

**LIAN Heng**  
Assistant Professor  
Division of Mathematical Sciences  
School of Physical & Mathematical Sciences  
B.Sc. (USTC), Ph.D. (Brown).

*Major research interests:* **Bayesian asymptotics in infinite-dimensional spaces, statistical genetics, functional data analysis**

Email: [henglian@ntu.edu.sg](mailto:henglian@ntu.edu.sg)  
<http://www.ntu.edu.sg/home/henglian/>  
Phone: (65) 6514 8373



### **Infinite-dimensional Bayesian Procedure.**

Recently there has been a growing interest in the asymptotics of nonparametric Bayesian estimation. In Barron et. al. (1999), the consistency of density estimation is proved by constructing appropriate sieves and bounding the metric entropy. Results on rates of convergence are obtained later by some other researchers. We proved a result along this line for the case that the underlying function are piecewise constant with a finite number of jumps, corrupted by i.i.d. Gaussian noise with known variance. A main difference from the previous problems studied is that we assume that the covariates are on a regular grid in the interval  $[0,1]$ , so the observations are independent but not identically distributed. We showed that if the prior is such that it sufficiently penalizes frequent jumps, the Bayesian procedure is consistent in the sense that the posterior probability mass will concentrate on a neighborhood of the true function, even when the true function is not stepwise constant, but satisfies some smoothness condition instead. We also constructed a simple prior which shows that the assumptions we made for consistency are easily satisfied in practice.

### **Functional Data Analysis**

In many experiments, functional data appear as the basic unit of observations. As a natural extension of the multivariate data analysis, functional data analysis provides valuable insights into these problems. We developed an estimation procedure for functional response models within the framework of RKHS. For nonlinear models, the parametric approach does not give satisfactory results. We show that within the RKHS framework, a simple estimate can be developed which reduces to linear regression computations very much like those derived in the parametric approach.

### **Analysis of Chip-Seq (Work In Progress)**

Over the last several years new technologies have emerged for the identification of the genome locations of proteins bound to DNA in vivo. Our focus here is on the ChIP-Seq technology that employs sequencing of the paired ends of precipitated fragments. This technology has already demonstrated great potential. However, at this time tools for planning these studies are not available. We developed generalized motif models that address the locations of specific functional sites in the genome. We found that there is a sensitivity vs. resolution tradeoff on the mean fragment lengths used in these studies, with good gains in sensitivity as fragment length increases and corresponding loss in resolution. We also found that not surprisingly sensitivity depends strongly on the enrichment factors of target sites. On the other hand, the proportion of expected sites in the genome that can be identified in these studies is insensitive to the number of sites assumed in a genome.

### **Selected Publications**

- H. Lian (2007), Consistency of Bayesian estimation of a step function, *Statistics & Probability Letters*.
- H. Lian (2007), On the consistency of Bayesian function approximation using step functions. *Neural Computation*.
- H. Lian (2007), Nonlinear functional models for functional responses in reproducing kernel Hilbert spaces. *Canadian Journal of Statistics*.