $\hat{q}$.

Future challenges, false discovery rates

- Estimator properties:
  - Optimal conservative estimates for $\widehat{FRD_\lambda}$ and $p\widehat{FDR_\lambda}$?
  - Convergence properties.
  - Operational properties of $\hat{q}$.
- Dependencies:
  - Finite size dependency behaviour.
  - Modelling dependency among hypothesis tests.

Future challenges, false discovery rates

- Estimator properties:
  - Optimal conservative estimates for $\widehat{FRD_\lambda}$ and $\widehat{pFDR_\lambda}$?
  - Convergence properties.
  - Operational properties of $\hat{q}$.
- Dependencies:
  - Finite size dependency behaviour.
  - Modelling dependency among hypothesis tests.
- Gain power from more information.
  - Assumption about the alternative distribution.

Future challenges, multiple hypothesis testing

Three reasons for using FDR in multiple hypothesis testing Benjamini & Hochberg (1995):

- Multiple end points problem :
  - Whether to recommend a new treatment or not.
  - Rejected null: Treatment better then standard for specific end point.

Future challenges, multiple hypothesis testing

Three reasons for using FDR in multiple hypothesis testing Benjamini & Hochberg (1995):

- Multiple end points problem
- Multiple separate decisions :
  - Two treatments compared for multiple subgroups.
  - Recommendations made for each subgroup.

Future challenges, multiple hypothesis testing

Three reasons for using FDR in multiple hypothesis testing Benjamini & Hochberg (1995):

- Multiple end points problem
- Multiple separate decisions
- Screening problems:
  - As in the microarray setting.
  - Validation in a more expensive 2nd phase, want to limit the cost.

Multiple end points and multiple separate decisions

- Multiple end points problem:
  - Whether to recommend a new treatment or not.
  - Rejected null: Treatment better then standard for specific end point.
- Multiple separate decisions:
  - Two treatments compared for multiple subgroups.
  - Recommendation made for each subgroup.
- Independent decisions.
- Why adjust significance?

Multiple end points and multiple separate decisions

- Multiple end points problem:
  - Whether to recommend a new treatment or not.
  - Rejected null: Treatment better then standard for specific end point.
- Multiple separate decisions:
  - Two treatments compared for multiple subgroups.
  - Recommendation made for each subgroup.
- Independent decisions.
- Why adjust significance?

Dependency

Decision theory and hypothesis testing

- Decision theory: Want to minimise expected loss.
- Single hyp. testing minimise $E(L_1)$;

	accept H	reject H
$H = 0$	0	0
$H = 1$	1	0

under the constraint $E(L_2) < \alpha$ with L_2

	accept H	reject H
$H = 0$	0	1
$H = 1$	0	0

Decision theory and hypothesis testing

- Decision theory: Want to minimise expected loss.
- Single hyp. testing minimise $E(L_1)$; under the constraint $E(L_2) < \alpha$ with L_2
- Using pFDR

	accept null	reject null
$H_i = 0$	0	$1 - \omega$
$H_i = 1$	ω	0

Decision theory and hypothesis testing

- Decision theory: Want to minimise expected loss.
- Single hyp. testing minimise $E(L_1)$;
  under the constraint $E(L_2) < \alpha$ with L_2
- Using pFDR

	accept null	reject null
$H_i = 0$	0	$1 - \omega$
$H_i = 1$	ω	0

- Natural choice of loss function?

Future challenges

- Dependency!

Future challenges

- Dependency!
- Is multiple hypothesis testing the right tool?
  - Exploration of dataset ⇒ estimation.
  - Make decision ⇒ loss function