Vergleich von Methoden zur Schätzung des Anteils wahrer Hypothesen bei multiplen Testverfahren

M. Walther, C. Hemmelmann, R. Vollandt, FSU Jena*

*GMDS-Jahrestagung 2006, Leipzig, 10.09.2006 - 14.09.2006

1. Einführung und Motivation

Ausgangspunkt:

simultanes Testen von m Nullhypothesen $H_1, \ldots, H_m$.

mögliche Situationen (Benjamini & Hochberg, 1995):

nicht bekannt

	nicht abgelehnt	abgelehnt	Σ
wahre Nullhypothesen	U	V	m_0
falsche Nullhypothesen	T	S	$m - m_0$
Σ	$m - R$ beobachtbar	R	m

FDR – Verfahren von Benjamini & Hochberg (BH)

- $p_{(1)} \leq \ldots \leq p_{(m)}$ geordnete Realisierungen der m unabhängigen p-Werte $P_1, \ldots, P_m$ für $H_1, \ldots, H_m$
- $k^* := max\left\{k : p_{(k)} \leq \alpha\, k/m\right\}$ für ein gegebenes $\alpha > 0$
- lehne alle Nullhypothesen H_i mit $p_i \leq p_{k^*}$ ab

➡

$$FDR := E\left\{(V/R)\, I(R>0)\right\} \leq \frac{m_0}{m}\alpha$$

Kennt man m_0 und wendet das Verfahren BH zum Niveau $\alpha' = m\alpha / m_0$ an, so gilt: $FDR \leq \alpha$

Ziel: Evaluation und Entwicklung von Schätzverfahren, für m_0 bzw. $\pi_0 := m_0/m$. Dabei sollte m_0 bzw. π_0 überschätzt und möglichst „gut" geschätzt werden.

Motivation:

- Festlegung oder Kontrolle multipler Fehlerraten, wie z.B. *FWE* oder *FDR;* Schätzung der *FDR*
- Schätzung des Anteils von Genen, welche (nicht) differentiell expremiert sind
- Proteomik, fMRT, Astrophysik
- m_0 bzw. π_0 sind Größen, die für sich allein betrachtet von großem Interesse sind

2. Ausgewählte Schätzverfahren

grundlegendes Modell:

$P_1, \ldots, P_m$ unabhängig und identisch verteilt mit Dichte

$$f(p) = \pi_0 + (1 - \pi_0) h(p), \quad 0 \le p \le 1.$$

Annahmen:

- P_i ist auf [0,1] gleichverteilt, falls H_i wahr ist
- P_i besitzt Dichte h, falls Alternativhypothese zu H_i wahr ist
- $P(H_i \text{ ist wahr}) = \pi_0$

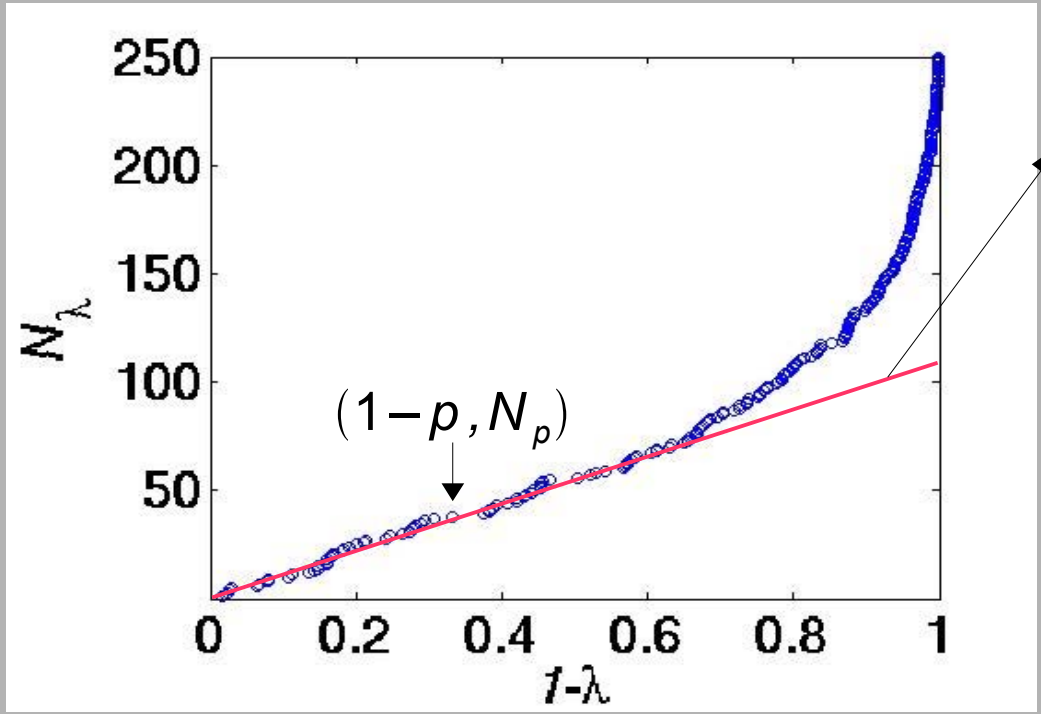
Verfahren basieren im wesentlichen auf Realisierungen $p_1,\ldots,p_m$ der p-Werte $P_1,\ldots,P_m$

- **Anstiegsmethode** (Schweder & Spjøtvoll, 1982)

$$N_\lambda := \#\{i : p_i > \lambda\}, \quad 0 < \lambda < 1$$

Annahme: *p*-Wert ist klein, falls Nullhypothese falsch

➡ $E(N_\lambda) \approx m_0(1-\lambda)$, sofern λ nicht zu klein



Anstieg ≡ Schätzung für m_0 bzw. $\pi_0 = m_0 / m$

$\Rightarrow \pi_0^{Anstieg}$

- **Bootstrap-Verfahren** (Storey, 2002)

Idee: Gleichverteilungsannahme $\Rightarrow$ im Intervall $(\lambda, 1)$ liegen ca. $m_0 \cdot (1-\lambda) = N_\lambda$ p-Werte $(0 < \lambda < 1)$

Ansatz:

wähle $\pi_0(\lambda) := \dfrac{N_\lambda}{m(1-\lambda)}$ als Schätzung für π_0

Wahl von λ mittels Bootstrapping

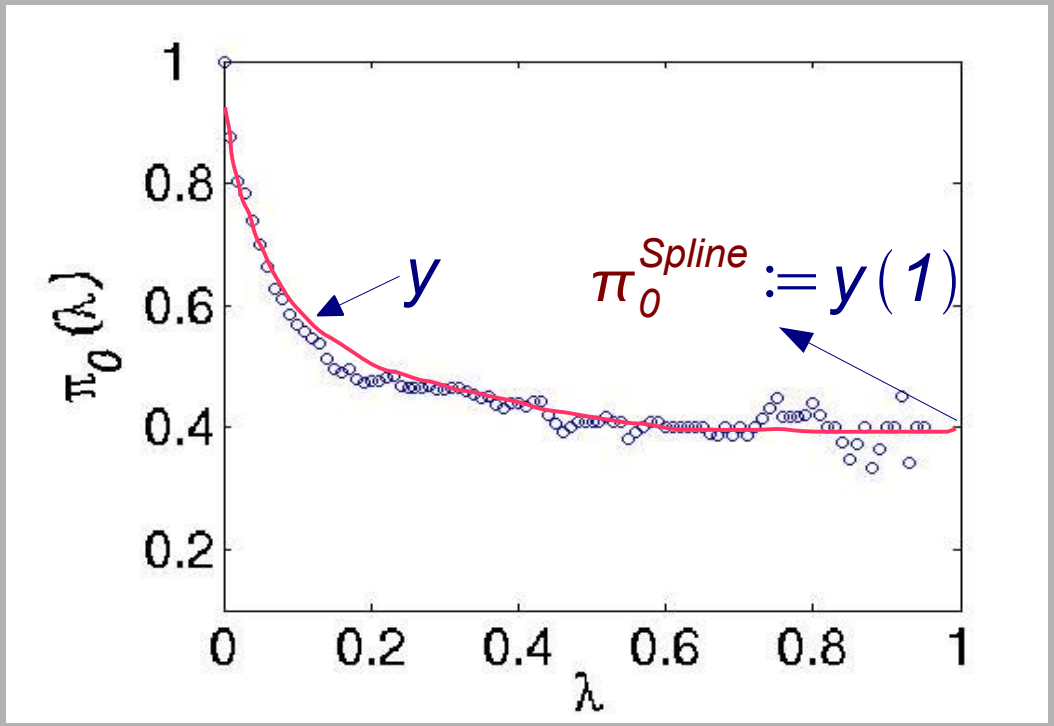
wähle $\pi_0^{boot} := \pi_0(\lambda^*)$ als Schätzung für π_0, wobei

$$\lambda^* = \underset{0 \le \lambda \le 1}{argmin} \frac{1}{K} \sum_{k=1}^{K} \left\{ \pi_0^k(\lambda) - \underset{0 \le \lambda \le 1}{min} \left\{ \pi_0(\lambda) \right\} \right\}^2$$

- **kubische Spline** (Storey & Tibshirani, 2003)

  oft gilt: $\lim_{\lambda \to 1} bias(\pi_0(\lambda)) = 0$

  Idee: schätze $\lim_{\lambda \to 1} \pi_0(\lambda)$ mittels gewichteter kubischer Spline



- **konvexe & monoton fallende Dichtefunktion** (Langaas, Lindqvist & Ferkingstad, 2005)

  Ziel: Schätzung von f durch monoton fallende & konvexe Dichtefunktion auf [0,1] (NPMLS)

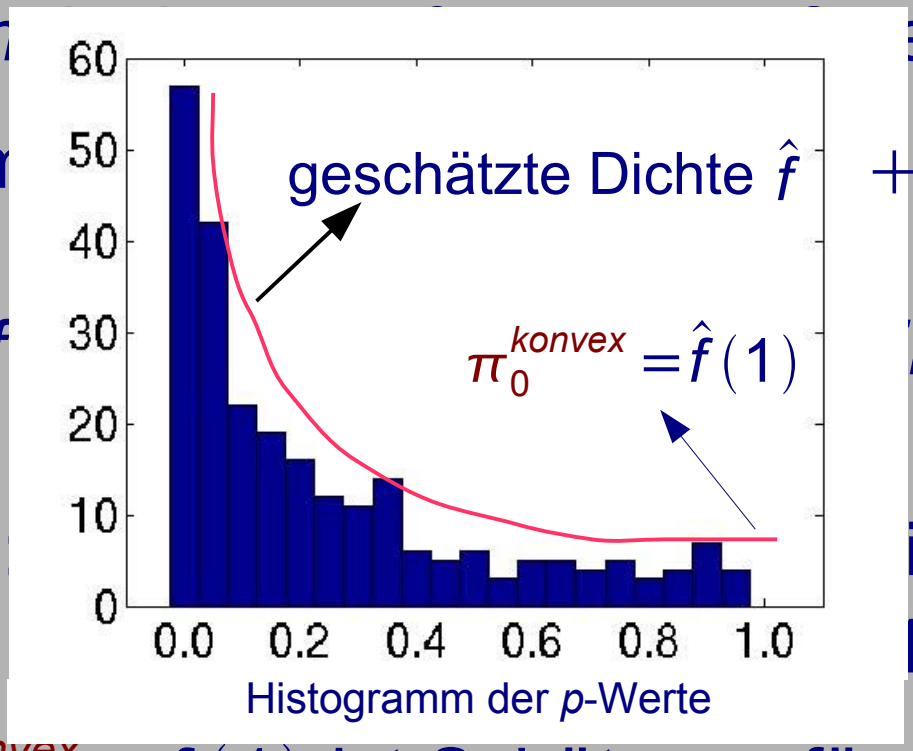
  Annahme: h … end auf [0,1]

  m … $+(1-\pi_0)h(1)$

  Idee: $\varphi(f$ … n

  NB: … ierbar, konvex …]

  ➡ $\pi_0^{konvex} := f(1)$ ist Schätzung für π_0



- **momentenerzeugende Funktion** (Broberg, 2004)

basiert auf momentenerzeugende Funktion der *p*-Werte

$$R(s)=E(e^{sP})=\int_0^1 e^{sx} f(x)\,dx$$

$$=\pi_0\frac{e^s-1}{s}+(1-\pi_0)\int_0^1 e^{sx} h(x)\,dx$$

$$=\pi_0 g(s)+(1-\pi_0)R_1(s), \qquad 0<s\le 1$$

unbekannt

$$\Rightarrow \qquad \pi_0 = \frac{R(s)-R_1(s)}{g(s)-R_1(s)}, \qquad 0<s\le 1 \qquad \textbf{(1)}$$

1. aus den p-Werten $p_1, \ldots, p_m$ schätze man R durch

$$R(s) = \frac{1}{m} \sum_{k=1}^{m} e^{sp_k} \qquad \textbf{(2)}$$

2. aus **(1)** erhält man für $0 < s_{n-1} < s_n \leq 1$ die Rekursion

$$R_1(s_n) = \frac{R(s_n) g(s_{n-1}) - R(s_{n-1}) g(s_n) + R_1(s_{n-1})(g(s_n) - R(s_n))}{g(s_{n-1}) - R(s_{n-1})} \qquad \textbf{(3)}$$

mit Startwert $R_1(s_1) = \frac{1 + R(s_1)}{2}$

Idee: Bestimme unter Verwendung von **(2)** und **(3)** den Quotienten **(1)** für $s = (0.01, 0.0101, 0.0102, \ldots, 1)$

➡ $\pi_0^{Moment} :=$ arithm. Mittel von **(1)** liefert Schätzer für π_0

- **weitere Schätzverfahren:** (nicht ausschöpfend)
  - BUM-Verfahren (Allison et al., 2002; Pounds & Morris, 2003)
  - LBE-Verfahren (Dalmasso, Broët & Moreau, 2004)
  - SEP-Verfahren (Scheid & Spang, 2004)
  - Permutationsverfahren (Meinshausen & Bühlmann, 2005)
  - BF-Verfahren (Meinshausen & Rice, 2006)
  - PRE-Verfahren (Efron, 2004)
  - SPLOSH-Verfahren (Pounds & Cheng, 2004)
  - Histogramm-Methode (Nettleton, Hwang et al., 2006)

3. Simulationsstudie

Erzeugen *n* *m*-dimensionale normalverteilte Zufallsvektoren

$$x_j \sim N(\mu, \Sigma) \quad (j = 1, \ldots, m)$$

mit Erwartungswertvektor

$$\mu_1 = \ldots = \mu_{m-m_0} = \Delta \quad \text{und} \quad \mu_{m-m_0+1} = \ldots = \mu_m = 0$$

und Kovarianzmatrix

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{21} & 1 & \cdots & \rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & 1 \end{pmatrix}$$

- n = 13 (Stichprobenumfang, Patienten oder Proben)
- m = 250 (Anzahl der Hypothesen bzw. Dimension)
- π_0 = **0.2**, 0.5, **0.8** (Anteil wahrer Hypothesen)
- Δ = **0.5**, 1, 2
- Korrelationsstruktur:
  - konstant: $\rho_{jk} = \rho \quad (j \neq k)$ mit $\rho \in \{0, 0.2, 0.5, 0.8\}$
  - gemischt:

$$\Sigma^{R1R2} = \begin{pmatrix} R1 & R2 & R2 \\ R2 & R1 & R2 \\ R2 & R2 & R1 \end{pmatrix}; \quad R1 = \begin{pmatrix} 1 & 2/3 & \cdots & 2/3 \\ 2/3 & 1 & \cdots & 2/3 \\ \vdots & \vdots & \ddots & \vdots \\ 2/3 & 2/3 & \cdots & 1 \end{pmatrix}, \quad R2 = (-1) \begin{pmatrix} 1/3 & 1/3 & \cdots & 1/3 \\ 1/3 & 1/3 & \cdots & 1/3 \\ \vdots & \vdots & \ddots & \vdots \\ 1/3 & 1/3 & \cdots & 1/3 \end{pmatrix}$$

- p-Werte basieren auf Einstichproben-t-Test
- Anzahl Simulationen: 10000

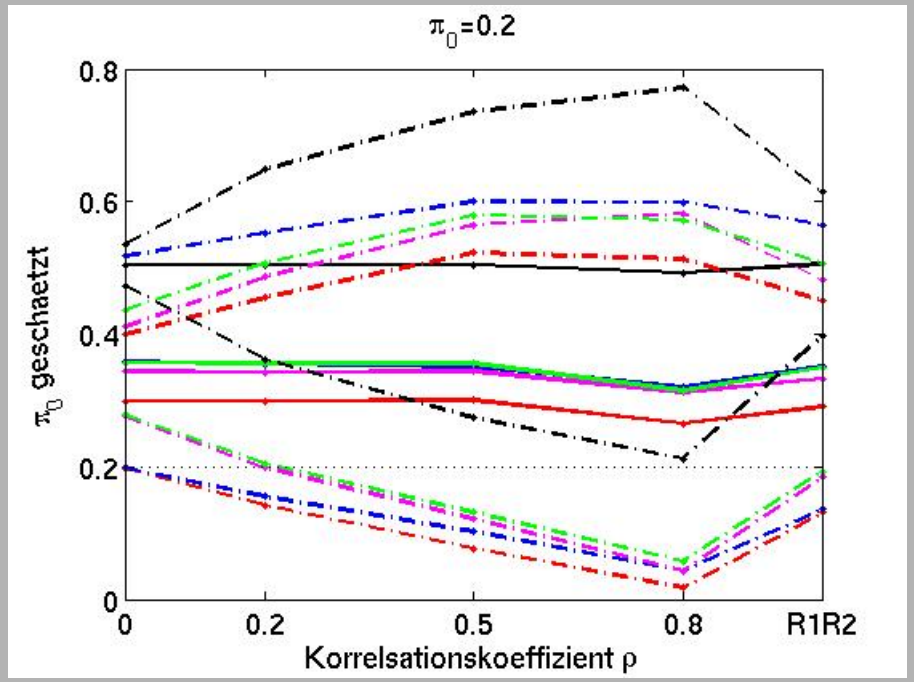
Ergebnisse

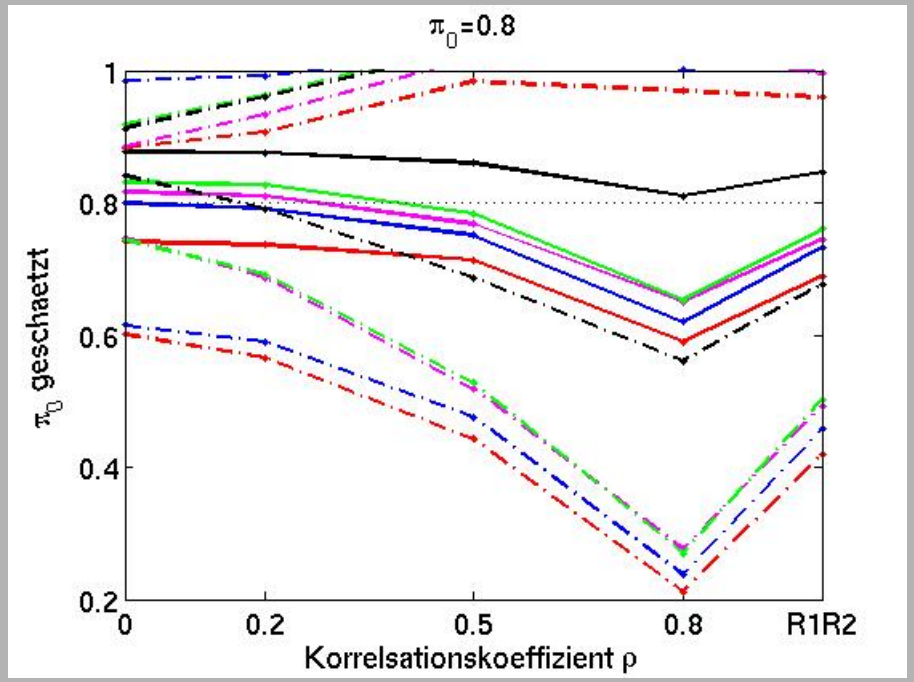
abhängig vom gewählten Verfahren wird für jede Simulation der Paramter π_0 geschätzt $\Longrightarrow$ $\hat{\pi}_0^{(k)}$

- Mittelwert und Standardabweichung
- $$MSE(\hat{\pi}_0) = \underbrace{\frac{1}{m-1}\sum_{k=1}^{m}(\hat{\pi}_0^{(k)} - \hat{\pi}_{0.})^2}_{Var(\hat{\pi}_0)} + \underbrace{(\hat{\pi}_{0.} - \pi_0)^2}_{Bias(\hat{\pi}_0)^2}$$

  arithm. Mittel
- Boxplots
- $P(\pi_0 \le \hat{\pi}_0 \le \pi_0 + \gamma(1-\pi_0)), \quad \gamma = 0.2,\ 1$

Mittelwert (—), Standardabweichung (– ·) & MSE: $\Delta = 0.5$

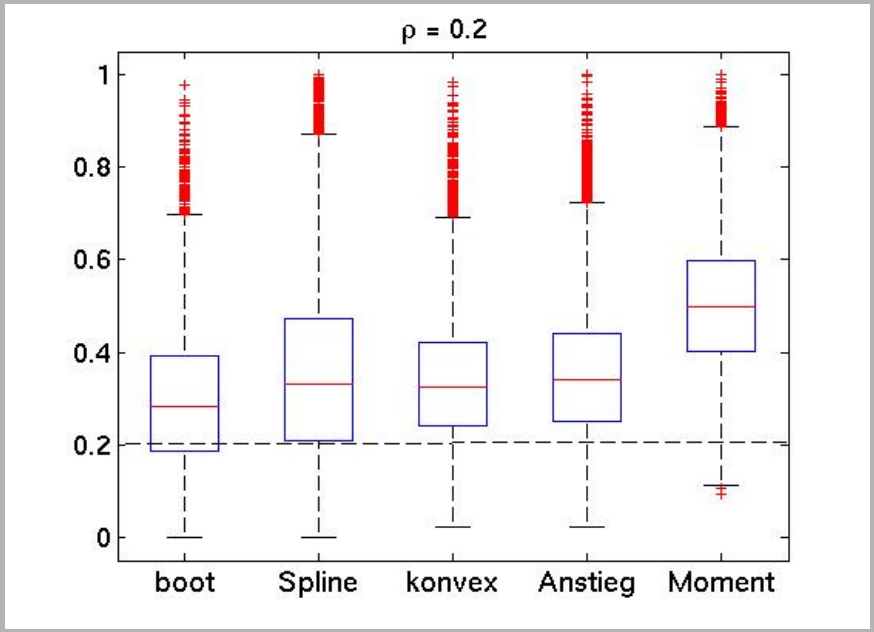


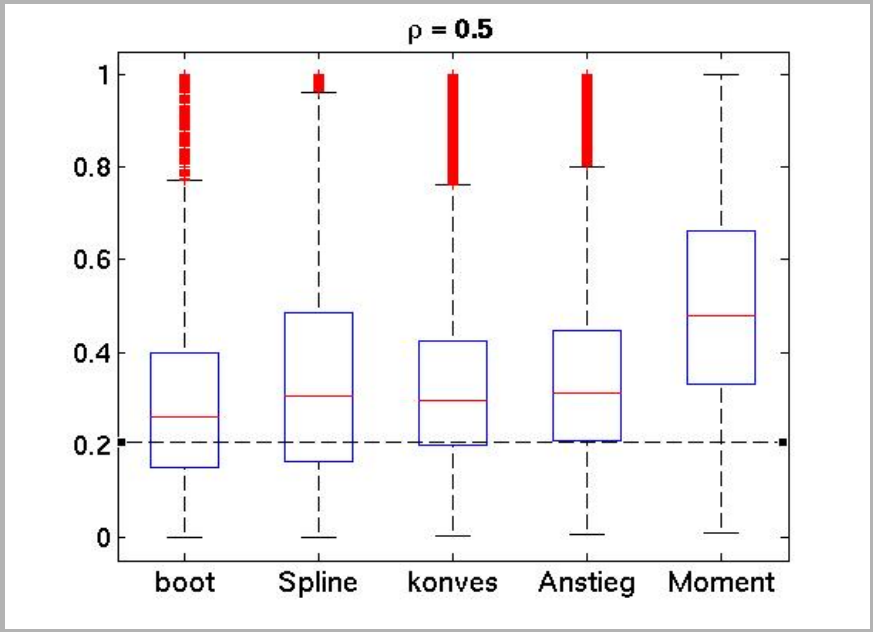


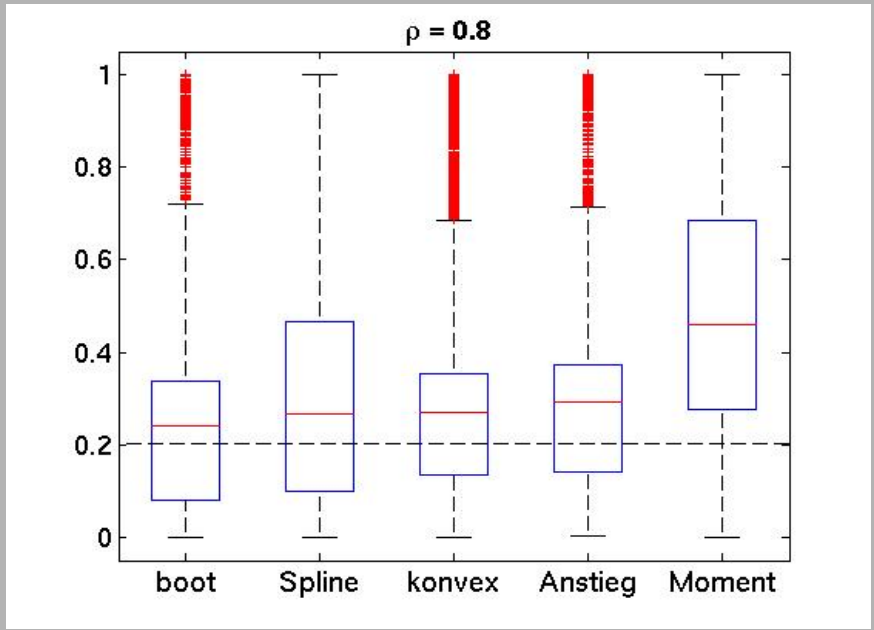
MSE	ρ = 0	ρ = 0.2	ρ = 0.5	ρ = 0.8	ρ = R1R2
boot	0.0199	0.0341	0.0596	0.0658	0.0337
Spline	0.0503	0.0634	0.0847	0.0919	0.0685
konvex	0.0252	0.0411	0.0698	0.0853	0.0399
Anstieg	0.0311	0.0472	0.0746	0.0796	0.0469
Moment	0.0934	0.1135	0.1457	0.1637	0.1051

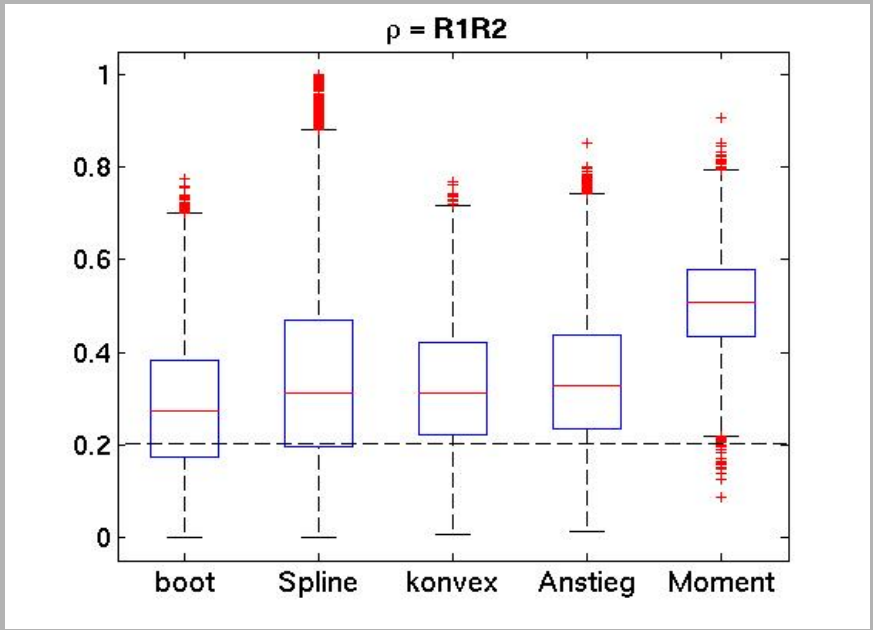
MSE	ρ = 0	ρ = 0.2	ρ = 0.5	ρ = 0.8	ρ = R1R2
boot	0.0233	0.0333	0.0807	0.1879	0.0847
Spline	0.0343	0.4050	0.0782	0.1787	0.0803
konvex	0.0050	0.0155	0.0634	0.1625	0.0662
Anstieg	0.0087	0.0189	0.6610	0.1682	0.0667
Moment	0.0072	0.0128	0.3360	0.0625	0.0311

Boxplots: $\pi_0 = 0.2$

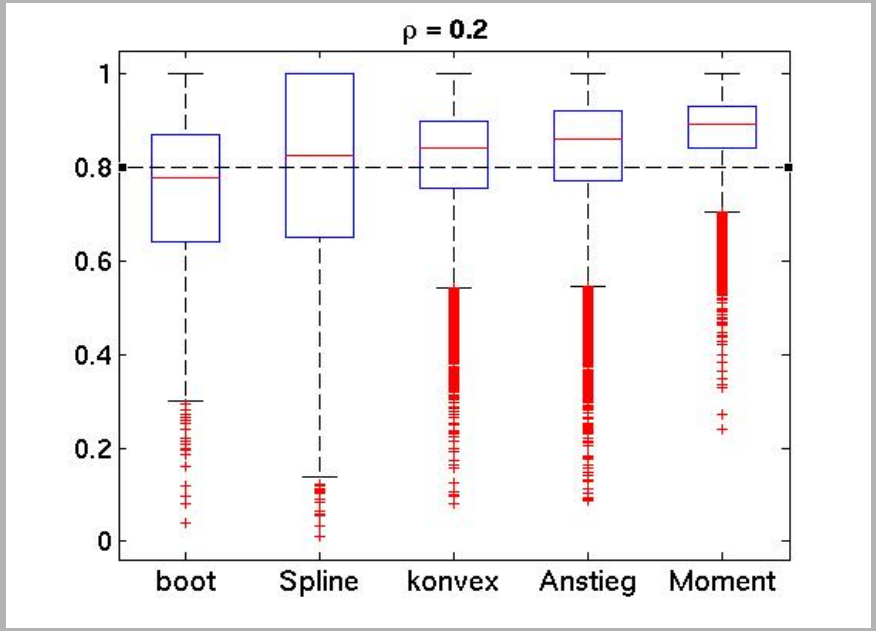


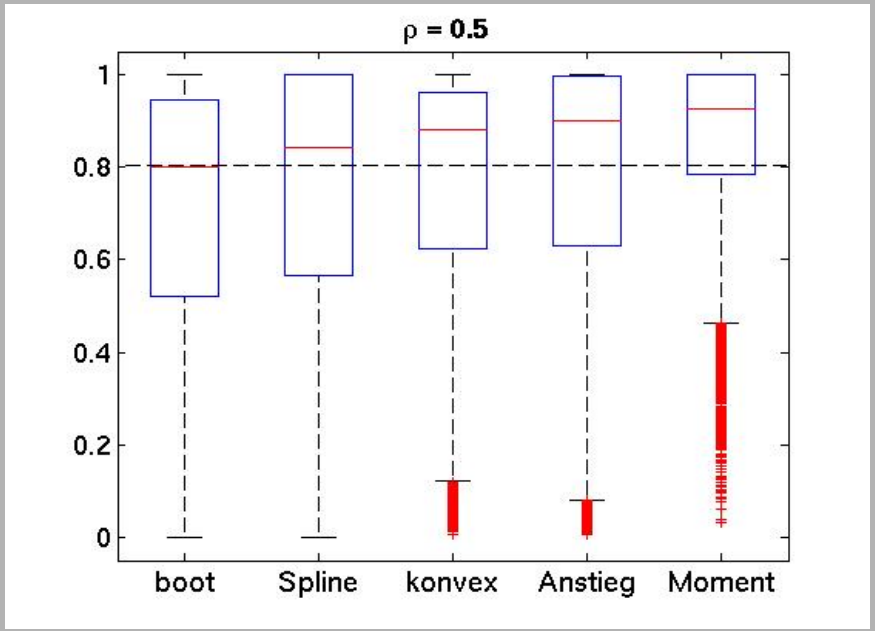


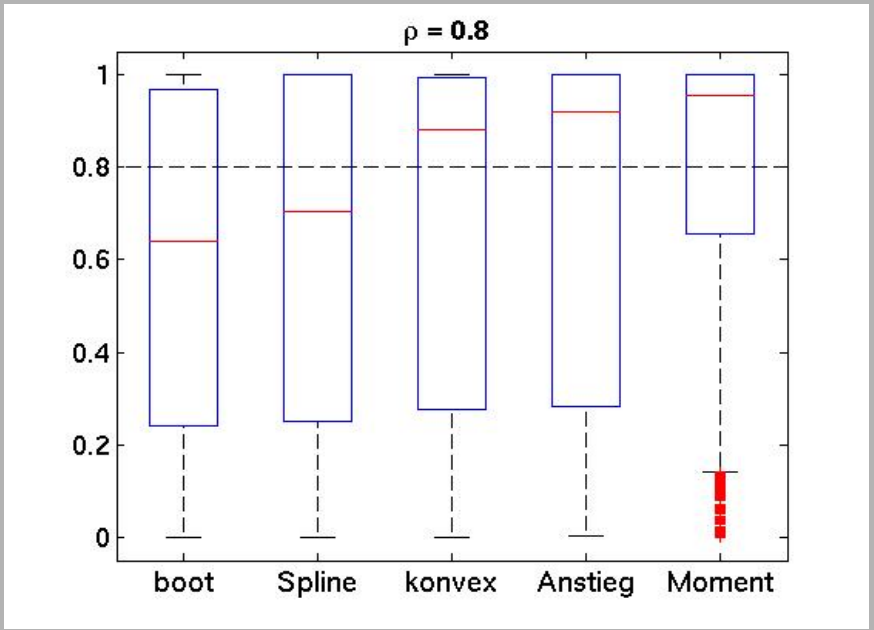


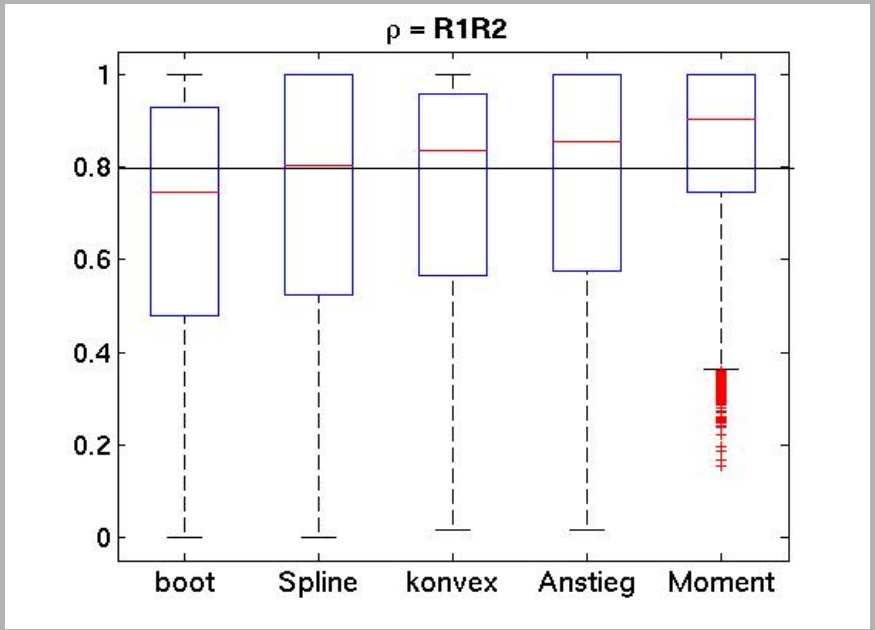


Boxplots: $\pi_0 = 0.8$



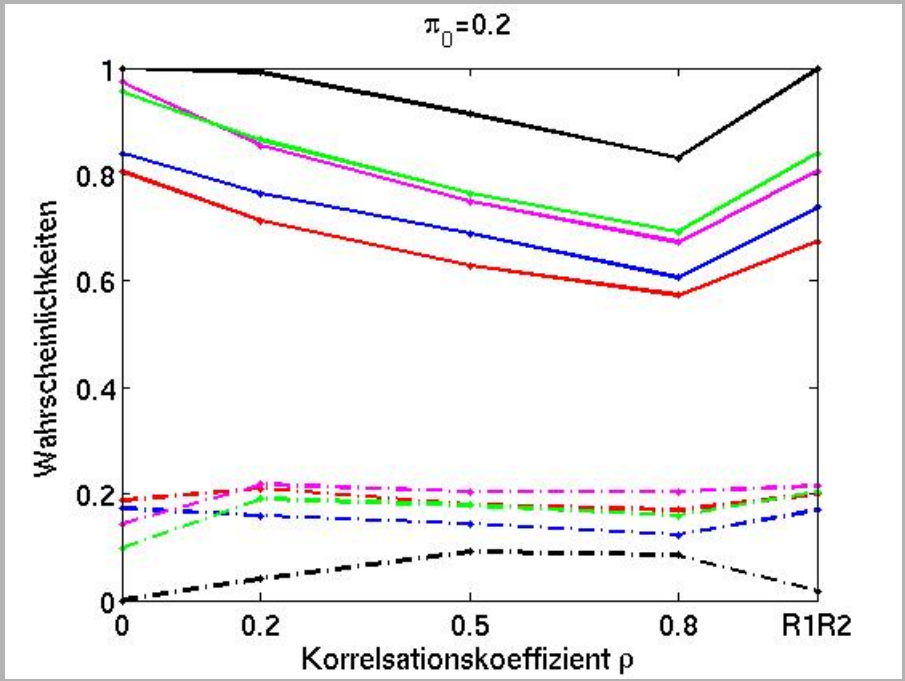


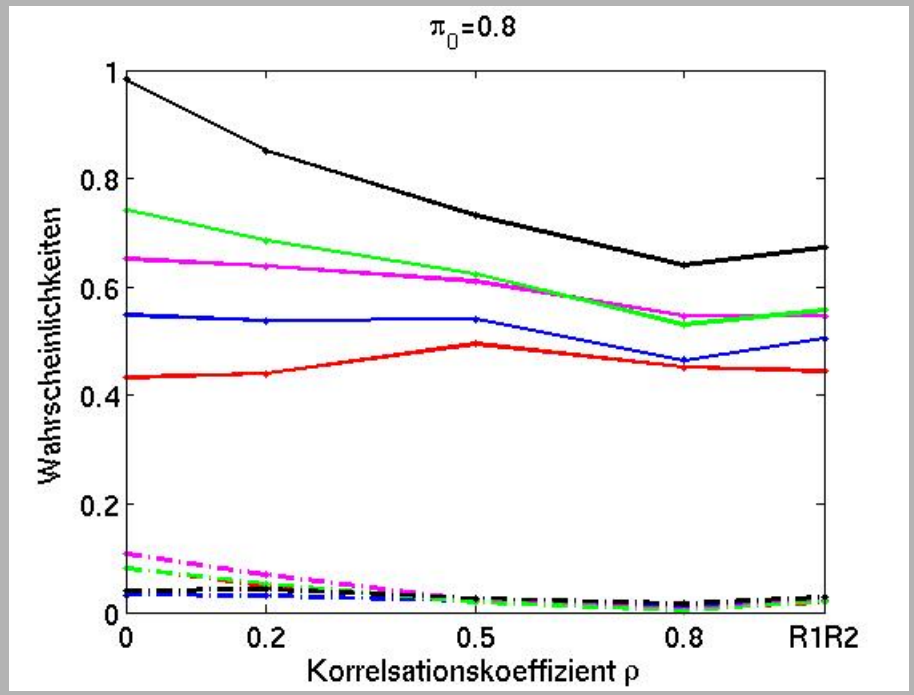




Wahrscheinlichkeiten:

boot – Spline – konvex – Anstieg – Moment





—— $P(\pi_0 \leq \hat{\pi}_0)$

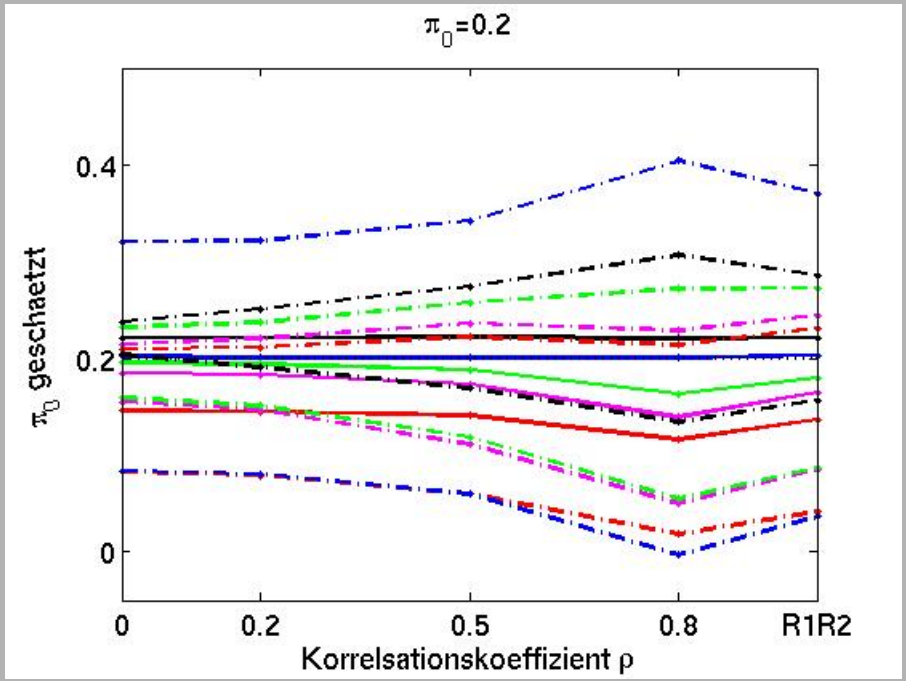
—• $P(\pi_0 \leq \hat{\pi}_0 \leq \pi_0 + 0.2(1-\pi_0))$

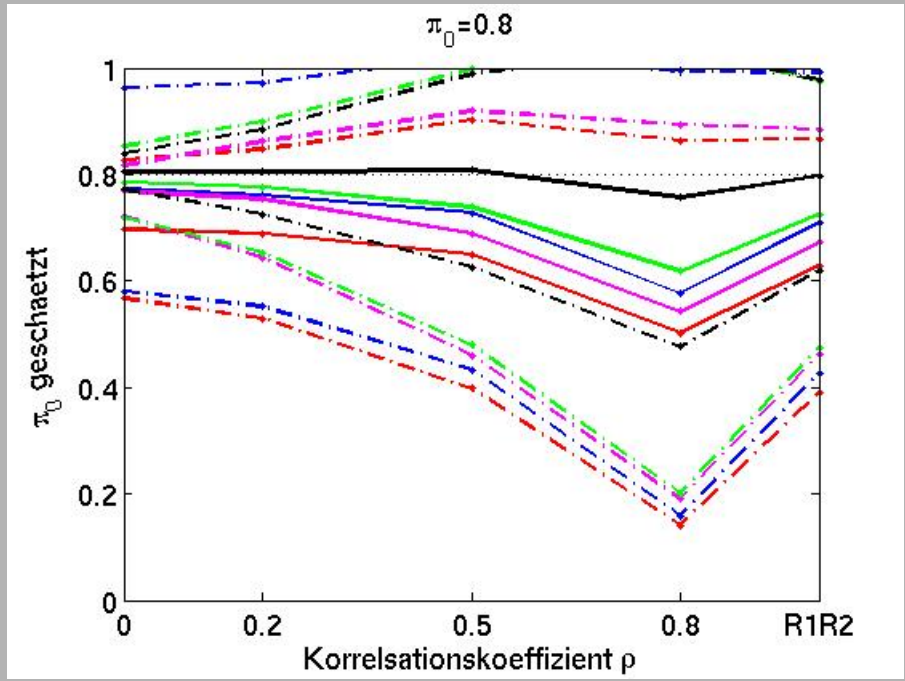
4. Fazit

- vorgestellten Verfahren liefern für kleine ρ „gute" Schätzungen für π_0 ; sofern der Anteil wahrer Hypothesen nicht zu groß ist
- Verfahren tendieren wahres π_0 zu unterschätzen
- einheitliches bzw. eigenständiges Gütekriterium für Schätzung noch offen
- Ausweg: Schätzung von π_0 nicht separat betrachten, sondern im Hinblick auf das verfolgte Ziel (z.B. *FDR*-Verfahren)

Vielen Dank für Ihre Aufmerksamkeit !

Mittelwert (—), Standardabweichung (–·) & MSE: $\Delta=1$

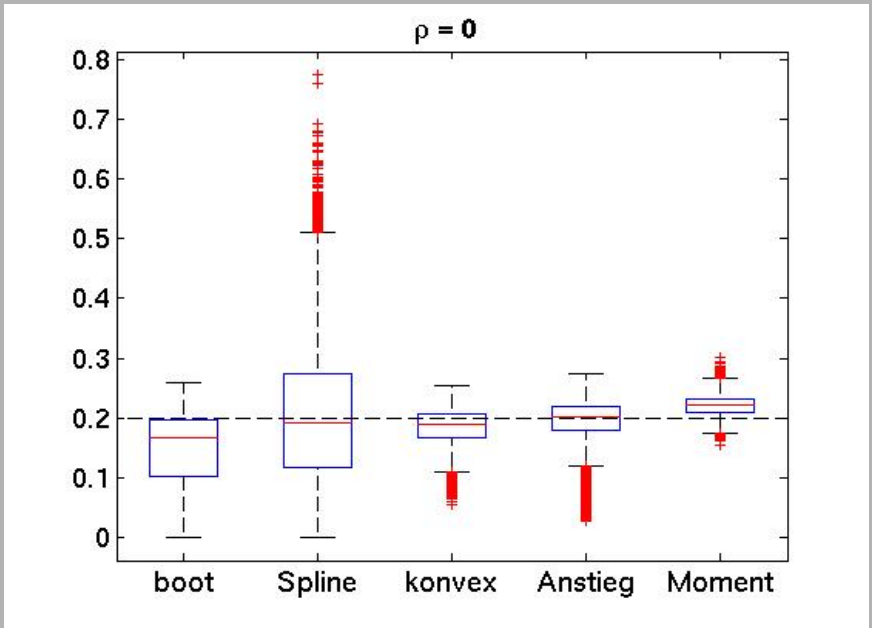


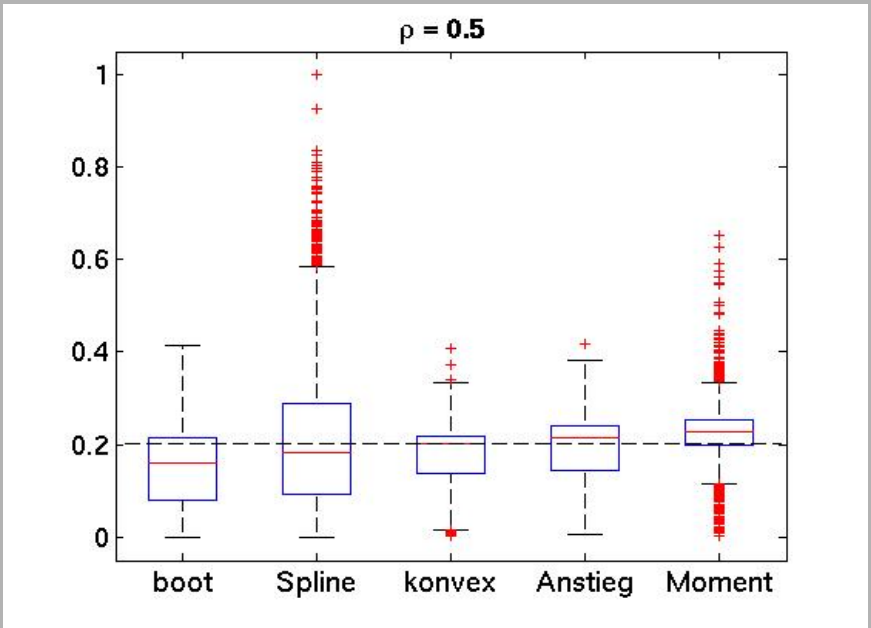


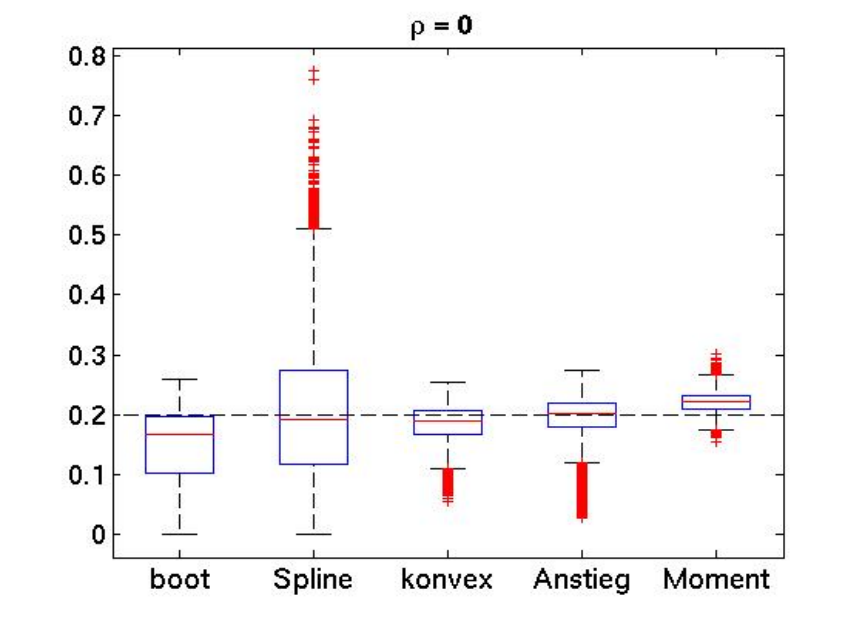
MSE	ρ = 0	ρ = 0.2	ρ = 0.5	ρ = 0.8	ρ = R1R2
boot	0.0069	0.0074	0.0100	0.0166	0.0131
Spline	0.0140	0.0145	0.0199	0.0415	0.0278
konvex	0.0011	0.0016	0.0046	0.0177	0.0076
Anstieg	0.0014	0.0019	0.0050	0.0131	0.0090
Moment	0.0007	0.0013	0.0032	0.0079	0.0046

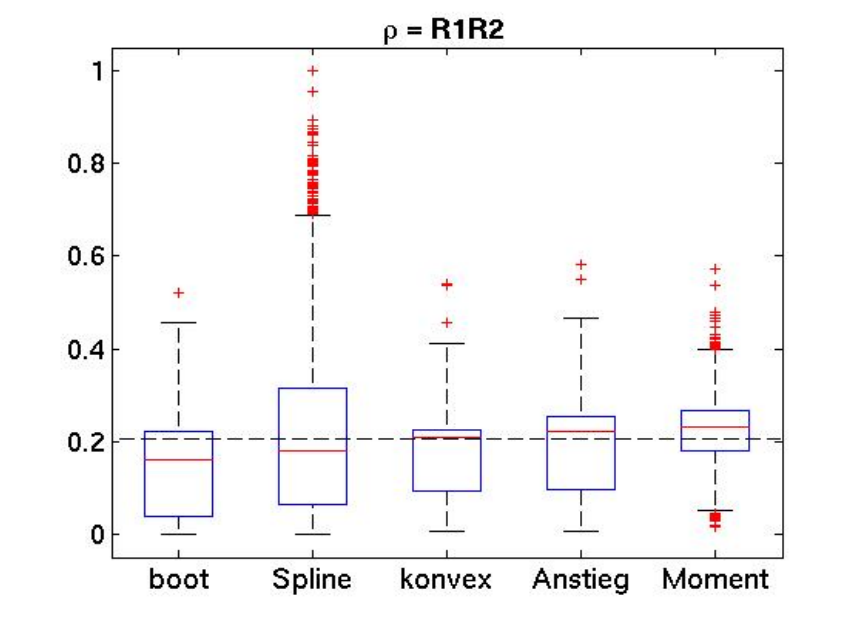
MSE	ρ = 0	ρ = 0.2	ρ = 0.5	ρ = 0.8	ρ = R1R2
boot	0.0274	0.0376	0.0859	0.2189	0.0863
Spline	0.0373	0.0454	0.0919	0.2246	0.0879
konvex	0.0032	0.0141	0.6510	0.1902	0.0610
Anstieg	0.0047	0.0157	0.0709	0.2070	0.0681
Moment	0.0012	0.0064	0.3290	0.0809	0.0318

Boxplots: $\pi_0 = 0.2$

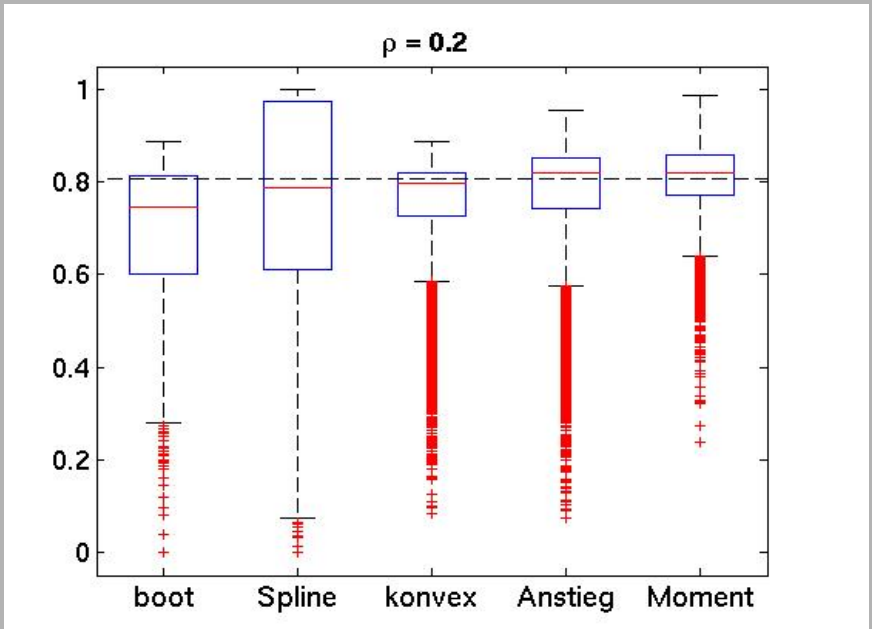


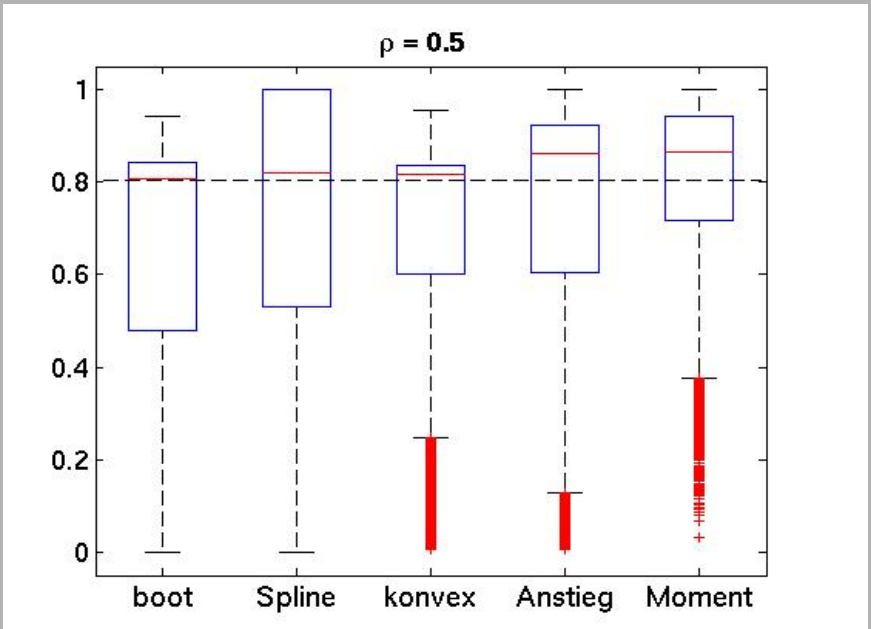


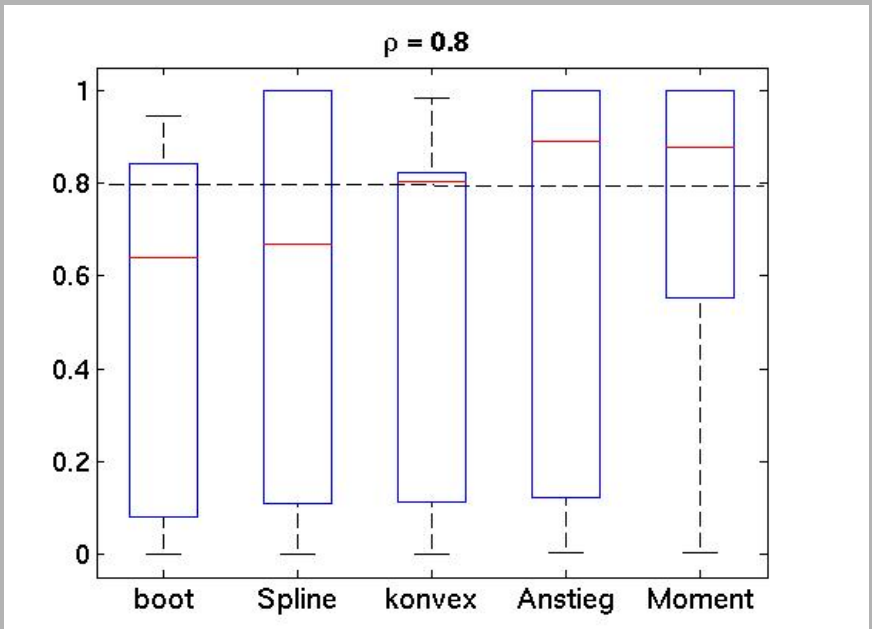


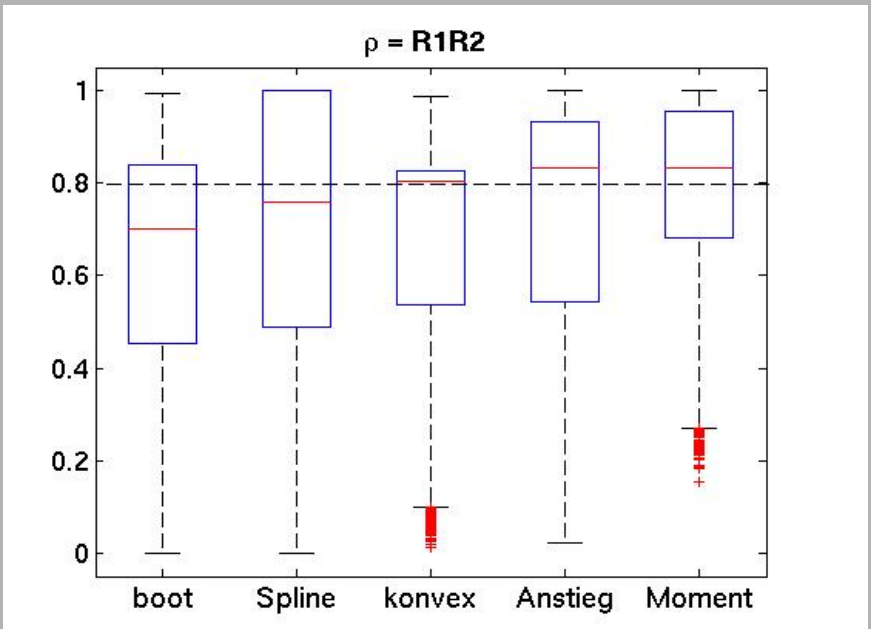


Boxplots: $\pi_0 = 0.8$



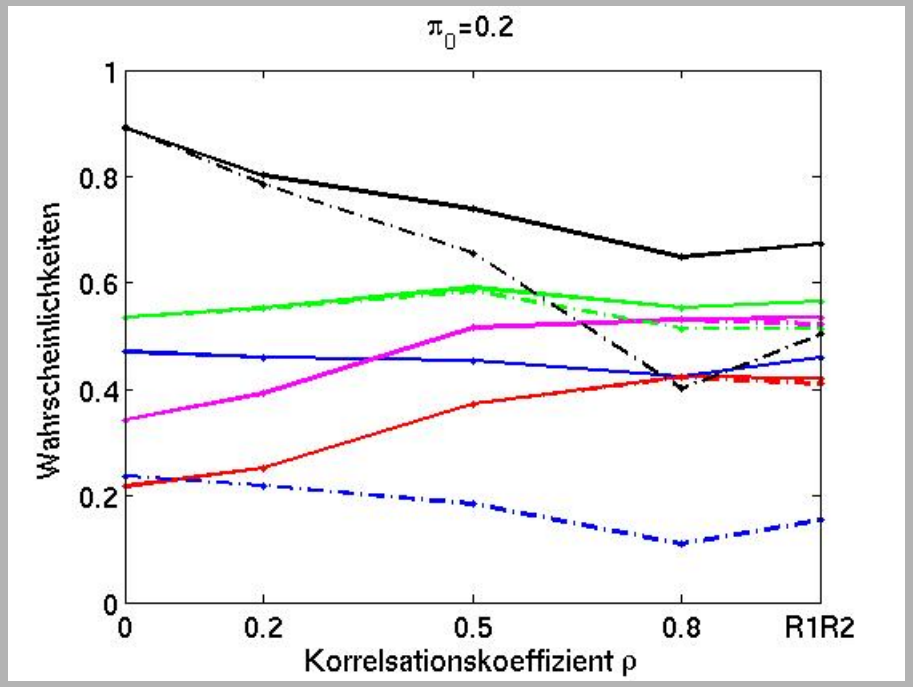


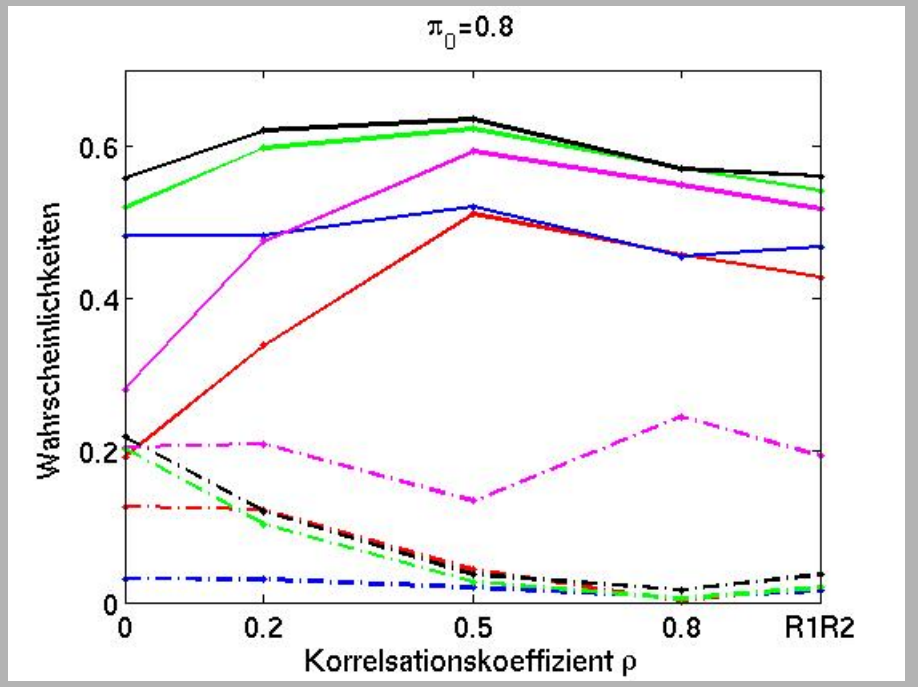




Wahrscheinlichkeiten:

boot – Spline – konvex – Anstieg – Moment





—— $P(\pi_0 \leq \hat{\pi}_0)$

—• $P(\pi_0 \leq \hat{\pi}_0 \leq \pi_0 + 0.2(1-\pi_0))$

Referenzen:

Allison, D.B., Gadbury, G.L., Heo, M., Fernandez, J.R., Lee, C.K., Prolla, T.A., Weindruch, R. (2002). A mixture model approach for the analysis of microarray gene expression data. *Computational Statistics and Data Analysis,* **39**1–20

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B,* **57** (1) 289-300

Benjamini, Y., Krieger, A.M., Yekutieli, D. (2005). Adaptive linear step-up procedures that control the false discovery rate, Department of Statistics and OR. Tel Aviv University, Tel Aviv, 2004

Broberg, P. (2004). A new estimate of the proportion unchanged genes in a microarray experiment, *Genome Biology,* **5**:p10

Dalmasso, C., Broët, P., Moreau, T. (2005). A simple procedure for estimating the fals discovery rate. *Bioninformatics,* **21** (5) 660-668

Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association,* **99** 96-104

Langaas, M., Lindqvist, B.H., Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data, *Journal of the Royal Statistical Society, Series B,* **67** (4) 555-572

Meinshausen, N., Bühlmann, P. (2005). Lower bounds for the number of false null hypotheses for multiple testing in associations under general dependence structures. *Biometrica,* **92** 893-907

Meinshausen, N., Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *The Annals of Statistics,* **34** (1) 373-393

Nettleton, D., Hwang, G.J.T., Caldo, R.A., Wise, R.P. (2006). Estimating the number of true null hypotheses from a histogram of *p*-values. *Journal of Agricultural, Biological and Environmental Statistics. In press.*

Pounds, S., Cheng, C. (2004). Improving false discovery rate estimation. *Bioinformatics,* **20** (11) 1737-1745

Pounds, S,. Morris, S.W. (2004) Estimating the occurence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distripution of *p*-palues. *Bioinformatics,* **19** (10) 1236-1242

Scheid, S., Spang, R. (2004). A stochastic downhill search algorithm for estimating the local false discovery rate. *IEEE Transactions on Computational Biology and Bioinformatics,* **1** 98-108

Schweder, T., Spjøtvoll, E. (1982). Plots of *p*-values to evaluate many tests simultaneously, *Biometrika*, **69** (3) 493-502

Storey, J.D. (2002). A direct approach to false discovery rates, *Journal of the Royal Statistical Society, Series B,* **65** (3) 479-498

Storey, J.D., Tibshirani, R. (2001). Estimating false discovery rate unter dependence, with applications to DNA mircoarray. *Technical Report 2001-2017.* Department of Statistics, Stanford University, Stanford

Storey, J.D., Tibshirani, R. (2003). Statistical significance for genomwide studies. *Proceedings of the National Academy of Sciences of the United States of America,* **100** (16) 9440-9445

Inhalt

1. Einführung und Motivation
2. Ausgewählte Schätzverfahren
3. Simulationsstudie
4. Fazit