Site occupancy and abundance models for analyzing multiple-visit detection/nondetection data

Huu-Dinh Huynh* (胡尼狄)

Institute of Statistics, National Chung Hsing University, Taiwan

Matthew Scofield

Department of Mathematics and Statistics, University of Otago, New Zealand

Wen-Han Hwang (黃文瀚)

Institute of Statistics, National Tsing Hua University

Abstract

We propose an enhanced site occupancy model for analyzing ecological detection/nondetection data obtained from multiple visits. The model distinguishes between abundance, occupancy, and detection probabilities. We allow for transient individuals through a community parameter c, that characterizes the proportion of individuals fixed across visits. This parameter seamlessly transitions from the standard occupancy model (c=0) to the N-mixture model (c=1), enabling a more accurate analysis of site occupancy data. Through theoretical developments and simulation studies, we demonstrate how this model effectively addresses biases inherent in conventional approaches, particularly for c is not at 0 or 1. We apply the model to various datasets of mammal and bird species, and compare it to current approaches.

回分組議程

An efficient design strategy for process optimization with categorical and numerical covariates

Yu-Jie Huang*, Mong-Na Lo Huang, ShengLi Tzeng
Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan

Kerby Shedden
Department of Statistics, University of Michigan, Ann Arbor, 48109, MI, USA

Abstract

In industrial optimization processes, managing plenty of categorical and numerical variables poses a challenge, particularly when certain variables are uncontrollable and strongly correlated. To address this issue, we propose a two-stage strategy aimed at improving product passing rates. First, we utilize a partitioning algorithm that combines the criteria of *Homogeneity* and *D-optimality* to group categorical covariates into more homogeneous blocks. Second, we develop statistical models for each block to optimize passing rates based on limited production line data. With the constructed blocked models, we refine passing rates by determining the optimal settings of the design fac- tors according to specific values of the uncontrollable covariates. Simulation studies illustrate the effectiveness of our partitioning algorithm across different subsets of categorical covariates. Furthermore, when applied to a real industrial process, the framework results in a 7.7% increase in the passing rate. This approach offers benefits by not only improving passing rates and simplifying optimal design selection but also addressing the complexities of heterogeneous materials or production lines in practical scenarios. The strategy can be extended to various industrial manufacturing processes, especially in situations involving diverse uncontrollable covariates, and holds potential for optimizing real-time production.

Keywords: parameter design, uncontrollable covariates, partition algorithm, heterogeneity, online production

回分組議程

A Geometric Algorithm for Contrastive Principal Component Analysis in High Dimension

*Rung-Sheng Lu
Department of Mathematics, National Taiwan University

Shao-Hsuan Wang
Graduate Institute of Statistics, National Central University

Su-Yun Huang
Institute of Statistical Science, Academia Sinica

Abstract

Principal component analysis (PCA) has been widely used in exploratory data analysis. Contrastive PCA (Abid et al., 2018), a generalized method of PCA, is a new tool used to capture features of a target dataset relative to a background dataset while preserving the maximum amount of information contained in the data. With high dimensional data, contrastive PCA becomes impractical due to its high computational requirement of forming the contrastive covariance matrix and associated eigenvalue decomposition for extracting leading components. In this presentation, we propose a geometric curvilinear-search method to solve this problem and provide a convergence analysis. Our approach offers significant computational efficiencies. Specifically, it reduces the time complexity from $\mathcal{O}\big((n \vee m)p^2\big)$ to a more manageable $\mathcal{O}\big((n \vee m)pr\big)$, where n,m are the sample sizes of the target data and background data, respectively, p is the data dimension and r is the number of leading components. Additionally, we streamline the space complexity from $\mathcal{O}(p^2)$, necessary for storing the contrastive covariance matrix, to a more economical $\mathcal{O}\big((n \vee m)p\big)$, sufficient for storing the data alone. Numerical examples are presented to show the merits of the proposed algorithm.

Keywords：Cayley retraction mapping; contrastive PCA; curvilinear-search; high dimension; principal component analysis; projected gradient; Stiefel manifold.

Complete Consecutive Order-Pairing Design and Its Distance-based Linear Model: Design Construction and Analysis for Order-of-Addition Experiments

Jing-Wen Huang（黃靖雯）*
National Tsing Hua University, Hsinchu, Taiwan

Frederick Kin Hing Phoa（潘建興）
Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

摘要

摘要內文 An order-of-addition (OofA) experiment aims to investigate how the order of the inputs affects the experimental response; this type of experiment has recently attracted great interest among practitioners in clinical trials and industrial processes. In this work, we introduce a new cost-efficient design called the Complete Consecutive Order-Pairing (CCOP) design, which not only considers the effects of the component order on the response, but also simultaneously considers the treatment effects due to the component levels. A new statistical model associated with the CCOP design is also proposed for identifying the optimal settings of both the component order and component level. The CCOP design method considers the effects of two successive treatments, and it takes the minimal number of runs (i.e. aims at cost efficiency), as each pair of level settings for two different factors appears exactly once. Compared to recent studies on OofA experiments, our designs handle pure order experiments and multi-level experiments well with a relatively small run size.

關鍵詞：Clinical Trials, Cost-Efficiency, Order-of-Addition Experiments