

**SOMS Friday Seminar**  
**January 18, 2008**  
**SMC 305**  
**2:30-3:25**

**MCMC Inference for  
a Model with Sampling Bias:  
An Illustration using SAGE Data**

**Russell Zaretzki**

Assistant Professor

Department of Statistics, Operations & Management Science  
University of Tennessee

**ABSTRACT**

This talk explores Bayesian inference for a biased sampling model in situations where the population of interest cannot be sampled directly, but rather through an indirect and inherently biased method. Observations are viewed as being the result of a multinomial sampling process from a tagged population which is, in turn, a biased sample from the original population of interest. This paper presents several Gibbs Sampling techniques to estimate the joint posterior distribution of the original population based on the observed counts of the tagged population. These algorithms efficiently sample from the joint posterior distribution of a very large multinomial parameter vector. Samples from this method can be used to generate both joint and marginal posterior inferences. We also present an iterative optimization procedure based upon the conditional distributions of the Gibbs Sampler which directly computes the mode of the posterior distribution. To illustrate our approach, we apply it to a tagged population of messenger RNAs (mRNA) generated using a common high-throughput technique, Serial Analysis of Gene Expression (SAGE). Inferences for the mRNA expression levels in the yeast *S. cerevisiae* are reported.

**Refreshments provided after seminar**