

STATISTICS COLLOQUIUM

MONDAY, OCTOBER 05, 2009

TALK: 4:00 PM — SCIENCE CENTER, ROOM 309

RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

“Supervised and Relational Topic Models”

David M. Blei

Assistant Professor of Computer Science
Princeton University

ABSTRACT

A surge of recent research in machine learning and statistics has developed new techniques for finding patterns of words in document collections using hierarchical probabilistic models. These models are called "topic models" because the discovered word patterns often reflect the underlying topics that permeate the documents. Topic models also naturally apply to data such as images and biological sequences.

In this talk I will review the basics of topic modeling, and discuss some recent extensions: supervised topic modeling and relational topic modeling. Supervised topic models allow us to use topics in a setting where we seek both exploratory and predictive power. Relational topic models---which are built on supervised topic models---consider documents interconnected in a graph. These models can be used to summarize a network of documents, predict links between them, and predict words within them.

Joint work with Jonathan Chang and Jon McAuliffe.