A/Prof. Olivier Thas

Associate Professor in Statistics
Department of Applied Mathematics, Biometrics, and Process Control
Ghent University
and
National Institute for Applied Statistics Research Australia
School of Mathematics and Applied Statistics
University of Wollongong

Title:

The Statistical Analysis of Microbiome Data

Abstract:

The human body is inhabited by billions of bacteria, viruses and fungi on all of its outer and inner surfaces, such as oral cavity, skin and gut. Each body site has its own unique community of micro-organisms adapted to its environmental conditions. The ensemble of these communities living in and on our body parts is called the human microbiome; it contains about a ten-fold as many cells as we have human cells. The human microbiome is sometimes referred to as the “forgotten organ”. Under normal circumstances these organisms are inoffensive and even useful as they contribute to food digestion and the maturation of the immune system. On the other hand, perturbations of the normal composition of the microbiome are often observed together with diseases such as gut inflammation and diabetes. With modern genome sequencing technologies it has become possible to measure the relative abundances of thousands of micro-organisms in a biological sample.

In this presentation I will give a brief overview of statistical data analysis methods (1) to visualise the biological variability in the human microbiome, and (2) to test for differential abundance between two or more conditions (e.g. healthy against diseased subjects). In particular, I will illustrate that the classical ordination methods from ecology are not appropriate for microbiome data visualisation and I will propose a new ordination method based on log-linear models and the negative binomial distribution. In the second part of my talk I will discuss some issues that arise when testing for differential abundance. Several authors have suggested that methods for testing for differential gene expression in RNASeq experiments, can be used. However, our results demonstrate that most of these methods do not succeed in controling the false discovery rate (FDR), or they show a very small power/sensitivity. I will end with showing some preliminary results of new methods that we are developing.