Optimality, Robustness and Fusion

(in computer experiments, factorial designs and beyond)

Boxin Tang

Simon Fraser University

How should we feel about ourselves and our design field?

- **we have a very proud history**

  (just take a look at the contributions of the founding father and early pioneers of statistics)

- **though not in the spotlight,**

  we are always on the main stage

  (we publish in and are on editorial boards of major statistics journals, math journals and engineering journals)

- **design genes live in every statistician**

  (you have design genes if you ever care about sample size determination, blinding, randomization, data quality, selection bias, active learning, etc)

- **"design" will outlast "big data"**

  (big data will be just data one day, not to diminish the significance of big data era.)

Optimality and Robustness

- optimality $\approx$ minimizing variance.
- robustness $\approx$ minimizing bias.
- two fundamental forces that drive the theory and applications of experimental design.

Factorial Designs

- non-standard parametrization
  (baseline parametrization; conditional effects)
- focus on Y or β ?
  (there are some advantages if focus is on Y)
- applications to other design areas
  (computer experiments, choice experiments and perhaps nonlinear models)

Computer Experiments

- what kind of space-filling designs to use?
  (require large-scale, well-designed, and impartial evaluations of various space-filling designs from low to high to very high dimensional problems)
- need stronger statistical justifications for space-filling designs
  (in terms of optimality and robustness for example)
- inferences based on a combination of randomization model and Gaussian process model? Or perhaps, nonparametric model like MARS?

Fusion

- Fusing theoretical construction and computer generation
- Fusing computer experiments and factorial designs
- Fusing design and analysis
  (more details next)

By fusing design and analysis (design-analysis), we obtain

designana

What is designana anyway?

Since Google has something for everything, let's google designana.

What I really mean is the following.

Consider a big data linear model

$$Y = X\beta + \epsilon,$$

- Y is $N \times 1$ and X is $N \times M$
- very large N and M.

Design:

is to identify an $n \times M$ submatrix of X, where $n \ll N$.

Analysis:

is to identify an $N \times m$ submatrix of X, where $m \ll M$.

Designana:

is to identify an $n \times m$ submatrix of X, where $n \ll N$ and $m \ll M$.

Thank you!