



# THE UNIVERSITY OF CHICAGO

Department of Statistics

## MASTER'S THESIS PRESENTATION

---

NELSON AUNER

Department of Statistics  
The University of Chicago

Combining Latent Topics with Document Attributes in Text Analysis

TUESDAY, May 13, 2014, at 2:00 PM  
Eckhart 110, 5734 S. University Avenue

### ABSTRACT

This paper introduces a variant to existing models of multinomial regression for text analysis. Using the base model introduced by Taddy (2013), we extend the data-generating model to incorporate topics not explained by metadata. In doing so, we seek to increase the prediction accuracy over existing techniques, bridge the gap between fast multinomial regression and standard topic models, and investigate methods for discovering new topics in a corpus. We explore computational aspects of our approach, provide software for parallelization of the algorithm, and conclude by proposing areas of future research.

---

For information about building access for persons with disabilities, please contact Matt Johnston at 773.702-0541 or send an email to [mhj@galton.uchicago.edu](mailto:mhj@galton.uchicago.edu).

If you wish to subscribe to our email list, please visit the following web site:  
<https://lists.uchicago.edu/web/arc/statseminars>.