



THE UNIVERSITY OF
CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

REBECCA C. STEORTS

Department of Statistical Science
Duke University

Entity Resolution with Societal Impacts
in Machine Learning

MONDAY, April 15, 2019 at 4:30 PM
Eckhart 133, 5734 S. University Avenue
Refreshments before the seminar at 4:00 PM in Jones 111

ABSTRACT

Very often information about social entities is scattered across multiple databases. Combining that information into one database can result in enormous benefits for analysis, resulting in richer and more reliable conclusions. In most practical applications, however, analysts cannot simply link records across databases based on unique identifiers, such as social security numbers, either because they are not a part of some databases or are not available due to privacy concerns. In such cases, analysts need to use methods from statistical and computational science known as entity resolution (record linkage or de-duplication) to proceed with analysis. Entity resolution is not only a crucial task for social science and industrial applications, but is a challenging statistical and computational problem itself. In this talk, we describe the past and present challenges with entity resolution. More specifically, I will discuss unsupervised Bayesian entity resolution models, which are able to identify duplicate records in the data, while quantifying uncertainty of the entity resolution process. In addition, one can prove tight theoretical bounds on the class of entity resolution models, which support the proposed approach. Finally, I present in progress work, where the class of models that I propose can scale to millions of records, illustrating the power of Bayesian methods for applications to both synthetic and real data.

For further information and inquiries about building access for persons with disabilities, please contact Jonathan Rodriguez at 773.702.8333 or send him an email at jgrodriquez@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following website:
<https://lists.uchicago.edu/web/subscribe/statseminars>.