Martingale approach for multiple testing and FDR control

Arnold Janssen

Lehrstuhl für Mathematische Statistik und Wahrscheinlichkeitstheorie
Mathematisches Institut
Heinrich-Heine-Universität Düsseldorf

Padua, 29.April 2016

partially joint work with Philipp Heesen

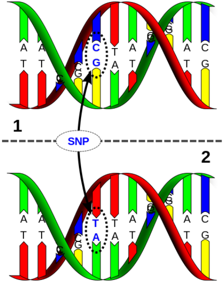
(supported by DFG grants)

Outline

Outline

GENOMICS,MEDICINE,COSMOLOGY...= A lot of data

n hypotheses have to be tested simultaneously.



n hypotheses H_i with p-values (H_i, p_i), $1 \leq i \leq n$

- H_i true if $i \in I_0$
- H_i false if $i \in I_1 = \{1, \ldots, n\} \setminus I_0$
- $n_0 := |I_0|$ and $n_1 := |I_1|$.
- order statistics $p_{1:n} \leq p_{2:n} \leq \ldots \leq p_{n:n}$

Linear step up procedure for $\alpha \in (0, 1)$

Benjamini/Hochberg (1995)

$$j^* := \max\{i : p_{i:n} \leq \alpha_{i:n}\}, \quad \alpha_{i:n} = \frac{i}{n}\alpha$$

"reject H_i if $p_i \leq \alpha_{j^*:n}$" (or $p_{i:n}$, $i \leq j^*$)

- $R = \#\{i : p_i \leq \alpha_{j^*:n}\}$ number of rejections
- $V = \#\{i :$ true null H_i with $p_i \leq \alpha_{j^*:n}\}$ number of false rejections
- $\frac{V}{R}$ false discovery proportion

Linear step up procedure for $\alpha \in (0, 1)$

Benjamini/Hochberg (1995)

$$j^* := \max\{i \,:\, p_{i:n} \leq \alpha_{i:n}\}, \quad \alpha_{i:n} = \frac{i}{n}\alpha$$

"reject H_i if $p_i \leq \alpha_{j^*:n}$" (or $p_{i:n}$, $i \leq j^*$)

- $R = \#\{i \,:\, p_i \leq \alpha_{j^*:n}\}$ number of rejections
- $V = \#\{i \,:\,$ true null H_i with $p_i \leq \alpha_{j^*:n}\}$ number of false rejections
- $\frac{V}{R}$ false discovery proportion

Basic independence (BI) assumptions:

- $(p_i)_{i \in I_0}$ and $(p_j)_{j \in I_1}$ are independent
- $(p_i)_{i \in I_0}$ are i.i.d. uniformly distributed on $(0, 1)$
- $(p_j)_{j \in I_1}$ arbitrary dependence is allowed

Theorem 1 (Benjamini/Hochberg 1995)

Under BI we have for the false discovery rate

$$FDR := E\left[\frac{V}{R \vee 1}\right] \leq \frac{n_0}{n}\alpha \quad \text{for} \quad \alpha_{i:n} = \frac{i}{n}\alpha.$$

Finner and Roters (2001) showed "=".

Benjamini and Yekutieli (2001), Storey (2002)

Basic independence (BI) assumptions:

- $(p_i)_{i \in I_0}$ and $(p_j)_{j \in I_1}$ are independent
- $(p_i)_{i \in I_0}$ are i.i.d. uniformly distributed on $(0, 1)$
- $(p_j)_{j \in I_1}$ arbitrary dependence is allowed

Theorem 1 (Benjamini/Hochberg 1995)

Under BI we have for the false discovery rate

$$FDR := E\left[\frac{V}{R \vee 1}\right] \leq \frac{n_0}{n}\alpha \quad \text{for} \quad \alpha_{i:n} = \frac{i}{n}\alpha.$$

Finner and Roters (2001) showed "$=$".

Benjamini and Yekutieli (2001), Storey (2002)

Number of citations (Google Scholar):

Benjamini Hochberg (1995)	29.418
Storey (2002)	3.342
Storey (2004)	896

Outline

2. Adaptive SU test procedures, stochastic process approach

Benjamini/Hochberg procedure: FDR$= \frac{n_0}{n}\alpha$ (linear SU test)

Idea of adaptive procedures: estimate n_0 by $\widehat{n}_0$

- One would like to replace α with $\frac{n}{\widehat{n}_0}\alpha$
- So far FDR $\approx \frac{n_0}{n}\frac{n}{\widehat{n}_0}\alpha \approx \alpha$ if $\frac{n_0}{\widehat{n}_0} \approx 1$

Definition

The adaptive SU procedure is a SU test with data dependent critical values $\widehat{\alpha}_{i:n}(\widehat{n}_0)$

$$\widehat{\alpha}_{i:n} = \frac{i}{\widehat{n}_0}\alpha.$$

2. Adaptive SU test procedures, stochastic process approach

Benjamini/Hochberg procedure: FDR$= \frac{n_0}{n}\alpha$ (linear SU test)

Idea of adaptive procedures: estimate n_0 by $\widehat{n}_0$

- One would like to replace α with $\frac{n}{\widehat{n}_0}\alpha$
- So far FDR $\approx \frac{n_0}{n}\frac{n}{\widehat{n}_0}\alpha \approx \alpha$ if $\frac{n_0}{\widehat{n}_0} \approx 1$

Definition

The adaptive SU procedure is a SU test with data dependent critical values $\widehat{\alpha}_{i:n}(\widehat{n}_0)$

$$\widehat{\alpha}_{i:n} = \frac{i}{\widehat{n}_0}\alpha.$$

Sarkar (2008): Condition on $\widehat{n}_0$ for finite sample FDR control.
For example: The condition is fulfilled under BI assumption for the estimator

$$\widehat{n}_0 = n \frac{1 - \widehat{F}_n(\lambda) + 1/n}{1 - \lambda} \quad \text{(see also Storey et al. (2004)),}$$

$\widehat{F}_n$ empirical distribution function of $p_1, \ldots, p_n$.

Disadvantage: The estimator only considers the empirical distribution function on one position, λ tuning parameter.

Now:

- new approach and new estimators $\widehat{n}_0$
- new procedures (not yet treated)

Sarkar (2008): Condition on $\widehat{n}_0$ for finite sample FDR control. For example: The condition is fulfilled under BI assumption for the estimator

$$\widehat{n}_0 = n\frac{1-\widehat{F}_n(\lambda)+1/n}{1-\lambda} \quad \text{(see also Storey et al. (2004))},$$

$\widehat{F}_n$ empirical distribution function of $p_1, \ldots, p_n$.

Disadvantage: The estimator only considers the empirical distribution function on one position, λ tuning parameter.

Now:

- new approach and new estimators $\widehat{n}_0$
- new procedures (not yet treated)

Motivation of the Storey estimator

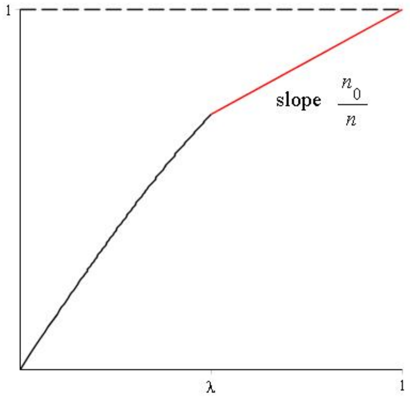
$$\frac{\widehat{n}_0(\lambda)}{n} = \frac{1 - \widehat{F}_n(\lambda) + \frac{1}{n}}{1 - \lambda}$$

Assume for a moment:

n_0 uniformly distributed p-values
n_1 p-values of false null, d.f. $F_1(t) \geq t$

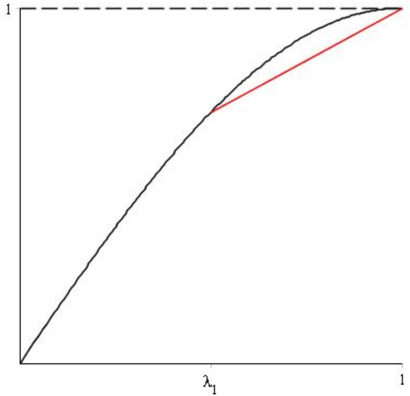
joint mixture: $F(t) = \frac{n_0}{n} t + \frac{n_1}{n} F_1(t),\ 0 \leq t \leq 1$

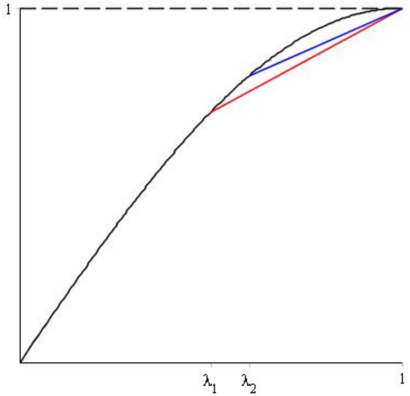
If $F_1(\lambda) = 1$ then $\frac{n_0}{n} = \frac{1-F(\lambda)}{1-\lambda}$.



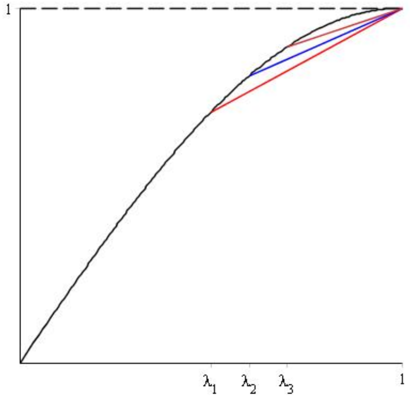
Storey estimator estimates the "slope" of $\widehat{F}_n$ on $[\lambda, 1]$

If $F_1(\lambda) < 1$ then $\frac{\widehat{n}_0(\lambda)}{n}$ is not accurate.





What is a good choice of λ?



What is a good choice of λ?

Reverse martingale model (Heesen/J. (2015))

$$t \mapsto \frac{\sum_{i \in I_0} \mathbf{1}_{[0,t]}(p_i)}{t} \qquad \text{reverse martingale}$$

w.r.t. backwards filtration

$$\mathcal{G}_t := \sigma\left(\mathbf{1}\left(p_i \leq s\right), s \geq t, 1 \leq j \leq n\right).$$

Examples (Heesen/J.), see the Appendix.

BI,

p-values for Marshall/Olkin type extreme value models,
blocks of identical independent p-values,
mixture of these models, ...

Reverse martingale model (Heesen/J. (2015))

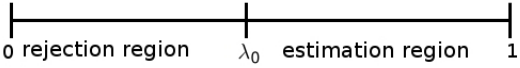
$$t \mapsto \frac{\sum_{i \in I_0} \mathbf{1}_{[0,t]}(p_i)}{t} \qquad \text{reverse martingale}$$

w.r.t. backwards filtration

$$\mathcal{G}_t := \sigma\left(\mathbf{1}\left(p_i \leq s\right), s \geq t, 1 \leq j \leq n\right).$$

Examples (Heesen/J.), see the Appendix.

BI,
p-values for Marshall/Olkin type extreme value models,
blocks of identical independent *p*-values,
mixture of these models, . . .



$$V(\lambda_0) := \#\{i \in I_0 : p_i \leq \lambda_0\}$$

Lemma (Heesen/J. (2015))

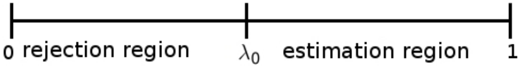
Assumptions: R-Martingale model

- *Rejection region* $[0, \lambda_0]$
  StepUp test with critical values $\alpha_{i:n} = \left(\frac{i}{\widehat{n}_0}\alpha\right) \wedge \lambda_0$
- *Estimation region* $[\lambda_0, 1]$
  Estimator $\widehat{n}_0 = f((\widehat{F}_n(t))_{t \geq \lambda_0})$ *and* $\widehat{n}_0 > 0$

Then the condition

$$E\left[\frac{V(\lambda_0)}{\widehat{n}_0}\right] \leq \lambda_0$$

implies FDR control, i.e. FDR $\leq \alpha$.



$$V(\lambda_0) := \#\{i \in I_0 : p_i \leq \lambda_0\}$$

Lemma (Heesen/J. (2015))

Assumptions: R-Martingale model

- *Rejection region* $[0, \lambda_0]$
  StepUp test with critical values $\alpha_{i:n} = \left(\frac{i}{\widehat{n}_0}\alpha\right) \wedge \lambda_0$
- *Estimation region* $[\lambda_0, 1]$
  Estimator $\widehat{n}_0 = f((\widehat{F}_n(t))_{t \geq \lambda_0})$ *and* $\widehat{n}_0 > 0$

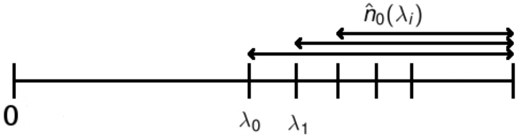
Then the condition

$$E\left[\frac{V(\lambda_0)}{\widehat{n}_0}\right] \leq \lambda_0$$

implies FDR control, i.e. FDR$\leq \alpha$.

Stationary approach: (under BI model)
The assumption is fulfilled for a linear combination of Storey estimators $\widehat{n}_0 = \sum_{i=1}^{k} \beta_i \widehat{n}_0(\lambda_i)$ for a couple of inspection points λ_i with $\lambda_0 \leq \lambda_1 < \ldots < \lambda_k < 1$, fixed $\beta_i > 0$ with $\sum \beta_i = 1$ and

$$\widehat{n}_0(\lambda_i) = \frac{1 - \widehat{F}_n(\lambda_i) + \frac{1}{n}}{1 - \lambda_i}.$$

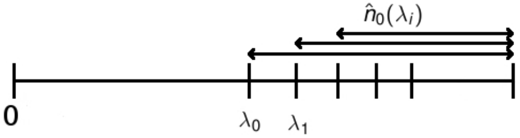


Practical guide (via conditional variance):

$$\beta_i = \frac{\sqrt{\frac{1}{\lambda_i} - 1}}{\sum_{j=1}^{k} \sqrt{\frac{1}{\lambda_j} - 1}}, \quad i = 1, \ldots, k.$$

Stationary approach: (under BI model)
The assumption is fulfilled for a linear combination of Storey estimators $\widehat{n}_0 = \sum_{i=1}^{k} \beta_i \widehat{n}_0(\lambda_i)$ for a couple of inspection points λ_i with $\lambda_0 \leq \lambda_1 < \ldots < \lambda_k < 1$, fixed $\beta_i > 0$ with $\sum \beta_i = 1$ and

$$\widehat{n}_0(\lambda_i) = \frac{1 - \widehat{F}_n(\lambda_i) + \frac{1}{n}}{1 - \lambda_i}.$$



Practical guide (via conditional variance):

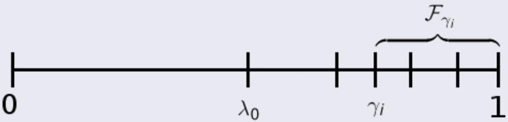
$$\beta_i = \frac{\sqrt{\frac{1}{\lambda_i} - 1}}{\sum_{j=1}^{k} \sqrt{\frac{1}{\lambda_j} - 1}}, \quad i = 1, \ldots, k.$$

Theorem (Adaptive procedure,dynamic approach, Heesen/J.)

Under the BI assumptions and

- $\lambda_0 < \gamma_1 < \ldots < \gamma_k \leq 1$,
- $\mathcal{F}_{\gamma_i} = \sigma((\widehat{F}_n(t))_{t \geq \gamma_i})$,
- $\widehat{\beta}_1, \ldots, \widehat{\beta}_k$ *data dependent weights with* $\sum \widehat{\beta}_i = 1$,
- $\widehat{\beta}_i$ *is* $\mathcal{F}_{\gamma_i}$ *measurable,*
- $\widehat{n}_0(\gamma_i) = f_i((\widehat{F}_n(t))_{t \geq \lambda_0})$ *with* $E\left[\frac{V(\lambda_0)}{\widehat{n}_0(\gamma_i)} \mid \mathcal{F}_{\gamma_i}\right] \leq \lambda_0$,

we get finite sample FDR control for the SU test procedure under $\widehat{n}_0 = \sum_{i=1}^{k} \widehat{\beta}_i \widehat{n}_0(\gamma_i)$.



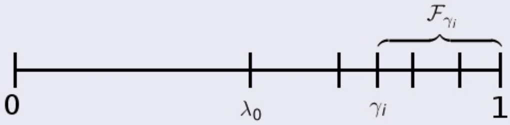
Proposal for the choice of $\widehat{\beta}_i$: Heesen/J. (2016)

Theorem (Adaptive procedure,dynamic approach, Heesen/J.)

Under the BI assumptions and

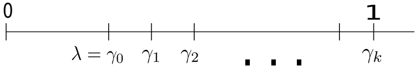
- $\lambda_0 < \gamma_1 < \ldots < \gamma_k \leq 1$,
- $\mathcal{F}_{\gamma_i} = \sigma((\widehat{F}_n(t))_{t \geq \gamma_i})$,
- $\widehat{\beta}_1, \ldots, \widehat{\beta}_k$ *data dependent weights with* $\sum \widehat{\beta}_i = 1$,
- $\widehat{\beta}_i$ *is* $\mathcal{F}_{\gamma_i}$ *measurable,*
- $\widehat{n}_0(\gamma_i) = f_i((\widehat{F}_n(t))_{t \geq \lambda_0})$ *with* $E\left[\frac{V(\lambda_0)}{\widehat{n}_0(\gamma_i)} \mid \mathcal{F}_{\gamma_i}\right] \leq \lambda_0$,

we get finite sample FDR control for the SU test procedure under $\widehat{n}_0 = \sum_{i=1}^{k} \widehat{\beta}_i \widehat{n}_0(\gamma_i)$.



Proposal for the choice of $\widehat{\beta}_i$: Heesen/J. (2016)

Motivation: via Storey estimator



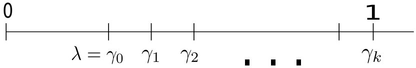
$$n\frac{1-\widehat{F}_n(\lambda)+\frac{1}{n}}{1-\lambda} \approx n\frac{1-\widehat{F}_n(\lambda)}{1-\lambda}$$

$$= \sum_{i=0}^{k-1} \underbrace{\frac{\gamma_{i+1}-\gamma_i}{1-\lambda}}_{=\beta_i} \underbrace{n\frac{\widehat{F}_n(\gamma_{i+1})-\widehat{F}_n(\gamma_i)}{\gamma_{i+1}-\gamma_i}}_{=\widehat{n}_0(\gamma_i)\ \mathcal{F}_{\gamma_i}\text{-measurable}}$$

choose dynamic $\mathcal{F}_{\gamma_i}$-measurable weights $\widehat{\beta}_i$

$$\widehat{n}_0 = \sum_{i=0}^{k-1} \widehat{\beta}_i \widehat{n}_0(\gamma_i)$$

Motivation: via Storey estimator



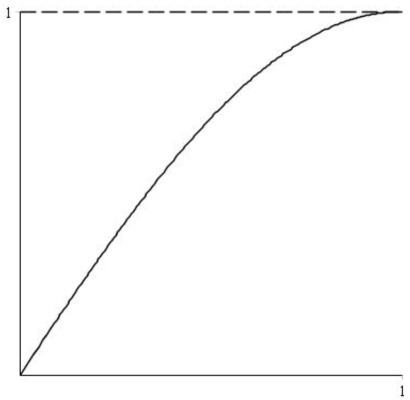
$$n\frac{1-\widehat{F}_n(\lambda)+\frac{1}{n}}{1-\lambda} \approx n\frac{1-\widehat{F}_n(\lambda)}{1-\lambda}$$

$$= \sum_{i=0}^{k-1} \underbrace{\frac{\gamma_{i+1}-\gamma_i}{1-\lambda}}_{=\beta_i} \; \underbrace{n\frac{\widehat{F}_n(\gamma_{i+1})-\widehat{F}_n(\gamma_i)}{\gamma_{i+1}-\gamma_i}}_{=\widehat{n}_0(\gamma_i)\ \ \mathcal{F}_{\gamma_i}\text{-measurable}}$$

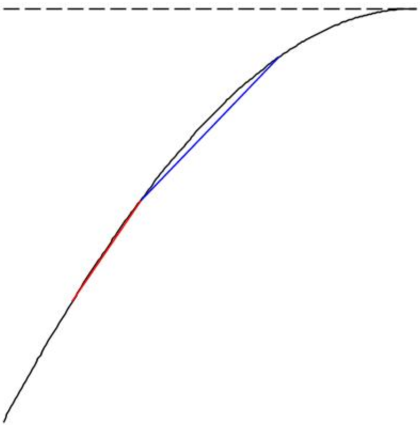
choose dynamic $\mathcal{F}_{\gamma_i}$-measurable weights $\widehat{\beta}_i$

$$\widehat{n}_0 \;=\; \sum_{i=0}^{k-1} \widehat{\beta}_i \widehat{n}_0(\gamma_i)$$

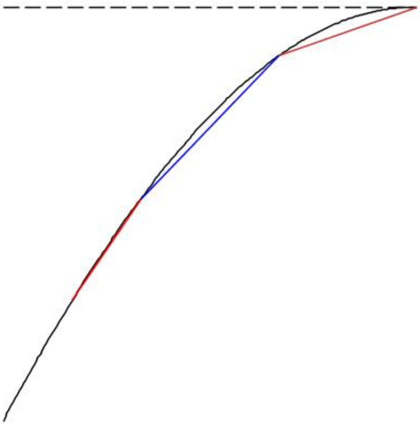
joint mixture



zoom: slope on subintervals $[\gamma_i, \gamma_{i+1}]$



backwards selection of contributing subintervals (data dependent)



Outline

Storey's procedure can be bad under dependence
Extreme dependence: $p_1 = p_2 = \ldots = p_n = U$ uniformly

$$\mathrm{FDR}_{\mathrm{Storey}} = \lambda\,(> \alpha) \quad \text{for large } n$$

R-martingale model

k independent chromosomes

R-martingale dependence within the chromosome
Use modified Storey estimators

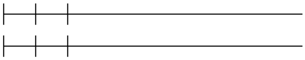
$$\widehat{n}(\kappa) := n\frac{1 - \widehat{F}_n(\lambda) + \frac{\kappa}{n}}{1-\lambda}, \quad \kappa > 1.$$

Heesen/J. (2015): Discussion about FDR control

Storey's procedure can be bad under dependence
Extreme dependence: $p_1 = p_2 = \ldots = p_n = U$ uniformly

$$\mathrm{FDR}_{\mathrm{Storey}} = \lambda\,(> \alpha) \quad \text{for large } n$$

R-martingale model



k independent chromosomes

...

R-martingale dependence within the chromosome
Use modified Storey estimators

$$\widehat{n}(\kappa) := n\frac{1 - \widehat{F}_n(\lambda) + \frac{\kappa}{n}}{1-\lambda}, \qquad \kappa > 1.$$

Heesen/J. (2015): Discussion about FDR control

Outline

References

- Benditkis, J., 2015. Martingale Methods for Control of False Discovery Rate and Expected Number of False Rejections. Dissertation. `http://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-37498/MartingaleMethods.Benditkis.2015.pdf`
- Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. Ser. B 57(1), 289–300.
- Benjamini, Y., Hochberg, Y., 2000. On the adaptive control of the false discovery rate in multiple testing with independent statistics. J. Educ. Behav. Statist. 25(1), 60–83.
- Finner, H. and Roters, M., 2001. On the false discovery rate and expected type I errors. Biom. J. 43(8), 985–1005.
- Finner, H., Dickhaus, T., and Roters, M., 2009. On the false discovery rate and an asymptotically optimal rejection curve. Ann. Statist. 37(2), 596–618.
- Gavrilov, Y., Benjamini, Y. and Sakar, S.K., 2009. An adaptive step-down procedure with proven FDR control under independence. Ann. Statist. 37(2): 619–629.
- Heesen, P. and Janssen, A., 2015. Inequalities for the false discovery rate (FDR) under dependence. Electron. J. Stat. 9, 679–716.
- Heesen, P. and Janssen, A., 2016. Dynamic adaptive multiple tests with finite sample FDR control. J. Stat. Plan. Inference 168, 38–51.
- Sarkar, S.K., 2008. On methods controlling the false discovery rate. Sankhyā 70(2, Ser.A), 135–168.
- Sakar, S.K., 2008. Two-stage stepup procedures controlling FDR. J. Statist. Plann. Infernece 138(4), 1072–1084.
- Storey, J.D., 2002. A direct approach to false discovery rate. J. R. Stat. Soc. Ser. B Stat. Methodol. 64(3). 479–498.
- Storey, J.D., Taylor, J.E. and Siegmund, D., 2004. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. J. R. Stat. Soc. Ser. B Stat. Methodol., 66(1),187–205.

Thank you for your attention!

Outline

Reverse martingale (Heesen/J.)

$$t \mapsto \frac{\sum_{i \in I_0} \mathbf{1}_{[0,t]}(p_i)}{t} \quad p_i \text{ uniform i.i.d.}$$

Martingale conditions (Benditkis (2015), Dissertation)

$$t \mapsto \frac{\sum_{i \in I_0} \mathbf{1}_{[0,t]}(p_i) - t}{1-t} \quad \mathcal{F}_t\text{-supermartingale,}$$

where $\mathcal{F}_t = \sigma\left(\mathbf{1}_{[0,s]}(p_i) : s \le t, 1 \le i \le n\right)$

Duality			
	p_i	$\longleftrightarrow$	$1 - p_i$
	t	$\longleftrightarrow$	$1 - t$
	$\mathcal{F}_t$	$\longleftrightarrow$	$\mathcal{G}_t$
	martingale	$\longleftrightarrow$	reverse martingale

Reverse martingale (Heesen/J.)

$$t \mapsto \frac{\sum_{i \in I_0} \mathbf{1}_{[0,t]}(p_i)}{t} \qquad p_i \text{ uniform i.i.d.}$$

Martingale conditions (Benditkis (2015), Dissertation)

$$t \mapsto \frac{\sum_{i \in I_0} \mathbf{1}_{[0,t]}(p_i) - t}{1 - t} \qquad \mathcal{F}_t\text{-supermartingale,}$$

where $\mathcal{F}_t = \sigma\left(\mathbf{1}_{[0,s]}(p_i) : s \le t, 1 \le i \le n\right)$

$$\begin{array}{lll} \text{Duality} & p_i & \longleftrightarrow 1 - p_i \\ & t & \longleftrightarrow 1 - t \\ & \mathcal{F}_t & \longleftrightarrow \mathcal{G}_t \\ & \text{martingale} & \longleftrightarrow \text{ reverse martingale} \end{array}$$

Appendix: Example of martingale models

- $U_1, \ldots, U_k$ i.i.d. uniform on $[0, 1]$ k-blocks of equal p-values

$$\begin{aligned} (p_1, \ldots, p_{n_1}) &= (U_1, \ldots, U_1) \\ \left(p_{n_1+1}, \ldots, \ldots, p_{n_1+n_2}\right) &= (U_2, \ldots, U_2) \quad \ldots \end{aligned}$$

- optional switching of martingales $(p_i)_{i \in I_0}$, $(\widetilde{p}_i)_{i \in I_0}$
  martingale dependent
  $\tau \in [0, 1)$ stopping time

$$p_i' = \begin{cases} p_i & \text{if } \mathbf{1}_{[0,\tau]}(p_i) = 1 \\ \widetilde{p}_i & \text{if } \mathbf{1}_{[0,\tau]}(\widetilde{p}_i) = 0 \end{cases}$$

Appendix: Example of martingale models

- $U_1, \ldots, U_k$ i.i.d. uniform on $[0,1]$ k-blocks of equal p-values

$$\begin{aligned} (p_1, \ldots, p_{n_1}) &= (U_1, \ldots, U_1) \\ \left(p_{n_1+1}, \ldots, \ldots, p_{n_1+n_2}\right) &= (U_2, \ldots, U_2) \quad \ldots \end{aligned}$$

- optional switching of martingales $(p_i)_{i \in I_0}$, $(\widetilde{p}_i)_{i \in I_0}$
  martingale dependent
  $\tau \in [0,1)$ stopping time

$$p'_i = \begin{cases} p_i & \text{if } \mathbf{1}_{[0,\tau]}(p_i) = 1 \\ \widetilde{p}_i & \text{if } \mathbf{1}_{[0,\tau]}(\widetilde{p}_i) = 0 \end{cases}$$

- $\{P$ on $[0,1]^{m_0}$: martingale measure $\}$ closed under convex combinations
- connection to martingale measures in mathematical finance
- Marshall/Olkin type dependence
  $X_1, \ldots, X_n$ i.i.d., Y independent of X's
  $Z_i = \min(X_i, Y)$ with continuous distribution function H

  $p_i := H(Z_i)$ martingale property

- $\{P$ on $[0,1]^{m_0}$: martingale measure $\}$ closed under convex combinations
- connection to martingale measures in mathematical finance
- Marshall/Olkin type dependence
  $X_1, \ldots, X_n$ i.i.d., Y independent of X's
  $Z_i = \min(X_i, Y)$ with continuous distribution function H

  $p_i := H(Z_i)$ martingale property