

# STATISTICS COLLOQUIUM

WEDNESDAY, MAY 8, 2013

TALK: 1:00 PM — SCIENCE CENTER RM. 705

RECEPTION: 12:30 PM — SCIENCE CENTER, 7TH FLOOR

**“Thank God that Regressing Y on X is Not the Same as  
Regressing X on Y: Direct and Indirect Residual Augmentations”**

**Xiaojin Xu  
Department of Statistics  
Harvard University**

## ABSTRACT

What does regressing Y on X versus regressing X on Y have to do with MCMC? It turns out that many strategies for speeding up data-augmentation type algorithms can be understood as fostering independence or “de-correlation” between a regression function and the corresponding residual, thereby reducing or even eliminating dependence among MCMC iterates.

There are two general classes of algorithms, those corresponding to regressing parameters on augmented data/auxiliary variables and those that operate the other way around. The interweaving strategy (Yu and Meng, 2011) provides a general recipe to automatically take advantage of both, and it is the existence of two different types of residuals that makes the interweaving strategy seemingly magical in some cases and promising in general.

The concept of residuals---which depends on actual data---also highlights the potential for substantial improvements when data augmentation schemes are allowed to depend on the observed data. At the same time, there is an intriguing phase transition type of phenomenon regarding choosing (partially) residual augmentation schemes, reminding us once more of the prevailing issue of trade-off between robustness and efficiency.

We will report on these latest theoretical investigations (using a class of normal/independence models) and empirical findings (using a posterior sampling for a probit regression) in the search for effective residual augmentations.