HARVARD UNIVERSITY
DEPARTMENT OF STATISTICS

Tel. 617-495-5496
Fax. 617-496-8057
statdept@stat.harvard.edu
http://www.stat.harvard.edu/

Science Center 7th Fl.
One Oxford Street
Cambridge, MA 02138

STATISTICS COLLOQUIUM

THURSDAY, OCTOBER 16, 2014
TALK: 2:45 PM — SCIENCE CENTER RM. 706
RECEPTION: 4:00 PM — SCIENCE CENTER, 7TH FLOOR

"Bayesian Regression Allelic Imbalance Models, or how to find the biased coins in the fountain of life"

Daniel Fernandez
Department of Statistics, Harvard University
Advisors: Jun S. Liu, Tirthankar Dasgupta, Joseph Blitzstein & Catherine Dulac

ABSTRACT

We are a diploid organism; we carry one copy of our genome from our mother, and one copy from our father. Nevertheless, not all the genes in our two genomes (two alleles), or all of our chromosomes are expressed in a balanced manner – some 'coins' are biased: the X chromosome is inactivated in females in a process called X-chromosome inactivation, some genes are inactivated randomly in a process called Stable autosomal monoallelic expression, some genes are expressed only in a parent-of-origin manner in a process called imprinting and many genes are regulated in an allele-specific manner in a process called cis-regulation of gene expression. But, how would one try to find the biased coins in the fountain of life?

In this talk we present a detailed search for this quest in the context of imprinting, and cis-eQTL that acts in an allele-specific manner, but the models are widely applicable to any scenario of allelic imbalance.

First we introduce the problem of estimating the expression of the two copies: the maternal and the paternal copy of a gene. Next, we propose frequentist and bayesian models of allele-specific expression, motivated by the central dogma of molecular biology and the data generating process (DGP) for RNA-Seq. We develop EM and Gibbs sampling approaches to estimate gene and transcript-specific expression from our proposed models, and pinpoint the key advantage of our models as being biologically interpretable and as having smaller number of false positives under realistic scenarios.

Second, we show how allele-specific estimates of gene expression, used in combination with a careful experimental design can help us look into the dynamics of imprinting across multiple BRAIN regions and at different developmental time points. Thus the motivation for our model name: BRAIM, Bayesian regression Allelic Imbalance model.

Finally, we show how similar models can be extended outside the bench, to study imprinting in human population studies for which we do not observe the chromosome of origin, and to also study cis-regulation of genes in an allele-specific manner.