Optimal designs for two-colour microarray experiments

Penny S. Sanchez

Discipline of Statistics

School of Mathematical Sciences

The University of Adelaide

February 2, 2010

Table of Contents

My PhD research focuses on the recommendation of optimal designs for two-colour microarray experiments. Two-colour microarrays are a technology used to investigate the behaviour of many thousands of genes in a single experiment. This technology has created the potential for making significant advances in the field of bioinformatics. Careful statistical design is crucial to realize the full potential of microarray technology. My research has focused on the recommendation of designs that are optimal in terms of precision for effects that are of scientific interest, making the most effective use of available resources. Based on statistical efficiency, the optimality criterion used is Pareto optimality. A design is defined to be Pareto optimal if there is no other design that leads to equal or greater precision for each effect of scientific interest and strictly greater precision for at least one. My PhD thesis was submitted in June and key aspects of my research are summarised below.

Pareto optimality enables the recommendation of designs that are particularly efficient for the effects that are of scientific interest. I have developed methodology to cater for effects of interest that correspond to contrasts rather than solely considering parameters of the statistical linear model. My approach also caters for additional experimental considerations such as contrasts that are of equal scientific interest. During my PhD, I have provided advice regarding the design of two-colour microarray experiments aimed at discovering the genetic basis of medical conditions.

For large experiments, it is not feasible to examine all possible designs in an exhaustive search for Pareto optimal designs. I have adapted the multiple objective metaheuristic method of Pareto simulated annealing to the microarray context. The aim of Pareto simulated annealing is to generate an approximation to the set of Pareto optimal designs in a relatively short time. At each iteration, a sample of generating designs is used to explore the design space in an efficient way. This involves the setting of a number of Pareto simulated annealing parameters and the development of appropriate quality measures. I have developed algorithms to search systematically for the optimal values of

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopy.

Signed,

Date:

Acknowledgments

There are many people I would like to thank for assisting me to achieve my potential during my PhD research. Firstly, I thank my PhD supervisors, Andrew Metcalfe and Gary Glonek, for their collective guidance and support. I thank Andrew for his enthusiasm, inspiration, dedication and being a wonderful role model. I thank Gary for his ingenuity, innovative thinking, attention to detail and high aspirations. I am grateful to Anna Tsykin for her additional support and advice in many areas including the biological aspects relevant to my research.

I am appreciative of the School of Mathematical Sciences for providing a supportive environment. I thank Liz Cousins for her encouragement and assistance during my PhD, particularly as Postgraduate Co-ordinator. I am also grateful to Charles Pearce for the wisdom and experience he has shared over many years. I thank Patty Solomon for assistance in the earlier stages of my PhD.

I am grateful to Telstra for additional support during my PhD. I thank Bob Richter and the team for our positive and motivating discussions, advice and support.

In terms of accessing materials in alternative formats, I have been grateful for support from the Disability Service at the University of Adelaide, postgraduate students Jason Whyte, Brian Webby and Josephine Varney as well as the statistical researchers I have communicated with in Australia and overseas.

I am appreciative for the financial grants and scholarships that I have received during my PhD, namely the Australian Postgraduate Award, June Opie Fellowship, Jean Gilmore Bursary,

International Biometrics Society Travel Award, AMSI/ICE-EM Winter School Travel Grant, AMSI Summer Symposium Travel Grant, AMATA Bursary as well as additional support from the School of Mathematical Sciences and Telstra.

Finally, I thank my family and friends that I hold dear to me for sharing this experience with me. I thank my husband Phil for his commitment, strength, advice, unconditional support and always believing in me. I also thank my parents, Val, Josie and Peggy-Anne for their encouragement and support.

Publications arising from this thesis

Bennett, P. S., Glonek G. F. V. and Solomon, P. J. (2005). Optimal Designs for Gene Expression Studies. *Statistical Solutions to Modern Problems: Proceedings of the 20th International Workshop on Statistical Modelling, Sydney, 10-15 July.* University of Western Sydney Press, Sydney.

Sanchez, P. S., and Glonek G. F. V. (2009). Optimal designs for two-colour microarray experiments.*Biostatistics 2009* 10(3), 561-574, 2009.