

# 國立中央大學

## 統計研究所

### 學術演講

主 講 人：黃愉閔 教授（東海大學統計系）

講 題：**Binary Surrogates with Stratification and Extra Variables**

時 間：108年11月12日（星期二） 上午11：00 ~ 12：00

地 點：中央大學鴻經館M429室

茶 會：上午 10：30 ~ 11：00      地 點：鴻經館 510 室

### ABSTRACT

The primary role of a surrogate variable is to serve as an indirect measure when it is difficult to use the major outcome variable  $X$  or to save on too many variables in assessing a disease status. In this work, we consider constructing an optimal binary surrogate  $Y$  to substitute such the feature variable  $X$ , accompanying with auxiliary information. We consider stratified samples where the strata are constructed using the disjoint intervals of  $X$ , from each of that, the data of binary status  $y$ , select input variable  $X$  and  $p$ -dimensional auxiliary variables are randomly gathered. The samples however are not IID, but come with unknown sampling weights disproportionally allocated. This talk has two parts: (1) We first present the construction of matching an optimal  $y$  (labeled  $\{0,1\}$ ) according to a cutoff value of  $X$ . We estimate the underlying distribution of  $X$  (both parametrically and nonparametrically), then the optimal cutoff is conducted from an ad-hoc estimation of the expected prediction errors with the zero-one loss. (2) The second part we study construction of a surrogate  $y$  labeled with  $\{-1,1\}$  based on the select input variable  $X$  and auxiliary variables. Our aim is to tune-up the original solved or generally recognized cutoff value by adding auxiliary information to the decision hyperplane, entailing on such knowledge. Under the stratification, we propose three probability models for the expected risk with the hinge loss. Based on these, we solve the support vector machine problems one with weighted conditional expected loss, the other two with corrections on biased selection. A QP-EM algorithm is built to solve the quadratic programming. We illustrate the methods with numerical examples.

◎敬請張貼

歡迎參加◎