Editorial
Volume 6:4, 2021

Journal of Molecular Histology and Medical Physiology

Open Access

In the Genomic Period, Statistics: Molecular Biology

Javed Khan[1*]

[1]*Editorial Office, Molecular Histology and Medical Physiology, Belgium*

Introduction

Breakthroughs in technology have considerably improved our capacity to comprehend the complicated realm of molecular biology in recent years. Rapid advances in genomic profiling techniques, like as high-throughput sequencing, have given computational biology and bioinformatics new opportunities and problems. Furthermore, several sophisticated methodologies (e.g., RNA-Seq, Chips-Seq, single-cell assays, and Hi-C) have been created by integrating genomic profiling tools with various experimental techniques in order to better investigate complicated biological systems.

The analysis of genomic datasets has become a significant problem as well as a topic of interest as more genomic datasets become available, both in terms of number and diversity. As a result, statistical approaches for addressing the issues raised by these newly created techniques are in great demand. Statistical Approaches for the Analysis of Genomic Data is a special issue of Genes that contains a number of papers that showcase state-of-the-art statistical methods for the analysis of genomic data and discuss potential avenues for development. One of the most researched subjects in genomics is gene expression. The expression levels of tens of thousands of genes may be evaluated concurrently using a variety of techniques ranging from microarrays to high-throughput transcriptome sequencing (RNA-Seq). Following the collection of such data, the first step is usually to identify genes whose expression levels are linked to experimental circumstances or results. Initial analysis can be done using two-group comparisons (also known as differential expression), linear or Cox regressions, or more sophisticated statistical models, depending on the kind of data. The initial differential expression analysis frequently reveals many potentially relevant genes due to the huge number of genes in a typical genome (e.g., 25,000 protein coding genes in the human genome). Unsupervised clustering analysis is frequently used to group genes with similar expression patterns together in order to better understand the underlying biology. The estimate mistakes in gene fold-changes during the first differential expression study are frequently disregarded in the downstream clustering analysis in current practise. The suggested model combines MCLUST's traditional Gaussian mixture clustering model with a random Gaussian measurement error assuming a given variance for each observation, and then fits the model using an extended Expectation–Maximization (EM) technique. The classification border of MCLUST-ME is determined by the distribution of measurement error for each observation, which has been demonstrated to enhance clustering efficiency in an RNA-Seq dataset on Arabidopsis thaliana.

The analysis of cancer biomedical information has long been plagued by the curse of dimensionality, since most cancer genomic studies have sample sizes of only a few hundred at most, despite the fact that tens of thousands of genetic characteristics are examined. A scientist proposes a Pathway-based Kernel Boosting (PKB) method for integrating gene pathway information for sample classification to leverage prior biological knowledge, such as pathways, and more effectively analyse cancer genomic data; the authors use kernel functions estimated from each pathway as base learners and learn the weights through an iterative optimization of the classification.

The PKB methodology utilises a second-order approximation of the loss function instead of the first-order approximation used in the traditional gradient descent boosting method, allowing for deeper descent at each step. Furthermore, the PKB incorporates two types of regularizations (L1 and L2) for selecting base learners in each iteration and outperforms other approaches when it comes to discovering routes that are important to the outcome variables.

How to cite this article: Khan Javed. In the Genomic Period, Statistics: Molecular Biology. *J Mol Hist Med Phys* 6 (2021) 20

***Address for Correspondence:** Javed K, Editorial Office, Molecular Histology and Medical Physiology, Belgium, E-mail: molecularhistologymedical@gmail.com*

Received 07 July 2021; **Accepted** 10 July 2021; **Published** 19 July 2021