THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

PETER LATHAM

Gatsby Computational Neuroscience Unit
University College London

Zipf's Law Arises Naturally from Hidden Structure

MONDAY, February 9, 2015, at 4:00 PM
Eckhart 133, 5734 S. University Avenue
Refreshments following the seminar in Eckhart 110

ABSTRACT

Both natural and artificial systems often exhibit a surprising degree of statistical regularity. One such regularity is Zipf's law. Originally formulated for word frequency, Zipf's law has since been observed in a broad range of phenomena, including city size, firm size, mutual fund size, amino acid sequences, and neural activity. Partly because it is so unexpected, a great deal of effort has gone into explaining it. So far, almost all explanations are either domain specific or require fine-tuning. For instance, in biology, one explanation for observations of Zipf's law is that biological systems sit at a special thermodynamic state, the critical point. Here we propose an alternative explanation, which exploits the fact that most real-world datasets can be understood as being generated from a latent variable model. We show that data generated from a such a model exhibits Zipf's law under very mild conditions. We provide the theoretical underpinnings of this result, illustrate it on words and neural data, and point out examples of Zipf's law in the literature for which we can identify a latent variable model. Finally, we show how our results can be used to inform models of neural data.