

# SEMINAR

**December 4, 2014 at 3:30pm**

**Sidney Smith Hall, Room 1083**

**\*Refreshments will be served at 3:15pm\***

**Rediscovery of Good-Turing Estimators via Bayesian  
Nonparametrics**

**Yee Whye Teh, University of Oxford**

---

The problem of estimating discovery probabilities has originally appeared in ecology, and plays an essential role across many applications arising from biology, bioinformatics, genetics, linguistics, designs of experiments, etc. A full range of statistical approaches, parametric and nonparametric as well as frequentist and Bayesian, have been proposed for estimating discovery probabilities.

In this talk we discuss the relationship between the celebrated Good-Turing approach, which is a frequentist nonparametric approach developed in the 1940s, and a Bayesian nonparametric approach recently introduced in the literature. Specifically, under a two parameter Poisson-Dirichlet prior, we show that for a large sample size the Bayesian nonparametric estimators for discovery probabilities are asymptotically equivalent to suitably smoothed Good-Turing estimators.

As a by-product of our results, we present a novel methodology for associating exact and asymptotic credible intervals to the Bayesian nonparametric estimators for discovery probabilities. This methodology is illustrated through the analysis of some benchmark Expressed Sequence Tags datasets.

Joint work with Stefano Favaro and Bernardo Nipoti.