Robustness Theory and Methodology: Recent Advances and Future Directions

Giseon Heo (University of Alberta)

September 2-4, 2016

1 Overview of the Field

While Box [1] used the word *robust* for the first time to mean *insensitivity* to the violation of assumptions, Tukey [2] and Huber [3], in the 1960s, made substantial contributions to establish robustness as a major sub-discipline of statistics. John Tukey at Princeton University invited Peter Bickel, Peter Huber and Frank Hampel in 1970-71 to make cooperative efforts for the further progression and development of robust statistics. This period was later called the *Princeton robustness year* and became known for its extensive Monte Carlo study of robust location estimates under symmetric long-tailed distributions [4]. The team focused on the theory of robust estimation based on specified properties and estimators under certain conditions. Since then, many review articles (e.g., Stigler [5], 2010) and books (e.g., Rieder [6], 2012) have been published and the definition of robustness largely refined. One more recent definition of *robustness* might be the "stability theory of statistical procedures [that] systematically investigates the effects of deviations from modeling assumptions on known procedures and, if necessary, develops new, better procedures" [7].

2 Recent Developments and Open Problems

Robustness has spread widely to other fields of statistics in the robust design of experiments, correction of misspecified regression functions, violation of error structures, and furthermore against wild observations or misspecified underlying distributions. Recently, robustness has expanded even further to the robust design of experiments for quantile regression and machine learning for robust active learning [8].

3 Presentation Highlights and Scientific Progress

The first speaker, Xiaojian Xu, presented on the historical development of robust statistics as a sub-discipline of statistics. Some landmark articles and studies for such developments were emphasized. In particular, the advances of robustification at the stage of experimental design were discussed and categorized. Xu also summarized Wiens' contributions to robust statistics, distribution theory and robustness, robust optimal designs [9], [10], robust sampling designs [11], [12], and robust active learning [8].

Xu's talk was followed by Julie Zhou, who discussed mimimax regression designs and challenges. Considered was a true regression model,

$$y_i = g(\mathbf{x}_i; \theta) + f(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \cdots, n,$$

where $f \in \mathcal{F}$ is the unknown disturbance function (a neighbourhood of functions), $E(\epsilon_i) = 0$, and $\mathrm{Cov}((\epsilon_1, \cdots, \epsilon_\mathrm{n})^\top) = \Sigma$. Let MSE$(\xi, f, \Sigma)$ be the mean squared matrix of an estimator $\hat{\theta}$ of θ and $\phi(\xi, f, \Sigma)$ a scalar function of MSE(ξ, f, Σ), where ξ is the design distribution of $\mathbf{x}_1, \cdots, \mathbf{x}_n$. Zhou formulated three minimax design problems,

$$\min_{\xi} \max_{f \in \mathcal{F}} \phi(\xi, f, \sigma^2 I),$$

$$\min_{\xi} \max_{\Sigma \in D_\Sigma} \phi(\xi, 0, \Sigma),$$

$$\min_{\xi} \max_{f \in \mathcal{F}} \max_{\Sigma \in D_\Sigma} \phi(\xi, f, \Sigma),$$

where D_Σ is a neighbourhood of covariance matrices. The challenges presented by these problems lie in the fact that their objective functions may not be convex nor smooth and it is thus difficult to construct minimax designs.

Rui Hu, the "youngest" of Wiens' PhD students, presented her research on "maxmin" designs (which are unlike the traditional minimix designs). She considered discriminating two rival models, f_0 and f_1, in hypothesis testing, namely,

$$H_0 : f_0(y|\mathbf{x}, \mu_0) \text{ versus } H_1 : f_1(y|\mathbf{x}, \mu_1),$$

where y is the response variable and $\mathbf{x}$ is the covariate vector. Hunter and Reiner [13], in 1965, constructed optimal designs when both f_0 and f_1 are fixed normal densities, and many researchers have since extended these classical optimal designs. In particular, López-Fidalgo *et al* [14], 2007, proposed a classical optimal design when the two rival models are non-normal. All of these models, however, assume that one of f_0 and f_1 is the true model. Hu and Wiens [15], in 2016, proposed a robust optimal design by assuming that the correct model lies within only an approximately known class (Hellinger neighborhood) of $f_0 \in \mathcal{F}_0$ or $f_1 \in \mathcal{F}_1$. Robust design is to find an optimal design measure ξ that maximizes the minimum power of the test over $\mathcal{F}_1$. Hu presented an analytical solution and the maximization portion was solved by proposing a sequential design. She also showed that this sequential design is indeed optimal. Hu and Wiens [15] further considered constructing a robust optimal design for discriminating between two models. A current open problem is an extension of their work to more than two rival models.

Zhou's student, Lucy Gao, proposed an extension of the distributionally robust logistic regression [16] to multinomial logistic regression [17]. Gao defined a distributionally robust optimization problem as minimizing a convex loss function over logistic regression coefficients on a family of probability measures within the Wasserstein metric ϵ. She also highlighted a method of estimating the misclassification rate of robust multinomial logistic regression by solving a tractable convex optimization problem and considering the best and worst risk scenarios.

Xiaojian Xu wrote in her abstract that "since statistics is the science of a process of data collecting, data analyzing, drawing conclusions, and (re)evaluating the process, the consideration of robustification can encompass any stage of a statistical process." Indeed, Matthew Pietrosanu, an undergraduate student of Heo, alluded to the effectiveness of robustness in estimating high dimensional surfaces from noisy point cloud data.

4 Outcome of the Meeting

Four academic generations participated the workshop, namely, John Collins (Wiens' PhD supervisor), Doug Wiens, Wiens' students, and their students. This fourth generation is not only continuing to solve fundamental research problems in robustness but is also branching its applications to shape analysis and statistical machine learning. Zhichun Zhai and Doug Wiens plan to lay groundwork in the applications of robust theory in machine learning. During the workshop, Xiaojian Xu, Linglong Kong, and Doug Wiens also discussed some new research problems on the robust designs of experiment for composite quantile regression.

In many statistical problems including robustness theory, it is common to fix certain parameter(s) before solving a given optimization problem. In recent years, it is becoming more popular to study the trend of solutions as the parameter(s) vary. In robustness theory, these problems are often formulated in a neighborhood of certain functions and/or spaces of probability measures. The size of the neighborhood ϵ must be set before solving a given loss function. However, the size ϵ can be built into the optimization problem itself. Rather than treating ϵ as fixed, one can view the pattern of optimization solutions as ϵ varies. During the workshop, a number of the researchers have observed this phenomenon. Julie Zhou and Giseon Heo plan to meet in February 2017 to discuss novel ways to view robust regression problems.

References

[1] G. Box, Non-normality and tests on variance. *Biometrika* **40** (1953), 318-335.

[2] J. Tukey, A survey of sampling from contaminated distributions. In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann, eds.), 448485. Stanford Univ. Press, 1960.

[3] P. Huber, Robust Estimation of a Location Parameter. *Annals* **35** (1964), 73-101.

[4] D. Andrews, P. Bickel, F. Hampel, P. Huber, W. Rogers and J. Tukey, Robust Estimates of Location: Survey and Advance, Princeton University Press, Princeton, N. J, 1972.

[5] S. Stigler, The Changing History of Robustness. Keynote address at *ICORES10*, Prague, 28 June 2010.

[6] H. Rieder, Robust asymptotic statistics. Springer Series in Statistics, 2012.

[7] F. Hampel, Robust Statistics: A brief introduction and overview. Research report, ETH Zurich, Switzerland, 2001.

[8] R. Nie and D. P. Wiens, Minimax Robust Active Learning for Approximately Linear Regression Models. Preprint 2016.

[9] D. P. Wiens, Minimax designs for approximately linear regression. *Journal of Statistical Planning and Inference*, **31** (1992), 353-371.

[10] H. Dette and D. P. Wiens, Robust Designs for Series Estimation. *Computational Statistics & Data Analysis*, **52** (2008), 4305-4324.

[11] Y. Rabhi and D. P. Wiens, Robust Sampling Designs for a Possibly Misspecified Stochastic Process. *Statistica Sinica*, 2016.

[12] Z. Zhai and D. P. Wiens, Robust Model-Based Stratified Sampling Designs. *Canadian Journal of Statistics*, **43** (2015), 554-577.

[13] W. G. Hunter and A.M. Reiner, Designs for Discriminating between Two Rival Models. *Technometrics* **7** (1965), 307-323.

[14] J. López-Fidalgo, C. Tommasi and P. C. Trandafir, An Optimal Experi-mental Design Criterion for Discriminating between Non-Normal Models. *Journal of the Royal Statistical Society Series B* **6** (2007), 231-242.

[15] R. Hu and D. Wiens, Robust discrimination designs over Hellinger neighbourhoods. Accepted by *Annals of Statistics*, in press (2016).

[16] S. Shafieezadeh-Abadeh, P. M. Esfahani, and D. Kuhn, Distributionally Robust Logistic Regression, *Advances in Neural Information Processing Systems*, **28** (2015), 1-9.

[17] L. L. Gao, Distributionally robust multinomial logistic regression, in preparation.