中央研究院統計科學研究所
學　術　演　講

講　題：Asymptotic Properties of
High-Dimensional Random Forests

演講人：紀建名 博士

(Institute of Statistics, National Tsing Hua University, Taiwan;
Data Sciences and Operations Department, Marshall School of
Business, USC, USA.)

時　間：**2022年1月17日（星期一）上午10:30-12:00**

地　點：中央研究院統計科學研究所 **B1F** 演講廳

※茶　會：上午10：10開始

※實體與線上視訊同步進行

※英文演講

Abstract

As a flexible nonparametric learning tool, random forests has been widely applied to various real applications with appealing empirical performance, even in the presence of high-dimensional feature space. Unveiling the underlying mechanisms has led to some important recent theoretical results on the consistency of the random forests algorithm and its variants. However, to our knowledge, all existing works concerning random forests consistency under the setting of high dimensionality were done for various modified random forests models where the splitting rules are independent of the response. In light of this, in this paper we derive the consistency rates for the random forests algorithm associated with the sample CART splitting criterion, which is the one used in the original version of the algorithm in Breiman (2001), in a general high-dimensional nonparametric regression setting through a bias-variance decomposition analysis. Our new theoretical results show that random forests can indeed adapt to high dimensionality and allow for discontinuous regression function. Our bias analysis characterizes explicitly how the random forests bias depends on the sample size, tree height, and column subsampling parameter. Some limitations of our current results are also discussed.