

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

Walther M, Hemmelmann C, Vollandt R

Institut für Medizinische Statistik, Informatik und Dokumentation, Friedrich-Schiller-Universität Jena, Deutschland
walmar@imsid.uni-jena.de

Einleitung

Ausgangspunkt unserer Betrachtungen sind multiple Testverfahren. In der Literatur findet man eine Vielzahl von Verfahren, die die FWE (familywise error rate) oder die FDR (false discovery rate) kontrollieren. Neben diesen klassischen Fehlerarten ist man bemüht neuere Verfahren zu entwickeln, die die so genannten gFWE (generalized familywise error rate) oder FDP (false discovery proportion) kontrollieren. Von mehreren dieser Verfahren ist aber bekannt, dass die kontrollierte Fehlerrate die vorgegebene Signifikanzschranke α gar nicht ausschöpfen und sogar durch $\pi_0\alpha$ begrenzt wird, wenn π_0 den Anteil der wahren Hypothesen bezeichnet (vgl. [1]). Die Power solcher Verfahren ließe sich also erhöhen, wenn π_0 bekannt wäre, insbesondere bei kleinem π_0 . Bei der Anwendung solcher Testverfahren ist es daher von Vorteil, vorher oder in einem ersten Schritt den Anteil der wahren Hypothesen zu schätzen oder zumindest nach oben zu begrenzen. Verschiedene derartige Vorgehensweisen existieren bereits in der Literatur (vgl. [2], [5], [6]). Diese und auch andere Methoden sollen verglichen werden. Es soll herausgefunden werden, in welchen Bereichen welche Schätzer am zuverlässigsten sind.

Methoden

Wir betrachten verschiedene Methoden zur Schätzung des Anteils π_0 wahrer Hypothesen in einem multiplen Testproblem, welche auf der Grundlage der p-Werte basieren. Dabei geht man stets von der Annahme aus, dass die p-Werte unabhängig und unter H_0 identisch verteilt sind. Eine dieser Methoden ist die von Storey (vgl. [5]). Diese beruht auf der Berechnung der relativen Anzahl an p-Werten, die oberhalb einer Schranke λ mit $0 < \lambda < 1$ liegen. Genauer bestimmt man für jedes $0 < \lambda < 1$ den Quotienten $\pi(\lambda) = \#\{p_i > \lambda\} / (m(1-\lambda))$ für $i=1, \dots, m$, wobei m die Anzahl der Hypothesen bezeichnet. Häufig wählt man $\lambda=0.5$ und schätzt den wahren Anteil indem man $\pi_0 = \pi(\lambda)$ setzt. Eine alternative Möglichkeit ist dasjenige λ^* zu bestimmen, für welches der mittlere quadratische Fehler $MSE(\lambda)$ zwischen wahren π_0 und $\pi(\lambda)$ minimal wird und wählt als Schätzwert für π_0 schließlich $\pi(\lambda^*)$. In vielen Fällen ist aber die Verteilung der p-Werte unbekannt und daher schätzt man den mittleren quadratischen Fehler $MSE(\lambda)$ unter Verwendung eines Bootstrap-Verfahrens.

Bei den Methoden von Benjamini und Hochberg (vgl. [1]) und Hsueh et al. (vgl. [3]) kommt man auf unterschiedliche Weise über den Größenunterschied zwischen den p-Werten zu Schätzern für den Anteil π_0 wahrer Hypothesen. Eine weitere Methode ist die Schätzung von π_0 mittels kubischer Splines (vgl. [6]). Hierbei wird ein kubischer Spline mit 3 Freiheitsgraden an die Werte $(\lambda, \pi(\lambda))$ angepasst und man wählt als Schätzwert für π_0 den Wert des kubischen Splines an der Stelle 1. Langaas et al. (siehe [4]) betrachteten folgende Schätzmethode für π_0 . Ausgangspunkt hierbei ist die Dichtefunktion f der Verteilungen der einzelnen p-Werte. Diese besitzt unter der Annahme der Unabhängigkeit und Gleichverteilung die Gestalt $f(p) = \pi_0 + (1-\pi_0)h(p)$ für $0 \leq p \leq 1$. Dabei ist h die unbekannte Dichte der Verteilung eines p-Wertes, wenn die Alternativhypothese wahr ist. Die in [4] vorgestellte Methode schätzt dann die Funktion f auf Grundlage der p-Werte durch eine konvexe und monoton fallende Dichtefunktion. Der Wert dieser Funktion an der Stelle 1 liefert dann einen Schätzwert für π_0 .

Die verschiedenen Schätzverfahren werden zunächst in Computersimulationen auf künstlich erzeugte m-dimensionale Daten angewandt, bei denen der Anteil π_0 der wahren Hypothesen bekannt ist. Hinsichtlich der erzeugten Daten gehen wir zunächst von normalverteilten Daten mit unterschiedlicher Korrelationsstruktur aus. Zum einen betrachten wir den Fall unabhängiger bis hochkorrelierter Daten und zum anderen den Fall, wenn die Korrelationsmatrix eine vorgegebene Blockstruktur besitzt. Als Testverfahren wählen wir beispielhaft den t-Test sowie den Wilcoxon-Test. Für unterschiedlich vorgegebene Anteile wahrer Hypothesen werden die Schätzmethoden miteinander verglichen.

Ergebnisse und Ausblick

Ein Vergleich zeigt, mit welchen Methoden der Anteil π_0 wahrer Hypothesen unterschätzt oder überschätzt wird. Ziel ist es zu erreichen, das wahre π_0 zu überschätzen, um somit die Power der einzelnen Testverfahren zu erhöhen. Aus einer großen Zahl von Schätzwerten lässt sich auch die Streuung der Schätzwerte als Maß für die Zuverlässigkeit der jeweiligen Methode ermitteln. Vor allem mit wachsender Korrelation sind die Streuungen groß und die Schätzgenauigkeiten niedrig (vgl. Abb.1). Ebenso verhält es sich mit wachsendem Anteil π_0 , wobei die Methode von Benjamini & Hochberg in unserem Falle eine Ausnahme ist. Dieses Verfahren beispielsweise nicht zu empfehlen, da die Streuung den wahren Anteil π_0 nicht überdeckt wie Abb. 1 zeigt.

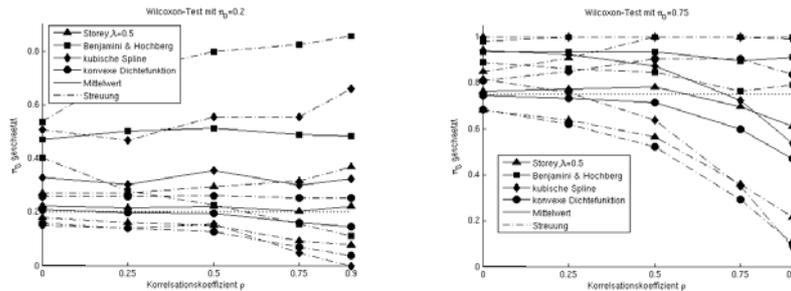


Abb.1 Mittelwert und Streuung der geschätzten Werte für π_0 in Abhängigkeit von konstanten Korrelationskoeffizienten ρ (Stichprobenumfang $n=10$, Anzahl der Hypothesen $m=100$, Anzahl der Simulationen $M=500$)

Durch Kombination verschiedener Ansätze soll versucht werden, verbesserte Schätzverfahren zu entwickeln, die wir hier vorstellen wollen. Die verschiedenen Schätzmethoden sollen auch auf reale Daten (EEG- und Proteinexpressionsdaten) angewandt und dabei auf ihre praktische Verwendbarkeit getestet werden.

Danksagung

Diese Arbeit wird unterstützt durch die Deutsche Forschungsgesellschaft (Projekt VO 683/2-1).

Literatur

- [1] Benjamini Y, Hochberg Y. On the adaptive control of false discovery rate in multiple testing with independent statistics. Journal of Educational and Behavioral Statistics 2000; 25; 06-83
- [2] Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. Research paper 01-03. Department of Statistics and OR, Tel Aviv University, Tel Aviv, Israel.
- [3] Hsueh H, Chen JJ, Kodell RL. Comparison of methods for estimating the number of true null hypothesis in multiple testing. Journal of Biopharmaceutical Statistics 2003; 13; 679-689.
- [4] Langaas M, Lindquist BH, Ferkingstad E. Estimating the proportion of true null hypotheses, with application to DNA microarray data. Journal of the Royal Statistical Society, Series B 2005; 67; 555-573.
- [5] Storey JD: A direct approach to false discovery rates. Journal of the Royal Statistical Society, Series B 2002; 64; 479-498.
- [6] Storey JD, Tibshirani R. Statistical significance for genome-wide experiments. Proceedings of the National Academy of Sciences USA 2003; 100; 3889-3894.