

Optimal Designs for Gene Expression Data using Model Averaging Estimators

In this thesis, optimal designs should be constructed for the simultaneous analysis of gene expression data. The analysis will be based on the VPA data, which contains gene expression data of valproic acid applied to human embryonic stem cells for thousands of different genes (see Krug et al. (2013) for details).

In a first step the data set will be analysed with respect to the Akaike information criterion (AIC) in order to select an appropriate concentration-response curve for each gene. Using these results, a design optimality criterion will be derived that addresses the simultaneous inference of all genes by also taking into account the distribution of the selected models (based on the AIC). More precisely, the results of Alhorn et al. (2021) shall be used to derive an appropriate optimality criterion for the problem at hand. The performance of the design, which optimizes the developed criterion, shall be compared to other designs by using a simulation study.

Literature:

Alhorn, K., Dette, H. & Schorning, K. Optimal Designs for Model Averaging in non-nested Models. *Sankhya A* **83**, 745–778 (2021). <https://doi.org/10.1007/s13171-020-00238-9>

Krug, A. K., Kolde, R., Gaspar, J. A., Rempel, E., Balmer, N. V., Meganathan, K., Vojnits, K., Baquie, M., Waldmann, T., Ensenat-Waser, R., Jagtap, S., Evans, R. M., Julien, S., Peterson, H., Zagoura, D., Kadereit, S., Gerhard, D., Sotiriadou, I., Heke, M., Natarajan, K., Henry, M., Winkler, J., Marchan, R., Stoppini, L., Bosgra, S., Westerhout, J., Verwei, M., Vilo, J., Kortenkamp, A., Hescheler, J. R., Hothorn, L., Bremer, S., van Thriel, C., Krause, K. H., Hengstler, J. G., Rahnenfuhrer, J., Leist, M. and Sachinidis, A. (2013): Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. *Arch Toxicol*, **87**, p. 123–143.

For further information please contact: schorning@statistik.tu-dortmund.de