

Optimale statistische Versuchsplanung

dreifaktorieller Zwei-Farben-cDNA-Microarray-Experimente

Dipl.-Stat. Sven Stanzel

Abstract

In der Medizin hat mit dem rasanten Wachstum der biotechnologischen Industrie in den letzten Jahrzehnten die Erforschung von mit Genmutationen assoziierten Ursachen für die Entstehung verschiedener Krankheitsbilder stark an Bedeutung zugenommen. Das Ziel der Durchführung von Genexpressionsanalysen mit Hilfe von Zwei-Farben-cDNA-Microarray-Experimenten [1] besteht in der Identifikation von Kandidatengenen, die für den Ausbruch einer bestimmten Erkrankung mitverantwortlich gemacht werden können. Langfristig sollen die hierbei gewonnenen Erkenntnisse zur Entwicklung von innovativen Medikamenten führen, die zur Therapie von genetisch assoziierten oder bedingten Erkrankungen verwendet werden können.

Landgrebe et al. haben ein spezielles genspezifisches lineares Modell mit festen Effekten für die statistische Auswertung der in Zwei-Farben-cDNA-Microarray-Experimenten generierten Genexpressionsdaten vorgeschlagen [2]. Durch eine geringfügige Modifikation dieses Modells kann zusätzlich zu den häufig betrachteten Einflussgrößen „Farbe“ und „Behandlung“ der Effekt des Faktors „Zelllinie“ auf die ermittelten Genexpressionsdaten analysiert werden.

Es wird gezeigt, dass der Parametervektor des Landgrebe-Modells in dieser speziellen dreifaktoriellen Designsituation nicht unverzerrt schätzbar ist. Anschließend werden ϕ_p -optimale Versuchspläne für das Schätzen aller Paarvergleiche von Behandlungen, für den Nachweis unterschiedlicher Behandlungsdifferenzen zwischen den Zelllinien, sowie für das Schätzen des aus diesen speziellen linearen Kontrasten zusammengesetzten Vektors, konstruiert. Für den Nachweis der ϕ_p -Optimalität wird dabei jeweils eine generalisierte Form der von Pukelsheim eingeführten Äquivalenztheoreme [3] verwendet.

Die erzielten Optimalitätsergebnisse sind nicht auf bestimmte Anzahlen von Arrays, Behandlungen und Zelllinien beschränkt und liefern damit allgemeingültige Optimalitätsaussagen für das Schätzen der hier untersuchten linearen Kontraste. Bei Betrachtung der linearen Kontraste für das Schätzen der Interaktionseffekte von Behandlungen und Zelllinien ist die gefundene Lösung allerdings abhängig von der vorliegenden Beziehung zwischen der Anzahl analysierter Behandlungen und der Anzahl verwendeter Zelllinien. Die Unabhängigkeit dieser Lösungen vom verwendeten Optimalitätskriterium lässt sich als Robustheitseigenschaft der konstruierten optimalen Blockpläne interpretieren. Insbesondere sind diese Versuchsanlagen damit A-, D-, E- und T-optimal für das Schätzen der betrachteten linearen Kontraste.

Für die Praxis ergeben sich aus den identifizierten optimalen Designs direkte Empfehlungen bezüglich der Wahl effizienter Versuchsanlagen für konkrete Zwei-Farben-cDNA-Microarray-Experimente mit vorgegebenen Anzahlen von Arrays, Behandlungen und Zelllinien. Durch die Verwendung entsprechender effizienter Versuchspläne lässt sich ein Zeitgewinn bei der praktischen Durchführung dieser Experimente und damit die Einsparung finanzieller Ressourcen erwarten.

Literatur

- [1] Brown, P.O., Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nature Genetics* 21 (1): 10-14.
- [2] Landgrebe J., Bretz F., Brunner E. (2006). Efficient design and analysis of two colour factorial microarray experiments. *Computational Statistics and Data Analysis* 50: 499-517.
- [3] Pukelsheim F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Schlagworte

Äquivalenztheorem – linearer Kontrast – lineares Modell mit festen Effekten –
Zwei-Farben-cDNA-Microarray-Experiment – ϕ_p -Optimalität