第 32 屆南區統計研討會

暨 2023 年中華機率統計學會年會及學術研討會

時　　間：2023 年 06 月 29 日至 06 月 30 日

地　　點：東華大學理工一館 (花蓮縣壽豐鄉大學路二段 1-12 號)

主辦單位：東華大學應用數學系

協辦單位：中華機率統計學會
中央研究院統計科學研究所
中國統計學社
國科會科學推展中心-數學組
東華大學
東華大學理工學院
東部區域運輸發展研究中心

議程委員：吳漢銘 教授　政治大學統計學系
沈宗荏 教授　中興大學應用數學系
徐南蓉 教授　清華大學統計學研究所
許湘伶 教授　高雄大學統計學研究所
陳春樹 教授　中央大學統計研究所
陳美如 教授　中山大學應用數學系
陳瑞彬 教授　成功大學統計學系
黃文瀚 教授　中興大學應用數學系
黃怡婷 教授　臺北大學統計學系
黃信誠 研究員 中研院統計科學研究所
黃冠華 教授　陽明交通大學統計學研究所
黃禮珊 教授　清華大學統計學研究所
楊欣洲 研究員 中研院統計科學研究所
銀慶剛 教授　清華大學統計所
鄭宗記 教授　政治大學統計學系
蕭守仁 教授　彰化師範大學數學系
蕭維政 教授　東吳大學財務工程與精算數學系
謝進見 教授　中正大學數學系
鍾冬川 教授　逢甲大學統計系
蘇佩芳 教授　成功大學統計學系
蘇南誠 教授　臺北大學統計學系
(姓名照筆畫排序)

目　錄

歷屆南區統計研討會主辦單位暨張文豹先生講座、周元燊院士講座、
趙民德教授講座主講者 …… I

致謝 …… II

南區統計三十二年(黃文璋教授撰) …… III

會議須知 …… 1

接駁車資訊 …… 2

場地平面圖 …… 3

議程表 …… 8

講座與論文獎資訊

周元燊院士講座主講人邢泰侖博士簡介 …… 17

周元燊院士講座演講摘要 …… 18

周元燊院士講座設置辦法及院士略歷 …… 19

趙民德教授講座主講人張源俊博士簡介 …… 21

趙民德教授講座演講摘要 …… 22

趙民德教授講座設置辦法 …… 23

昔年種柳(趙民德教授撰) …… 24

中華機率統計學會魏慶榮統計論文獎設置辦法 …… 27

中華機率統計學會魏慶榮統計論文獎歷屆獲獎名單 …… 28

全球人壽演講摘要 …… 31

分組論文時間表/摘要

分組議程 SI　時間表/摘要: 2023 年 6 月 29 日 (星期四)　13:30 - 14:50 …… 33

分組議程 SII 時間表/摘要: 2023 年 6 月 29 日 (星期四)　15:20 - 16:40 …… 99

分組議程 SIII 時間表/摘要: 2023 年 6 月 30 日 (星期五)　11:00 - 12:20 …… 175

分組議程 SIV 時間表/摘要: 2023 年 6 月 30 日 (星期五) 13:30 - 14:50 ····· 237
分組議程 SV 時間表/摘要: 2023 年 6 月 30 日 (星期五) 15:00 - 16:20 ····· 301
與會人員名單 ····· 367
贊助單位廣告 ····· 379

歷屆南區統計研討會主辦單位
暨張文豹先生講座、周元燊院士講座、趙民德教授講座主講者

屆次	主辦單位	舉辦日期	張文豹先生 講座主講者	周元燊院士 講座主講者	趙民德教授 講座主講者
一	國立中山大學應用數學系	81 年 6 月 25-26 日			
二	國立中山大學應用數學系	82 年 7 月 12-13 日			
三	國立成功大學統計學系	83 年 7 月 18-19 日			
四	國立中正大學數學系	84 年 7 月 17-18 日			
五	國立中山大學應用數學系	85 年 7 月 15-16 日			
六	國立成功大學統計學系	86 年 7 月 11-12 日			
七	國立中正大學數學系	87 年 6 月 05-06 日			
八	陸軍軍官學校數學系	88 年 7 月 15-16 日	趙民德博士		
九	國立成功大學統計學系	89 年 5 月 27-28 日	魏慶榮博士		
十	國立中正大學數學系	90 年 6 月 23-24 日	韋端(韋伯韜)博士		
十一	國立中山大學應用數學系	91 年 6 月 27-28 日	李克昭博士		
十二	國立高雄大學應用數學系 (高雄金典酒店)	92 年 6 月 26-27 日	鄭清水博士		
十三	國立政治大學統計學系 (台北亞太會館)	93 年 6 月 24-25 日	黃文璋博士		
十四	國立成功大學統計學系	94 年 6 月 25-26 日	梁賡義博士		
十五	國立中正大學數學系	95 年 6 月 24-25 日	陳珍信博士		
十六	國立高雄大學統計學研究所	96 年 6 月 22-23 日	方開泰博士		
十七	國立東華大學應用數學系	97 年 6 月 27-28 日	銀慶剛博士		
十八	國立中山大學應用數學系	98 年 6 月 26-27 日	蔡偉彥博士		
十九	國立成功大學統計學系	99 年 7 月 06-07 日	刁錦寰院士		
二十	國立中正大學數學系	100 年 6 月 24-25 日	趙蓮菊博士		
二十一	輔仁大學統計資訊學系	101 年 6 月 29-30 日	蔡瑞胸院士		
二十二	國立高雄大學統計學研究所	102 年 6 月 28-29 日	陳毅恆博士		
二十三	國立東華大學應用數學系	103 年 6 月 27-28 日	程毅豪博士		
二十四	國立彰化師範大學數學系、 統計資訊研究所	104 年 6 月 27-28 日	寇星昌博士	黃顯貴博士	
二十五	國立中山大學應用數學系	105 年 6 月 24-25 日	羅小華博士	姚怡慶博士	
二十六	國立臺北大學統計學系	106 年 6 月 23-24 日	盧鴻興博士	郁彬博士	
二十七	國立成功大學統計學系	107 年 6 月 29-30 日	石瑜博士	曾勝滄博士	
二十八	國立中興大學統計學研究所	108 年 6 月 21-22 日	黃信誠博士	林共進博士	
二十九	國立中正大學數學系	109 年 8 月 20-21 日		王乃昕博士	熊昭博士
三十	國立高雄大學統計學研究所	110 年10 月 30-31 日		劉月琴博士	陳素雲博士
三十一	逢甲大學統計學系、應用數學系	111 年 7 月 28-29 日		黃瓊玉博士	王清雲博士
三十二	國立東華大學應用數學系	112 年 6 月 29-30 日		邢泰侖博士	張源俊博士
三十三	國立中山大學應用數學系	113 年 6 月			
三十四	國立台北大學統計學系	114 年 6 月			
三十五	國立成功大學統計學系	115 年 6 月			

備註：

1. 前兩屆名稱分別為"第一屆高雄區統計研討會"，"第二屆高雄區統計研討會"；有幾屆有其他研討會或學術組織共同掛名。
2. 第十二屆及第十三屆不在主辦單位校園內舉行，括號內為舉辦地點。
3. 陸軍軍官學校數學系現名為管理科學系。

致謝

東華大學應用數學系很榮幸能夠舉辦 第三十二屆南區統計研討會。 這是東華大學應用數學系第三次辦理南區統計研討會。南區統計研討會為國內統計界最盛大年度盛事。東華大學應用數學系很感謝能一直參與其中。

承蒙歷屆主辦學校辛苦的付出讓南區統計研討會能順利地傳承。特別地、因新冠疫情的緣故以至過去兩屆南區統計研討會無法如期舉行辦理。在此我們要感謝前兩屆承辦單位國立高雄大學統計學研究所與逢甲大學統計學系、應用數學系的辛勞。今年南區統計研討會共規劃 63 演講場次，議程主題相當多元。同時，大會很榮幸能邀請邢泰侖教授與張源俊教授擔任周元燊院士講座與趙民德教授講座。相信各位參與者定能滿載而歸。

主辦單位在此要感謝所有與會者的參與，東華大學、輔仁大學謝邦昌副校長與各級單位對於第 32 屆南區統計研討會的支持與補助。最後感謝系上老師 、行政人員與同學辛苦的幫忙才能讓研討會順利舉行。

最後，敬祝會議順利，各位暑假愉快。

東華大學 應用數學系敬上

南區統計三十二年

東華大學成立於民國 83 年，第 1 年便有應用數學研究所碩士班。因緣際會，東華應數籌備時期，我曾參與些微的工作，因而之後不時會留意其發展。此研究所由統計學起家，隨著大學部(民國 85 年)及博士班(民國 87 年)的快速成立，教師員額也隨之增加，領域遂逐漸增多。97 學年，東華大學與花蓮教育大學整併後，應數系更如魚得水。目前學士班分為數學科學組及統計組兩班，碩士班亦有應數及統計兩班，兵多將廣。學校位在花蓮，不負花蓮是台灣 20 縣市中面積最大之名，東華校地也是大得不得了。251 公頃已很驚人了，但這只是壽豐校區，還有在花蓮市的美崙校區，及墾丁海生館園區內的屏東校區。全校學生將近 1 萬人，以規模而言，即使在現今國內大學眾多競爭激烈下，仍頗能有番作為。

南區統計研討會，是民國 81 年中山大學所發起，初期主要是由南部的中山應數、成大統計及中正數學輪流主辦。之後參與的學校逐漸增加，早已不限南部，每年輪流在台灣各地舉行。也是成立於民國 81 年的中華機率統計學會，其年會及學術研討會，第 1 次與南區合辦，是在中山的第 11 屆(民國 91 年)。而自第 20 屆(民國 100 年)中正南區起至今，兩會每年都合辦了。這是國內統計相關領域的工作者，一年一度發表研究成果及交流的大好機會。

雖地處台灣的後山，但東華應數師生，並不只想遠離塵囂、看山看海，享受花東縱谷所孕育，豐富的取之無禁、用之不竭的自然資源。自成立起，便積極參與學術活動。東華應數首度舉辦較大規模的統計研討會，是在創所將滿 3 年，民國 86 年 7 月初的“第一屆東區統計研討會”，那是傅權教授當系主任的時代。該研討會吸引不少統計界的朋友參與，也有幾位遠從北美來的。那時東華校園已有些雛型，與會人員大部分是第 1 次到東華，莫不為東華應數所在的理工一館前，那片美不勝收、波光粼粼的湖面所吸引，對那一望無際的校園更是嘆為觀止。這是個適合思考、做學問的地方。原來東華不僅有詩人楊牧(1940–2020，於前一年(民國 85 年)來此擔任人文社會科學學院院長)，亦有一群有心的統計學者，在這裡快樂地耕耘，人文薈萃，已在東台灣營造出一統計據點。東華應數自此在統計界打出名號了。

東華應數第一次主辦南區，是民國 97 年 6 月下旬的第 17 屆。前一年周元燊院士(1924–2022)參加高雄大學統計所主辦的第 16 屆南區，眾人邀他隔年到東華，他欣然同意，也依約前往。於是東華第 1 次的南區，便有了院士的認證，這個台灣東部的統計據點更加閃亮了。周先生對台灣統計活動一向極為支持，可惜已於去年 3 月鶴駕西歸了。在南區行之有年的“張文豹先生講座”，那年主講者為花蓮子弟銀慶剛教授。東華開風氣之先，在會議的第 2 天，安排大陸院士陳木法教授為大會主講，這是南區首次有兩位主講。而自第 24 屆(民國 104 年)彰師大統資所的南區起，每年便都固定有兩位講座了。東華應數第 2 次主辦南區，是民國 103 年 6 月下旬的第 23 屆。那年的“張文豹先生講座”是程毅豪教授。有趣的是，銀、程兩位年輕教授，均於擔任東華南區講座的隔年，即獲國科會(科

技部)傑出研究獎，也算是佳話。前兩次主辦南區，運籌帷幄的皆是曹振海教授。今年是東華應數的第 3 次南區，世代交替，已輪由吳韋瑩教授掌舵了。

統計雖有很長一段欣欣向榮的時期，唯近年來對於碩士班，有若干時尚領域興起，且產業界的需求亦有些改變；至於博士班，則可能因教職覓得不易，且大學新進教師壓力日益增大，導致統計碩博士班對學生的吸引力，強度似不若以往，相信東華應數也感受到了。過去 50 餘年，在華人機率統計界，長期居領導地位的學者，除周元燊院士外，尚有刁錦寰院士(1933–)。他於今年 4 月出版回憶錄“生逢其時”，描述其由英國、大陸、台灣至美國的成長及求學經過，與學術生涯事蹟，大院士的真情告白，對今日研究所之經營頗具參考價值。

刁先生大學就讀台大經濟系，服完 1 年兵役後，民國 45 年(1956)前往美國紐約大學(New York University)的經濟系深造，但去後考慮到將來的出路，即轉到商學院就讀。以兩年取得碩士學位後，在紐約的銀行上班。又為了能多存些錢，曾有 4 個月晚上到餐館兼差，直至太太生產後才停止。

當時 MBA 的待遇不到理工碩士之半，讓刁先生深感挫折。在友人建議下，決定去攻讀博士學位。於是工作 1 年後，1959 年秋天，他到中西部威斯康辛大學麥迪遜分校(University of Wisconsin-Madison)商學院國際金融系的博士班。由於系上要求副修第 2 專長，雖只曾在大二時修過 1 門統計學，在學長建議下，他選擇統計，並至數學系修習數理統計，第 2 年又修高等微積分。又由於第 1 年所修的幾門經濟學方面之課程，表現均極為突出，在開課教授之鼓勵下，第 2 年他轉到老本行經濟系，且在這年課程結束後，通過經濟系的資格考。當時的系主任 H. Edwin Young 教授(1917–2012，1968–1977 年擔任威大校長)，還給他一封信，稱讚他是系上 20 年來(1941–1961)最優秀的學生。

刁先生在第 2 年選修剛來威大創辦統計系的 George E. P. Box (1919–2013)教授開設之兩門統計課程中，認識了機械系博士生，大他 10 歲的吳賢銘(1923–1992，1982 年當選中研院院士)，因其大力建議，故雖在經濟系遊刃有餘，仍於第 3 年跟隨 Box 教授寫博士論文。自此跨足統計界，且於 1 年後(1962)便完成博士學位，隨即留校在統計系任教。第 1 次教書，且就講授博士班的重頭戲—數理統計，刁先生如履薄冰，相當用心地準備，後來他的講義，成為系上此門課之教材。因半路出家，為快速補充統計知識，教書第 1 年，刁先生還去旁聽多變量分析、變異數分析與實驗設計，及決策理論等課程。而因曾待過經濟、商學及統計等學系，並轉換自如，凡走過便留下痕跡，使刁先生易於與不同領域的人合作，且勇於嘗試，即使主題陌生的產業界計畫，也敢大膽承接。

刁先生樂於服務，1987 年 8 月，促成“泛華統計協會”的成立，並擔任首任會長；國際一流統計期刊“中華統計學誌”(Statistica Sinica)，首期於 1991 年出版，刁先生為首任主編。不僅如此，刁先生畢生於美國、台灣及大陸，完成很多開創性的工作、成立或發起多個組織及研討會，也經常受邀赴世界各地開設統計課程。他曾獲得很多榮譽，包括講座、會士，及院士頭銜，獎章也不少，還有兩個榮譽博士學位，在學術界備受推崇。而雖工作異常忙碌，但刁先生家庭亦照

顧得很好，回憶錄中特闢一章，談“六口之家：婚姻生活與子女教養”。刁先生以旺盛的精力，永不止息地做他認為該做的事，頗值得吾等後生晚輩效法。

回憶錄中亦提到，民國 7、80 年代，他如何規畫設計，使統計碩士訓練更具競爭力。經過 3、40 多年，不論產業及學術界，各方面都有很大的變化，今日當然得有新的作法。在刁先生的自述中，求學及工作過程中，曾多次面臨抉擇，那時待遇及出路都可能是考量，初任教時且感到學然後知不足。連大院士年輕時都無法不考慮現實問題，此對今日國內統計所須重視給學生更好的學習環境，包括課程設計、獎助學金之提供，及職前訓練等，應能有若干啟發。刁先生向來不躑躅不前，他坦承關鍵時刻都有貴人指點迷津。此既顯示刁先生樂於察納雅言，亦顯示年輕人在成長過程中，不時需有前輩供諮詢，這也是辦學者宜留意的。

刁先生雖成就非凡，但回憶錄之書名卻謙虛地取為“生逢其時”。只是由他做成那麼多事觀之，較可能是他擅長感受到時機的來臨，且知把握，當然亦可能是他善於創造時機。今日統計系所的經營，早已不能守株待兔，而是到了需與時俱進，隨時掌握時機，必要時並得創造時機的階段。

往昔博學強記、博古通今都是文人的極大優點。皇帝欣賞畫作時，即使僅含混地講出幾個字，身邊大臣常能快速念出皇帝想要的詩詞。但今天只需知道幾個關鍵詞，一 Google，整首詩詞或文章便都出來了。另外，5 世紀時，我國數學家祖沖之(429−500)將圓周率 π 計算到小數點後 7 位，這在當時是獨步世界的成就。但隨著計算機技術之快速發展，至民國 111 年 6 月，已能給出 π 到小數點後 100 兆位了。換句話說，今日記性好或擅長計算，可能都不是什麼了不起的長處了，只要懂得善用工具卻可。而原本 Google 後，有時得從出現的一長串資訊中，加以挑選所要，去年 11 月 ChatGPT 橫空出世，其功能更強大了。回答問題、編故事、寫文章、寫程式，甚至提供交友方案等，功能強大。雖帶來不小便利，但也屢引起恐慌，如是否有些工作會被取代？教學現場該如何應變？其實歷史上，新工具就是不斷產生，而自 18 世紀工業革命興起後，就已出現以機器取代人了。因此與其擔憂，還不如努力使自己生逢其時。

東華應數明年將迎來成立 30 周年。在“說文解字”裡，“三十年為一世”。似乎才剛成立，但轉眼已快一世了。少年子弟江湖老，本屆南區的與會人員中，在東華第 1 年便曾來過的，應相當稀少。而曾來參加“第一屆東區統計研討會”者，雖也不會太多，但應有一些。看著今日東華已成為如此具規模的一所大學，撫今追昔，這其中不知有多少人的心血！我們該對東華大學，尤其是應數系歷來師生表示敬意。

最後預祝本研討會圓滿成功，且東華應數系持續向上提昇，讓東部這個統計據點，成為一顆統計明珠。

國立高雄大學統計學研究所

黃文璋敬上 民國 112 年 6 月

會議須知

1. 研討會原則上不可使用自備電腦，請發表人將發表內容存為 PowerPoint 檔或 pdf 檔，存放於隨身碟中帶至現場。電腦作業系統為 Windows，存放隨身碟請注意檔案格式，並提早交給現場工作人員測試，若發表人有特殊需求，請提早告知，以利提早準備。

2. 大會提供行李寄放服務，行李房（理工一館 B101）開放時間為：
   6/29（四）8：30 - 16：50 ； 6/30（五）8：30 - 16：30。
   ** 不提供**隔夜**及**貴重物品**寄放，貴重物品請務必自行保管 **

3. 6/29（四）晚宴資訊：
   飯店：翰品酒店花蓮 B1 。
   地址：花蓮市永興路 2 號 （電話：03-8235388）。
   ◎ 晚宴請以名牌內之**晚餐餐券**作為晚宴入場資格，請務必攜帶。
   ◎ 因會議人數眾多，場地有限，**請務必依晚餐餐券所標示之廳別入座**。

4. 請有獲得交通及住宿費補助者，於 6/29（四）及 6/30（五） 11:00～14:00 時段，攜帶符合規定之「單據」至 理工一館 A 棟 1F 指定櫃檯辦理。

5. 交通資訊：
   ◎校內交通：
   (1) 車輛行駛本校校園，請遵循校區速限 30 公里，減速慢行並優先禮讓行人。
   (2) 本校校門進出車輛以「車牌辨識系統」管理，行經校門經系統辨識車牌後進出，行駛專用車道並減速慢行、注意路口左右來車及行人。未於期限內申請者依本校「車輛出入校園管理及清潔維護費收費辦法」進行收費。
   (3) 訪客機車依規定不得進入校園，請將車輛停放於校門口之機踏車停車場；自行車可直接入校，請行駛專用車道。

   ◎計程車：
   從花蓮火車站搭計程車跳表價格約 400-500 元不等，車程約 20-30 分鐘。東華大學特約計程車行：(1)中美計程車：(03)8234432、(03)8237011；(2)名星汽車行：0800432432；(3)國風汽車行：0800097097；(4) 仟合計程車：(03)8353600。

   ◎本會交通接駁車資訊，請參閱手冊內「接駁車資訊」。
   ◎更多交通資訊請參閱

如何到東華

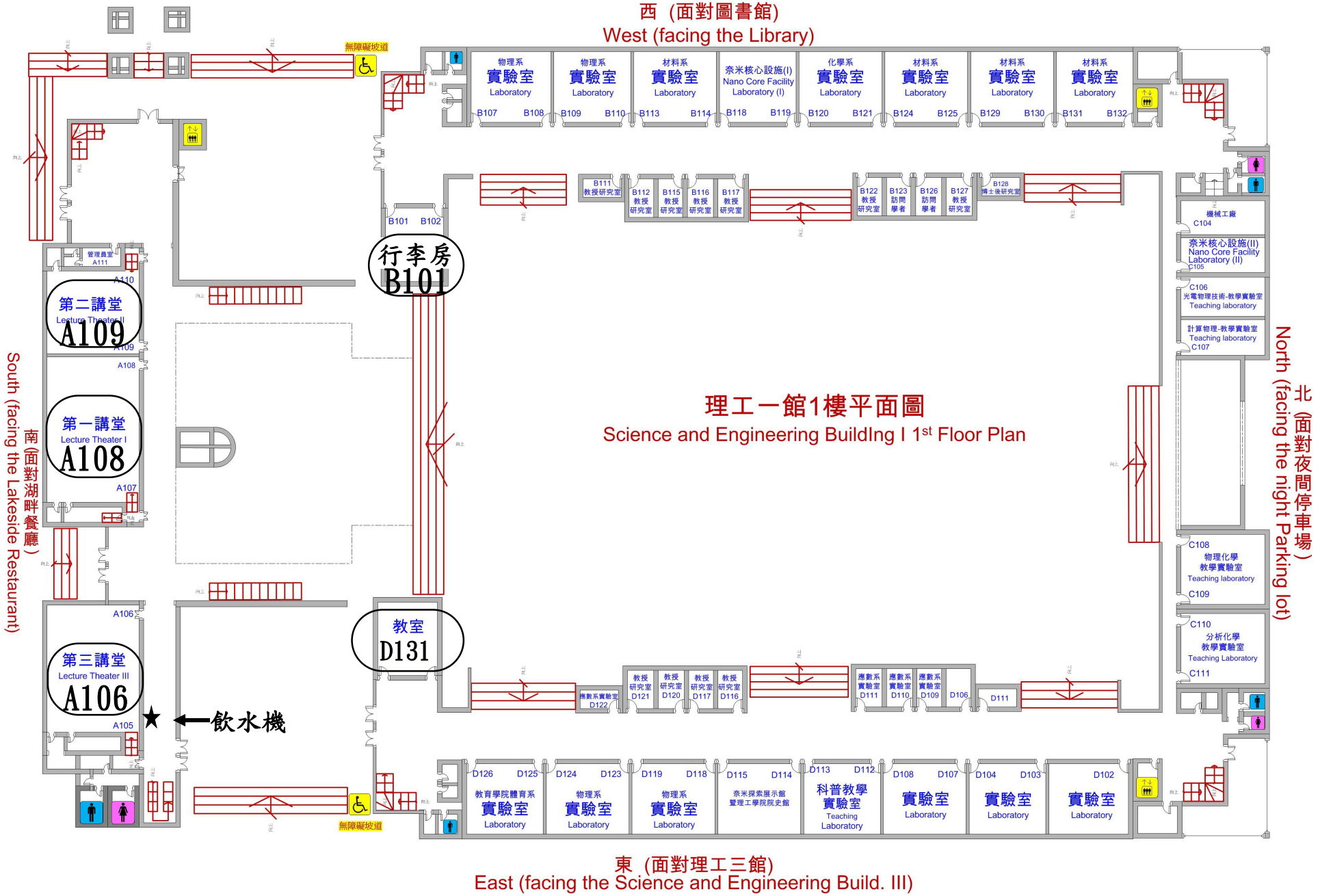
接駁車資訊

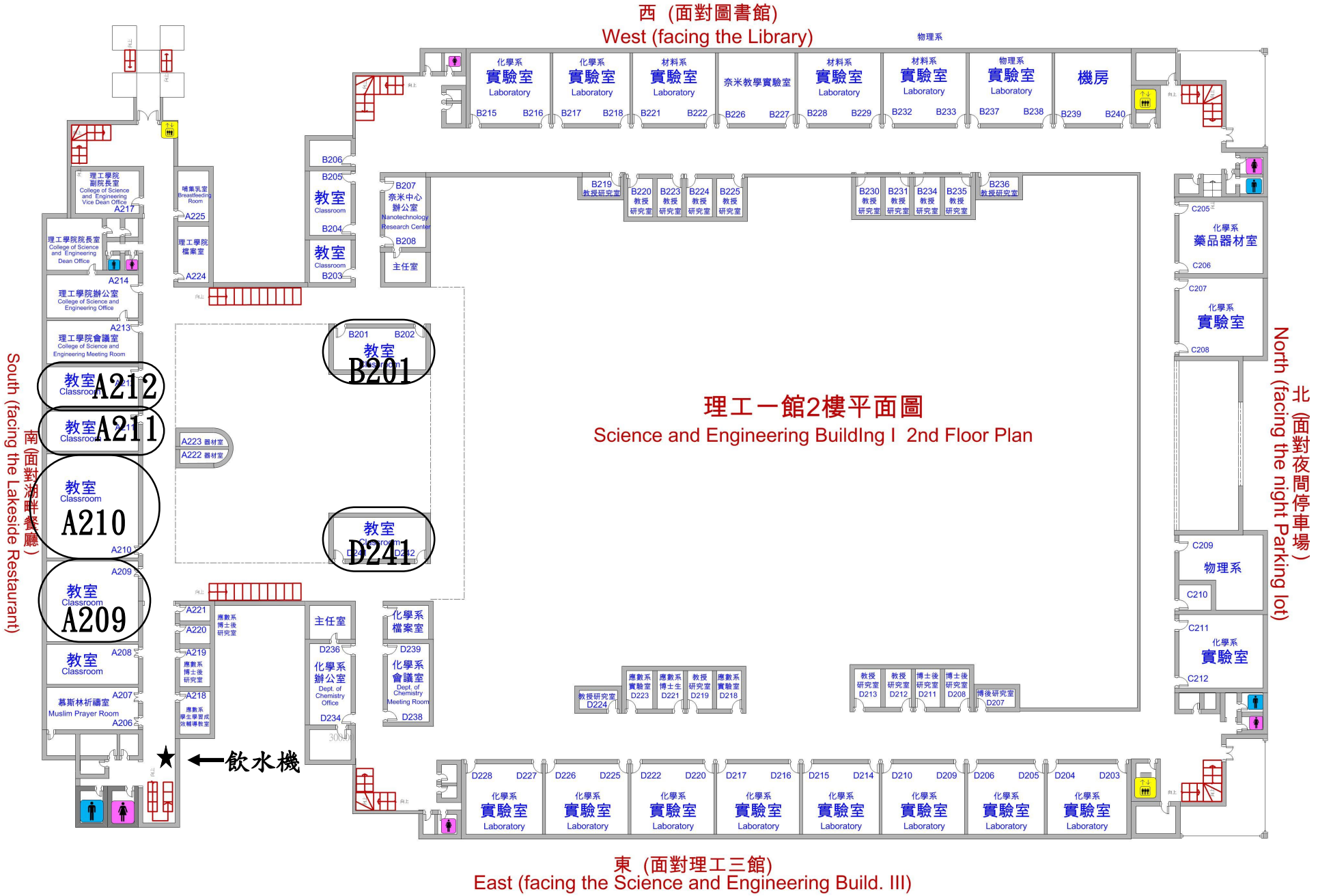
日期	發車時間	起點	終點
6/28(三)	16：20	花蓮火車站前站	大會特約飯店 (福容、麗軒、翰品、美侖)
6/29(四)	08：15	大會特約飯店 (麗軒、翰品)	東華大學 (理工一館旁大型停車場)
6/29(四)	08：15	大會特約飯店 (福容、美侖)	東華大學 (理工一館旁大型停車場)
6/29(四)	08：30	花蓮火車站前站	東華大學 (理工一館旁大型停車場)
6/29(四)	17：00	東華大學 (理工一館旁大型停車場)	晚宴會場(翰品酒店)
6/29(四)	20：50	晚宴會場 (翰品酒店)	東華大學 (理工一館旁大型停車場)
6/29(四)	20：50	晚宴會場 (翰品酒店)	大會特約飯店 (美侖、福容)
6/30(五)	08：15	大會特約飯店 (麗軒、翰品)	東華大學 (理工一館旁大型停車場)
6/30(五)	08：15	大會特約飯店 (福容、美侖)	東華大學 (理工一館旁大型停車場)
6/30(五)	16：00	東華大學 (理工一館旁大型停車場)	志學火車站
6/30(五)	16：40	東華大學 (理工一館旁大型停車場)	大會特約飯店 (福容、麗軒、翰品、美侖)
6/30(五)	16：40	東華大學 (理工一館旁大型停車場)	花蓮火車站前站

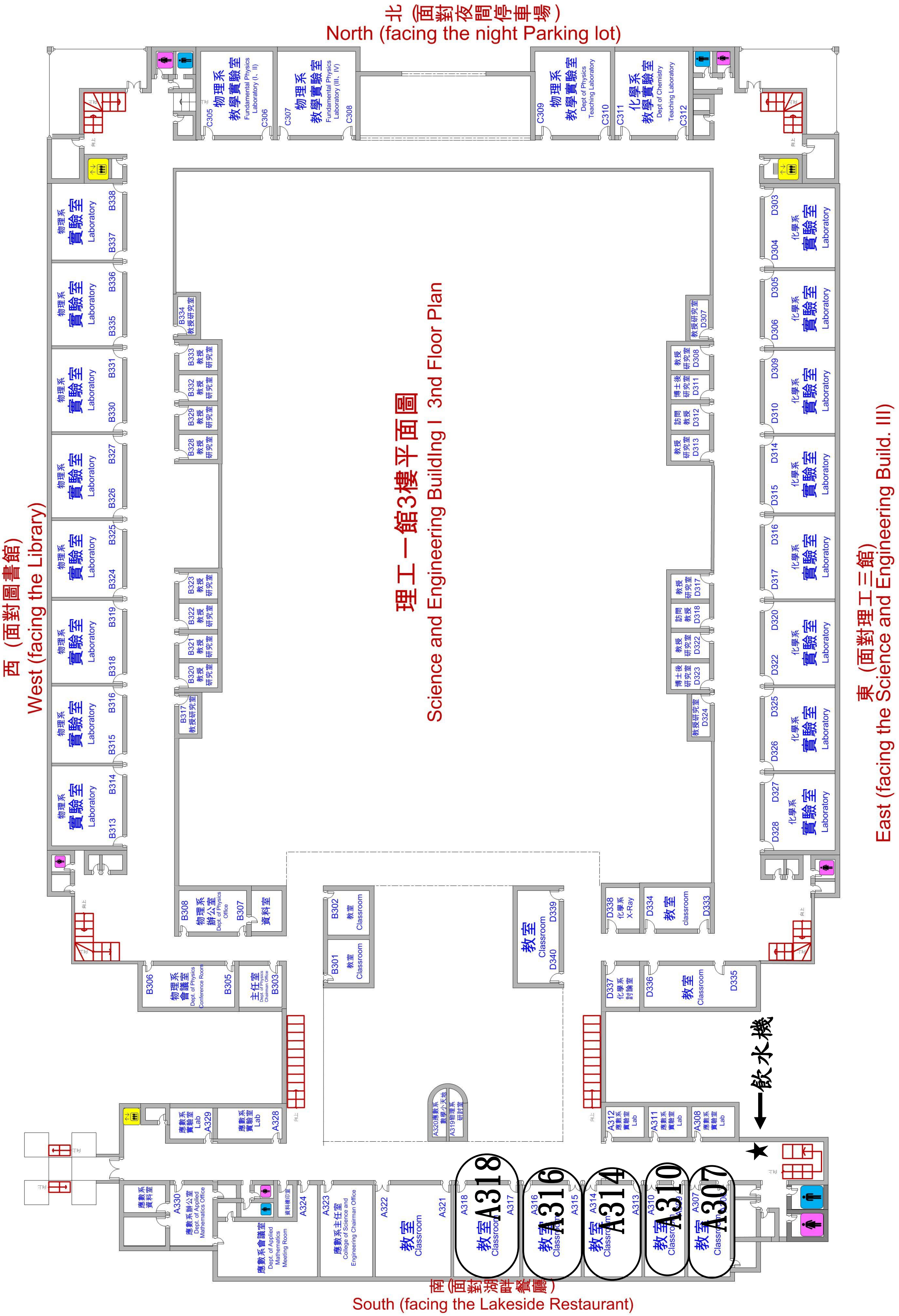
備註：

1. 東華大學距花蓮火車站(含花蓮市區、特約飯店)約 30-40 分鐘車程。
2. 大會特約飯店行駛路線之順序，將依當天交通實際況狀遶行。
3. 請提前至指定地點等候，逾時不候，敬請諒解。

國立東華大學校園導覽
行政區 Administrative Zone
服務區 Amenity
教學區 Academic Zone
宿舍區 Dormitory
校園警衛室 Campus Security Office
綜合停車場 Parking
廁所 Toilet
餐廳 Restaurant
主要方向指標 Sign Post
校園位置平面圖 Campus Map
公車停靠站 Bus Shop
景觀區 Scenic Zone
接駁車
報到地點
開幕式 大會演講
行政大樓
藝術學院
人社三館
人社一館
人社二館
花師教育學院
管理學院
學生活動中心
圖書館
資訊中心
湖畔餐廳
理工一館
理工二館
理工三館
育成中心
東湖
華湖
小華湖
體育館
運動場
游泳池
向晴莊
涵星莊
仰山莊
多容館
擷雲莊
行雲莊
集賢館
沁月莊
迎曦莊
藝術工坊
東華會館
原住民民族學院
環境暨海洋學院
綜合球場 1
綜合球場 2
壘球場
棒球場
棒、壘球場
志學門
志學門機車入口
往壽豐鄉(壽豐火車站)
往志學村(火車站)
機車停車場
汽車停車場
停車場
志學街
芳華道
翠林道
吳全道
平和道
伊人橋
景觀橋
天月橋
原住民民族學院
社區中心
社區球場 1
附設幼兒園
擺 賽心築


西（面對圖書館）
West (facing the Library)
理工一館1樓平面圖
Science and Engineering BuildIng I 1st Floor Plan
南（面對湖畔餐廳）
South (facing the Lakeside Restaurant)
北（面對夜間停車場）
North (facing the night Parking lot)
東（面對理工三館）
East (facing the Science and Engineering Build. III)
第一講堂 Lecture Theater I A108
第二講堂 Lecture Theater II A109
第三講堂 Lecture Theater III A106
行李房 B101
教室 D131
飲水機
無障礙坡道
物理系實驗室 Laboratory
材料系實驗室 Laboratory
奈米核心設施(I) Nano Core Facility Laboratory (I)
化學系實驗室 Laboratory
奈米核心設施(II) Nano Core Facility Laboratory (II)
機械工廠 C104
光電物理技術-教學實驗室 Teaching laboratory C106
計算物理-教學實驗室 Teaching laboratory C107
物理化學教學實驗室 Teaching laboratory C108
分析化學教學實驗室 Teaching Laboratory C110
科普教學實驗室 Teaching Laboratory
教育學院體育系實驗室 Laboratory
奈米探索展示館暨理工學院院史館


西（面對圖書館）
West (facing the Library)
物理系
化學系 實驗室 Laboratory B215 B216
化學系 實驗室 Laboratory B217 B218
材料系 實驗室 Laboratory B221 B222
奈米教學實驗室 B226 B227
材料系 實驗室 Laboratory B228 B229
材料系 實驗室 Laboratory B232 B233
物理系 實驗室 Laboratory B237 B238
機房 B239 B240
B206
B205 教室 Classroom
B204 教室 Classroom B203
B207 奈米中心 辦公室 Nanotechnology Research Center
B208 主任室
B219 教授研究室
B220 教授 研究室
B223 教授 研究室
B224 教授 研究室
B225 教授 研究室
B230 教授 研究室
B231 教授 研究室
B234 教授 研究室
B235 教授 研究室
B236 教授研究室
C205
化學系 藥品器材室
C206
C207
化學系 實驗室
C208
C209
物理系
C210
C211
化學系 實驗室
C212
理工學院 副院長室 College of Science and Engineering Vice Dean Office A217
哺集乳室 Breastfeeding Room A225
理工學院檔案室 A224
理工學院院長室 College of Science and Engineering Dean Office A214
理工學院辦公室 College of Science and Engineering Office A213
理工學院會議室 College of Science and Engineering Meeting Room A212
教室 Classroom A212
教室 Classroom A211
教室 Classroom A210
教室 Classroom A209
教室 Classroom A208
慕斯林祈禱室 Muslim Prayer Room A207 A206
B201 B202 教室 Classroom B201
教室 Classroom D241
A223 器材室
A222 器材室
A221
A220 應數系 博士後 研究室
A219 應數系 博士後 研究室
A218 應數系 學生學習成效輔導教室
主任室 D236
化學系 辦公室 Dept. of Chemistry Office D234
化學系 檔案室 D239
化學系 會議室 Dept. of Chemistry Meeting Room D238
理工一館2樓平面圖
Science and Engineering BuildIng I 2nd Floor Plan
教授研究室 D224
應數系 實驗室 D223
應數系 博士生 D221
教授 研究室 D219
應數系 實驗室 D218
教授 研究室 D213
教授 研究室 D212
博士後 研究室 D211
博士後 研究室 D208
博後研究室 D207
D228 D227 化學系 實驗室 Laboratory
D226 D225 化學系 實驗室 Laboratory
D222 D220 化學系 實驗室 Laboratory
D217 D216 化學系 實驗室 Laboratory
D215 D214 化學系 實驗室 Laboratory
D210 D209 化學系 實驗室 Laboratory
D206 D205 化學系 實驗室 Laboratory
D204 D203 化學系 實驗室 Laboratory
★ ←飲水機
南（面對湖畔餐廳）
South (facing the Lakeside Restaurant)
北（面對夜間停車場）
North (facing the night Parking lot)
東（面對理工三館）
East (facing the Science and Engineering Build. III)


北（面對夜間停車場）
North (facing the night Parking lot)
理工一館3樓平面圖
Science and Engineering Building I 3nd Floor Plan
西（面對圖書館）
West (facing the Library)
東（面對理工三館）
East (facing the Science and Engineering Build. III)
南（面對湖畔餐廳）
South (facing the Lakeside Restaurant)
飲水機
C305 物理系 教學實驗室 Fundamental Physics Laboratory (I、II) C306
C307 物理系 教學實驗室 Fundamental Physics Laboratory (III、IV) C308
C309 物理系 教學實驗室 Dept of Physics Teaching Laboratory C310
C311 化學系 教學實驗室 Dept of Chemistry Teaching Laboratory C312
B338 物理系 實驗室 Laboratory B337
B336 物理系 實驗室 Laboratory B335
B331 物理系 實驗室 Laboratory B330
B327 物理系 實驗室 Laboratory B326
B325 物理系 實驗室 Laboratory B324
B319 物理系 實驗室 Laboratory B318
B316 物理系 實驗室 Laboratory B315
B314 物理系 實驗室 Laboratory B313
B334 教授研究室
B333 教授研究室
B332 教授研究室
B329 教授研究室
B328 教授研究室
B323 教授研究室
B322 教授研究室
B321 教授研究室
B320 教授研究室
B317 教授研究室
B308 物理系辦公室 Dept of Physics Office
B307 資料室
B306 物理系會議室 Dept of Physics Conference Room
B305 主任室 Dept of Physics Chairman Office B303
B302 教室 Classroom
B301 教室 Classroom
D303 化學系 實驗室 Laboratory D304
D305 化學系 實驗室 Laboratory D306
D309 化學系 實驗室 Laboratory D310
D314 化學系 實驗室 Laboratory D315
D316 化學系 實驗室 Laboratory D317
D320 化學系 實驗室 Laboratory D322
D325 化學系 實驗室 Laboratory D326
D327 化學系 實驗室 Laboratory D328
D307 教授研究室
D308 教授研究室
D311 博士後研究室
D312 訪問教授
D313 教授研究室
D317 教授研究室
D318 訪問教授
D322 教授研究室
D323 博士後研究室
D324 教授研究室
D338 化學系 X-Ray
D334 教室 classroom D333
D339 教室 Classroom D340
D337 化學系討論室
D336 教室 Classroom D335
A312 應數系實驗室 Lab
A311 應數系實驗室 Lab
A308 應數系實驗室 Lab
A329 應數系實驗室 Lab
A328 應數系實驗室 Lab
A320應數系 實習小天地 A319物理系研討室
A330 應數系辦公室 Dept. of Applied Mathematics Office
應數系資料室
應數系會議室 Dept. of Applied Mathematics Meeting Room
資料櫃印室
A324
A323 應數系主任室 College of Science and Engineering Chairman Office
A322 教室 Classroom A321
A318 教室 Classroom A317
A316 教室 Classroom A315
A314 教室 Classroom A313
A310 教室 Classroom
A307 教室 Classroom

第三十二屆南區統計研討會暨
2023中華機率統計學會年會及學術研討會

議程表

112 年 6 月 29 日(星期四)

09:10 | 09:50	報到 (理工一館-第一講堂)					
10:00 | 10:40	開幕式 (學生活動中心-演講廳A101)					
10:40 | 11:40	周元燊院士講座 講者: 邢泰侖 教授 (University of Michigan) 主持人: 陳春樹 教授 (學生活動中心-演講廳A101)					
11:40 | 12:00	碩士論文短講 (學生活動中心-演講廳A101)					
12:20 | 13:20	午餐&中華機率統計學會會員大會(理工一館-第二講堂A109)					
場次 場地	SI-1 理A310	SI-2 理B201	SI-3 理A212	SI-4 理A109	SI-5 理A209	SI-6 理A211
SI 13:30 | 14:50	魏慶榮 統計論文獎	空間統計 方法與應用	Statistical Quality Management	生物統計	計算統計	計量分析的 新進展
		Invited session	Invited session	Invited session	Invited session	Invited session
	Chair : 潘建興	Organizer/ Chair : 鄭宗記	Organizer/ Chair : 楊素芬	Organizer/ Chair : 蘇佩芳	Organizer/ Chair : 林宗儀	Organizer/ Chair : 銀慶剛
	Speaker: 1. 董奕賢 2. 邱詠惠 3. 潘彥碩	Speaker: 1. 余清祥 2. 陳怡如 3. 張軒瑜	Speaker: 1. 周珮婷 2. 呂明哲 3. 李俊毅	Speaker: 1. 曾議寬 2. 張玉媚 3. 溫淑惠	Speaker: 1. 陳泰賓 2. 王婉倫 3. 楊郁成	Speaker: 1. 楊睿中 2. 郭漢豪 3. 俞淑惠

第三十二屆南區統計研討會暨
2023中華機率統計學會年會及學術研討會

議程表

112 年 6 月 29 日(星期四)

09:10 | 09:50	報到 (理工一館-第一講堂)						
10:00 | 10:40	開幕式 (學生活動中心-演講廳A101)						
10:40 | 11:40	周元燊院士講座 講者: 邢泰侖 教授 (University of Michigan) 主持人: 陳春樹 教授 (學生活動中心-演講廳A101)						
11:40 | 12:00	碩士論文短講 (學生活動中心-演講廳A101)						
12:20 | 13:20	午餐&中華機率統計學會會員大會(理工一館-第二講堂A109)						
場次 場地	SI-7 理A318	SI-8 理A314	SI-9 理A316	SI-10 理A210	SI-11 理A106	SI-12 理D131	SI-13 理A307
SI 13:30 | 14:50	生物醫學統計	跨領域(大氣科學)	生醫統計方法	International session - Probability and Statistical Mechanics	生物演化與疾病預測	Contributed session - A	Contributed session - B
	Invited session	Invited session	Invited session	Invited session	Invited session		
	Organizer /Chair：張中	Organizer/ Chair：鄭芳怡	Organizer/ Chair：鄒宗山	Organizer/ Chair：Lung-Chi Chen	Organizer/ Chair：Wei-Wen Hsu	Chair：高正雄	Chair：黃逸輝
	Speaker: 1. 馬瀰嘉 2. 張升懋 3. 李宗翰	Speaker: 1. 柯建盈 2. 林遠見 3. 陳柏孚 4. 徐佩君	Speaker: 1. 戴安順 2. 王偉銘 3. 簡莉珠	Speaker: 1. Cheuk-Yin Lee 2. Yuki Chino 3. Yoshinori Kamijima	Speaker: 1. 張憶壽 2. Wei-Wen Hsu 3. 鍾冬川	Speaker: 1.趙伯昇 2.郭冠麟 3.林家弘 4.馬繡嬪 5.王薇淳	Speaker: 1.黃逸輝 2.葉詠富 3.鄭亦甯 4.羅祖佑 5.王嘉顥 6.莊沅蓉

14:50 | 15:20	茶會						
場次 場地	SII-1 理A310	SII-2 理B201	SII-3 理A212	SII-4 理A109	SII-5 理A209	SII-6 理A211	SII-7 理A318
SII 15:20 | 16:40	網絡分析	應用統計新發展	醫療大數據之深度學習與統計分析	統計智慧與精準醫療	統計方法與應用	非參數和半參數統計方法	應用機率
	Invited session	Invited session	Invited session	Invited session	Invited session	Invited session	Invited session
	Organizer/ Chair： 郭鋸霖	Organizer/ Chair： 沈宗荏	Organizer/ Chair： 黃冠華	Organizer/ Chair: 陳君厚 Co-Organizer: 李丕強	Organizer/ Chair: 謝進見	Organizer/ Chair： 黃禮珊	Organizer/ Chair： 蕭守仁
	Speaker: 1. 潘建興 2. 李政德 3. 郭鋸霖	Speaker: 1. 林宗儀 2. 袁子倫 3. 洪宗乾	Speaker: 1. 陳泰賓 2. 許巍嚴 3. 黃冠華	Speaker: 1. 何之行 2. 楊欣洲 3. 蕭自宏	Speaker: 1. 周孟穎 2. 蔡嘉仁 3. 謝進見	Speaker: 1. 江金倉 2. 張馨文 3. 林嘉韡	Speaker: 1. 王家禮 2. 陳美如 3. 洪芷漪
17:00	接駁車發車(東華大學前往晚宴會場)						
18:00 | 20:40	晚宴 （花蓮翰品酒店）						
20:50	接駁車發車(晚宴會場前往住宿飯店及東華大學)						

14:50 | 15:20	茶會						
場次 場地	SII-8 理A314	SII-9 理A316	SII-10 理A210	SII-11 理A106	SII-12 理D241	SII-13 理D131	SII-14 理A307
SII 15:20 | 16:40	電信資料的相關研究與應用	Statistical Methods for Functional Observations	衛星資料	International session (CIPS&KSS)	統計於產業實務暨政策運用	Contributed Session - C	Contributed Session - D
	Invited session	Invited session	Invited session	Invited session	Invited session		
	Organizer: 黃怡婷 Chair: 王鴻龍	Organizer/ Chair： 江其衽	Organizer/ Chair： 黃文瀚	Organizer: CIPS Chair： Tsai-Hung Fan	Organizer/ Chair: 陳馨蕙	Chair： 許湘伶	Chair： 施銘杰
	Speaker: 1. 賴威宇 2. 柯怡瑄 3. 王鴻龍	Speaker: 1. 李百靈 2. 李國榮 3. 陳裕庭	Speaker: 1. 劉小菁 2. 黃成勇 3. 劉正彥	Speaker: 1. Chun-Shu Chen 2. Sheng Li Tzeng 3. Chae Young Lim	Speaker: 1. 林金龍 2. 彭俊能 3. 陳馨蕙	Speaker: 1.林登右 2.江彥頡 3.邱士軒 4.林亮言 5.蔡茜婷 6.曾羿文	Speaker: 1.褚育誠 2.陳芃辰 3.許育禎 4.陳彥霖 5.蘇家瑩 6.黃意婷
17:00	接駁車發車(東華大學前往晚宴會場)						
18:00 | 20:40	晚宴 （花蓮翰品酒店）						
20:50	接駁車發車(晚宴會場前往住宿飯店及東華大學)						

便當及茶會領取地點：理工一館 A 棟一樓。
用餐空間：理 A108、理 A106、理 D131、理 A307。
接駁車乘車處：理工一館旁大停車場。

第三十二屆南區統計研討會暨
2023中華機率統計學會年會及學術研討會

議程表

112 年 6 月 30 日(星期五)

09:00 | 09:20	報到 (理工一館-第一講堂)					
09:30 | 10:30	趙民德教授講座 講者: 張源俊 教授 (Academia Sinica) 主持人: 曹振海 教授 (學生活動中心-演講廳A101)					
10:40 | 11:00	茶會					
場次 場地	SIII-1 理A310	SIII-2 理B201	SIII-3 理A212	SIII-4 理D241	SIII-5 理A209	SIII-6 理A211
SIII 11:00 | 12:20	可靠度衰變資料分析	Detection of implicit patterns in data	實驗設計	空間統計與資訊	International session (CIPS&KSS)	International session
	Invited session	Invited session	Invited session	Invited session	Invited session	Invited session
	Organizer/ Chair： 李宜真	Organizer/ Chair： 曾聖澧	Organizer/ Chair： 陳瑞彬	Organizer/ Chair： 詹大千	Organizer/ Chair： Cathy W. S. Chen	Organizer/ Chair： Henry Horng-Shing Lu
	Speaker: 1. 王義富 2. 董弘平 3. 鄭雅珊	Speaker: 1. 黃信誠 2. 黃世豪 3. 鍾思齊	Speaker: 1. 陳瑞彬 2. 蔡風順 3. 蔡碧紋	Speaker: 1. 唐嘉宏 2. 陳建州 3. 林柏丞	Speaker: 1. Cathy W. S. Chen 2. Yondai Kim 3. Keun Baik Lee	Speaker: 1. Shaw-Hwa Lo 2. Inchi Hu 3. Jacky Chung-Hao Wu
12:20 | 13:30	午餐 & 國科會統計學門(理工一館-第二講堂A109) & 全球人壽演講 (理工一館-第三講堂A106) 講者 ： 謝明華 博士 主持人：孫立憲 教授					

第三十二屆南區統計研討會暨
2023中華機率統計學會年會及學術研討會

議程表

112 年 6 月 30 日(星期五)

09:00 | 09:20	報到 (理工一館-第一講堂)						
09:30 | 10:30	趙民德教授講座 講者: 張源俊 教授 (Academia Sinica) 主持人: 曹振海 教授 (學生活動中心-演講廳A101)						
10:40 | 11:00	茶會						
場次 場地	SIII-7 理A318	SIII-8 理A314	SIII-9 理A316	SIII-10 理A210	SIII-11 理A106	SIII-12 理A109	SIII-13 理A307
SIII 11:00 | 12:20	工業統計	財務計量	象徵性資料分析與應用	應用統計	Contributed Session - E	Contributed Session - F	Contributed Session - G
	Invited session	Invited session	Invited session	Invited session			
	Organizer/ Chair： 蘇南誠	Organizer/ Chair： 黃士峰	Organizer/ Chair： 吳漢銘	Organizer/ Chair： 鍾冬川	Chair： 鄭順林	Chair： 林聖軒	Chair： 鄭宗琳
	Speaker: 1. 許湘伶 2. 李名鏞 3. 黃偉恆	Speaker: 1. 邱海唐 2. 陳亭甫 3. 翁新傑	Speaker: 1. 吳漢銘 2. 高君豪 3. 林良靖	Speaker: 1. 陳秀雯 2. 劉峰旗 3. 郭珈妤	Speaker: 1.鄭順林 2.吉怡宗 3.林楷崙 4.吳紫豪 5.劉家銘	Speaker: 1.柯鋅瑀 2.張韶恩 3.陳昱豪 4.謝綸豪 5.高興翰	Speaker: 1.賴暐婷 2.古裕彥 3.Giap Van Su
12:20 | 13:30	午餐 & 國科會統計學門(理工一館-第二講堂A109) & 全球人壽演講 (理工一館-第三講堂A106) 講者 ： 謝明華 博士 主持人：孫立憲 教授						

場次 場地	SIV-1 理A310	SIV-2 理B201	SIV-3 理A212	SIV-4 理D241	SIV-5 理A209	SIV-6 理A211.
SIV 13:30 | 14:50	計量經濟	計量經濟學 理論與應用	空間統計方法 及應用	生物統計	應用機率	高維度資料 分析
	Invited session	Invited session	Invited session	Invited session	Invited session	Invited session
	Organizer/ Chair： 黃彥棕	Organizer/ Chair： 冼芻蕘	Organizer/ Chair： 陳春樹	Organizer: 程毅豪 Chair： 王秀瑛	Organizer/ Chair： 陳美如	Organizer: 呂恆輝 Chair： 林建同
	Speaker: 1. 陳樂昱 2. 顏佑銘 3. 劉祝安	Speaker: 1. 陳由常 2. 莊雅婷 3. 李金全 4. 冼芻蕘	Speaker: 1. 余清祥 2. 詹大千 3. 沈宗荏	Speaker: 1. 王秀瑛 2. 林惠文 3. 施嘉翰	Speaker: 1. 李志煌 2. 林偉傑 3. Kyung-Youn Kim	Speaker: 1.陳素雲 2.陳定立 3.黃名鉞
14:50 | 15:00	休息時間					
場次 場地	SV-1 理A310	SV-2 理B201	SV-3 理A212	SV-4 理D241	SV-5 理A209	SV-6 理A211
SV 15:00 | 16:20	空間統計	High- dimensional data analysis	工業統計 及應用	財務金融	Causal Inference	財務工程
	Invited session	Invited session	Invited session	Invited session	Invited session	Invited session
	Organizer/ Chair： 張雅梅	Organizer/ Chair： 黃灝勻	Organizer/ Chair： 許湘伶	Organizer/ Chair： 蕭維政	Organizer/ Chair： 林聖軒	Organizer/ Chair： 孫立憲
	Speaker: 1.林培生 2.楊洪鼎 3.張雅梅	Speaker: 1. 彭柏翔 2. 黃學涵 3. 賴驥緯	Speaker: 1. 張明中 2. 李宜真 3. 林長鋆	Speaker: 1. 鄭宏文 2. 陳虹吟 3. 匡顯吉	Speaker: 1. 楊子霆 2. 林聖軒 3. 溫啟仲	Speaker: 1. 王俞凱 2. 吳牧恩 3. 孫立憲
16:40	賦歸 接駁車發車					

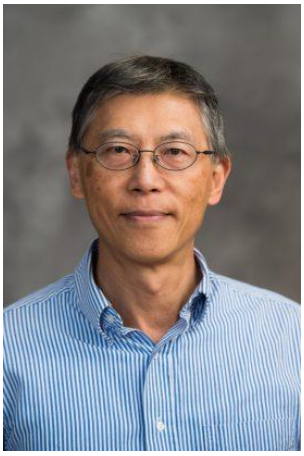
場次 場地	SIV-7 理A318	SIV-8 理A316	SIV-9 理A210	SIV-10 理A106	SIV-11 理A109
SIV 13:30 | 14:50	統計方法 Invited Session	Contributed Session - H	Contributed Session - I	Contributed Session - J	Contributed Session - K
	Organizer/ Chair： 李燊銘	Chair： 孫立憲	Chair： 張志浩	Chair： 林聖軒	Chair： 施銘杰
	Speaker: 1. 謝淑惠 2. 盧馬汀 3. 李燊銘	Speaker: 1.楊竣皓 2.錢映伶 3.陳虹君 4.羅育聖 5.陳諾霖 6.劉銘軒	Speaker: 1.龔一鴻 2.林宣廷 3.蔡羽涵 4.游宗諺 5.李翊瑄 6.邱崴泰	Speaker: 1.李易儒 2.江皓瑋 3.林珈卉 4.解振成 5.石亞喬 6.呂亮葳	Speaker: 1.賴恩語 2.黃怡甄 3.林奕勳 4.謝譯瑭 5.陳冠憲 6.黃乾哲
14:50 | 15:00	休息時間				

場次 場地	SV-7 理A318	SV-8 理A314	SV-9 理A316	SV-10 理A210	SV-11 理A106	SV-12 理A109
SV 15:00 | 16:20	臨床試驗之統計方法與應用 Invited session	統計方法 Invited session	Contributed Session - L	Contributed Session - M	Contributed Session - N	Contributed Session - O
	Organizer/ Chair： 林資荃	Organizer/ Chair： 張志浩	Chair： 曾玉玲	Chair： 陳忠和	Chair： 曹振海	Chair： 蘇南誠
	Speaker: 1. 林資荃 2. 陳瑱芳 3. 林奕廷 4. 黃智揚	Speaker: 1. 顏佐榕 2. 朱基祥 3. 張志浩	Speaker: 1.鄭竣元 2.林建璋 3.王律涵 4.陳柏豪 5.林君儀 6.林俊源	Speaker: 1.陳忠和 2.范姜翔星 3.劉佳穎 4.洪煒傑 5.李曉妮 6.温筑涵	Speaker: 1.呂一昕 2.陳君毅 3.蔡明憲 4.管彥鳴 5.何郁涵 6.何婉瑜	Speaker: 1.廖靖芸 2.陳慧霜 3.施承翰 4.黃文顥 5.薛皓澤 6.黃凱琪
16:40	賦歸 接駁車發車					

便當及茶會領取地點：理工一館 A 棟一樓。
用餐空間：理 A108、理 D131、理 A307。
接駁車乘車處：理工一館旁大停車場。

周元燊院士講座

主講者: 邢泰侖 (Tailen Hsing) 教授



邢教授為美國密西根大學統計系的講座教授 (Michael B. Woodroofe Collegiate Professor of Statistics),主要研究領域包括 Extreme value problems、functional data、long-range dependence、space-time models 等。

邢教授在統計與機率方面研究工作十分傑出。許多研究成果發表在國際頂尖期刊如 Annals of Statistics、Annals of Applied Statistics、Journal of the Royal Statistical Society、Series B, Biometrika、Annals of Probability、 Statistica Sinica、Extreme 等。邢教授為第一位台灣統計學者擔任 Annals of Statistics 的 co-editor。同時亦擔任多種頂尖統計與期刊編輯委員如 Statistical Science、Bernoulli、Statistica Sinica 等。邢教授於 2010-2015 擔任美國密西根大學統計系系主任。目前亦擔任中研院統計科學研究所諮詢委員。

近年選擇重要出版

- Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley.
- Hsing, T., Brown, T., and Thelen, B. (2016). Local intrinsic stationarity and its inference, *Annals of Statistics* **44** , 2058-2088.
- Chiou, J.-M., Chen, Y.-T., and Hsing, T. (2019). Identifying multiple changes for a functional data sequence with application to freeway traffic segmentation, *Annals of Applied Statistics* **13** , 1430-1463.
- Shen, J. and Hsing, J. (2020). Hurst index estimation, *Annals of Statistics* **48** , 838-862.
- Li, J., Li, Y. and Hsing, T. (2021). On functional processes with multiple discontinuities, *JRSSB.* **84** , 933-972.
- Yagers, D., Stoev, S., and Hsing, T. (2022). A functional-data approach to the Argo data, *Annals of Applied Statistics* **16** , 216-246.
- Shen, J., Stoev, S., and Hsing, T. (2022). Tangent fields, intrinsic stationarity, and self similarity, *Electronic Journal of Probability,* **27** 1-56.
- Yarger, D., Korte-Stapff, M., Stoev, S., and Hsing, T. (2022). A multivariate functional-data mixture model for spatio-temporal data: inference and cokriging.
- Kartsioukas, R., Stoev, S., and Hsing, T. (2023). Spectral density estimation of function-valued spatial processes.

周元燊院士講座

主講者: 邢泰侖 (Tailen Hsing) 教授

Department of Statistics, University of Michigan

A functional-data perspective of the Argo data

Abstract

The Argo data is a modern oceanography dataset that provides unprecedented global coverage of temperature and salinity measurements in the upper 2,000 meters of depth of the ocean. In this talk, we present our approaches on two specific problems from the perspective of functional data analysis. First, we develop spatio-temporal functional kriging methodology for mean and covariance estimation to predict temperature and salinity as a smooth function of depth. Second, we discuss the functional co-kriging problem of predicting oxygen concentration based on temperature and salinity data. By combining tools from functional and spatial data analysis, including smoothing splines, local regression, and multivariate spatial modeling and prediction, our approaches provide advantages over current methodology in oceanography research that consider pointwise estimation/prediction at fixed depths.

周元燊院士講座設置辦法

103年5月17日第八屆第四次理監事會議議決通過

一、設置目的：中華機率統計學會(以下簡稱本會)為感謝周元燊院士創立本會及長期以來對國內統計界之卓越貢獻，特設置周元燊院士講座(以下簡稱本講座)，以肯定對提昇國內統計學術水準有傑出貢獻之國內外學者。

二、推薦方式：本講座每年一名，由南區統計研討會主辦單位負責遴選，並推薦至本會理監事會備查。

三、獎勵方式：由南區統計研討會主辦單位邀請給一大會演講，並由本會致贈獎金六萬元整。

四、經費來源：由黃文璋教授捐助。

五、本辦法經本會理監事會議通過後實施，修正時亦同。

周元燊院士略歷

周元燊教授，1974 年當選中央研究院院士，原籍湖北襄樊市，與當代機率學家鍾開萊教授並稱為二十世紀華人最偉大的機率學家。中文機率一詞，便是他第一個創立的新詞。周院士著作等身，影響並提攜華人世界中眾多一流的機率學家及統計學家， 號稱國寶級大師亦當之無愧。周院士對兩岸統計界貢獻厥偉，其著作擲地有聲，是吾等後輩難望其項背的。不止於此，周院士於 1992 年發起成立中華機率統計學會，他並擔任第一屆理事長。另外，海峽兩岸統計學研討會也是周院士促成舉辦的。周院士現居於台灣新北市八里區，常來往兩岸，為提昇華人統計影響力努力不怠。

趙民德教授講座

主講者: 張源俊 (Yuan-chin Ivan Chang) 教授

張教授的主要研究領域包含了序列分析(Sequential Analysis)、廣義線性模型(Generalized Linear Model), 教育統計(Educational Statistics)、機器/統計學習(Machine/Statistical Learning)、長期追蹤資料分析(Longitudinal Data Analysis)等。主題看似不多，但其實都不相近。這反應了他研究的選擇與態度。張教授有興趣的問題與涉獵領域十分豐富，其中沉潛多年當屬音樂。除了是古典音樂的深度愛好者，他也曾參與合唱團，學習/演奏長笛，音樂素養不在話下。在數學/統計科普推廣上，他一直不遺餘力，從調音與數學，到孔子談網路教育，以及新近的咖啡、音樂與學術品味快閃演講，樂此不疲。如此豐富的生活中，他仍擔任了不同時期 Sequential Analysis Journal, Journal of Japanese Society of Computational Statistics 的 Associated Editor; Statistica Sinica, JCSA 的 Members of Chief-Editors, Managing Editor 等職務。在正式或非正式的場合，他常與其他研究者，特別是年輕一輩的研究者交換研究想法與新方向。提攜鼓勵新一代學者上，他投注了相當的心力。早在 AI 還在常溫階段的 2004 年，他就推動了 Statistical and Machine Learning Workshop; 資料科學統計合作社自然也與他關係密切… 林林總總，未完待續...

趙民德教授講座

主講者: 張源俊 (Yuan-chin Ivan Chang) 教授

中央研究院統計科學研究所

Distributed sequential estimation procedures – Something old is new again.

Abstract

Data collected from distributed sources or sites commonly have different distributions or contaminated observations. Active learning procedures allow us to assess data when recruiting new data into model building. Thus, combining several active learning procedures together is a promising idea, even when the collected data set is contaminated. In this talk, we will briefly describe how we can apply sequential analysis in Statistics, a concept raised by Wald (1945) during WWII, to modern machine learning such that we cannot only save both the sampling and computing costs, and stay robust while dealing with the contaminated data scenarios commonly found analyzing big data sets, and a demonstration with Python is available through Github at https://github.com/zhuojianc/dsep.

趙民德教授講座設置辦法

109年2月21日第十屆第四次理監事聯席會議議決通過

一、趙民德先生為中央研究院統計科學研究所之創所所長，有鑒於趙先生對國內統計學界的發展具卓越貢獻，中華機率統計學會（以下簡稱本會）特設置趙民德教授講座（以下簡稱本講座），以表彰對於提昇國內統計學術水準與發展具傑出貢獻之國內外學者。

二、本講座之經費由本會接受社會各界捐贈，設置專戶儲存備用，或由本會支付。

三、本講座教授應於統計與機率學術領域有傑出貢獻或具卓著聲望，每年由該年度之南區統計研討會主辦單位負責遴選一名，並推薦至本會理事會備查。

四、本講座教授由南區統計研討會主辦單位邀請至大會擔任專題演講，本會致贈獎金新台幣五萬元整。

五、本辦法經本會理事會議決議通過後實施，修正時亦同。

昔年種柳

趙民德

今年(民國七十一年)七月一日，中央研究院正式成立了統計學籌備處。這件事情至少有兩種意義：第一：這是政府對於過去二十多年統計界同仁的貢獻的一種正式認可：統計學自理、工、農、商這些學院的羽翼下跳出來，開始與別的獨立科學如物理、化學、數學等分庭抗禮；第二：我們意識到國內的統計學方面的研究、發展、應用和推廣，還有許多路要走。我們應該研究什麼主題？發展那些系統？如何推展應用？怎麼將統計學所表現出來的近代的格物致知的方法，廣泛地深植到社會的思想形態裡？我們應該認真地開始回答這些問題，而統計所的成立，給我們一個機會。

我想藉這篇文章，介紹一下統計學，國內的統計學以及我們想做的工作。

統計是一種科學方法和思想方式，是一種入世的、積極的哲學態度，是一種進步社會中自然的現象。

二

狹義的統計學，是泛指一套圍繞著以數量為主的資料的科學的格物致知方法。這套方法的一個最大特徵是：統計學家深切地體認到誤差的存在，並積極的面對可能的誤差，而使得經過這套方法所導出的結論，其因誤差而產生的曖昧減少。統計學的方法並不能無中生有，但它的確致力於儘量濾去誤差，而得到傳統方法所不能得到的結論。誤差如水，真象若石。水落，所以石出。如果水中原本無石，水落當然也仍然無石。統計的方法之所以大行其道，是因為誤差的是近代生活的一部分。一切人類所蒐集的數量資料中，其不包含誤差者百不得一。社會愈進步，則所需要蒐集與分析的(包含著誤差的)資料也愈多。而我們愈嘗試的去以無誤差的概念去推演結論，所得的結論裡便愈充滿誤差。而統計方法，因為能正視誤差的存在，反而可以得到更合理的結論。因為這個原故，我們巧妙的自台灣地區人中抽樣，以求得平均國民所得或失業率；我們自少數病人的追蹤記錄裡研究一種新藥的效力；我們用精心設計的品管圖來看生產線上的狀況；我們小心地分配肥料與農藥的成份，而來看那一種配方使稻米的產量最多。這些都是統計方法：圍繞著包含了誤差的數字所作的種種精巧的努力。這些包括了蒐集資料以及分析資料。資料，是統計學的食糧,氧氣和水。資料，是現代的文字，是工業社會的文字，是科學的文字。統計學只是幫你去認字、去寫字，並用這樣的文字寫作品。而這些作品所表達的是另一度空間的現代社會。它非詩、非畫；但更精、更簡，更觸到痛處。因為它的目的是探測真象，而它的手段則是濾去誤差及噪音。一方面，它像一個楔而不捨的偵探，搜尋線索，避開兇手所佈的疑陣，抓住了問題的核心；另一方面，它也像一個嚴正法官，考核證據，一絲不苟，既要不枉，也要不縱。我們不能無動於衷，因為統計是入世的科學，所以統計工作者也心在紅塵。我們觀察、整理、估計、檢定、模擬、預測...我們主要的對象是這個社會的形形色色的一切可以用數量測度的事物。我們看到的山，是山；我們見到的水，是水。而且因為

我們是用有情的眼光來看的，所以我們看到的山，不免嫵媚；見到的水，不免清秀。

三

國內目前的統計，可概分為人文、生物及數理三類。這三支主流，以一個開發中的國家而言，都算得上相當源遠流長。而其他計量科學的支脈，如計量經濟、工商管理、工業工程、品質管制及保證、自動控制...也並不是沒有人才。自經濟發展的過程來說，我們的社會已繁榮到必需以接近天文的數字來描述的程度。大到政策的制定，小到一種商品的行銷，都離開不了計量方法。然而這些以統計學為基礎的近代計量方法，(還有以電子計算為輔的近代系統)，並沒有成為社會裡中上級官員或經理人員的思想方式的一部分。國內在統計學及其相關科學方面，固然人才不足，但並不是基本的問題。國內的問題在理論與實務的不配合，以及統計的工作環境的缺乏。國內並非不能培養統計人才；只是不易培養立即能夠上陣的統計人才；即使培養了，也不易將他們置於易於發揮功效的位置。就算這兩者都做到，也因為缺乏適當的環境，這些人才的努力，皆不免流於事半功倍。因此，我們有近五十年的政府統計資料，但其在民間的威信，尚待建立；我們的生物統計源遠流長，但至今尚推行不了一套完善的藥物管制制度，我們有將近三十名統計博士。但多散處各大專院校。論文雖然時有佳作，但對整個的統計甚至於計量科學的發展，並沒有發揮出風起雲湧的作用。但是我們不必氣餒。我們自己不肯一顧的政府統計資料，有外國人在分析檢定，以期孤立出我國發展迅速的原因；十餘年來，品管工作者默默的建立起類似日本的品管制度；政府機構及民間企業裡也漸漸地增加了計量分析方面的人員。而在國外，不論在怎樣的統計學會的集會裡,都可找的到中國人在發熱發光。

這樣的成績固然堪差幸慰，但是如果我們看一看統計學及統計工作者在別的國家的經濟發展裡所作的貢獻，我們還有許多路要走。如果我們不能將近代的計量方法，注射到我們國家的經濟結構的重要關節；為果我們不能影響這個社會的中堅份子，讓他們以客觀的、計量的、容忍並適度控制誤差的態度和方法來處理現代的企業經營，那麼我們即使有一二位名師大匠,五六十名有佳作的研究人員，我們仍然在實務上趕不上日本，理論上比不上印度，更遑論英美等國家矣。中國的統計，並不是不需要名師大匠。中國也並不是沒有相當傑出的名師大匠。但我們如果向這個方向走，我想統計研究所在十五年內，還趕不上史丹福大學的統計系。中國的統計學家並不是寫不出高深的論文，但是統計學本身，在這個社會裡還沒有堅實的基礎，那些高高在上的論文，有若用極度的平衡技巧的支持的高空走鋼絲的表演者。表演雖然精彩，掌聲也未見稀疏，但是我們知道每年畢業的新的計量科學方面的碩士，有相當困難去找待遇優厚的工作；更不論那數百名那些相關科系畢業的大學生了。只有在這些計量分析人才，在國內成為各界爭相邀聘的對象時，我們才可以勉強說統計學的基礎，初初建立。

四

遠在民國十八年的時候，所謂的甘莫爾設計委員會，便建議國民政府於其下之主計總監部。這項建設經過立院會議之後，主計總監部的名稱換成了主計處，直屬國民政府，以下設歲計、會計、統計三局，分掌全國歲計、會計、統計事務。這是我國正

式有統計制度的開始。其後主計處在民國三十七年改隸行政院，名稱曾一度稱為主計部，旋即改回。五十三年以來，內憂外患、不絕如縷，然而國府的統計工作如財務統計，基本國勢調查，國民所得統計、物價勞工統計、公務統計、公務人員及其工作之統計等，均在超然統計制度之下，維持至今。我國在民國二十一年便訂定了統計法(目前的三法修正，為民國六十一年五月二十六日公佈施行)。在各國政府所作的統計資料蒐集工作而言，我國是走在相當前面的。據六十三年二月全計處編印的「各項統計方法專輯」，目前我國至少有氣象、人文...到社會衛生，其他公務等十五項，九十六目統計資料。我們不能說政府沒有做統計工作，我們只能說我們沒有好好的利用這些寶貴資料。在學術界，我國的統計學者也頗為傲人。台大的農藝研究所開始用電子計算機作逐次迴歸分析的時候，國外所謂的「統計程式庫」也不過在萌芽階段，並沒有如今日一般地風起雲湧。好幾個農業實驗所的實驗設計，也為國內的品種改良做了註腳。相對地說，我國的數理統計起步的反而相當的晚。

五

我們不要把自己分得太細太嚴：這是統計，那不是統計，而另一個分明便是作業研究...。如果我們這樣分，那麼每一個領域便的非常窄，我們不能自前人得到什麼幫助，也不能有效的幫助別人。所有的近代計量科學，都和統計有關。我們無意併吞他們，應該如對待芳鄰一樣：我們借給他們油鹽，也希望有時候能自他們那裡去借柴米；我們鼓勵孩子們和家的來往，而希望有一天或能發展進一步關係...我們也不要把理論和實務分得太嚴。理論如花，實用如果。不會開刀的醫生，和不會分析資料的統計學者一樣。花和果互相關連。有的果、種子發不成芽；有的花，結不了果。但是花和果都一試、再試。這是一個機率的世界，沒有失敗，便也不會有成功。如果我們在花和果之間，畫一條界線；如果我們在統計的理論與實務中間劃一條鴻溝，那麼花雖然香，總離不了溫室；理論雖佳，也了無意義。讓我們去試著了解並尊敬到目前為止國內的計量科學所做的工作。他們的門，也許正鎖，但我們何嘗去用心敲過？他們的園地，也許貧瘠，但我們何嘗去用心幫著灌溉？當我們有百萬大軍的時候，再去分海、陸、空；再去分軍、師、旅、團。而現在，我們規模粗備，雖然望盡天涯路，卻幸而不是高樓獨上。我們應該試著去做一些大規模的團隊工程：不論是統計的理論還是實務，我們應該學會以二十個人合做一個計畫、三四個人合作一篇論文。我們還不是趙雲關羽，單槍匹馬，在千軍萬馬裡尚殺不出重圍。而在國外，比我們強得多的研究發展單位，用比我們多的人力物力，做比我們野心小得多的研究計劃。

後記

這是1982年的舊作，今日讀來，心情猶在。國內的統計研究工作，十八年來成長了十二倍，但學者們「單兵攻擊」的毛病依舊。

標題是後來加上的。取自「昔年種柳，依依漢南；今看搖落，萋愴江潭。木猶如此，人何以堪。」2000某日又讀，2020/8/19重印。

中華機率統計學會魏慶榮統計論文獎設置辦法

100 年 3 月 19 日訂定
102 年 1 月 12 日第七屆第七次理監事會議修訂議決通過
106 年 11 月 18 日第九屆第五次理監事會議修訂議決通過
108 年 6 月 21 日第十屆第一次會員大會會議修訂議決通過

一、 目的：為紀念魏慶榮教授，特設置魏慶榮統計論文獎（簡稱本獎）。本獎設置之目的除為了表彰魏教授在統計與機率研究上的卓越成就外，亦希望持續發揚其樂於提攜後進之精神，獎勵優秀年輕學者，促進台灣機率統計的研究與發展。

二、申請資格：凡就讀於國內博士班之研究生，或取得國內博士學位未滿一年者（由申請截止日回溯）皆可申請，但其送審論文須屬於統計、機率或計量經濟等領域。

三、申請程序：申請者需填寫申請表，並由論文指導教授一名推薦。申請者將論文及申請表寄至中華機率統計學會。

四、審查程序：本學會邀請數名教授組成審查委員會(本會學術委員為當然委員）進行論文審查。經初審擇優邀請申請人至本學會主辦或協辦之學術研討會發表演講，並進行複審評分，最後決選出至多三位獲獎者（可從缺)。獲獎以一次為限。

五、獎勵方式：特優獎至多一名發給獎金三萬元及獎狀，優等獎各發給獎金一萬五千元及獎狀，並進行公開表揚。

六、申請截止日：每年四月三十日。

七、經費來源：由銀慶剛教授負責籌措。

八、本辦法經本學會理監事會議討論通過後實施，修正時亦同。

中華機率統計學會魏慶榮論文獎歷屆獲獎名單

年度	獲獎獎項	姓名	獲獎時就讀學校
97 年度	特優獎	黃士峰	中山大學應用數學系
	優等獎	李全濱	中山大學應用數學系
		張婉琦	成功大學統計學系
	佳作	田銀錦	中央大學統計研究所
		彭建育	清華大學統計學研究所
		謝敏雄	東海大學統計學系
98 年度	特優獎	陳慎健	政治大學統計學系
	優等獎	謝淑惠	東海大學統計學系
		劉峰旗	逢甲大學應用統計學系
	佳作	陳致綱	政治大學國際經營與貿易學系
	評審獎	洪弘	臺灣大學數學系
99 年度	特優獎	江達生	逢甲大學統計學系
	優等獎	王婉倫	中央大學統計研究所
		葉倚任	臺灣科技大學資訊工程學系
	佳作	洪宛頻	中山大學應用數學系
		陳馨蕙	東華大學經濟學系
		潘家群	交通大學統計學研究所
100 年度	特優獎	張志浩	交通大學統計學研究所
	優等獎	許湘伶	中山大學應用數學系
		鄭宏文	臺灣大學財務金融學系
	佳作暨最佳演說獎	陳馨蕙	東華大學經濟學系
	佳作	詹嘉豪	成功大學統計學系
		劉姿杏	中興大學應用數學系
101 年度	特優獎	林良靖	中山大學應用數學系
	優等獎	高竹嵐	中央大學統計研究所
		楊巧意	中央大學統計研究所
	佳作	顏汝芳	臺灣大學財務金融學系
		羅盛豐	臺灣大學財務金融學系

年度	獲獎獎項	姓名	獲獎時就讀學校
102 年度	特優獎	陳穎頻	中央大學統計研究所
	優等獎	黃世豪	中山大學應用數學系
		謝宗震	清華大學統計學研究所
	佳作	謝宛茹	交通大學統計學研究所
		王价輝	清華大學統計學研究所
		陳淑瑜	逢甲大學統計學系
103 年度	特優獎	黃名鉞	臺灣大學數學系
	優等獎	邱海唐	中山大學應用數學系
		王泰期	政治大學統計學系
	佳作	李章益	政治大學金融學系
		徐紫婷	臺灣大學流行病學與預防醫學研究所
104 年度	特優獎	王彥雯	臺灣大學流行病學與預防醫學研究所
	優等獎	曾聖澧	中國醫藥大學公共衛生學系
		王文廷	交通大學統計學所
	佳作	林瑞祥	臺灣大學流行病學與預防醫學研究所
		蕭維政	國立中央大學統計研究所
105 年度	特優獎	王紹宣	國立台灣大學數學所統計組
	優等獎	蔡宗憲	國立成功大學統計學研究所
		張明中	國立清華大學統計學研究所
	佳作	楊洪鼎	彰化師範大學數學系
		陳秉洋	成功大學統計學研究所
		姚禹丞	清華大學統計學研究所
106 年度	特優獎	李宜真	國立清華大學統計學研究所
	優等獎	游翔	國立清華大學統計學研究所
		施嘉翰	國立中央大學統計學研究所
	佳作	梁佑任	臺灣大學生醫電子或資訊學研究所
		鐘俊江	國立成功大學統計學研究所
		張瑄凌	國立台灣大學財務金融所
107 年度	特優獎	紀建名	國立台灣大學經濟學研究所
	優等獎	Suneel Babu Chatla	國立清華大學服務科學研究所
		翁新傑	國立中央大學統計研究所

108 年度	特優獎	林建同	國立清華大學統計學研究所
	優等獎	鍾長恕	國立政治大學金融學系
109 年度	特優獎	黃學涵	國立清華大學統計學研究所
	優等獎	董弘平	國立清華大學統計學研究所
110 年度	特優獎	彭柏翔	國立清華大學統計學研究所
	優等獎	賴暐婷	國立成功大學統計學系
111 年度	特優獎	林蔚安	國立成功大學統計學研究所
	優等獎	Aljo Clair Pingal (高裘雷)	逢甲大學統計系
	優等獎	黃煜鈞	國立台灣大學流行病學與預防醫學研究所

全球人壽演講

ChatGPT 在金融業的可能應用

June 30, 13:20 – 13:20

理工一館 A106

◆Chair：

孫立憲 教授

◆Speaker:

謝明華 博士

政治大學風險管理與保險學系教授
政治大學數位金融創新實驗室執行長 (https://digitalfin.tw)
政治大學風險與保險研究中心主任(https://www.rirc.tw)
台灣亞太監理科技協會理事長(https://www.arta.tw)
台灣風險與保險學會秘書長(https://tria.asia)
期交所結算委員會召集人 (https://www.taifex.com.tw)
中華民國退休基金協會理事 (https://www.pension.org.tw)
政治大學金融科技研究中心副主任(http://www.ftrc.nccu.edu.tw)
政治大學 Amundi 鋒裕匯理投資創新研究中心副主任
(https://www.aiirc.nccu.edu.tw)
全球人壽獨立董事 (https://www.transglobe.com.tw)
台灣金融服務業聯合總會境外結構型商品審查委員
GARP 財務風險分析師 (FRM)

分組議程 SI 時間表/摘要

分組議程 SI : 2023 年 6 月 29 日 (星期四) 13:30 - 14:50

SI-1	**魏慶榮統計論文獎**	**Chair：潘建興**	**理 A310**	**p.37**

董亦賢 - 國立中央大學統計研究所

The first-passage-time moments for the Hougaard process and its Birnbaum-Saunders approximation

邱詠惠 - 國立中央大學統計研究所

ASemiparametricApproach for Estimating Fixed Effects with Spatial Confounding

潘彥碩 - 國立清華大學統計學研究所

MPCGA: A Tree-Based Chebychev's Greedy Algorithm

SI-2	**空間統計方法與應用**	**Chair：鄭宗記**	**理 B201**	**p.41**

余清祥 - 國立政治大學統計系

Using Geographically Weighted Regression for Spatial Data Visualization

陳怡如 - 國立政治大學統計系

Geographically weighted regression analysis for continuous nonnegative response data

張軒瑜 - 中華經濟研究院綠色經濟研究中心

Testing trending time-varying coefficient spatial panel data models

SI-3	**Statistical Quality Management**	**Chair：楊素芬**	**理 A212**	**p.45**

周珮婷 - 國立政治大學統計系

Data-driven Anomaly Detection

呂明哲 - 朝陽科技大學

An EWMA Control Chart for the Median of Birnbaum-Saunders

李俊毅 - 國立成功大學統計學系

Control charts for profile monitoring of within- profile correlations using the Tweedie exponential dispersion process model

SI-4	**生物統計**	**Chair：蘇佩芳**	**理 A109**	**p.49**

曾議寬 -國立中央大學統計所

Semiparametric models for the design on three-arm noninferiority trials

張玉媚 - 東海大學統計系

Adaptive Randomization Design for Multiarm Response - Survival Clinical Trials

溫淑惠 - 慈濟大學公共衛生學系

A phenome-wide association study of postoperative infection in patients after total knee arthroplasty

分組議程 SI : 2023 年 6 月 29 日 (星期四) 13:30 - 14:50

SI-5 **計算統計** **Chair：林宗儀** **理 A209** **p.54**

陳泰賓 - 國立陽明交通大學生物醫學暨工程學院

The Algorithm for HRRT PET Reconstruction with Inverse Problem

王婉倫 - 國立成功大學統計學系

A Selection Model for Multivariate Longitudinal Data with Censored and Nonignorable Intermittent Missing Outcomes

楊郁成 - 國立中興大學應用數學系

Estimation and Prediction in t Linear Mixed Models with Non-ignorable Dropout

SI-6 **計量分析的新進展** **Chair：銀慶剛** **理 A211** **p.59**

楊睿中 - 國立臺灣大學經濟學系

Synthetic Historical Control for Policy Evaluation without Cross-sectional Controls

郭漢豪 - 國立臺灣大學經濟學系

Nonlinear Least Squares, Model Selection, and Model Averaging for Social Interaction Models

俞淑惠 - 國立高雄大學統計學研究所

New Results on Generalized Dynamic Factor Model: Prediction and Model Selection

SI-7 **生物醫學統計** **Chair 張中** **理 A318** **p.63**

馬瀰嘉 - 國立成功大學統計學系

誤差平方和最小化條件下的因子分析

張升懋 - 國立臺北大學統計學系

Modeling Multiple-Criterion Diagnoses by Heterogeneous-Instance Logistic Regression

李宗翰 - 國立中正大學數學系

Exact Confidence Coefficients of Confidence Interval for a Difference of the Binomial Proportions

SI-8 **跨領域(大氣科學)** **Chair：鄭芳怡** **理 A314** **p.68**

柯婕盈 - 國立中央大學大氣科學學系

Unsupervised Machine Learning for Analyzing Ensemble Simulations of Extreme Heavy Rainfall Event in NorthernTaiwan

林遠見 - 國立中央大學土木工程學系

Investigation of the nonlinear relationship between PM2.5 and ten leading causes of death based on risk analysis approach

陳柏孚 - 國立臺灣大學氣候天氣災害研究中心

Deep Learning Satellite Imagery Analysis to Reconstruct Historical Typhoon Data

徐佩君 - 國立中興大學循環經濟研究學院

2021 Texas Cold Snap: Manifestation of Natural Variability and a Recent Warming Trend

分組議程 SI : 2023 年 6 月 29 日 (星期四) 13:30 - 14:50

SI-9 **生醫統計方法** **Chair：鄒宗山** **理 A316** **p.74**

戴安順 - 國立成功大學統計學系

Causal mediation analysis for recurrent event using additive rate models: generalizations from the Baron and Kenny method

王偉銘 - 國立成功大學

A Modified Single-stage Sampling Procedure for Heteroscedasticity Analysis of Means

簡莉珠 - 高雄醫學大學基礎科學教育中心

Combining dependent p-values by gamma distributions

SI-10 **International session - Probability and Statistical Mechanics** **Chair:Lung-Chi Chen** **理 A210** **p.78**

Cheuk-Yin Lee - 國立清華大學數學系

Hitting probabilities of Gaussian random fields and SPDEs

Yuki Chino - 國立陽明交通大學應用數學系

Affection for Critical Point and Asymptotic Behaviour of SAW from Random Environment

Yoshinori Kamijima - 國家理論科學中心數學組

A local limit theorem for the long-range self-avoiding walk

SI-11 **生物演化與疾病預測** **Chair：Wei-Wen Hsu** **理 A106** **p.82**

張憶壽 - 國家衛生研究院

Adapting the lung cancer PLCO model for Taiwan

Wei-Wen Hsu - Department of Environmental and Public Health Sciences, University of Cincinnati

A Covariate-Adjusted Classification for Multiple Longitudinal Biomarkers

鍾冬川 - 逢甲大學統計學系

Effect of Polytomy on the Parameter Estimation and Goodness of Fit of Phylogenetic Linear Regression Models for Trait Evolution

分組議程 SI : 2023 年 6 月 29 日 (星期四) 13:30 - 14:50

SI-12 **Contributed session - A** **Chair：高正雄** **理 D131** **p.86**

趙伯昇 - 國立中山大學應用數學系

On the Asymptotic Behavior of Allocation Scheme with Geometric Distributions

郭冠麟 - 國立高雄大學統計學研究所

A Novel Mechanism for Describing Zero-inflated Data:
An Application with Zero-Inflated Positive Skellam Distribution

林家弘 - 國立東華大學應用數學系統計碩士班

Statistical Learning in the Universal Kriging

馬繻嬪 - 國立中央大學統計研究所

Unsupervised Generalized Learning-based NEM Algorithm for Clustering

王薇淳 - 國立臺北大學資訊與工程學系

利用統計方法與集群分析預測暨評估平衡度

SI-13 **Contributed session - B** **Chair：黃逸輝** **理 A307** **p.92**

黃逸輝 - 淡江大學數學系

加速壽命模型於測量誤差下的無母數回歸校正

葉詠富 - 國立陽明交通大學數據科學與工程研究所

Locating the Ampulla of Vater and Estimating the Direction of Bile Duct by
Deep Learning Methods

鄭亦甯 - 國立臺北大學統計學系

風險對比特幣、比特幣與金融資產連動之影響

羅祖佑 - 國立中山大學應用數學系

基於動態網路模型的虛擬貨幣投資組合

王嘉顥 - 國立臺北大學統計學系

CEO 可信度對加密貨幣價格之影響

莊沅蓉 -國立臺灣大學統計與數據科學研究所

Nonparametric Mediation Analysis of Survival Outcomes

Session I-1

June 29, 13:30 – 14:50

理工一館 A310

魏慶榮論文獎場次

◆ **Organizer/Chair：**

潘建興 - 中央研究院 統計科學研究所

◆ **Speaker:**

1. 董亦賢 – 中央大學統計研究所

2. 邱詠惠 – 中央大學統計研究所

3. 潘彥碩 – 清華大學統計學研究所

The first-passage-time moments for the Hougaard process and its Birnbaum–Saunders approximation

Chien-Yu Peng (彭健育)[1], Yi-Shian Dong (董奕賢)*[2], Tsai-Hung Fan (樊采虹)[2]

[1]Institute of Statistical Science, Academia Sinica

[2]Graduate Institute of Statistics, National Central University

Abstract

Hougaard processes, which include gamma and inverse Gaussian processes as special cases, as well as the moments of the corresponding first-passage-time (FPT) distributions are commonly used in many applications. Because the density function of a Hougaard process involves an intractable infinite series, the Birnbaum–Saunders (BS) distribution is often used to approximate its FPT distribution. This article derives the finite moments of FPT distributions based on Hougaard processes and provides a theoretical justification for BS approximation in terms of convergence rates. Further, we show that the first moment of the FPT distribution for a Hougaard process approximated by the BS distribution is larger and provide a sharp upper bound for the difference using an exponential integral. The conditions for convergence coincidentally elucidate the classical convergence results of Hougaard distributions. Some numerical examples are proposed to support the validity and precision of the theoretical results.

Keywords: Characteristic function, Contour integration, Exponential dispersion model, Residue, Stirling numbers.

A Semiparametric Approach for Estimating Fixed Effects with Spatial Confounding

Yung-Huei Chiou（邱詠惠*）, Chun-Shu Chen（陳春樹）

Graduate Institute of Statistics, National Central University

Abstract

In spatial regression analysis, confounding between fixed effects and random effects adversely affects the estimates of regression coefficients. This paper proposes a novel estimation method for regression coefficients based on the fixed rank kriging approach which does not require specifying any parametric covariance structures for response variables and hence is more practical. The proposed method involves selecting the number of basis functions which affects the bias and the variance of estimators. Two methods based on resampling and error loss are respectively proposed to control the mean squared errors of estimators, resulting in the Bagging estimator and γ-estimator for estimating regression coefficients. Theoretical properties related to the proposed methodology are justified. The simulations under the settings of spatial regression models with spatial confounding reveal that the proposed estimation methods of regression coefficients perform well regardless ofthe underlying correlation structure is stationary, nonstationary, isotropic, or anisotropic. Finally, an application to precipitation data in Colorado is presented.

Keywords: Basis function, Bias reduction, Fixed rank kriging, Mean squared error, Restricted spatial regression

MPCGA: A Tree-Based Chebychev's Greedy Algorithm

Yan-Shuo Pan*(潘彥碩)[1], Ching-Kang Ing(銀慶剛)[1], Guan-Hua Huang(黃冠華)[2]

[1]Institute of Statistics, National Tsing Hua University

[2]National Yang Ming Chiao Tung University

Abstract

Prediction and feature selection are essential topics for statistical and machine learning (ML) methods when dealing with high-dimensional data. However, they come with limitations: statistical methods may exhibit lower predictive ability than ML methods, while ML methods are often criticized as black boxes. In this paper, we introduce a tree-based algorithm, the Multipath Chebyshev Greedy Algorithm (MPCGA), which enhances the predictive performance of statistical methods and feature selection capabilities under model misspecification. This algorithm extends the Chebychev's Greedy Algorithm (CGA) and High Dimensional Information Criterion (HDIC) into a tree-expanded structure, allowing for the simultaneous consideration of multiple models. MPCGA outperforms traditional statistical methods when models are misspecified, while maintaining high feature selection precision. Furthermore, we propose accelerated algorithms to boost the computational speed of MPCGA handling indicator features in binary outcome cases. The paper includes a case study on a lung cancer dataset, demonstrating that utilizing ML methods with the feature set selected by MPCGA leads to suitable results.

Keywords: Chebyshev's Greedy Algorithms, high-dimensional information criterion, machine learning, feature selection, logistic regression model

Session I-2

June 29, 13:30 – 14:50

理工一館 B201

◆Organizer/Chair：

鄭宗記 – 國立政治大學統計學系

◆Speaker:

1. 余清祥 – 國立政治大學統計學系
2. 陳怡如 – 國立政治大學統計學系
3. 張軒瑜 – 中華經濟研究院綠色經濟研究中心

地理加權迴歸在視覺化分析之探討

Using Geographically Weighted Regression for Spatial Data Visualization

Jack C. Yu (余清祥)[1], Yin-Yee Leong (梁穎誼)[2], Rou-Yun Kuo (郭柔芸)[3]

[1] 國立政治大學統計系

[2] 逢甲大學風險管理與保險系

[3] 國立政治大學統計所

摘要

近年大數據蓬勃發展，統計分析的應用更為廣泛，各領域資料以不同型態出現，資料視覺化（Data Visualization）成為探索性資料分析（Exploratory Data Analysis）的核心。視覺化對於空間資料尤為重要，藉由圖表等工具可有效地呈現資料主要特性，包括空間異質性（Spatial Inhomogeneity）、空間自相關（Spatial Autocorrelation），做為後續研究進行的依據。地理加權迴歸（GWR，Geographically Weighted Regression）可視為空間資料的迴歸分析，描述目標變數與解釋變數間的局部關係，用於展示變數關係隨地理位置的變化。本文探討地理加權迴歸的適用時機，透過電腦模擬說明 GWR 的限制及可能問題，測試修正方法是否有效，同時提出這個方法的使用建議。

關鍵詞：地理加權迴歸、空間統計、視覺化、探索性資料分析、電腦模擬

Geographically weighted regression analysis for continuous nonnegative response data

Vivian Yi-Ju Chen[1], Yun-Ciao Yang[2]

[1]Department of Statistics, National Chengchi University

[2]Department of Statistics, Tamkang University

Abstract

Geographically weighted regression (GWR) is widely used in various disciplines to analyze spatial heterogeneity or nonstationarity with respect to data relationships. However, GWR is typically limited to analyzing continuous dependent variables assumed to follow a symmetric normal distribution. In many fields, nonnegative continuous data are frequently observed and may come with substantial amounts of zeros followed by a right-skewed distribution of positive values. When dealing with such type of outcomes, GWR may not provide adequate insights regarding spatially varying regression relationships. This study intends to extend GWR based on compound Poisson distribution, thus allowing for not only the exploration of relationship heterogeneity but also the accommodation of spatial nonnegative continuous response variables. We first present the model specification of the proposed method and then discuss the associated modeling issues, such as bandwidth selection and testing for spatial nonstationarity in regression relationship. The developed technique is further evaluated through simulations. Lastly, an empirical example is given to illustrate the application of our approach.

Keywords: Geographically weighted regression, Spatial heterogeneity, Nonnegative continuous data, Compound Poisson distribution

Testing trending time-varying coefficient spatial panel data models

Hsuan-Yu Chang[1], Xiaojun Song[2], Jihai Yu[2]

[1]Chung-Hua Institutionfor Economic Research

[2] Peking University

Abstract

This paper proposes a nonparametric test for trending time-varying coefficient spatial panel data models. The residual-based U-statistic tests whether the coefficients of the spatial panel data models are time-varying. The proposed test includes three cases: (i) whether the spatial coefficient, the regression coefficient, and the time effects are all time-varying; (ii) whether the spatial coefficient and the regression coefficient are time-varying while the time effects follow a trend function; (iii) whether the spatial coefficient changes over time while the regression coefficient and the time effects are time-varying. All statistics follow an asymptotic standard normal distribution under the null hypothesis of parameter constancy and diverge to infinity in probability under the corresponding alternatives.

Keywords: Nonlinear time trend, Spatial autoregressive models, U-statistic, Time-varying coefficient

Session I-3

June 29, 13:30 – 14:50

理工一館 A212

◆Organizer/Chair：

楊素芬 – 國立政治大學統計學系

◆Speaker:

1. 周珮婷 – 國立政治大學統計學系
2. 呂明哲 – 朝陽科技大學
3. 李俊毅 – 國立成功大學統計學系

Data-driven Anomaly Detection

Pei-Ting Chou

Department of Statistics, National Chengchi University

Abstract

Data-driven learning approaches, which rely on data to learn and make decisions, have gained significant attention in anomaly detection research. In this study, we focus on data-driven learning for anomaly detection and propose a novel method that utilizes a distance matrix defined through random walk. By employing random walk to explore data geometry, we record the recurrence time of the random walk results. This enables us to derive a same-cluster-sharing probability, which serves as a new distance measure. Additionally, we investigate the application of Siamese Neural Networks for similarity prediction between paired data as an alternative approach to anomaly detection. To evaluate the effectiveness of our proposed methods, real datasets will be employed. The outcomes of this study will provide valuable insights into the strengths and limitations of the proposed approaches for handling anomaly detection. The findings will be beneficial for practitioners and researchers engaged in anomaly detection problems.

An EWMA Control Chart for the Median of Birnbaum-Saunders Distribution

Ming-Che Lu[1], Su-Fen Yang[2]

[1]Department of Accounting, Chaoyang University of Technology

[2]Department of Statistics, National Chengchi University

Abstract

The Birnbaum-Saunders (BS) distribution is a lifetime distribution describing fatigue failures and general random wear failures. By means of monitoring the median of BS distribution, the quality control of lifetime of products subject to failures can be performed. This study explores the statistical process control of the median of BS distribution, and proposes an exponentially weighted moving average (EWMA) control chart to detect whether the distribution median shifts. We obtain the control limits for various small samples by a serious of simulations. The simulated in-control average run lengths (ARLs) of the proposed EWMA control chart are consistent with the preset levels of false alarm rates, documenting that the chart is stable and reliable. Comparing to the existing control charts for monitoring the median of BS distribution, the proposed EWMA control chart presents smaller out-of-control ARLs, indicating that the proposed EWMA chart performs better than the existing charts.

Keywords: Average run length, Birnbaum-Saunders distribution, exponentially weighted moving average, fatigue life, statistical process control

Control charts for profile monitoring of within- profile correlations using the Tweedie exponential dispersion process model

Chung-I Li[1], Meng-Rong Tsai[2]

[1] Department of Statistics, National Cheng Kung University

[2] Department of Statistics, National Cheng Kung University

Abstract

In many circumstances, the quality of a process or product is best characterized by the functional relationship between a response variable and one or more explanatory variables that is typically referred to as profile. Profile monitoring is used to test the stability of this relationship over time. The most relevant profile monitoring methods are usually based on the assumptions that the observations within each profile are independent of each other and normality distributed. However, in some applications, those assumptions could be violated in many quality control applications. In this study, we use the Tweedie process models to describe the within-profile correlation. Two exponentially weighted moving average control charts are proposed for monitoring phase II process when the within-profile data are correlated. The performance of the proposed methods is evaluated via simulation studies. Furthermore, the proposed methods are applied to a real data set.

Keywords: profile monitoring, Tweedie process, within-profile correlation. Phase II control chart

Session I-4

June 29, 13:30 – 14:50

理工一館 A109

◆Organizer/Chair：

蘇佩芳 – 國立成功大學統計學系

◆Speaker:

1. 曾議寬 – 國立中央大學統計所
2. 張玉媚 – 東海大學統計系
3. 温淑惠 – 慈濟大學公共衛生學系

Semiparametric models for the design on three-arm noninferiority trials

Yi-Kuan Tseng

Graduate Institute of Statistics, National Central University

Abstract

A gold standard design for three-arm noninferiority trials, which includes an experimental group, a reference group and a placebo group using parametric and semiparametric models was investigated for censored survival outcome. In the literature, so far, the sample size determination for three-arm design with survival endpoints is limited. Therefore, to complement the literature, the purpose of this study is to develop a statistical method to calculate the sample size for three-arm inferiority trials. Three semiparametric models, Cox, accelerated failure time and transformation models were considered for survival data. Minimum required sample size and testing power with optimal allocation of the three groups were considered as the criteria of performance of both parametric and semiparametric methods. Simulation studies were conducted to find out the performance of all methods. Moreover, treating the three groups as covariates, the semi-parametric hazards regression model can be employed to the three-arm testing procedure. The validity of the proposed method will be investigated by simulation studies. We illustrate the usefulness of the proposed methods through a bladder cancer data of three-arm design

Keywords: Noninferiority trials, Semi-parametric regression model, Optimal allocation

Adaptive Randomization Design for Multiarm Response – Survival Clinical Trials

Yu-Mei Chang

Department of Statistics, Tunghai University

Abstract

The outcome-adaptive randomization design (ARD) has the advantage of assigning more patients to superior treatment arms by updating assignment probabilities according to accumulating trial results. However, ARD implementation can be difficult when survival times are the primary endpoints, as it takes a long time to observe results. Surrogate endpoints, such as tumor shrinkage, are easier to implement since they are observable shortly after treatment. Huang et al. (2009) proposed a Bayesian ARD for randomized phase II trials that builds a connection between short-term response and long-term survival time through seamless design. In this study, we obtain the posterior mean of survival time for multi-arm under the Weibull distribution and compare the performance of three ARDs, AR_MAX, AR_MR, and DBCD, through simulation. AR_MAX calculates the probability of each treatment arm having the best therapeutic effect and performs Bayesian adaptive randomization, while AR_MR is a new design proposed by Yuan and Yin (2011) that reduces variability and improves efficiency. DBCD takes into account both the proportion of subjects assigned to each treatment and the probability that each treatment has the best therapeutic effect. Simulation results indicate that AR_MR requires a smaller number of total subjects, thus saving the cost of human resources, and DBCD assigns more subjects to the superior treatment, increasing their chances of receiving the best treatment compared to the other designs.

Keywords: adaptively randomization, doubly adaptive biased coin design, response rate, Bayesian method

A phenome-wide association study of postoperative infection in patients after total knee arthroplasty

温淑惠

慈濟大學公衛系

Abstract

We aim to use phenome-wide association study (PheWAS) to investigate potential causal relationship between preoperative comorbidities and postoperative infection in patients after total knee arthroplasty (TKA). To identify significant associated pre-operative comorbidities, both the National Health Insurance Research Databases (NHIRD), and UK Biobank (UKB) were used. Phecode system was adopted to measure comorbidities. Logistic regression was used to examine which preoperative Phecode are significantly associated with infection after TKA. Next, we used UKB to perform a PheWAS to examine the association between genetic risk scores (GRS) of associated comorbidity and a total of 736 Phecodes. The infection rate was 1.30% in NHIRD. A total of 17 Phecodes were significantly associated with postoperative infection. The majority of the associated Phecodes were related to cellulitis in the Dermatologic system. In UKB, the infection rate after TKA was 1.28%. We repetitively found that cellulitis-related Phecodes (OR=2.28, $p=1.92\times10^{-5}$) were associated with infection. Hence, we conducted a PheWAS to test the association of each Phecode with cellulitis-GRS. Four PheCodes including cellulitis and local infections (OR=1.10, $p=2.08\times10^{-5}$), varicose veins of lower extremity (OR=1.12, $p=5.52\times10^{-5}$), venous complications embolism in pregnancy (OR=1.24, $p=1.38\times10^{-4}$) and vitamin B-complex deficiencies (OR=1.21, $p=1.92\times10^{-4}$) were significantly associated with cellulitis-GRS. However, the cellulitis-GRS was not associated with postoperative infection (OR=1.00, p=0.99). For individual SNP analysis, we found that rs13195420 (OR=1.21, p=0.034) and rs7728043 (OR=1.19, p=0.044) were correlated with postoperative infection at nominal significance. Among preoperative comorbidities, cellulitis-related Phecodes are associated with postoperative infection after TKA. However, the genetic effect of cellulitis on the occurrence of postoperative infection in UKB was not evident. Further research of the PheWAS using Taiwan Biobank is needed.

Keywords: Postoperative infection, Phenome-wide association study, Phecode, UK Biobank

Session I-5

June 29, 13:30 – 14:50

理工一館 A209

◆Organizer/Chair：

林宗儀 – 國立中興大學應用數學系

◆Speaker:

1. 陳泰賓 – 國立陽明交通大學生物醫學暨工程學院
2. 王婉倫 – 國立成功大學統計學系
3. 楊郁成 – 國立中興大學應用數學系

The Algorithm for HRRT PET Reconstruction with Inverse Problem

Tai-Been Chen

Department of Medical Imaging and Radiological Sciences, I-Shou University

Institute of Statistics, National Yang Ming Chiao Tung University

Abstract

The reconstructed positron emission tomography (PET) is one of the applications in the field of inverse problem in this talk. A state-of-the-art high resolution research tomography (HRRT) provides high resolution PET images with full 3D human brain scanning. But, a short time frame in dynamic study causes many problems related to the low counts in the acquired data. The PDS-OSEM (**Ordered Subsets Expectation Maximization**) algorithm was proposed to reconstruct the HRRT image with a high signal-to-noise ratio that provides accurate information for dynamic data. The new algorithm was evaluated by simulated image, empirical phantoms, and real human brain data. Meanwhile, the time activity curve was adopted to validate a reconstructed performance of dynamic data between PDS-OSEM and OP-OSEM algorithms. According to simulated and empirical studies, the PDS-OSEM algorithm reconstructs images with higher quality, higher accuracy, less noise, and less average sum of square error than those of OP-OSEM. The presented algorithm is useful to provide quality images under the condition of low count rates in dynamic studies with a short scan time.

Keywords: Inverse Problem, Positron Emission Tomography, OSEM

A Selection Model for Multivariate Longitudinal Data with Censored and Nonignorable Intermittent Missing Outcomes

Wan-Lun Wang [1]*, Tsung-I Lin [2]

[1] Department of Statistics and Institute of Data Science, National Cheng Kung University

[2] Institute of Statistics, National Chung Hsing University

Abstract

The analysis of multivariate longitudinal data could encounter some complications due to censorship induced by detection limits of the assay and non-response occurring when participants missed scheduled visits intermittently or discontinued participation. This paper establishes a generalization of the multivariate linear mixed model that can accommodate censored responses and nonignorable missing outcomes simultaneously. To account for the non-ignorable missingness, the selection approach which decomposes the joint distribution as a marginal distribution for the primary outcome variables and a model describing the missing process conditional on the hypothetical complete data is used. A computationally feasible Monte Carlo expectation conditional maximization (MCECM) algorithm is developed for parameter estimation with the maximum likelihood (ML) method. Furthermore, a general information-based approach is presented to assess the variability of ML estimators. The techniques for the prediction of censored responses and imputation of missing outcomes are also discussed. The methodology is motivated and exemplified by a real dataset concerning HIV-AIDS clinical trials. A simulation study is conducted to examine the performance of the proposed method compared with other traditional approaches.

Keywords: AIDS clinical trials; MCECM algorithm; Missing not at random; Multivariate linear mixed model; Truncated multivariate normal distribution

Estimation and Prediction in *t* Linear Mixed Models with Non-ignorable Dropout

Yu-Chen Yang [1](Presenter), Wan-Lun Wang [2], Tsung-I Lin [3]

[1] Department of Applied Mathematics, National Chung Hsing University[2]

Department of Statistics and Institute of Data Science, National Cheng Kung University

[3] Institute of Statistics, National Chung Hsing University

Abstract

One common missing pattern in longitudinal data analysis is *dropout* in the sense that some subjects may withdraw prematurely. In general, dropout is regarded as a non-ignorable mechanism when the probability of missing outcomes is related to the unobserved data variable. The *t* linear mixed (tLME) models have been established as a robust tool for modeling continuous longitudinal data in situations of containing outliers or heavy-tailed noise. This paper investigates the use of the tLME models to address the problem of non-ignorable missing dropout. We consider the selection modeling strategy by incorporating a logistic response function to describe the probability of the dropout process. A Monte Carlo expectation conditional maximization (MCECM) algorithm is developed to simultaneously compute maximum likelihood (ML) estimates of the missingness indexing parameters of the logistic response function and mixed-effects parameters of full-data tLME model that are of scientific interest. Additionally, the standard errors of the ML estimates can be calculated using the Monte Carlo version of the information matrix obtained by the Louis formula. Simulation studies are conducted to assess the capability of tLME model in the presence of non-ignorable dropout and outliers. Meanwhile, the performance is compared to its normal counterpart in terms of information criteria indices, the accuracy of estimation for fixed effects and missingness indexing parameters, and prediction of missing responses. The proposed methodology is demonstrated through two real-world examples from an AIDS clinical trial and a schizophrenia study, both of which provide practical implications when the

non-ignorable dropout is considered in the analysis.

Keywords: Dropout, Incomplete longitudinal data, MCECM algorithm, Multivariate *t* distribution, Sensitivity analysis, Selection model

Session I-6

June 29, 13:30 – 14:50

理工一館 A211

◆Organizer/Chair：

銀慶剛 – 國立清華大學統計研究所

◆Speaker:

1. 楊睿中 – 國立臺灣大學經濟學系
2. 郭漢豪 – 國立臺灣大學經濟學系
3. 俞淑惠 – 國立高雄大學統計學研究所

Synthetic Historical Control for Policy Evaluation without Cross-sectional Controls

[1]陳宜廷，[2]楊睿中，[3]楊子霆
[1]國立臺灣大學財務金融學系
[2]國立臺灣大學經濟學系
[3]中研院經濟所

Abstract

We propose a synthetic historical control (SHC) method for policy evaluation in the absence of cross-sectional controls. This method predicts the counterfactual of the treated unit by replacing the infeasible cross-sectional controls of the synthetic control method with a set of historical controls based on a flexible time-series model. This model accounts for not only observed control variables but also a latent time-varying component in addition to the time-varying intervention effect. We illustrate the usefulness of our method using a Monte Carlo simulation and empirical applications.

Nonlinear Least Squares, Model Selection, and Model Averaging for Social Interaction Models

Hon Ho Kwok
Department of Economics, National Taiwan University

Abstract

This paper develops a unified theory of consistent estimation, model selection (MS), and model averaging (MA) for social interactions models (which include regression models as special cases).

First, the unified theory is based on prediction errors. The consistent and asymptotically normal estimators are nonlinear least squares (NLS) estimators. The MS and MS criteria are derived from prediction errors. The proposed NLS estimators are new; the commonly used consistent estimators are generalized methods of moments (GMM) and maximum likelihood (ML).

Second, the proposed MS and MA criteria are generalizations of Mallows Cp. We derive a set of penalty term based on prediction errors.

Third, we develop two types of MA: taking averages over predicted values (prediction averaging) and over parameters (parameter averaging).

Fourth, we provide in-depths discussions of the estimation quality and model selection problems in social interaction models. The MS methods are motivated by the problems in selecting networks, variables, and functional forms in the literature of social interactions. The MA methods are motivated by obtaining predicted values and parameter estimates of lower variances and mean squared errors.

New Results on Generalized Dynamic Factor Model: Prediction and Model Selection

Yang Lo[1], Shu-Hui Yu[2]

[1]National Tsing Hua University

[2] National University of Kaohsiung,

Abstract

Dynamic factor models are frequently used to model high-dimensional time series. The underlying concept can be said to be the methodology of dimension reduction. In this article, we propose two innovative prediction methods for dynamic factor models of high-dimensional time series. These approaches incorporate three ideas. First, based on spectral density analysis, we compute the lag *h* covariance matrix used for the estimation of the loading matrix. Second, based on the eigenvalues of a non-negative definite matrix, we determine the number of factors and use the weighted least squares and projection methods to estimate the common factors. Finally, we propose a new approach to determine the order of the vector autoregressive (VAR) model of dynamic factors and use the Bayesian information criterion (BIC) for model selection. In support of the performance and usefulness of the proposed methods, simulations and applications to financial data are provided.

Keywords: Dynamic factor model, dimension reduction, spectral density analysis, lag *h* covariance matrix, number of factors, model selection

Session I-7

June 29, 13:30 – 14:50

理工一館 A318

◆Organizer/Chair：

張中 – 國立中山大學應用數學系

◆Speaker:

1. 馬瀰嘉 – 國立成功大學統計學系
2. 張升懋 – 國立臺北大學統計學系
3. 李宗翰 – 國立中正大學數學系

誤差平方和最小化條件下的因子分析

蔡茜婷, 馬瀰嘉

國立成功大學統計學研究所

摘要

在現今大數據的世代中為了因應各種不同的資料型態，新的資料處理方法不斷推陳出新。而統計、電腦科學，以及資訊等各領域專家對於研發新方法的背景與對於其他領域既有方法的理解不完全，因此可能導致引用上的錯誤並使得結果不可靠。在現代研究結果高度流通的環境中，如此的錯誤結果可能在未經驗證的情況下再度被誤用。Sigg & Buhmann (2008) 提出應用 EM 演算法(Expectation – Maximization Algorithm)的非負稀疏主成份分析，能夠將高維度的資料進行快速的降維處理，在網路上被廣泛引用。但作者可能不熟悉統計領域的多變量分析理論，故混淆因子分析模型與主成份分析的觀念導致文獻數學公式的推導上有誤。本研究將對於作者錯誤的數學推導進行修正，並應用多變量統計分析(Johnson & Wichern，2007; Applied Multivariate Statistical Analysis)書中的傳統因子分析方法，提出使重建誤差平方和最小化的方法。接著使用修改 Sigg & Buhmann (2008)文獻後的方法與本研究所提出的因子分析 EM 演算法，分別使用模擬數據和實例作比較，以評估何者誤差平方和較小。在統計模擬中，我們生成維度 5 維、10 維和 15 維的資料，來觀察不同維度下對於兩種不同方法所得出的誤差平方和有何影響。在實例中，我們針對臉部影像的高維度資料進行操作，以了解我們修改過後的方法在高維度資料降維的表現。

關鍵詞：因子分析、主成份分析、資料降維

Modeling Multiple-Criterion Diagnoses by Heterogeneous-Instance Logistic Regression

Chun-Hao Yang[1] Ming-Han Li[2] Shu-Fang Wen[2] Sheng-Mao Chang*[3]

[1] Institute of Applied Mathematical Sciences, National Taiwan University,

[2] Biomedical Technology and Device Research Laboratories, Industrial Technology Research Institute

[3]Department of Statistics, National Taipei University

Abstract

The diagnoses of Alzheimer's disease (AD) and its prodromal stage mild cognitive impairment (MCI) are examples of multiple-criterion diagnoses. Clinically, the diagnosis of MCI/AD is determined by the impairment statuses of five cognitive domains. If one of these cognitive domains is claimed impaired, the patient is diagnosed with MCI, and if two out of the five domains are impaired, the patient is diagnosed with AD. This diagnostic procedure relates the MCI/AD status modeling to multiple-instance learning [1], where each domain resembles an instance. However, traditional multiple-instance learning assumes common predictors among instances, but in our case, each domain is associated with different neuropsychological questionnaires [2] as predictors. In this paper, we generalized the multiple-instance logistic regression to accommodate the heterogeneity in predictors among different instances. The proposed model is dubbed Heterogeneous-Instance Logistic Regression (HILoR) and the expectation-maximization algorithm[3] is applied to solve the estimation problem arising from the missing instance statuses. We also derived two variants of HILoR from modeling both the MCI diagnosis and AD diagnosis. Because HILoR fully considers the probability structure of a multiple-criteria diagnosis, the proposed model additionally provides the probabilities of being impaired in each criterion/domain. Take our data analysis, for example. We predicted one's MCI (AD) status after 3 (5) years of the initial visit and then provided the impairment probabilities (or percentiles) in each of the five domains. Treatments can therefore be applied

according to more deteriorated domains. This merit contributes to personalized disease prevention. Finally, the proposed HILoR model is validated in terms of its estimation accuracy, latent status prediction, and robustness via extensive simulation studies and real data analysis.

Keywords: EM Algorithm, Logistic Regression, Multiple-Criteria, Diagnosis, Multiple-Instance Learning

Exact Confidence Coefficients of Confidence Interval for a Difference of the Binomial Proportions

Yu-An Huang（黃于安）and Chung-Han Lee*（李宗翰）

Department of Mathematics, National Chung Cheng University

Abstract

For a confidence interval of a parameter in two independent binomial distributions, the coverage probability is a variable function of the parameter. The confidence coefficient is the infimum of the coverage probabilities. Since we do not know which point in the parameter space the infimum coverage probability occurs at, the exact confidence coefficients are unknown. In this article, we propose a method for calculating the exact confidence coefficients of confidence intervals for a difference of the binomial proportions. Using this methodology, we illustrate the performance of existing intervals and provide recommendations.

Keywords: Binomial distribution; Confidence coefficient; Confidence interval; Coverage probability

Session I-8

June 29, 13:30 – 14:50

理工一館 A314

◆Organizer/Chair：

鄭芳怡 – 國立中央大學大氣科學學系

◆Speaker:

1. 柯婕盈 – 國立中央大學大氣科學學系
2. 林遠見 – 國立中央大學土木工程學系
3. 陳柏孚 – 國立臺灣大學氣候天氣災害研究中心
4. 徐佩君 – 國立中興大學循環經濟研究學院

Unsupervised Machine Learning for Analyzing Ensemble Simulations of Extreme Heavy Rainfall Event in NorthernTaiwan

Chieh-Ying Ke (柯緁盈), Pei-Jung Tsai (蔡沛蓉), Kao-Shen Chung (鍾高陞)

Department of Atmospheric Sciences, National Central University

Abstract

This study explores the application of various unsupervised machine learning methods to analyze ensemble simulations of an extreme heavy rainfall event in Northern Taiwan. We employed clustering algorithms, such as K-means, Spectral Clustering, OPTICS and BIRCH, to identify distinct rainfall patterns. The 128-member ensemble simulations, which were obtained and perturbed from National Centers for Environmental Prediction (NCEP) Final operational global analysis (FNL), are generated by WRF (Weather Research and Forecasting Model). By applying clustering algorithms and dimensionality reduction techniques, distinct rainfall patterns and dominant meteorological variables are identified. These findings contribute to a deeper understanding of the event and can aid in enhancing forecasting and mitigation strategies for similar events worldwide.

Keywords: unsupervised machine learning, ensemble simulations, extreme heavy rainfall, clustering.

Investigation of the nonlinear relationship between $PM_{2.5}$ and ten leading causes of death based on risk analysis approach

Yuan-Chien Lin[1*], Hua-San Shih[1], Chun-Yeh Lai[1]

[1]Department of Civil Engineering, National Central University

Abstract

Air pollution is a growing global concern that has raised significant alarm. Numerous epidemiological studies have demonstrated connections between fine particulate matter ($PM_{2.5}$) and various diseases. However, these studies often rely on short-term data and models that focus on specific groups, making it challenging to comprehend the long-term impact, non-linear associations, and spatial-temporal health risks associated with chronic diseases. Therefore, this study proposes an analysis and mapping of the likelihood of $PM_{2.5}$ exceeding acceptable levels using long-term spatial-temporal monitoring data through radial basis function estimation. Additionally, we construct and compare multiple linear regression and generalized additive models to explore both linear and non-linear relationships between the long-term average $PM_{2.5}$ concentration, the probability of $PM_{2.5}$ surpassing the standard, and standardized mortality for the top ten causes of death across all towns and villages in Taiwan from 2010 to 2017.

The findings from the linear models indicate that increasing $PM_{2.5}$ concentration leads to elevated mortality rates for malignant neoplasms, pneumonia, chronic lower respiratory diseases, chronic liver diseases, and cirrhosis. Conversely, mortality rates for heart diseases and esophageal cancer show a decrease. As for the results from the non-linear model, they reveal significant non-linear associations between $PM_{2.5}$ concentration and mortality rates for malignant neoplasms, heart disease, diabetes, and cancers of the trachea, bronchus, lung, liver, intrahepatic bile duct, and esophagus. Consequently, long-term exposure to $PM_{2.5}$ may pose a substantial risk factor for numerous acute and chronic diseases. The implications of this study extend beyond the local context, offering valuable air quality and health management references for governments worldwide and crucial information regarding long-term health risks for residents in the study area.

Keywords: air pollution, chronic diseases, ten leading causes of death, risk analysis, generalized additive model, $PM_{2.5}$ potential map

Deep Learning Satellite Imagery Analysis to Reconstruct Historical Typhoon Data

Buo-Fu Chen

Center for Weather Climate and Disaster Research, National Taiwan University

Abstract

The anthropogenic influences have been linked to Tropical cyclone (TC) poleward migration, extreme precipitation, and an increased proportion of major hurricanes. Understanding past trends and variability in TC circulation and structure is critical for projecting future TCs in the changing climate. However, the subjective-analyzed and spatiotemporal heterogeneous historical datasets lead to uncertainties and reduced confidence in the assessed responses of TC to climate change. Also, the lack of literature discussing TC structure in the changing climate highlights the need for more data.

Here, we use deep learning to create "observations" in the past and construct a objective global TC structure dataset since 1981. Training on a uniquely-labeled dataset integrating historical "best-tracks" and numerical model reanalysis of 2004–2016 TCs, our model converts multichannel satellite imagery to a 0-750-km wind profile of the axisymmetric surface winds. The model performance is verified by comparing it to independent satellite radar observation of surface winds for 2017−19 TCs.

Based on our homogenized dataset, we show a decreasing trend of annual-mean TC circulation but also a increasing exposure risk to extremely large TCs. Notably, our data suggest that the revealed trends regarding TC intensity based on the best-track data are exaggerated. We anticipate this new deep learning method/dataset opens a door to further improve operational forecasting and climate modeling of TC structure evolution.

Keywords: Deep learning, Satellite data, Typhoon, Tropical cyclone structure

2021 Texas Cold Snap: Manifestation of Natural Variability and a Recent Warming Trend

Pei-Chun Hsu[1,2], Huang-Hsiung Hsu[1], Hao-Jhe Hong[1], Ying-Ting Chen[1], Yu-Luen Chen[1], and Wan-Ling Tseng[1]

[1]Research Center for Environmental Changes, Academia Sinica

[2] Academy of Circular Economy, National Chung Hsing University

Abstract

We demonstrate that the extreme cold snap in Texas in February 2021 resulted from a large-scale background circulation anomaly over the North Pacific and North America that was a manifestation of both natural variability (namely, a negative Pacific–North American pattern, PNA) and a warming trend that began in the late 1990s. Numerical experiments revealed that the negative PNA pattern could be driven by negative (La Niña-like) sea surface temperature (SST) anomalies in the equatorial eastern Pacific on an interannual–decadal timescale and the SST warming trend in the subtropical North Atlantic on a longer timescale. Our study demonstrated the key role of extratropical wave activity and the intrinsic mode in causing the extreme cold in North America. The arguments favoring polar or tropical influences may partially explain the source of variability, considering that extratropical weather and climate variability could be modulated by not only polar or tropical perturbations but also the enhanced tropical–extratropical interaction and mid-latitude instability. A systematic approach involving both empirical diagnostics and a suite of carefully designed numerical experiments is required to determine the relative influence of various sources of variability.

Keywords: Extreme cold event, Natural variability, Warming trend, Pacific–north American pattern, Sea surface temperature

Session I-9

June 29, 13:30 – 14:50

理工一館 A316

◆Organizer/Chair：

鄒宗山 – 國立中央大學統計研究所

◆Speaker:

1. 戴安順 – 國立成功大學統計學系
2. 王偉銘 – 國立成功大學
3. 簡莉珠 – 高雄醫學大學基礎科學教育中心

Causal mediation analysis for recurrent event using additive rate models: generalizations from the Baron and Kenny method

An-Shun Tai (戴安順)

國立成功大學 統計學系

Abstract

With recent advances in causal mediation models, a growing number of medical researchers are thinking about investigating the mediation mechanism. In practice, the Baron and Kenny method, also called the product method, has been broadly used in causal mediation analysis. However, the Baron and Kenny method can produce biased estimates for the causal mediation effect in the context of recurrent event analysis. To address this, this study proposes a novel mediation model for recurrent event based on additive rate models. The proposed method generalizes the Baron and Kenny estimator as it allows time-to-event outcomes and non-additive covariates structure. A robust semiparametric locally efficient estimator is proposed for estimation. In a real-world application, the proposed method is used to evaluate the average treatment effects of two diabetes drugs on cardiovascular disease recurrence and investigate the potential mediation role of clinical factors in this process.

A Modified Single-stage Sampling Procedure for Heteroscedasticity Analysis of Means

Wei-Ming Wang[1], Miin-Jye Wen[1, 2]

[1]Department of Statistics and Institute of Data Science, College of Management, National Cheng Kung University

[2]Institute of International Management, College of Management, National Cheng Kung University

Abstract

The analysis of means (ANOM) is a method that can compare the mean of each treatment to the overall mean. According to the graphical result of a statistical data analysis, we can specify which one is different from another. One of the assumptions of the classical ANOM model is that the variances are equal. However, it is not always true for the practice. We applied two inference procedures, single-stage and modified single-stage sampling, to solve heteroscedastic analysis of means (HANOM). We provided a real application for illustrating these two procedures clearly. In addition, we build a user-friendly interface to show the results of the HANOM by using *R* Shiny.

Keywords: heteroscedasticity, single-stage sampling procedure, analysis of means.

Combining dependent *p*-values by gamma distributions

Li-Chu Chien

Center for Fundamental Science, Kaohsiung Medical University

Abstract

Combining correlated *p*-values from multiple hypothesis testing is a most frequently used method for integrating information in genetic and genomic data analysis. In this work, we propose an empirical method based on the gamma distribution (EMGD) for combining dependent *p*-values from multiple hypothesis testing. The proposed test, EMGD, allows for flexible accommodating the highly correlated *p*-values from the multiple hypothesis testing into a unified *p*-value for examining the combined hypothesis that we are interested in. The EMGD retains the robustness character of the empirical Brown's method (EBM) for pooling the dependent *p*-values from multiple hypothesis testing. Moreover, the EMGD keeps the character of the method based on the gamma distribution that simultaneously retains the advantages of the z-transform test and the gamma-transform test for combining dependent *p*-values from multiple statistical tests. The two characters lead to the EMGD that can keep the robust power for combining dependent *p*-values from multiple hypothesis testing. The performance of the proposed method EMGD is illustrated with simulations and real data applications by comparing with the existing methods, such as the EBM and the harmonic mean *p*-value method.

Session I-10

June 29, 13:30 – 14:50

理工一館 A210

◆Organizer/Chair：

Lung-Chi Chen – 國立政治大學應用數學系

◆Speaker:

1. Cheuk-Yin Lee – 國立清華大學數學系
2. Yuki Chino – 國立陽明交通大學應用數學系
3. Yoshinori Kamijima – 國家理論科學中心數學組

Hitting probabilities of Gaussian random fields and SPDEs

Cheuk Yin Lee

National Tsing Hua University

Abstract

Gaussian random fields and SPDEs play important roles in probability and statistical physics. In this talk, I will discuss hitting probabilities of these stochastic processes. A basic problem is to determine whether a d-dimensional process hits a given set with positive probability. For Gaussian random fields and SPDEs, there have been some known upper and lower bounds for hitting probabilities, which allow us to determine if hitting can occur in non-critical dimensions, but the bounds give no conclusion in critical dimensions. Usually, this problem is more difficult to solve in the critical case. I will present some recent results on hitting probabilities for Gaussian random fields and systems of SPDEs in their critical dimensions.

Keywords: Hitting probabilities, stochastic processes, Gaussian random fields, systems of stochastic PDEs, Gaussian noise

Affection for Critical Point and Asymptotic Behaviour of SAW from Random Environment

千野由喜(Yuki Chino)

國立陽明交通大學應用數學系

Abstract

Self-Avoiding Walk (SAW) is one of statistical-mechanical models, which has been studied to understand the behavior of polymers. In general we treat SAW models on various lattices but through this talk we treat square and triangle lattices. We consider random conductances attached to each bond on the lattice and call SAW on random conductors. The model shows different critical point and behavior than the model on homogeneous lattices. Similar model was studied by H. Lacoin [2, 3] in 2014 and has already known some results, which is related the difference for the critical point on a super-critical percolation cluster. The speaker already showed the existence and some estimate of the critical point in the collaboration with A. Sakai [1] in 2016. In this talk we would like to discuss on the critical behavior on square lattice in each dimension.

References

[1] Y. Chino and A. Sakai. The quenched critical point for self-avoiding walk on random conductors. J. Stat. Phys. 163 (2016): 754–764.

[2] H. Lacoin. Non-coincidence of quenched and annealed connective constants on the supercritical planar perco- lation cluster. Probab. Theory Relat. Fields 159 (2014): 777–808.

[3] H. Lacoin. Existence of a non-averaging regime for the self-avoiding walk on a high-dimensional infinite percolation cluster. J. Stat. Phys. 154 (2014): 1461–1482.

A local limit theorem for the long-range self-avoiding walk

上島芳倫(Yoshinori Kamijima)

Mathematics Division, National Center for Theoretical Sciences

Abstract

The self-avoiding walk (SAW) is a model added self-avoidance interaction to the random walk. In other words, each path does not visit the same vertex on a graph more than once. It is known that the spread-out short-range SAW enjoys the central limit theorem [van der Hofstad and Slade (2003) *AAM*]. Taking an average on a ball, they also proved a certain type of a local limit theorem. For the spread-out long-range SAW, the power-law decay of the two-point function was shown in [Chen and Sakai (2019) *CMP*]. In this talk, I will explain an attempt to prove a local limit theorem for the spread-out long-range SAW in the original sense. Our motivations come from combining the results of the previous researches. This is a joint work with Lung-Chi Chen (National Chengchi University) and Yuki Chino (National Yang Ming Chiao Tung University).

Session I-11

June 29, 13:30 – 14:50

理工一館 A106

◆Organizer/Chair：

Wei-Wen Hsu – Department of Environmental and Public Health Sciences, University of Cincinnati

◆Speaker:

1. 張憶壽 – 國家衛生研究院
2. Wei-Wen Hsu – Department of Environmental and Public Health Sciences, University of Cincinnati
3. 鍾冬川 – 逢甲大學統計學系

Adapting the lung cancer PLCO model for Taiwan

I-Shou Chang

National Health Research Institutes

Abstract

The lung cancer PLCO model for smokers was developed using the control arm of the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial in the USA. Selecting individuals for low-dose CT lung cancer screening based on PLCO 2012 model has certain advantages over the NLST criteria that consider only people aged between 55 and 74 and having smoked 30+ pack-years and smoked in the past 15 years. PLCO 2012 model was prospectively validated in the USA, Canada, Germany, Australia, UK, Brazil, and Poland. It is beneficial to validate or adapt it for Taiwan. Because there is no suitable prospective cohorts for its validation/adaptation, we performed the adaptation by synthesizing the SPES2010 (Synthesized Taiwanese population of cancer-free ever-smokers in 2010) and AMCCSE (Age-matched case-control study of lung cancer among ever-smokers) and by estimating the NESLP2010 (number of ever-smoking lung cancer patients diagnosed in 2011—2016). Our adaptation considered overall calibration slope and calibration-in-the-large issue. Evaluation of the adapted model PLCOT using AMCCSE, SPES2010, and Taiwan Cancer Registry shows that PLCOT had an AUC of 0.78 and excellent calibration on subgroups defined by age and smoking experience, and better clinical usefulness than criteria based on age limits and smoking experience. Starting July 1, 2022, the Taiwan Administration of Health Promotion started the Lung Cancer Early Detection Project that offers free low-dose CT lung cancer screening to individuals satisfying the NLST criteria. This work may be useful for counseling. (Part of this work is based on Chien et al. 2022)

Keywords: Absolute risk prediction models, Data synthesis, Lung cancer, Model updating

A Covariate-Adjusted Classification for Multiple Longitudinal Biomarkers

Suizhi Yu[1], Pai-Ling Li[2], Wei-Wen Hsu[3]

[1] Department of Statistics, Kansas State University, USA

[2] Department of Statistics, Tamkang University, Taiwan

[3] Division of Biostatistics and Bioinformatics, Department of Environmental and Public Health Sciences, University of Cincinnati, USA

Abstract

The classification methods based on a linear combination of multiple biomarkers have been widely used to improve disease screening and diagnosis accuracy. However, their extensions to multiple longitudinal biomarkers are often computationally expensive and rely on restrictive assumptions such as equally spaced time points. Oftentimes, important covariates that are associated with biomarkers or patient outcomes are not incorporated into these classification procedures due to additional complexity. In this paper, we propose a simple classification method that is particularly for multiple longitudinal biomarkers with an adjustment for important covariates. With the technique of natural cubic spline basis, each longitudinal biomarker can be characterized by spline coefficients with a significant dimension reduction. The proposed method is a nonparametric two-stage method with the first step combining all spline coefficients obtained from every longitudinal biomarker and then adjusting for important covariates as the second step. Specifically, the optimal linear combination of those spline coefficients can be acquired using an AUC-based stepwise method without any distributional assumption. Afterward, covariates are included for additional improvement in classification. The asymptotic properties can be shown with the maximum rank correlation estimators. For illustration, the proposed method is applied to Alzheimer's Disease data and the primary biliary cirrhosis data. We also conduct an extensive simulation study to assess the finite-sample performance of the proposed method for multiple longitudinal biomarkers.

Keywords: Alzheimer's Disease, AUC, Nature Cubic Spline, Disease Detection

Effect of Polytomy on the Parameter Estimation and Goodness of Fit of Phylogenetic Linear Regression Models for Trait Evolution

鍾冬川

逢甲大學統計學系

Abstract

Phylogenetic regression models are powerful tools for studying trait evolution across related species by incorporating phylogenetic trees into the analysis of comparative data. Yet, the accuracy of these trees, particularly polytomy trees, can impact the validity of comparative analyses. This study conducts extensive simulations to evaluate the performance of Gaussian process-based regression models (Brownian motion, Ornstein--Uhlenbeck process, early burst) and branch-stretching models (Pagel's λ, δ, κ) when faced with soft polytomies at the root or an inadequately defined clade. Although polytomy generally doesn't significantly alter model fit or parameter estimate, differences are observed between trees and models. Notably, the Pagel λ and OU models yield more accurate estimates and superior fits compared to other models. Thus, while correcting phylogeny is crucial, choosing suitable models is equally important when handling polytomies.

Keywords: phylogenetic comparative analysis; regression analysis; polytomy; Gaussian process; trait evolution

Session I-12

June 29, 13:30 – 14:50

理工一館 D131

◆Chair：

高正雄 - 國立中正大學數學系

◆Speaker:

1. 趙伯昇 – 國立中山大學應用數學系
2. 郭冠麟 – 國立高雄大學統計學研究所
3. 林家弘 – 國立東華大學應用數學系統計碩士班
4. 馬繻嬪 – 國立中央大學統計研究所
5. 王薇淳 – 國立臺北大學資訊與工程學系

On the Asymptotic Behavior of Allocation Scheme with Geometric Distributions

Po-Sheng Chao (趙伯昇)

Department of Applied Mathematics, National Sun Yat-sen University

Abstract

Consider an allocation scheme of *n* indistinguishable balls independently distributed into *N* different boxes with equal probability. Then the contents of the boxes follow the multinomial distribution, which is a usual allocation scheme. A generalization of this scheme was introduced by Kolchin (See Kolchin [1] and Pavlov [2]), allowing the allocation probability to be unequal. In this talk, we will focus on Kolchin's model with the allocation probability driven by geometric distributions. We will show the asymptotic behavior of the order statistics of the numbers of balls allocated to boxes.

Keywords: allocation scheme, Discrete Gamma Distribution, Converge in Distribution

Reference

[1] V. F. Kolchin, *Random Graphs*, Encyclopedia Math. Appl. 53, Cambridge Univ. Press, Cambridge, 1999.

[2] Yu. L. Pavlov, Random forests. *Probabilistic Methods in Discrete Mathematics* (Petrozavodsk, 1996), VSP, Utrecht, 1997, pp. 11–18.

A Novel Mechanism for Describing Zero-inflated Data: An Application with Zero-Inflated Positive Skellam Distribution

Kuan-Lin Kuo (郭冠麟)

National University of Kaohsiung

Abstract

The Skellam distribution models the difference between two independent latent Poisson variables that represent opposing effects on count responses, referred to as the "assistance" and "resistance" effects and modeled by the two latent variables, respectively. In cases where the resistance effect dominates the assistance effect, the observed count data may be zero due to the absence of positive contributions from the assistance effect. Therefore, only the positive values of the Skellam distribution are observable, and the negative values contribute to zero counts. We refer to this model as the zero-inflated positive Skellam (ZIPSK). In this article, we apply the ZIPSK regression model to analyze count data with excess zeros, where the regression component models the effects of predictors on the assistance and resistance parameters. We estimate the ZIPSK regression model using the maximum likelihood (ML) method and impose a forward selection procedure to specify each predictor to the assistance and resistance part, respectively. The numerical results demonstrate satisfactory performance of the proposed methodology in parameter estimation and model selection. We compare the predictive performance of the proposed methodology with the hurdle Poisson and zero-inflated Poisson models, which are commonly applied to zero-inflated data.

Keywords: maximum likelihood estimation, model selection, Skellam distribution, zero-inflated data, ZIPSK regression model

Statistical Learning in the Universal Kriging

Chia-Hung Lin

Department of Applied Mathematics, National Dong Hwa University, Taiwan

Abstract

Ensemble learning consists of multiple algorithms to enhance predictive performance. However, such superiority may lose for the dependent data such as the geographical data.Combining the ensemble learning with the variogram approach, a learning procedure is proposed to look for the adequate model. Further, kriging based on the selected model is applied in the forecast issue. The simulation study demonstrates that the proposed procedure works well for the forecast issue. The analysis and comparison with an existing method for PM 2.5 data show better performance of the proposed method.

Keywords: ensemble learning, dependent data, variogram approach, kriging, PM 2.5 data.

Unsupervised Generalized Learning-based NEM Algorithm for Clustering

Ru Pin Ma

Graduate Institute of Statistics, National Central University

Abstract

Cluster analysis is a common method to preview the underlying structure of the data in machine learning, image analysis, and ecological researches. Most of unsupervised learning clustering methods such as the EM algorithm and the NEM algorithm were developed based on the mixture likelihood ratio and the number of clusters needs to be specified in advance. In addition, the accuracy of the NEM method is sensitive to the specification of the initial value and it also suffers the singular issues in the computation of the matrix especially for the three-dimensional data sets. In this thesis, a generalized learning-based NEM (GLB-NEM) algorithm is proposed which not only solves the sensitivity of the initial value and the singular issues of the matrix, but also the proposed algorithm can automatically converge to the correct number of clusters. Statistical inferences about the proposed GLB-NEM cluster algorithm are justified both in theories and simulation studies. Also, two real data examples are analyzed for illustration. (This is a joint work with Chun-Shu Chen)

Keywords: Cluster analysis; Gaussian mixture model; Machine learning; NEM algorithm; Unsupervised learning

利用統計方法與集群分析預測暨評估平衡度

王薇淳[1]、黃怡婷[1]、林伯星[2]

[1]國立臺北大學統計學系

[2]國立臺北大學資訊與工程學系

摘要

每個人行動平衡度常會因爲受傷、生病或是老化等方面的原因造成下降，臨床都會採用五個尺度量表，如柏格氏平衡量表（Berg Balance Scale; BBS），進行一次評估，但非常耗時，且評估會因測量者的經驗而有差異，再者，有些受試者執行方式明顯不協調，但 BBS 卻得分一樣，限制 BBS 的應用廣度。 Lin 等人（2022）設計一個穿戴式慣性感測裝置，讓受試者執行 BBS 所設計的動作，感測裝置有加速度與角速度儀，會記錄動作的加速度與角速度，經由深度學習模型，也驗證系統可以有效預測 BBS 分數。但該系統僅在約 150 位健康受試者測試，深度學習模型有可能有過度配適的疑慮。

本研究提出使用樣條函數來配適動作對應加速度與角速度的軌跡，以樣條函數對應的係數估計當成受試者的特徵變數，再採用相關分析（Correlation Analysis)、主成分分析（PCA) 與 獨立成分分析（Independent Component Analysis, ICA ）合併切片逆迴歸法（Sliced Inverse Regression, SIR）等方法進行維度縮減，最後以支持向量機進行預測，並計算平均絕對誤差（Mean Absolute Error; MAE）作為模型表現評斷標準。最後，以找出的特徵變數，本研究以 K-means 與 PAM（Partition Around Medoids）兩種非監督集群法，將分數一樣的受試者分群，優化 BBS 的判讀。

研究發現有些任務的相異性以及重點穿戴位置，僅需在右上臂與兩大腿穿戴慣性感測裝置即可，動作僅需雙腳併攏維持站姿一分鐘、站著轉頭向後看、轉圈走 360 度、做到站共四項簡單的測試動作即有所區別。除此之外，任務的執行時間長度亦與測驗相關，在精簡測驗時間與任務的條件下，可區別真正健康的受試者與潛在跌倒風險的受試者，以作為研究者探討功能性平衡的分析之參考。

關鍵詞：Berg Balance Scale, PCA, Correlation Analysis, ICA, SIR, SVR, K-means , Partition Around Medoids

Session I-13

June 29, 13:30 – 14:50

理工一館 A307

◆Chair：

黃逸輝 - 淡江大學數學系

◆Speaker:

1. 黃逸輝 – 淡江大學數學系
2. 葉詠富 – 國立陽明交通大學數據科學與工程研究所
3. 鄭亦甯 – 國立臺北大學統計學系
4. 羅祖佑 – 國立中山大學應用數學系
5. 王嘉顥 – 國立臺北大學統計學系
6. 莊沅蓉 – 國立臺灣大學統計與數據科學研究所

加速壽命模型於測量誤差下的無母數回歸校正

黃逸輝
淡江大學數學系

Abstract

Accelerated failure time (AFT) models are appealing due to their intuitive interpretation. However, when covariates are subject to measurement errors, the naive estimation can lead to severe bias.

To address this issue, the regression calibration (RC) approach is a widely applicable and effective method.

The RC method requires a good estimate of the conditional mean of the true covariate, which can be obtained through a parametric distribution or a validation dataset.

Here, we propose a novel method that utilizes error augmentation to duplicate independent covariates to facilitate
nonparametric estimation of the conditional mean.

Our approach does not require a validation set or make parametric distribution assumptions about the true covariate. We demonstrate through simulation studies that our approach is robust in many aspects and is also less impacted by heavy censoring than traditional approaches.

Locating the Ampulla of Vater and Estimating the Direction of Bile Duct by Deep Learning Methods

Author Yung-Fu Yeh[1], Cheuk-Kay Sun[2], Henry Horng-Shing Lu[3]

[1]Institute of Data Science and Engineering, National Yang Ming Chiao Tung University
[2] Shin Kong Wu Ho-Su Memorial Hospital
[3] Institute of Statistics, National Yang Ming Chiao Tung University

Abstract

The identification of the ampulla of Vater and the direction of the bile duct during endoscopic retrograde cholangiopancreatography (ERCP) is a challenging procedure in gastroenterology and hepatology. Failure to accurately identify the direction of the bile duct can result in treatment failure and fatal complications. However, anatomical variations in the ampulla of Vater can make it difficult to accurately locate it, leading to significant delays in ERCP. In fact, if the delay exceeds 5 minutes, more than 5 cannulation attempts, or more than 2 pancreatic guidewire passages, it will result in ERCP failure.

To address this issue, we collect 186 ERCP video films from patients treated by Dr. Sun at Shin Kong Wu Ho-Su Memorial Hospital in Taiwan between 2021 and 2022. We processed and learned the data from each frame of the video films. Using the transfer learning technique on the pre-trained EfficentNetB0 model, we classified whether the ampulla of Vater was included in every endoscopic image with 97.73% accuracy. Then we used the YOLOv4 object detection model to locate the ampulla of Vater and achieved an AP of 55.64%, an AP50 of 92.36%, and an AP75 of 59.79%. Finally, we estimated the direction of the bile duct using image processing techniques, with a mean angle error (MAE) of 15.67º and a mean angle error rate (MAER) of 8.71% in the test set.

Based on these models, we established a computer-assisted system for locating the ampulla of Vater and estimating the direction of the bile duct during ERCP procedures. This system can assist physicians in shortening the examination time and effectively reducing the complications caused by the inspection.

Keywords: Transfer learning, Object detection, ERCP, Ampulla of Vater, Bile duct

風險對比特幣、比特幣與金融資產連動之影響

鄭亦甯
國立臺北大學統計學系

摘要

本研究欲探討風險(包含地緣政治風險與恐慌指標)如何影響比特幣的報酬及波動性，和風險如何影響比特幣與其他金融資產(包含股市、匯市、原物料)的動態相關性。利用一般化自我迴歸條件異質變異模型(GARCH 模型)、不對稱 GARCH 模型分別分析比特幣的波動性是否具不對稱性；再透過動態條件相關係數模型(DCC 模型)分析比特幣與其他資產價格的動態相關性，探討風險如何影響比特幣與其他資產的關聯性。根據實證結果顯示，恐慌指標對比特幣的報酬有顯著負面的影響，而地緣政治風險對比特幣的報酬有顯著的正面影響。波動性方面，兩者對比特幣則是顯著的正面影響。再者，VIX 對於比特幣和其他資產的動態相關性，顯著性會因資產不同而有所差異，其中，石油、英國股票代表和亞洲地區的股票代表及外匯有較明顯的影響。

關鍵詞：比特幣、VIX 指數、地緣政治風險、GARCH 模型、DCC 模型

基於動態網路模型的虛擬貨幣投資組合

作者： 羅祖佑
指導教授：郭美惠
國立中山大學應用數系

摘要

在金融市場中，投資組合的目於對產進行多元化分配以達到降低風險的效果。本研究中，我們提出了一個基於動態網路模型虛擬貨幣投資組合方法。首先，我們使用加密貨幣的報酬率計算之間皮爾森相關係數，接著使用最小生成樹和大過濾圖建立加密貨幣 相鄰矩陣及金融 網絡。再 來，我們計算 相鄰矩陣的 特徵中心性，並使用高、 低 ρ-dependent 以及 partial ρ-dependent 等方法選擇投資組合中的加密貨幣。在權重 等方法選擇投資組合中的加密貨幣。在權重 配置方面，我們 討論利用等權重（ equal weight）、切線投資組合 權重 （tangency portfolio weight）及階層式風險平價（ hierarchical risk parity）作為投資策略。 作為投資策略。 最後，我們使用動態更新的方式調整每個 周期 的投資策略。 研究結果顯示，不 同的更新 周期和投資策略會對組合的表現產生不同影響。 若每日更新投 資組合 ，我們建議追求高報酬的投資 者使用以最小生成樹搭配動態更新策略； 若每月更新投資組合 ，則建議追求高報酬投資者使用最大過濾圖並搭配 partial ρ-dependent 及切線投資組合 切線投資組合 權重 的投資策略進行。

關鍵詞： 金融網路、 最小生成樹大過濾圖階層式風險 平價 、 投資組合特徵中心性

CEO 可信度對加密貨幣價格之影響

王嘉顥

國立臺北大學統計學系

摘要

在本篇論文中，主要在探討加幣貨幣創辦人或者 CEO 的第一印象對投資者是否決定投資一間公司的影響，當一間公司剛成立時，投資者有著嚴重的資訊不對稱，尤其在加密貨幣這個缺乏監管的領域，是否要投資一家公司對投資者而言是一個艱難的決定，在缺乏資訊的情況下，投資者往往只能透過創辦人或 CEO 來判斷，這個時候 CEO 給人的第一印象就顯得非常重要了，在本文中透過 CEO 的三個臉部特徵來將第一印象的評分依據具體化。本文中將 CEO 的臉部特徵分為三個部分，分別是可信度(Trust)、吸引力(Attr)和支配性(Dom)，為了量化可信度、吸引力和支配性的特徵，我們使用機器學習方法的結構來預測這些臉部特徵分數。研究結果顯示，CEO 的可信度，會給投資者帶來信心，讓他們相信擁有這些特質 CEO 的公司是可靠的，因此投資者可能會願意購買可信度高的 CEO 所在公司發行的貨幣，在這種情況下，也會讓加密貨幣的價格更高，為 CEO 本人及公司帶來收益。

關鍵詞：臉部特徵、金融科技、價格表現

Nonparametric Mediation Analysis of Survival Outcomes

莊沅蓉、黃彥棕、王建玲

國立臺灣大學統計與數據科學所、中央研究院統計所、Department of Statistics, University of California, Davis

Abstract

Most mediation analyses for survival outcomes depend on models, which are subject to misspecification. In this work, we propose a nonparametric estimation method for the direct and indirect effects of mediation analysis. We aim to estimate the cumulative distribution functions of the effects, so for the counterfactual distribution of the mediator and time-to-event outcome, we apply the (adjusted) Nadayara-Watson estimator and Beran estimator, respectively. Theoretical properties as well as simulation results are also provided. We applied our method to two studies. One is the influenza vaccination data collected from children in Hong Kong in 2009-2010 (Cowling et al., 2019). The analysis showed that most vaccination effect is mediated by HAI (hemagglutination inhibition) titer in older children, indicating that HAI titer can be a suitable surrogate marker, particularly for older children. The other analysis is on TCGA (The Cancer Genome Atlas project) genomic data. We found that the PIK3CA gene mutation affects the survival time by mRNA expression of the GPC1 gene by a special pattern that the effect decreases the mortality during the early follow-up but then increases the mortality during the late follow-up.

Keywords: Nonparametric Estimation, Time-to-event Outcome, Mediation Analysis

分組議程 SII 時間表/摘要

分組議程 SII : 2023 年 6 月 29 日 (星期四) 15:20 - 16:40

SII-1 **網絡分析** **Chair：郭鋸霖** **理 A310** **p.103**

潘建興 - 中央研究院統計科學研究所

A Mixture Model of Truncated Zeta Distribution with Applications to Scientific Collaboration Networks

李政德 - 國立成功大學統計學系

Graph Machine Learning and Its Applications

郭鋸霖 - 國立高雄大學統計學研究所

網絡資料重抽樣：以管窺天

SII-2 **應用統計新發展** **Chair：沈宗荏** **理 B201** **p.108**

林宗儀 - 國立中興大學應用數學系

A robust factor analysis model based on the canonical fundamental skew-t distribution

袁子倫 - 東海大學應用數學系

短期用戶端電力負載預測

洪宗乾 - 國立屏東科技大學工業管理系

統計在精實現場改善的一些應用

SII-3 **醫療大數據之深度學習與統計分析** **Chair：黃冠華** **理 A212** **p.112**

陳泰賓 - 義守大學醫學影像暨放射科學系

Using deep learning to automatically delineate tumor boundaries in multi-parametric MRI liver images

許巍嚴 - 國立中正大學資訊管理學系暨研究所

Motor Imagery EEG Classification using Deep Learning and Statistical Analysis

黃冠華 - 國立陽明交通大學統計學研究所

利用基於深度學習之物件偵測模型於 X 光影像的常見胸部疾病偵測

SII-4 **統計智慧與精準醫療** **Chair：陳君厚** **理 A109** **p.117**

何之行 - 中央研究院歐美研究所

Big Data, Precision Medicine and Bias: Ethical and Legal Consideration

楊欣洲 - 中央研究院統計科學研究所

AI-Enhanced Integration of Genomic and Medical Imaging Analysis for Early Detection and Risk Assessment of Type 2 Diabetes

蕭自宏 - 台中榮民總醫院醫學研究部

Taiwan Precision Medicine Initiative: The Experience from Taichung Veterans General Hospital

分組議程 SII : 2023 年 6 月 29 日 (星期四) 15:20 - 16:40

SII-5	**統計方法與應用**	**Chair：謝進見**	**理 A209**	**p.121**

周孟穎 - 淡江大學數學系

比較既有分類方法對茶葉種類分群的初果

蔡嘉仁 - 輔仁大學統計資訊學系

On statistical power in linear relationships with heteroscedastic measurement errors in both axes

謝進見 - 國立中正大學數學系

Proportional Hazard Model and Proportional Odds Model under Dependent Truncated Data

SII-6	**非參數和半參數統計方法**	**Chair：黃禮珊**	**理 A211**	**p.125**

江金倉 – 國立臺灣大學應用數學研究所

A General Semi-parametric Elliptical Distribution Model for Semi- supervised Learning

張馨文 - 中央研究院統計科學研究所

A non-smoothing framework for inference on functional means

林嘉韡 - 國立清華大學統計所

Subdata Selection for Big-Data Nonparametric Regression

SII-7	**應用機率**	**Chair：蕭守仁**	**理 A318**	**p.129**

王家禮 - 國立東華大學應用數學系

The Strategic Entering Time of a Commerce Platform

陳美如 - 國立中山大學應用數學系

On the limiting spectral distributions of stochastic block models

洪芷漪 - 國立政治大學應用數學系

A random spread model with frozen types

SII-8	**電信資料的相關研究與應用**	**Chair：王鴻龍**	**理 A314**	**p.133**

賴威宇 - 國立中山大學應用數學系

內政部電信信令人口統計資料建置、分析與應用

柯怡瑄 - 國立臺北大學

運用電信資料與空間統計探討雙北地區居住特徵

王鴻龍 - 國立臺北大學統計學系

象徵性資料分析法於電信信令資料的矩陣視覺化與分群

分組議程 SII : 2023 年 6 月 29 日 (星期四) 15:20 - 16:40

SII-9	**Statistical Methods for Functional Observations**	**Chair：江其衽**	**理 A316**	**p.137**

李百靈 - 淡江大學統計系
Generalized linear model with functional covariate and its derivatives

李國榮 - 國立成功大學統計系
Spatiotemporal Bayesian Varying Coefficient Models for Poisson Areal Data

陳裕庭 - 國立臺灣大學統計與數據科學研究所
Multiple Changepoints Detection in a Functional Data Sequence

SII-10	**衛星資料**	**Chair：黃文瀚**	**理 A210**	**p.141**

劉小菁 - 國家太空中心地面研發處
Application of Statistics and AI methods in Satellite Images

黃成勇 - 國家太空中心
福衛七號之衛星數據科學應用

劉正彥 - 國立中央大學太空科學與工程學系
地震電離層前兆與電離層暴之衛星數據科學應用
Statistical Analyses on the Ionospheric Ion Density and Total Electron Content during the M7.3 Iran-Iraq Border Earthquake and Magnetic Storms Occurring in November 2017

SII-11	**International session (CIPS&KSS)**	**Chair:Tsai-Hung Fan**	**理 A106**	**p.147**

Chun-Shu Chen - Graduate Institute of Statistics, National Central University
A Spatial ZIP-like Model

ShengLi Tzeng - Department of Applied Mathematics, National Sun Yat-sen University
Adaptive Calibrations of Spatially Misaligned IoT Data

Chae Young Lim - Department of Statistics, Seoul National University
Spatio-Temporal Analysis of Dependent Risk with an Application to Cyberattacks Data

SII-12	**統計於產業實務暨政策運用**	**Chair：陳馨蕙**	**理 D241**	**p.151**

林金龍 - 國立東華大學財務金融學系
台灣金融網路風險模型

彭俊能 - 國立台灣科技大學資通安全與教育中心
運用機器學習於行銷策略

陳馨蕙 - 中華經濟研究院
Dynamics of youth's overseas employment during pandemic and geopolitical risks : New insights from Taiwan's evidence

分組議程 SII : 2023 年 6 月 29 日 (星期四) 15:20 - 16:40

SII-13 **Contributed Session - C** **Chair：許湘伶** **理 D131** **p.157**

林登右 - 國立中山大學應用數學系

Exploring Conformational Landscape of Cryo-EM Using Energy-aware Path Finding Algorithm

江彥頡 - 國立臺北大學統計學系

結合 Jackknife 模型平均與 Almon 多項式法之異質噪聲 MIDAS 模型探討

邱士軒 - 國立臺北大學統計學系

二維偏斜常態資料之兩階段管制圖經濟設計

林亮言 - 國立中山大學應用數學系

Portfolio construction based on dynamic networks, centrality and weights

蔡茜婷 - 國立成功大學

誤差平方和最小化條件下的因子分析

曾羿文 - 國立清華大學統計學研究所

Change-point detection EWMA control charts for monitoring Weibull scale parameter

SII-14 **Contributed Session - D** **Chair：施銘杰** **理 A307** **p.164**

褚育誠 - 國立陽明交通大學

Causal mediation analysis in longitudinal studies -
Integrated multi-mediational g-formula with censoring, death truncation and competing

陳芃辰 - 國立陽明交通大學統計學研究所

Application of Hierarchical Classification Model Based on Deep Learning
in COVID-19 and Different Fields of Chest Diseases

許育禎 - 國立中山大學應用數學系

Integrated Analysis of DNA Methylation and its Correlation with
Gene Expression and Clinic in Endometrial Cancer

陳彥霖 - 國立陽明交通大學

Robust Inference for Causal Mediation Analysis of Recurrent Event Data

蘇家瑩 - 中央研究院統計科學研究所

Exploring Gene Expression Variability in Pan-Cancer Analysis:
The Influence of TP53 Mutation on Survival Outcomes via Mediation Analysis

黃意婷 - 中央研究院統計科學研究所

Positive association of ambient ozone exposure and
pancreatic cancer in a European ecological study and a Taiwanese population study

Session II-1

June 29, 15:20 – 16:40

理工一館 A316

◆Organizer/Chair：

郭鋸霖 - 國立高雄大學統計學研究所

◆Speaker:

1. 潘建興 – 中央研究院統計科學研究所
2. 李政德 – 國立成功大學統計學系
3. 郭鋸霖 – 國立高雄大學統計學研究所

A Mixture Model of Truncated Zeta Distribution with Applications to Scientific Collaboration Networks

Hohyun Jung[1], Frederick Kin Hing Phoa[2]

[1]Department of Statistics, Sungshin Women's University, Seoul, South Korea

[2] Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

Abstract

The degree distribution has attracted considerable attention from network scientists in the last few decades to have knowledge of the topological structure of networks. It is widely acknowledged that many real networks have power-law degree distributions. However, the deviation from such a behavior often appears when the range of degrees is small. Even worse, the conventional employment of the continuous power-law distribution usually causes an inaccurate inference as the degree should be discrete-valued. To remedy these obstacles, we propose a finite mixture model of truncated zeta distributions for a broad range of degrees that disobeys a power-law behavior in the range of small degrees while maintaining the scale-free behavior. The maximum likelihood algorithm alongside the model selection method is presented to estimate model parameters and the number of mixture components. The validity of the suggested algorithm is evidenced by Monte Carlo simulations. We apply our method to five disciplines of scientific collaboration networks with remarkable interpretations. The proposed model outperforms the other alternatives in terms of the goodness-of-fit..

Keywords: Degree Distribution, Power Law, Finite Mixture Model, Truncated Zeta Distribution, Scientific Collaboration Networks

Graph Machine Learning and Its Applications

李政德

國立成功大學統計學系暨數據科學研究所

Abstract

Graph structures can be used to describe various relationships and interactions among individuals, with graph data-based machine learning downstream tasks, including essential node classification and link prediction, belonging to the realm of Graph Machine Learning (GML). The advancement of deep learning techniques has enabled the effective learning of node feature representation vectors on graph data, known as Graph Representation Learning (GRL), and even training node feature generation and downstream tasks simultaneously, forming Graph Neural Networks (GNNs). In this keynote speech, we will introduce the most crucial GRL and GNN techniques in the field of graph machine learning and showcase the superior and extensive feature representation learning capabilities of GML through research achievements. We will demonstrate how graph machine learning can be gracefully integrated into various data types that do not have a pre-existing graph structure. This includes tabular data classification, regression, missing value processing, anomaly detection, and natural language processing tasks such as fake news detection and text classification. We will also explore time series prediction tasks and provide examples of how GNN can be applied to recommendation systems, computer vision, and drug development. Our goal is to help the audience understand and attempt to incorporate graph machine learning into their research fields and project tasks. Furthermore, we will examine how graph machine learning can be utilized in different scenarios, such as tabular data, natural language processing, and time series prediction. By providing various examples of GNN applications in recommendation systems, computer vision, and drug development, we hope to inspire the audience to consider the potential of graph machine learning in their own work. If time permits, we will also discuss the development direction of graph machine learning in the context of trustworthy artificial intelligence. By shedding light on the future of GML and its potential role in ensuring the trustworthiness of AI systems, we aim to provide a comprehensive outlook on the evolving landscape of graph-based learning techniques.

Keywords: Graph Machine Learning、Graph Neural Networks、Graph Data Science

網絡資料重抽樣：以管窺天

郭錕霖

國立高雄大學統計學研究所

摘要

本研究僅針對無向且無權重之網絡。一般而言，靜態網絡資料通常只有一筆，即一張圖，假設此觀察網絡來自一個未知的機率空間，我們如何對網絡特徵（如三角形個數）進行統計推論？資料重抽樣應該是一個可行的解方。在本研究中，我們將提出一個網絡資料的重抽樣方法，透過模擬實驗的結果，我們認為此方法是可被應用的。

關鍵詞：網絡特徵、重抽樣、clique

Session II-2

June 29, 15:20 – 16:40

理工一館 B201

◆Organizer/Chair：

沈宗荏 - 國立中興大學應用數學系

◆Speaker:

1. 林宗儀 – 國立中興大學應用數學系
2. 袁子倫 – 東海大學應用數學系
3. 洪宗乾 – 國立屏東科技大學工業管理系

A robust factor analysis model based on the canonical fundamental skew-*t* distribution

Tsung-I Lin [1] (Presenter), Wan-Lun Wang [2] and I-An Chen [1]

[1] Institute of Statistics, National Chung Hsing University
[2] Department of Statistics and Institute of Data Science, National Cheng Kung University

Abstract

The traditional factor analysis rested on the assumption of multivariate normality has been extended by considering the restricted multivariate skew-*t* (rMST) distribution for the unobserved factors and errors jointly. However, the rMST distribution has limited use for characterizing skewness that concentrates in a single direction. This paper is devoted to introducing a more flexible robust factor analysis model based on the broader canonical fundamental skew-*t* (CFUST) distribution, called the CFUSTFA model. The proposed new model can account for more complex features of skewness toward multiple directions. An efficient alternating expectation conditional maximization algorithm fabricated under several reduced complete-data spaces is developed to estimate parameters under the maximum likelihood (ML) perspective. To assess the variability of parameter estimates, we present an information-based approach to approximating the asymptotic covariance matrix of the ML estimators. The effectiveness and applicability of the proposed techniques are demonstrated through the analysis of simulated and real datasets.

Keywords: AECM algorithm; Canonical fundamental skew-t distribution; Factor scores; Truncated multivariate t distribution; Unrestricted multivariate skew-t distribution

短期用戶端電力負載預測

袁子倫[1]、朱基祥[2]、韓意勤[3]

[1]東海大學應用數學系

[2]東海大學統計學系

[3]財團法人資訊工業策進會/數位轉型研究院/能源數據科技中心/能源數據服務組

摘要

本研究擬探討具有半參數基底函數迴歸模型 （semi-parametric regression modeling）應用於不同用戶間電力負載量短期預測之成效。因不同用戶的用電行為不同，導致用戶的用電習慣不同。在開發負載預測模型前，本研究將針對用戶間的用電行為做初步分析，其可了解自身用電習慣，也可減少模型變異誤差。

建立模型時，須考慮解釋變數之重要影響，且過多的解釋變數會導致模型配適耗時較長，並且有模型過度配適(overfitting)結果。所以在本文中，除了檢視用戶間的用電行為外，並考慮重要的解釋變數之間關係。最後，藉由交叉驗證(cross validation）之自適應 Lasso 迴歸方法（adaptive lasso）來篩選重要因子進行預測。評估模型之準則為預測結果和實際負載量的平均絕對比例誤差(Mean Absolute Percentage Error, MAPE)，並比較用戶用電行為是否會受體感溫度影響。

關鍵詞：半參數迴歸模型、週期性函數基底、Lasso 變數選擇、平均絕對比例誤差。

統計在精實現場改善的一些應用

洪宗乾

國立屏東科技大學 工業管理系

摘要

精實管理為豐田汽車生產的製造核心理念，並於近 10 幾年陸續導入台灣某些產業或生產鏈，也都見到卓越的績效。精實生產主要強調製造過程中的價值流動，降低前置時間來滿足客戶需求。若以價值流動為觀察點較能提供具全體運營單位考量的改善策略，但由於不同部門變異性質大，如何掌握一致性、縮小差異及現場執行的可行性，來促進流動，就需要統計的工具及思維。

此篇分享兩個實證案例，分別為大型食材物流業的庫存管理及冒口保溫套的生產，第一個案例除了有效降低庫存也讓管理人員作業更輕鬆；第二個案例則是顯著降低烘乾過程的不良率，同時也提升了產能。

Session II-3

June 29, 15:20 – 16:40

理工一館 A212

◆Organizer/Chair：

黃冠華 - 國立陽明交通大學統計學研究所

◆Speaker:

1. 陳泰賓 – 義守大學醫學影像暨放射科學系
2. 許巍嚴 – 國立中正大學資訊管理學系暨研究所
3. 黃冠華 – 國立陽明交通大學統計學研究所

Using deep learning to automatically delineate tumor boundaries in multi-parametric MRI liver images

Tai-Been Chen[1], Ming-Hung Hsiao[2,3]

[1]Department of Medical Imaging and Radiological Science, I-Shou University

[2]Department of Information Engineering, I-Shou University

[3]Department of Radiology, Chang Gung Memorial Hospital-Kaohsiung Medical Center

Abstract

Motivation: Hepatic tumor segmentation has been a widely studied and challenging field for many years. Accurately identifying liver cancer and assessing its size is crucial for treatment decisions. This study aims to use deep learning network architectures for segmenting Magnetic Resonance Imaging (MRI) liver cancer image models.

Objectives: The main objective of this study is to understand the performance of segmentation models for segmenting liver cancer in different MRI sequences, with the expectation of improving the accuracy of segmentation models applied to MRI liver cancer images.

Methods and Materials: This retrospective study used MRI liver cancer images from Kaohsiung Chang Gung Memorial Hospital, approved by the IRB (No: 202100818B0), collected from multi-sequence MRI abdominal liver cancer images confirmed as liver cancer from January 2015 to April 2021, and used deep learning image segmentation models as the primary research method. All the MRI sequences used in our study included in-out phase, DWI, TE160, T2FS, T2WI, and LAVA pulse sequences. Liver cancer images were extracted and annotated from each patient's images; five Fully Convolutional Network (FCN) models were used, including Xception, InceptionResNetV2, MobileNetV2, ResNet18, and ResNet50, combined with three optimizers (SGDM, ADAM, RMSProp). The performance of liver cancer segmentation results was evaluated using five metrics: Global Accuracy, Mean Accuracy, Intersection over Union (IoU), Weighted IoU, and Mean Boundary F-1 Score (BF).

Results: The Mean Accuracy and Mean IoU scores of MobileNetV2 and ResNet50 were both above 0.52, with ResNet50 slightly outperforming MobileNetV2. Among all FCN models' Mean Accuracy and Mean IoU performance metrics in this study, ResNet50 performed the best with Global Accuracy, Mean Accuracy, IoU, Weighted IoU, and BF of 0.991, 0.741, 0.646, 0.986, and 0.612, respectively. Five Fully FCN models were employed, with InceptionResNetV2 applied to LAVA C+, T2WI, Heavy T2 (TE160), In phase and Out phase; MobileNetV2 applied to LAVA C-; ResNet50 applied to T2FS; and ResNet18 applied to DWI.

Conclusions: Different MRI sequences should be applied to different FCN models. When the appropriate FCN model is applied to the corresponding sequence, higher accuracy can be achieved.

Keywords: Liver cancer, Deep learning, FCN, MRI sequences.

Motor Imagery EEG Classification using Deep Learning and Statistical Analysis

Wei-Yen Hsu[1,2], Ya-Wen Cheng[1], Michelle Liou[3] and Arthur Chih-Hsin Tsai[3]

[1] Department of Information Management, National Chung Cheng University

[2] Center for Innovative Research on Aging Society (CIRAS), National Chung Cheng University

[3] Institute of Statistical Science, Academia Sinica

Abstract

In brain-computer interface (BCI) work, how correctly identifying various features and their corresponding actions from complex Electroencephalography (EEG) signals is a challenging technology. However, most current methods do not consider EEG feature information in spatial, temporal and spectral domains, and the structure of these models cannot effectively extract discriminative features, resulting in limited classification performance. To address this issue, we propose a novel motor-imagery (MI) EEG discrimination method, namely wavelet-based temporal-spectral-attention correlation coefficient (WTS-CC), to simultaneously consider the features and their weighting in spatial, EEG-channel, temporal and spectral domains in this study. The initial Temporal Feature Extraction (iTFE) module extracts the initial important temporal features of MI EEG signals. The Deep EEG-Channel-attention (DEC) module is then proposed to automatically adjust the weight of each EEG channel according to its importance, thereby effectively enhancing more important EEG channels and suppressing less important EEG channels. Next, the Wavelet-based Temporal-Spectral-attention (WTS) module is proposed to obtain more significant discriminative features between different MI tasks by weighting features on two-dimensional time-frequency maps. Finally, a simple discrimination module is used for MI EEG discrimination.

Keywords: Brain-computer interface, Motor-imagery EEG, EEG-channel attention, Temporal-spectral attention, Wavelet transform

利用基於深度學習之物件偵測模型於 X 光影像的常見胸部疾病偵測

洪禕婕、黃冠華*

國立陽明交通大學統計學研究所

摘要

本研究旨在利用深度學習方法，標示出胸部 X 光影像中呈現異常的部位，並註記此異常部位所反映的可能胸部疾病。我們使用義大醫院所提供的 1635 張經由專業醫生所標註疾病標籤及邊界框的胸部 X 光影像，來進行深度學習模型的建立。我們設計四種不同的胸部異常檢測方案。這些方案的差別在於是對所有疾病進行一次性地偵測或者單獨偵測個別疾病，還有是否只針對二元分類預測有病的影像才進行物件偵測。四種方案皆使用三個基於深度學習之 Faster R-CNN、Mask R-CNN 與 RetinaNet 物件偵測模型，預測疾病範圍與位置，以及使用 EfficeintNet-B2 模型進行二元分類。

關鍵詞：胸部 X 光影像、深度學習、物件偵測、二元分類

Session II-4

June 29, 15:20 – 16:40

理工一館 A109 第二講堂

◆Organizer/Chair：

陳君厚 - 中央研究院統計科學研究所

◆Co-organizer

李丕強 - 啟碁科技

◆Speaker:

1. 何之行 – 中央研究院 歐美研究所
2. 楊欣洲 – 中央研究院 統計科學研究所
3. 蕭自宏 – 台中榮民總醫院醫學研究部

Big Data, Precision Medicine and Bias: Ethical and Legal Consideration

Chih-hsing Ho[1]

[1]Associate Research Fellow, Institute of European and American Studies & the Research Center for Information Technology Innovation (joint appointment), Academia Sinica

Abstract

Big data and precision medicine have the potential to revolutionize healthcare by enabling personalized treatment approaches based on individual patient characteristics. However, the utilization of big data in precision medicine also brings concerns regarding the potential presence of biases that can influence clinical decision-making and exacerbate health disparities. This abstract explores the intersection of big data, precision medicine, and bias, highlighting the challenges and implications for equitable healthcare delivery. Big data analytics in precision medicine involve the collection and analysis of vast amounts of diverse patient data, including demographic information, genomic profiles, and clinical outcomes. While this wealth of information holds promise for tailoring treatments to individual patients, it also presents the risk of bias. Biases can emerge at various stages of data collection, preprocessing, analysis, and interpretation, resulting in unequal representation and potentially leading to disparities in healthcare outcomes. In conclusion, big data and precision medicine offer significant opportunities for personalized healthcare. However, the presence of biases within these approaches poses challenges to achieving equitable outcomes. Efforts to improve data representativeness, algorithmic transparency, diversity in research, and clinician education are vital in mitigating bias and promoting equitable access to the benefits of precision medicine. By addressing these challenges, big data and precision medicine can contribute to improved health outcomes for all individuals, irrespective of their background or demographic characteristics.

Keywords: Big Data, Precision Medicine, Bias, Algorithm, Transparency

AI-Enhanced Integration of Genomic and Medical Imaging Analysis for Early Detection and Risk Assessment of Type 2 Diabetes

Yi-Jia Huang[1], Chun-houh Chen[2], Hsin-Chou Yang[1,2*]

[1]Institute of Public Health, National Yang Ming Chiao Tung University

[2]Institute of Statistical Science, Academia Sinica

Abstract

Type 2 Diabetes (T2D) is a global public health concern due to its increasing prevalence. Early detection of T2D plays a vital role in improving individuals' health and enhancing well-being. This study leverages artificial intelligence, specifically eXtreme Gradient Boosting (XGBoost), to develop predictive models for T2D based on genomic and medical imaging data. The models aim to establish a prediction model and identify high-risk subgroups for T2D within a cohort of 68,769 Taiwan Biobank participants. The approach integrates Polygenic Risk Score (PRS) and Multi-image Risk Score (MRS) with demographic factors and environmental exposures to assess T2D risk. The model's performance is evaluated using the Area Under the Receiver Operating Curve (AUC). Results demonstrate that genomic information alone is insufficient for accurate T2D prediction (AUC = 0.73), whereas medical imaging data, including abdominal ultrasonography, vertebral artery ultrasonography, bone mass density, and electrocardiography, significantly improves prediction accuracy (AUC = 0.89). The best-performing model integrates genomic, medical imaging, and demographic variables (AUC = 0.94), successfully identifying subgroups at high risk of developing T2D. The study also presents an online risk assessment website for T2D. In summary, this research represents the first integration of whole-genome and medical imaging data for T2D risk assessment. By utilizing artificial intelligence to analyze genomic, medical imaging, and demographic factors, this study contributes to early detection, precision health, and prevention of T2D.

Keywords: Type 2 Diabetes, Polygenic Risk Score, Multi-image Risk Score, eXtreme Gradient Boosting

Taiwan Precision Medicine Initiative: The Experience from Taichung Veterans General Hospital

蕭自宏

台中榮民總醫院醫學研究部

Abstract

Precision medicine is based on the combination of genomics, medication, environmental exposure, and different lifestyle among individuals, pushing medical diagnosis, treatment and prediction, from traditional cohort study to personalized medicine. Precision medicine will integrate large numbers of clinical practice, medical information, genomics information, and life styles of patients to conduct integrated analysis, and comprehensively explore the impact of individualized differences on medical care, also achieve comprehensive personalized medical treatment; moreover, it can improve health care and enhance treatment development. The "Taiwan Precision Medicine Initiative" project in Taichung Veterans General Hospital aims to collect more than 100,000 genetic data from patients. The program used clinically applicable genetic loci to explore the strategy of clinical disease risk assessment, drug dose, and modified medication, etc. The project also utilizes the genome-wide association analysis to analyze variant sites, combining the participants' clinical data, past medical history, lifestyle, biochemical values in order to find more applications of health care. For example, we identified that the prevalence of genetically defined familial hypercholesterolemia (FH) was 1.13% in the hospital-based population in Taiwan, which was higher than the rate observed in individuals with clinically defined FH. We also found the risk of CAD and AMI was increased to varying degrees in subjects with different FH risk alleles. Through the project, We expected to utilize precision medicine to achieve personalized medicine, increase the quality and welfare of our patients, and also establish a unique Han nationality precision medicine genetic database. By integrating the electronic medical records and clinical information, we will create novel precision medicine research, improve research quality and passion, and promote medical development.

Session II-5

June 29, 15:20 – 16:40

理工一館 A209

◆Organizer/Chair：

謝進見 - 國立中正大學數學系

◆Speaker:

1. 周孟穎 – 淡江大學數學系
2. 蔡嘉仁 – 輔仁大學統計資訊學系
3. 謝進見 – 國立中正大學數學系

比較既有分類方法對茶葉種類分群的初果

周孟穎

淡江大學數學學系

摘要

對資料分類與分群的時候，只要有適當的特徵變數，目前既有的分類方法可以訓練適當的模型並且會有好的分類結果。在資料量足夠的情況，可以藉由類神經網路方法來進行特徵變數的自動擷取和模型訓練。如果分群狀況在平面或空間中不理想時，可以藉由 kernel 函數把資料映射到維度較高的空間後再進行分類模型的訓練。以上既有的方法在資料量充足之下的分類效果通常結過會相當好。然而當資料量與我們想要分類的數目接近的時候，除了需要處裡模型訓練資料量不足的問題，如果要訓練比較好的分類模型，需要重新尋找資料的特徵變數。在這裡我們將應用以上的方法，來處裡來自一筆實驗室茶葉的資料。這筆資料大約有 170 筆，分別記錄了茶葉的產地、種類、風味、包裝、五種離子濃度等變數。茶葉的風味可以由離子濃度來決定，不同的產地也會對茶葉的離子組成有所影響。甚至在同種類的茶葉當中，離子濃度可能稍微有所差異。因此我們將用既有的分類方法，對離子濃度去分類以上的類別變數。比較各個方法從當中找出分類結果較好的原因，或是由資料視覺化的方法去找出可解釋的特徵變數。最後會給予在資料量足夠、可以增加分類準確率、可能適用的其它分類方法等情況，給予實驗室未來增加實驗資料量和變數種類的初步建議。

關鍵詞：茶葉的離子濃度、分類方法、資料視覺化、特徵變數。

On statistical power in linear relationships with heteroscedastic measurement errors in both axes

Jia-Ren Tsai

Department of Statistics and Information Science, Fu Jen Catholic University

Abstract

Sample size is a key factor in statistical estimation. An overly small sample results in low statistical power, but large samples can be difficult to gather. The requisite minimum sample size is affected by numerous factors, including measurement errors, that must be accounted for in its estimation. Such estimation is prone to deviations due to the effects of measurement errors in the observation. In this paper, differences in sample size estimation under different definitions of reliability ratios were investigated in a heterogeneous linear measurement error model. Moreover, we established statistical power functions for (1) a single slope parameter and (2) the joint parameters of the intercept term and parameter term. A simulation study was conducted to evaluate the relationship and influence of the average reliability ratio, sample size, and power function under the variation of heterogeneous measurement errors and a real data example is given to illustrate the application of the model. (This paper jointed with Yi-Jia Huang)

Keywords: Average reliability ratio, Heterogeneous error, Measurement error model, Power of test, Sample size.

Proportional Hazard Model and Proportional Odds Model under Dependent Truncated Data

Jin-Jian Hsieh

Department of Mathematics, National Chung Cheng University

Abstract

Truncation data arise when the interested event time can be observed only if it satisfies a certain condition. Most of the existing approaches analyze this kind of data by assuming the truncated variable is quasi-independent of the interested event time. But, in many situations, the quasi-independence assumption may be not suitable. In this article, the authors consider the copulas to relax the quasi-independence assumption. Additionally, the survival function of the interested event time is estimated by Chaieb et al. (2006). Then, the authors extend the related methods by Hsieh et al. (2008) and Hsieh and Lai (2019) to propose two estimation procedures for the proportional hazard (PH) model and the proportional odds (PO) model, which can be applied to the right truncated data, and the left truncated and right censoring data. Subsequently, the performances of the proposed estimation approaches are assessed via simulation studies. Finally, the proposed methodologies are applied to analyze two real datasets (the retirement center dataset and the transfusion-related AIDS dataset).

Keywords: Censoring, Copula model, Proportional hazard model, Proportional odds model, Dependent truncated data.

Session II-6

June 29, 15:20 – 16:40

理工一館 A211

◆Organizer/Chair：

黃禮珊 - 國立清華大學統計所

◆Speaker:

1. 江金倉 – 國立台灣大學應用數學研究所
2. 張馨文 – 中央研究院統計科學研究所
3. 林嘉韓 – 國立清華大學統計所

A General Semi-parametric Elliptical Distribution Model for Semi-supervised Learning

江金倉

國立臺灣大學數學系

Abstract

This research presents a novel semi-parametric elliptical distribution model with broad applications in semi-supervised learning tasks. We develop a pseudo maximum likelihood method to estimate parameters and classify groups using both labeled and unlabeled data. The proposed estimator achieves local asymptotic maximin-efficiency with a suitable size of unlabeled data, outperforming the estimator that relies solely on labeled data. By employing appropriately chosen bandwidth selectors and initial parameter estimates that ensure numerical stability and convergence, we efficiently maximize the objective function using low-dimensional groupwise pseudo-likelihood functions in a block coordinate descent manner. Additionally, the study comprehensively analyzes the effects of labeled and unlabeled data on the pseudo maximum likelihood estimator and classifier. Simulation studies and empirical data applications illustrate the superiority of our methodology.

A non-smoothing framework for inference on functional means

Hsin-wen Chang[1], Ian W. McKeague[2]

[1]Academia Sinica

[2] Columbia University

[2] City University of Hong Kong

Abstract

This talk introduces a nonparametric inference framework that is applicable to occupation time curves derived from wearable device data. Motivated by the right-continuity of these curves, we develop a non-smoothing approach that involves weaker conditions than existing conditions imposed when using smoothing to estimate functional means under a fixed dense design. Notably, our procedure allows discontinuities in the functional covariances while accommodating discretization of the observed trajectories. Under this non-smoothing framework, we devise an empirical likelihood method to construct confidence bands for the functional means. Our method utilizes the known optimality of empirical likelihood. It also respects range and monotonicity constraints on occupation time curves. A simulation study shows that the proposed procedures outperform competing functional data procedures. We illustrate the proposed methods using wearable device data from an NHANES study.

Keywords: Accelerometry, Bootstrap, Functional data analysis, Nonparametric likelihood ratio, Occupation measures

Subdata Selection for Big-Data Nonparametric Regression

林嘉韓

國立清華大學統計學研究所

摘要

各個領域內資料的數量隨著科技的發展正以爆炸級的成長速度增加，現在的資料量在過去發展的統計模型中經常會面臨計算上的困難。基於這個問題，選取一組有代表性的 subdata 來降低運算使用的資料量便是一種解決方法。

在這篇文章中我們針對的資料為一維的變數 X 與一維的反應變數 Y。我們不對(X, Y)間的 mean function 進行過多限制並使用 local linear regression 進行估計，提出基於殘差平均與標準差抽取一組子資料集的演算法。

過去研究中已有提出在 minimize 與 minimax MISE 這兩種條件下變數 X 分布的 optimal designs，後續稱為 optimal 與 minimax design。

跟傳統實驗設計不同的是我們的方法有運用到 Y 的資訊，接著跟這兩種比較我們的演算法有以下優點:

從理論推導我們發現使用較小的 bandwidth 時會收斂到 optimal design，使用較大 bandwidth 得到的 subdata 所估計出的 mean function 則會對過大的 bias 做修正；從模擬的結果來看，當 mean function 有明顯較其他區域崎嶇時我們的演算法會有更好的估計效果。

關鍵詞：Subdata selection、Nonparametric regression、Local linear

Session II-7

June 29, 15:20 – 16:40

理工一館 A318

◆Organizer/Chair：

蕭守仁 - 國立彰化師範大學數學系

◆Speaker:

1. 王家禮 – 國立東華大學應用數學系
2. 陳美如 – 國立中山大學應用數學系
3. 洪芷漪 – 國立政治大學應用數學系

The Strategic Entering Time of a Commerce Platform

Chia-Li Wang

Department of Applied Mathematics, National Dong Hwa University

Abstract

The surge of service and commerce platforms, such as e-commerce and internet-of-things, have rapidly changed our lives. How to avoid the congestion and get the job done in the platform is now a common problem that many people encounter every day. This requires platform users to make decisions about when to enter the platform.

To that end, we investigate the strategic entering time of a simple platform containing random numbers of buyers and sellers of some item. Upon a trade, the buyer and the seller gain respective profits, yet they pay the cost of waiting in the platform. To maximize their expected payoffs from trading, both buyers and sellers can choose their entering times. This creates an interesting and practical framework of a game that is played among buyers, among sellers, and between them. That is, a strategy employed by a player is not only against players of its type, but also a response to those of the other type, and, thus, a strategy profile is composed of strategies of buyers and sellers.

The players' best response, the Nash equilibrium (NE) strategy profile is derived by a pair of differential equations, which, in turn, are used to establish its existence and uniqueness. More importantly, its structure sheds valuable insights of how the entering strategy of one side (buyers or sellers) is affected by the entering behavior of the other side. These results provide a base for the study of dynamic pricing for stochastic demand-supply imbalances. Finally, comparisons between the social welfares (the sum of the payoffs incurred by individual participants) obtained by the optimal strategy and by the NE strategy are conducted for showing the efficiency loss relative to the socially optimal solution. That should help to manage the platform better

Keywords: Double-Sided Queue, Non-cooperative Game, Nash Equilibrium, Price of Anarchy

On the limiting spectral distributions of stochastic block models

May-Ru Chen[1] and Giap Van Su[1,2]
[1] National Sun Yat-sen University
[2] Thai Nguyen University of Education

Abstract

Erdős-Rényi graph is a random graph in which the probability of a connection between two nodes follows a Bernoulli distribution independently. The stochastic block model (SBM) is an exten-sion of the Erdős-Rényi graph by dividing nodes into K subsets,

known as blocks or communities. Let $\widetilde{A}_N = (\widetilde{A}_{ij}^{(N)})$ be an $N \times N$ normalized adjacency matrix of an SBM with K blocks of any sizes, and let $\mu_{\widetilde{A}_N}$ be the empirical spectral density of $\widetilde{A}_N$.

In this talk, we first showed that if the connecting probabilities between nodes of different blocks are zero, then $\lim_{N\to\infty} \mu_{\widetilde{A}_N} = \mu$ exists almost surely, and we gave the explicit formulas for μ and its Stieltjes transform, respectively. Second, we showed under a suitable condition on the maximum of connecting probability between nodes in different blocks, say by ζ_0, $\mu_{\widetilde{A}_N}$ converges both in probability and expectation as $N \to \infty$, followed by $\zeta_0 \to 0$.

A random spread model with frozen types

Jyy-I Hong
Department of Mathematical Sciences, National Chengchi University

Abstract

When a type has been ``frozen'' (namely, a type of which an individual only produces one individual of the same type), its spread pattern will be changed and this change will affect the long-term behavior of the whole system. However, in a frozen system, the offspring mean matrix is no longer primitive so that the Perron-Frobenius theorem can’t be applied directly when predicting the spread rates. In this talk, we will characterize the offspring mean matrix of the frozen system and analyze the relationship among the growth rate, the spread rate and the eigenvectors associated with the maximal eigenvalues of the offspring mean matrix and its component submatrices.

Session II-8

June 29, 15:20 – 16:40

理工一館 A314

◆Organizer：

黃怡婷 - 國立臺北大學統計學系

◆Chair：

王鴻龍 - 國立臺北大學統計學系

◆Speaker:

1. 賴威宇 – 國立中山大學應用數學系
2. 柯怡瑄 – 國立臺北大學
3. 王鴻龍 – 國立臺北大學統計學系

內政部電信信令人口統計資料建置、分析與應用

賴威宇

內政部統計處

摘要

過去我國常見人口統計有「戶籍統計」與「普查資料」2 種，各具特色與用途。但由於工商經濟發展、交通便利、民眾跨區通勤通學移動愈趨頻繁，各地區之平/假日、日/夜間之活動差異愈來愈大，亟需建立一套能反映日間活動人口的全國性「人流統計」供各界應用。

本資料集係我國首次，以最完整及細緻化的空間統計單元（縣市、鄉鎮市區、村里及最小統計區），結合 3 大電信業者電信信令資訊，建置全國的電信信令人口統計資料。除揭露各地區不同時段人口分布特色，更以多項新創指標（人流密度、外來人口率、外人吸引力倍數、人潮坪效等），綜整出鄉鎮市區及村里發展亮點。各空間統計單元之電信信令人口統計資料，均置於內政部社會經濟資料服務平台（SEGIS，網址：https://segis.moi.gov.tw）提供各界下載參用。

關鍵詞：電信信令、人口統計、內政部

運用電信資料與空間統計探討雙北地區居住特徵

柯怡瑄[1]、嚴潔翎[1]、黃怡婷[1]、王鴻龍[1]、詹士樑[2]

[1]國立臺北大學統計學系

[2]國立臺北大學不動產與城鄉學系

摘要

內政部統計處（2020）運用整合的手機訊號資料推估夜間停留、日間活動人口與日間特定時間旅次，與戶籍資料比較，兩者的人數有縣市別的差異。除人口推估，台灣民眾持有手機的涵蓋率高達 9 成，但手機資料僅有點位與時間手機訊號資料，可以有哪些應用呢？

黃怡婷等人（2022）擷取研究資料的期間路過新北市手機用戶訊號資料，經由資料處理，整理出用戶特徵變數，利用集群分析發現高達 25% 的用戶僅有一個主要活動區域。本研究進一步整合次級資料，探討主要活動區域僅有一個的用戶，其居住地區特徵。

本研究以雙北地區二級經濟發布區為區域劃分依據，結合皇后式與距離方法建立空間權重矩陣，確認主要活動區域數為一人數分布存在空間自相關性，並建構卜瓦松迴歸模型、空間自迴歸模型、條件自相關模型，比較三種模型配適結果，探討影響該區域主要活動區域數為一人數的因子與效果。

研究結果顯示該地區托老中心家數、地區居民年齡組成、超級市場家數等因子會影響該地區人口的移動情形，盼此項研究的發現可以為城市空間規劃和社會經濟發展等方面的研究提供參考，使其更加合理且有效地規劃城市空間，提高交通運輸效率，促進城市經濟和社會的發展。

關鍵詞：皇后空間權重矩陣、卜瓦松迴歸模型、空間自迴歸模型、條件自相關模型、電信資料、雙北居住特徵

象徵性資料分析法於電信信令資料的矩陣視覺化與分群

陳逸瑄[1]，王鴻龍[2]，吳漢銘[*3]

[1, 2] 國立臺北大學統計學系

[3] 國立政治大學統計學系

摘要

電信信令資料記錄行動裝置的使用者於某時間點的地理位置資訊，具有空間分佈及非線性移動軌跡等多重特徵。相較於戶籍資料或旅運調查資料，運用電信信令資料的分析結果，可提供更即時的人流資訊。本研究結合矩陣視覺化之技術與象徵性資料分析方法於電信信令資料，針對行動裝置使用者的移動軌跡進行二階段的分析與探索，以了解人群與人群間的時空交互作用與群聚移動關係。我們首先採用動態時間校正法作為使用者間軌跡距離之度量指標，計算出使用者距離矩陣後，以矩陣視覺化的技術呈現分群的結果，初步找出移動模式相似的人群。為了解決大數據計算上的困難，我們應用區間型象徵性資料分析法，將相似人群之經緯度數值資料摘要成區間型經緯度資料，並以二維色階作為區間型經緯度資料矩陣視覺化的依據，呈現出實際地理位置的遠近關係與人群移動範圍之大小。藉由本研究針對電信信令資料所提出的人群移動視覺化分析程序，相信可以有效地分析處理大規模即時或長期時空資料，協助專家更精確的預測人流，了解人群移動特徵及其關聯性，並有助於各場域的應用。

關鍵字: 資料視覺化、時空資料、軌跡分群、人群移動樣態，

Session II-9

June 29, 15:20 – 16:40

理工一館 A316

◆Organizer/Chair：

江其衽 – 國立臺灣大學統計與數據科學研究所

◆Speaker:

1. 李百靈 – 淡江大學統計系
2. 李國榮 – 國立成功大學統計系
3. 陳裕庭 – 國立臺灣大學統計與數據科學研究所

Generalized linear model with functional covariate and its derivatives

Pai-Ling Li

Department of Statistics, Tamkang University

Abstract

We propose a generalized functional linear regression model by considering a functional covariate and its derivatives as functional predictors. The unobserved derivatives of a random function may carry useful information and need to be estimated. We apply the notion of functional principal component analysis to modeling functional predictors. The proposed regression model is parameterized in various ways to investigate the effect of each functional predictor. The performance of the proposed method is demonstrated through a data example. This is a joint work with Professor Jeng-Min Chiou.

Keywords: Functional data analysis, generalized linear model, principal components.

Spatiotemporal Bayesian Varying Coefficient Models for Poisson Areal Data

李國榮老師

成功大學統計系

Abstract

A Bayesian spatiotemporal generalized linear regression model with varying coefficients is proposed for identifying the existence of spatial pattern in the treatment for Ischaemic stroke, and to understand better the relationship between selected risk factors, such as comorbidity, medication and environmental and social interactions with the rate of mortality to patients of Ischaemic stroke. We develop a Metropolis-withinGibbs algorithm to conduct fully Bayesian inference and adapt a version of the deviance information criterion to choose the optimal number of latent heterogeneous generative mechanisms. When applied to a data set extracted from the National Health Insurance Research Database (NHIRD) of Taiwan, which was collected from a retrospective cohort study across time period 2004 through 2013, a spatial-temporal model allows one to explore the spatial variation in the risk to Ischaemic stroke under the influence of temporal random effects. We discover that the variation in mortality rates of those who suffer from Ischaemic stroke across 349 townships in Taiwan is also attributed to spatial heterogeneity in the coefficients for important factors. Composition of drug use, as well as the comorbid conditions each has and ties to his or her environmental and social factors at various times, are in relation to the mortality rate of the population at risk of Ischaemic stroke in each geographical area. Making such relation clear can serve as a guide to the pre-allocation of limited resources in hope to reduce life-threatening damage possibly caused by the uneven distribution of treatment options for unanticipated needs.

Multiple Changepoints Detection in a Functional Data Sequence

Yu-Ting Chen[1], Jeng-Min Chiou[1], Tzee-Ming Huang[2]

[1] Institute of Statistics and Data Science, National Taiwan University, Taiwan

[2] Department of Statistics, National Cheng Chi University, Taiwan

Abstract

We present a new approach known as greedy segmentation (GS) to identify multiple changepoints for a functional data sequence. The proposed multiple changepoint detection criterion links detectability with the projection onto a suitably chosen subspace and the changepoint locations. With the derived consistency property, the GS estimator can identify the true changepoints efficiently in a greedy manner. Meanwhile, a GS estimator-based test statistic is proposed to help estimate the number of changepoints. These results lead to a novel algorithm capable of identifying both the number and locations of changepoints. The finite sample performance of the proposed algorithm is investigated through intensive simulation studies. As an application, we apply our method to temporal changepoint detection in weather datasets.

Keywords: Covariance operator, Functional principal component, Mean shift

Session II-10

June 29, 15:20 – 16:40

理工一館 A210

◆Organizer/Chair：

黃文瀚 – 國立清華大學統計所

◆Speaker:

1. 劉小菁 – 國家太空中心地面研發處
2. 黃成勇 – 國家太空中心
3. 劉正彥 – 國立中央大學太空科學與工程學系

Application of Statistics and AI methods in Satellite Images

劉小菁

國家太空中心

Abstract

Satellite images are vital for numerous applications such as agriculture, urban planning, and environmental monitoring. Accurate image analysis can provide decision-making information, and the growing demand for high accuracy image information has led to the application of statistical methods and artificial intelligence (AI) techniques in satellite image classification.

Statistical applications, including Image Classification & Segmentation of Landscape, and some critical Index or Measurement generated by Statistic or Probability method, enabling the extraction of valuable information. The Classification methods introduced in this presentation contain Linear Regression with Noise Removal, Gaussian Maximum Likelihood Classifier, and indicator kriging method. FORMOSAT-5 image and some open SAR image (Sentinel-1) are used for demonstrating the image application. We also apply the Change of Taiwan West Coast example for critical Measurement generated by Statistic or Probability method.

AI methods have demonstrated considerable potential in satellite image application, particularly in object recognition and classification. The IPS-TASA team employs AI techniques in applications such as:

(1) The Change of Taiwan West Coast: monitoring coastal changes to assess environmental impacts and identify trends for future planning.
(2) Rice Detection: identifying rice fields and monitoring growth stages to facilitate efficient agricultural practices.
(3) Aircraft or Ship Detection: detecting and tracking aircraft and ships for security, surveillance, and traffic management.
(4) Segmentation of Landscape and Cities: categorizing images into roads, urban and non-urban areas, and trees for urban planning, resource management, and environmental monitoring.

As a result, the integration of Statistics methods and AI techniques has effectively improved satellite image application, leading to enhance image analysis accuracy and valuable information extraction. The continuous development of these methods will contribute significantly to the understanding and management of our planet's resources and environment.

福衛七號之衛星數據科學應用

黃成勇

國家太空中心衛星操控組

摘要

福爾摩沙衛星七號，為臺美雙方大型國際合作案，由六顆衛星組成之觀測星系，執行氣象和太空天氣觀測任務。自 2019 年 6 月 25 日發射以至 2023 年 5 月底，已經提供大氣 700 萬筆，電離層 500 萬筆以上的觀測資料。國家太空中心自主衛星操控系統 (Cross-Platform Satellite Operation Control : XPSOC)，同時利用地表 10 個地面信號接收站，執行 10 萬次以上之衛星通聯，下載福衛七號觀測資料，平均能在 30 分鐘內，提供氣象單位進行氣象預報作業。

在使用福衛七號大氣觀測資料，可普遍減少氣象預報誤差 6%，改進颱風路徑預報誤差 6-10%。比起傳統颱風生成預報，加入福衛七號觀測更能提早 12-72 小時就預報出颱風的生成。在電離層觀測方面，傳統的太空天氣預報多仰賴太空天氣模式，福衛七號提供了大量的電離層觀測資料，促成了利用資料同化預報太空天氣技術的發展，使利用觀測資料預報太空天氣的夢想成真，使台灣成為除美國外，第二個具有同化預報太空天氣的國家。

關鍵詞：福爾摩沙衛星七號、颱風預報、太空天氣

地震電離層前兆與電離層暴之衛星數據科學應用

Statistical Analyses on the Ionospheric Ion Density and Total Electron Content during the M7.3 Iran-Iraq Border Earthquake and Magnetic Storms Occurring in November 2017

J. Y. Liu (劉正彥)[1,2,3]*, Y. I. Chen (陳玉英)[4], F. Y. Chang (張富淵)[1,2],

Y. C. Wen (文允晟)[2], and C. K. Chao (趙吉光)[1,2]

[1]Center for Astronautical Physics and Engineering, National Central University

[2]Department of Space Science and Engineering, National Central University

[3]Center for Space and Remote Sensing Research, National Central University

[4]Graduate Institute of Statistics, National Central University

Abstract

The mission of Advanced Ionospheric Probe (AIP) onboard FORMOSAT-5 (F5) satellite is to detect seismo-ionospheric precursors (SIPs) and observe ionospheric space weather. F5/AIP plasma quantities in the nighttime of 22:30 LT (local time) and the total electron content (TEC) of the global ionosphere map (GIM) are used to study SIPs of an M7.3 earthquake in the Iran-Iraq Border area on 12 November 2017, as well as signatures of two magnetic storms on 7 and 21-22 November 2017. Statistical analyses of the median base and one sample test are employed to find the characteristics of temporal SIPs in GIM TEC over the Iran-Iraq area. The anomalous increases of the GIM TEC and F5/AIP ion density over the epicenter area on 3-4 November (day 9-8 before the M7.3 earthquake) agrees with the temporal SIP characteristics that the significant TEC increase frequently appears on day 14-6 before 53 M≥5.5 earthquakes in the area during 1999-2016. The receiver operating characteristic (ROC) curve is further applied to confirm the statistical significance of the characteristic, which can be used to detect temporal SIPs at the area. The spatial analyses show that the SIPs frequently appear specifically over the epicenter day 9-8 before the M7.3 earthquake and day 10-9 before a M6.1 earthquake on 1 December,

while proponent TEC increases occur at worldwide high-latitudes on the two magnetic storm days. Odds and odds ratios are employed to validate the SIPs being observed and to find possible locations of forthcoming large earthquakes. The box-and-whisker plots and Mann-Whitney U test are used to see if F5/AIP ion quantities are significantly changed during the SIP and storm periods. The F5/AIP ion velocity uncovers that the SIPs of the two earthquakes are caused by eastward seismo-generated electric fields, and the two positive storm signatures are due to the prompt penetration electric fields.

Keywords: seismo-ionospheric precursors, ionospheric storm, ion quantities, total electron content, statistical analyses

Session II-11

June 29, 15:20 – 16:40

College of Science and Engineering I

A106 Lecture Theater III

◆Organizer/Chair：

CIPS

Tsai-Hung Fan - Graduate Institute of Statistics, National Central University

◆Speaker:

1. Chun-Shu Chen - Graduate Institute of Statistics, National Central University

2. ShengLi Tzeng - Department of Applied Mathematics, National Sun Yat-sen University

3. Chae Young Lim - Department of Statistics, Seoul National University

A Spatial ZIP-like Model

Chun-Shu Chen

Graduate Institute of Statistics, National Central University

Abstract

Spatial count responses with an excessive number of zeros and a set of covariates are common. To alleviate deviations from model assumptions, we propose a spatial zero-inflated Poisson-like methodology to model this type of data which only relies on the first two moments of responses. We design an iterative estimation procedure to estimate regression coefficients and variograms under the generalized estimating equation framework. Also, the stabilization of estimators is evaluated via a block jackknife method. After parameter estimation, a criterion based on the mean squared error of the estimated mean structure is proposed to select covariates. Numerical results show the effectiveness of the proposed methodology. This is a joint work with Professor Chung-Wei Shen.

Adaptive Calibrations of Spatially Misaligned IoT Data

ShengLi Tzeng[1], Chi-Wei Lai[2], Hsin-Cheng Huang[3]

[1] Department of Applied Mathematics, National Sun Yat-sen University

[2]Institute of Statistics, National Tsing Hua University

[3]Institute of Statistical Science, Academia Sinica

Abstract

We consider a challenging problem of calibrating geo-referenced data that suffers from spatial misalignment due to the use of multiple instruments to measure the same variables. Some instruments are precise but expensive, while low-cost instruments are less accurate but more widespread. Data fusion techniques are often employed to combine these different sources and extract more information, but spatial misalignment hinders the direct application of usual fusion methods. Before data fusion, we need careful calibration for those untrustworthy observations. Otherwise, without reliable models, more data can introduce bias and noise.

To address this, we propose a strategy to calibrate fine particulate matter ($PM_{2.5}$) data in Taiwan. We have two sources of $PM_{2.5}$ concentration measurements from traditional monitoring stations and low-cost IoT devices called AirBoxes. AirBoxes are unreliable but easily deployed and form a large network. A one-size-fits-all calibration procedure for all AirBoxes does not work well, because the relationship between measurements from AirBoxes and traditional monitoring stations is not homogeneous in space and many outliers exist. We develop a fast, robust method to model the $PM_{2.5}$ processes and use a spatially varying coefficient regression framework to calibrate AirBox measurements. The calibration significantly improves $PM_{2.5}$ prediction performance, reducing the rooted mean-squared prediction error by 37% to 67% compared to those without calibration.

Keywords: Geostatistics; Robust estimation; Spatially varying coefficient model; Heterogeneous variance; Misalignment.

Spatio-Temporal Analysis of Dependent Risk with an Application to Cyberattacks Data

Chae Young Lim

Department of Statistics, Seoul National University

Abstract

We propose a spatial-temporal model that accounts for dependencies among cyberattacks. The proposed model can be considered as a spatial extension of a GARCH model. The estimation is carried out using a Bayesian approach, which is demonstrated to work well in simulation study. A real data analysis is conducted on publicly available honeypot data, for which we create a dependency measure for cyberattack data. Analyzing honeypot data by groups of the attackers based on their attack patterns reveals distinct characteristics in their variability in time.

Session II-12

June 29, 15:20 – 16:40

理工一館 D241

◆Organizer/Chair：

陳馨蕙 – 中華經濟研究院

◆Speaker:

1. 林金龍 – 國立東華大學財務金融學系
2. 彭俊能 – 國立台灣科技大學資通安全與教育中心
3. 陳馨蕙 – 中華經濟研究院

台灣金融網路風險模型

林金龍

國立東華大學財務金融系榮譽教授

摘要

美國於2008年爆發次房貸危機，繼而引發全球金融危機，明確彰顯金融體系緊密的相互聯結關係。如何衡量網路風險及瞭解傳染途徑乃當務之急，它已經吸引經濟、統計學家，金融業者與政府主管機構的高度重視。本文應用網路模型研究台灣的金融網路風險，有兩個明確目標：（一）我們以台灣銀行的資產負債表建構網路模型，進而建立金融網路風險指標；（二）我們模擬當金融體系中的銀行違約或遭受財務困境時，其傳染途徑。

更明確的說，本計畫分別運用最大熵法與最小機率密度法以金融機構的資產負債表，設算金融機構間的兩兩借貸（interbank lending）矩陣，進而計算三種網路風險指標，即（1）衝擊敏感度（impact susceptibility）；（2）衝擊流動性（impact fluidity）；（3）衝擊擴散度（impact diffusion）。除了應用這些指標繪制金融網路圖，我們分別在 DebtRank的線性傳染機制下（即當某一銀行損失1%至25%的自有資本時，它會依比例反應給它的債主，後者的債權也會損失1%至 25%）銀行遭受損失的傳染途徑；及傳統的門檻機制（銀行只有在違約的情形下，才會傳染給它的債主）的假設下，模擬銀行風險的傳染途徑。

用網路模型來分析金融風險最大的優點在於能夠直接的分析個體金融機構與總體金融體系間複雜的傳遞與回饋關係，而不需依賴過份簡化的金融聯結結構，係當前研究總個體金融風險最先進的研究方法。

實證分析台灣24家銀行2019年的財報資料，發現台灣銀行的網路風險非常小，整體銀行的衝擊敏感度，與各銀行的衝擊敏感度與衝擊擴散度都不顯著異於0。

究其原因係源有二。首先，台灣銀行間的借貸規模非常小，很難構成風險傳染管道。另外，台灣近年來銀行資金非常充沛，而銀行貸款核放保守，因而風險不大。唯上述分析僅是基於2019年的財報資料，有必要進一步追溯過去20年的財報資料進一步分析。此工作看似簡單，但財報資料品質卻是一個問題。

另外，若能獲得各銀行持有金融資產的詳細資料則能進一步分析透過共同資產

的網路風險，而後者可能更適合台灣的現況。

關鍵詞： 系統性風險，金融網路， 結算算則，資產負債表，借貸矩陣，重疊的資產組合，最大熵法，最小機率密度法，衝擊敏感度，衝擊流動性，衝擊擴散度，傳染途徑，違約，壓力， 銀行間市場，DebtRank

運用機器學習於行銷策略

彭俊能

財團法人電信技術中心

摘要

目前零售業的產業環境受電商產業興盛，使經營現況受到相當大的衝擊。此外，零售業科技面應用無法趕上人工智慧的發展，亦阻礙智慧零售的發展。本研究透過文獻的研究與訪談，發現其中的關鍵的問題點在於統計模型的建置難度過高，使得一般業者無法操作坊間現有的建模軟體來落實智慧零售的各種場景。據此，本研究協助業者執行智慧零售的概念，讓業者了解大數據與人工智慧對於行銷是有相當大的幫助。鑒於此，本研究將以國內某購物中心的消費者購物資料作為研究的依據，並透過大數據分析來達成智慧零售的目的。

首先，計畫團隊透過文獻與實地訪談發現，業者雖然有使用數位工具做行銷策略的擬定，但是仍只是將圖表數位化，行銷方案仍依據經驗來擬定。本研究則透過統計與資料科學的學理去找出解決方案，並且藉由大數據分析的四個主要功能(分群、分類、關聯、預測) 來協助業者擬定行銷策略，使業者在客群擴散、DM 發送、搭售策略與行銷成本等方案上，皆能有所助益。因此，本研究將專注在大量減少業者在行銷成本支出，但亦可獲得大量人潮與營收為目標。

關鍵詞：購物中心；人工智慧；市場營銷策略；智能零售；數據應用

Dynamics of youth's overseas employment during pandemic and geopolitical risks: New insights from Taiwan's evidence

Shin-Hui Chen[1], Huan-Kai Tseng[2], Thu-Ha Thi An[1]
Ya-Chi Lin[3], Yu-tzung Chang[2], Kuo-Chun Yeh[4], Kai-Wei Liang[1], Wai-Ru Khoo[11]
Division of Taiwan Economy, Chung-Hua Institution for Economic Research
[2] Department of Political Science, National Taiwan University
[3] Department of Economics, Feng Chia University
[4] Graduate Institute of National Development, National Taiwan University

Abstract

Taiwan's brain drain has been a significant concern due to China's strong economic pull, and the situation may have worsened over the past decade. Besides, the pandemic and geopolitical conflicts might have added to this problem in recent years. This paper explores the dynamics of Taiwan's youth overseas employment using the questionnaires conducted at two separate times under different ruling parties in Taiwan, allowing comparison analysis between the two survey results. The 2015 survey focused on reasons for youth working abroad, through which we captured the pull and push factors for this movement. Meanwhile, the survey in 2022 provided updates and extensions to the data, reflecting policy changes under the current DPP government. Further, we consider the impacts of the pandemic, geopolitical risks and global supply chain restructuring in the second survey. Significantly, in the 2022 questionnaire, to better capture the dynamics of Taiwanese migrant labor flows, we move a step further to observe the intentions and motivations of Taiwanese in both working overseas and returning home. Our empirical

results using the 2015 data show no evidence for the current DPP government's critics that highly educated and high-tech workers have been attracted to China. In contrast, personal considerations, including salary, were significant reasons. However, the preliminary statistics from the 2022 data show a substantial shift in the workers' choice and the pull factors from destination areas. Though most of Taiwan's overseas employees experienced working in China, the preference for working in China decreased due to US-China trade conflicts and the pandemic. The intention to enter the U.S. increased. Over one-third of the respondents wish to work abroad, while the highest proportion is among highly educated workers. In the opposite trend of moving back to Taiwan, most workers with bachelor's degrees want to return, while more workers with higher education levels do not. Significantly, the top percentage of workers with no return plan in the long term belongs to those with master's degrees. Salary and personal purposes might be the main pull and push factors for labor outflows in the short term. Long-term goals and the vision of structural changes in global supply chains have become strong pull factors for overseas working willingness. Taiwan's brain drain alerts us that more policy efforts should be made to pull back talent.

Keywords: Brain drain, Geopolitical risks, Youth overseas employment

Session II-13

June 29, 15:20 – 16:40

理工一館 D131

◆Chair：

許湘伶 – 國立高雄大學統計學研究所

◆Speaker:

1. 林登右 – 國立中山大學應用數學系
2. 江彥頡 – 國立臺北大學統計學系
3. 邱士軒 – 國立臺北大學統計學系
4. 林亮言 – 國立中山大學應用數學系
5. 蔡茜婷 – 國立成功大學
6. 曾羿文 – 國立清華大學統計學研究所

Exploring Conformational Landscape of Cryo-EM Using Energy-aware Path Finding Algorithm

Teng-Yu Lin

Department of Applied Mathematics, National Sun Yat-sen University

Abstract

Single particle cryo-electron microscopy (cryo-EM) is a powerful tool for determining the structure of dynamic macromolecular machines at near-atomic resolution. Recently, various reconstruction models have successfully extracted structural heterogeneity from noisy images, potentially allowing for the identification of kinetically preferred sequences of transitions between two conformational states in dynamic processes. Typically, these sequences are addressed using pathfinding algorithms on a two-dimensional energy landscape. However, it is not clear how to construct a high-dimensional energy landscape and search for a kinetically preferred path on it. Therefore, it is a challenge to implement previous algorithms on recent leading models like cryoDRGN, which represent structural heterogeneity using a latent space higher than two dimensions. Here, we introduce a novel method for finding the kinetically preferred path in the high-dimensional latent space. This method searches for the shortest path in a graph that defines the edge weights based on free-energy difference for each point. We demonstrate that the proposed method can reveal the correct transition state in synthetic data with continuous conformational changes and can find the stable pathway in different noise levels.

Keywords: cryo-EM, free-energy landscape, graph traversal, path finding

結合 Jackknife 模型平均與 Almon 多項式法之

異質噪聲 MIDAS 模型探討

江彥頡

國立臺北大學統計學系

摘要

本研究討論混頻結構模型在模型不確定性之下的一些議題探討，模型參考於 Ghysels 等人 (2007) 所提出的混合數據抽樣模型(Mixed DAta Sampling, MIDAS) ，然而我們預設加權函數為未知，故使用 Almon (1965) 提出之 Almon 滯後多項式聚合，以半參數的形式逼近該函數。使用的研究方法為基於普通最小平方法 (OLS) 下的模型平均方法，在此架構下因為本研究有使用異質性模型，考慮以 Hansen (2012) 提出的 Jackknife 模型平均法 (JMA) 在蒙地卡羅模擬中各種配置下與 Mallows 模型平均法 (MMA)互相比較。

本論文主要探討兩個議題：第一，Almon 滯後多項式的參數，也就是降維的程度，希望找到選擇的規律與較良好取值範圍。第二，針對不同情境下，進行各研究方法的比較，並證明 JMA 的出色數值性能。研究結果顯示：(1) Almon 多項式法的參數選擇範圍主要與滯後期衰減模式有關，與噪聲關係較小。(2) JMA 在噪聲波動越劇烈的情況下越優於 MMA。(3) 最複雜模型於模型平均法的權重佔比與樣本數呈正比。

關鍵詞：半母數 MIDAS 模型、Almon 滯後多項式、Jackknife 模型平均法、異質性模型

二維偏斜常態資料之兩階段管制圖經濟設計

邱士軒、蘇南誠

國立臺北大學統計研究所

摘要

本篇論文延伸探討 Su, Chiang, Chen, Tsai and Shyr (2014) 的兩階段管制圖經濟設計，考慮製程變數服從不同型態的偏斜常態分配，其與 Su and Xie (2022) 中考慮的廣義偏斜常態分配的情況有別。我們將改善 Su and Xie (2022) 計算上的困難，且兼顧更廣義的偏斜狀況，以及降低計算成本。在這樣的考量下進行同樣的監控流程，探討其最佳利潤的經濟設計，並與文獻上的情況比較。另外，我們也將與 Su and Yu (2022) 所提出的使用加權變異法以及偏度校正法建構出的管制圖，比較相同架構下的經濟設計。

關鍵詞：經濟設計、馬可夫鏈逼近法、偏斜常態分配、替代變數、兩階段管制圖

Portfolio construction based on dynamic networks, centrality and weights

Author : Liang-Yen Lin Advisor : Prof. Mei-Hui Guo

Department of Applied Mathematics at National Sun Yat-sen University

Abstract

The main focus of this study is to explore the use of financial network models to construct dynamic investment portfolios. We employ two methods, the correlation coefficient and the Gaussian kernel function, to calculate the similarity matrix between stocks. The financial network model was constructed using the Minimum spanning tree and the Planar Maximally Filtered Graph. The network's centrality is calculated using the eigenvectors of the network adjacency matrix. When selecting stocks for investment portfolios, we consider multiple methods, including high and low centrality stock selection, ρ-dependent stock selection, and partial ρ-dependent stock selection. Asset allocation can also be done using different ways, such as the basic equally weight, the tangency portfolio weight based on modern portfolio theory, the Hierarchical Risk Parity, and weight optimization for the Sortino ratio and the upside/ downside risk ratio. Based on the above stock selection and weighting methods, we employ a dynamically updated approach to adjust the investment strategy for each period. We construct suitable network models, investment portfolios, and weight allocation methods for different periods. An empirical analysis was conducted using the stocks in the S&P100, which included adjusted prices of each stock from 2014/01/03 to 2021/12/30. The results indicate that a dynamic portfolio strategy con- structed with the Gaussian kernel function network obtains the best cumulative returns in the test set.

Keywords: Asset allocation, Centrality, Financial network, Minimum spanning tree, Planar Maximally Filtered Graph, Portfolio, Sharpe ratio, Sortino ratio, upside/ downside ratio

誤差平方和最小化條件下的因子分析

蔡茜婷

國立成功大學統計學研究所

指導教授：馬瀰嘉 博士

摘要

在現今大數據的世代中為了因應各種不同的資料型態，新的資料處理方法不斷推陳出新。而統計、電腦科學，以及資訊等各領域專家對於研發新方法的背景與對於其他領域既有方法的理解不完全，因此可能導致引用上的錯誤並使得結果不可靠。在現代研究結果高度流通的環境中，如此的錯誤結果可能在未經驗證的情況下再度被誤用。Sigg & Buhmann (2008) 提出應用 EM 演算法(Expectation – Maximization Algorithm)的非負稀疏主成份分析，能夠將高維度的資料進行快速的降維處理，在網路上被廣泛引用。但作者可能不熟悉統計領域的多變量分析理論，故混淆因子分析模型與主成份分析的觀念導致文獻數學公式的推導上有誤。本研究將對於作者錯誤的數學推導進行修正，並應用多變量統計分析(Johnson & Wichern，2007; Applied Multivariate Statistical Analysis)書中的傳統因子分析方法，提出使平均重建誤差最小化的方法。接著使用修改 Sigg & Buhmann (2008)文獻的方法後，衍伸出四種不同的因子分析組合，並使用模擬數據做比較，評估何種組合下的平均誤差平方和較小。統計模擬中，假設資料來自於多變量常態分配。在固定母體期望值下，設定三組不同維度大小 5、10 與 15，以及設定因子負荷量矩陣為共同因子較明顯與不明顯兩組。來觀察四種組合所得出的平均誤差平方和與平均重建誤差。在實例分析中，我們針對臉部影像的高維度資料進行操作，以了解我們修改過後的方法在高維度資料降維的表現。

關鍵詞：多變量分析、因子分析、主成份分析、資料降維

Change-point detection EWMA control charts for monitoring Weibull scale parameter

曾羿文
國立清華大學統計學研究所

ABSTRACT

Monitoring the lifetime of Weibull distribution is a crucial task in the field of reliability analysis. Traditionally, the in-control parameters are often assumed to be known on Phase II online monitoring of Weibull process. Otherwise, a large amount of Phase I in-control data are required to estimate them. In this article, based on the standardized likelihood ratio statistic, we propose two types of exponentially weighted moving average (EWMA) chart to monitor the Weibull scale parameter (or lifetime) assuming that the shape parameter is fixed and known. The proposed charts can perform well without knowing the in-control scale parameter with only a few Phase I in-control data. Moreover, the proposed charting schemes can estimate the unknown change-point and parameter at the same time when the proposed charts trigger a signal. Simulation results show that the proposed charts outperform the self-starting EWMA chart of Dogu and Noor-ul-Amin (2023) in most of the out-of-control scenarios considered. Furthermore, we can extend our charting methods to Type-II censored data to accommodate the censoring mechanism in lifetime experiments. Finally, we use an example to demonstrate the applicability of the proposed charts.

Keywords：Control chart, likelihood ratio test, change-point detection, EWMA, Weibull lifetime, Type-II censoring

Session II-14

June 29, 15:20 – 16:40

理工一館 A307

◆Chair：

施銘杰 – 國立東華大學應用數學系

◆Speaker:

1. 褚育誠 – 國立陽明交通大學
2. 陳芃辰 – 國立陽明交通大學統計學研究所
3. 許育禎 – 國立中山大學應用數學系
4. 陳彥霖 – 國立陽明交通大學
5. 蘇家瑩 – 中央研究院統計科學研究所
6. 黃意婷 – 中央研究院統計科學研究所

Causal mediation analysis in longitudinal studies – Integrated multi-mediational g-formula with censoring, death truncation and competing

Yu-Cheng Chu[1], Sheng-Hsuan Lin[1]

[1]National Yang Ming Chiao Tung University

Abstract

Longitudinal data is one of the most common data frameworks in public health and biomedical research. Causal mediation analysis is one of the most important topics in causal inference, which can further quantify the detailed causal relationship mechanisms that have been identified. However, there are two methodological limitations and application shortcomings when applying causal mediation analysis to longitudinal data: (1) how to correct for selection bias caused by death and handle incomplete causal effects due to death truncation or competition; (2) how to analyze the mechanism of time-varying multiple mediator factors. This paper proposes a methodological framework and equations that can cover the above issues, including: (1) the multi-mediational g-formula, which can handle time-varying multiple mediator factors; (2) the survival mediation formula, which can accommodate death truncation or competition; and (3) the survival multi-mediational g-formula, which integrates the above two. In addition, we also use the popular separation method to provide corresponding causal mediation analysis methodology and prove that under the suitable assumption, the separation effect and natural effect have the same definition and identification result.

Keyword: Causal Inference, Mediation Analysis, Death Truncation, Death competing, Time-varying structure

Application of Hierarchical Classification Model Based on Deep Learning in COVID-19 and Different Fields of Chest Diseases

Peng-Chen Chen

National Yang Ming Chiao Tung University

Abstract

With the increasing demand for accurate and efficient analysis of medical images, deep learning techniques have shown great potential in medical image analysis. We aim to develop a robust model which is capable of effectively classifying chest X-rays into multiple categories, including COVID-19 and other specific chest diseases. To achieve this, we collect a diverse dataset of chest X-rays, including COVID-19-positive cases and different chest diseases. Preprocessing techniques are applied to improve image quality and extract relevant features. A hierarchical approach is used to address the multi-class classification problem, allowing the model to classify images in different stages. It includes stages for COVID-19 vs. non-COVID-19, normal vs. abnormal radiographs, lung field vs. other field diseases, and classification of inflammatory, obstructive, and neoplastic diseases with fibrosis. In summary, our hierarchical classification model shows strong performance over several stages. In the first stage, we achieved an accuracy of 99.67% and a macro average F1-score of 98.56% for the classification of COVID-19 and non-COVID-19 cases. In the second stage, the model discriminated between normal and abnormal chest X-rays with an accuracy rate of 82.86% and a macro average F1-score of 82.40%. In the third stage, the model classified lung field disease and disease in other areas with an accuracy of 79.58% and a macro average F1-score of 72.56%. Finally, in the last stage, an accuracy of 68.86% was achieved for classifying different types of lung field disease. This result demonstrates the effectiveness and potential of our model in categorizing various chest diseases which has potential applications in healthcare settings, assisting radiologists in COVID-19 diagnosis and screening, as well as identifying other chest abnormalities.

This is the joint work with Prof. Henry Horng-Shing Lu at the National Yang Ming

Chiao Tung University.

Keywords: hierarchical classification, deep learning, COVID-19, chest X-ray, medical image analysis.

Integrated Analysis of DNA Methylation and its Correlation with Gene Expression and Clinic in Endometrial Cancer

許育禎

Yu Chen, Hsu

國立中山大學應用數學系

Abstract

Endometrial cancer (EC) is ranked 6th among the most common gynecologic cancer worldwide in 2020. Many studies indicate that aberrant DNA methylation can regulate gene expression and causes various diseases, especially cancer. Although in the past, most attention was on the promoter hypermethylation of genes that expressions are silenced in cancer (e.g., tumor-suppressor genes), the subsequent studies of genome-wide DNA methylation have highlighted hypomethylation as another signature of cancer cells. Besides, DNA methylation within gene body more frequently appears than in promoter, and it has been observed to have a positive correlation with the expression levels of certain cancer-associated genes.

This study downloads the methylation and expression data from The Cancer Genome Atlas (TCGA) database. Compared with most of the methylation studies in EC, our study innovatively selects two novel probe bias correction methods: Regression on Correlated Probes (RCP) and REgression on Logarithm of Internal Control probes (RELIC) for methylation data preprocessing. After that, R package: Limma and DESeq2 were used for exploring molecular targets such as differentially methylated CpG sites (DMCs) and differentially expressed genes (DEGs) between tumor and normal tissue in EC. According to the trend of methylation mentioned above, both hypermethylated and hypomethylated genes are paired in this talk with up-regulated or down-regulated genes based on the location of CpG sites to calculate the correlations, and genes with a significant correlation are identified.

Cox proportional hazards model was then applied to identified genes to analyze their

relationships with survival. Meanwhile, to examine the joint contribution of selected genes, pathway analysis was performed, and protein-protein interaction (PPI) network was constructed as well. Moreover, the associations of methylation for the identified genes with clinical variables and with the information regarding molecular subtypes seldom revealed in EC methylation study were investigated. In summary, this study provides some biomarkers related to methylation in EC, which can be further analyzed in clinical research.

Keywords: Endometrial cancer, Methylation, Expression, Survival, Clinic

Robust Inference for Causal Mediation Analysis of Recurrent Event Data

Yan-Lin Chen[1], Yan-Hong Chen[2], Pei-Fang Su[3], Huang-Tz Ou[4], and An-Shun Tai[3]

[1] Institute of Statistics, National Yang Ming Chiao Tung University, Hsin-Chu, Taiwan

[2] Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

[3] Department of Statistics, National Cheng Kung University, Tainan, Taiwan.

[4] Department of Pharmacy, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Abstract

Recurrent events, including cardiovascular events, are commonly observed in biomedical studies. Researchers must understand the effects of various treatments on recurrent events and investigate the underlying mediation mechanisms by which treatments may reduce the frequency of recurrent events are crucial. Although causal inference methods for recurrent events have been proposed, they cannot be used to assess mediation. This study proposed a novel methodology of causal mediation analysis that accommodates recurrent outcomes of interest in a given individual. A formal definition of causal estimands (direct and indirect effects) within a counterfactual framework is given, empirical expressions for these effects are identified. To estimate these effects, both parametric and semiparametric methods, including a triply robust estimator, were developed. The proposed methodology was demonstrated in a real-world application. The method was applied to measure the effects of two diabetes drugs on the recurrence of cardiovascular disease and to examine the mediating role of kidney function in this process.

Keywords: Causal inference, inverse probability weighting, mediation analysis, recurrent events, robust inference, triply robust estimation.

Exploring Gene Expression Variability in Pan-Cancer Analysis: The Influence of *TP53* Mutation on Survival Outcomes via Mediation Analysis

Jia-Ying Su* (蘇家瑩)
Bioinformatics Program, TIGP , Academia Sinica
Institute of Biomedical Informatics, National Yang Ming Chiao Tung University

Yen-Tsung Huang (黃彥棕)
Institute of Statistical Science, Academia Sinica

The *TP53* tumor suppressor gene mutation is highly prevalent in various human cancers, leading to alterations in gene expression. Most studies have focused on its effect on average gene expression levels, neglecting the impact on gene expression variability. This study aimed to investigate whether *TP53* mutations contribute to abnormal variance in gene expression among cancer patients. The gene expression and somatic mutations data, and the relevant clinical information were obtained from The Cancer Genome Atlas (TCGA) for 33 cancer types. We conducted F-tests (adjusting for age, gender, race, stage, and treatment) for each cancer to assess the significance of the effect of *TP53* mutation on the variance of average transcriptome-wide gene expression levels in individuals. Our findings revealed that *TP53* mutation significantly influenced gene expression variance (FDR < 0.05) in breast invasive carcinoma (BRCA), pancreatic adenocarcinoma (PAAD), stomach adenocarcinoma (STAD), and brain lower-grade glioma (LGG). Furthermore, a comprehensive analysis identified specific genes significantly affected by *TP53* in terms of variance rather than mean expression, accounting for 11%, 27%, 9%, and 16% of protein-coding genes in BRCA, PAAD, STAD, and LGG, respectively. Mediation analyses demonstrated that *TP53* mutations increased the risk of death in LGG ($p = 0.008$) by altering gene expression variability. These findings shed light on the mechanisms through which *TP53* influences cancer prognosis and warrant validation in external datasets.

Key words: gene expression variability, mediation analysis, pan-cancer analysis

Positive association of ambient ozone exposure and pancreatic cancer in a European ecological study and a Taiwanese population study

Yi-Ting Huang (黃意婷)*, Yen-Tsung Huang (黃彥棕),

Institute of Statistical Science, Academia Sinica; Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

Abstract

Introduction

Pancreatic cancer is one of the leading causes of cancer-related death, but our understanding of the etiology of this devastating disease remains very limited. The aim of this study was to explore the relationship between air pollutants and pancreatic cancer.

Method

An ecological study of European countries and a nationwide nested case-control study using the National Health Insurance Database in Taiwan were conducted. In the case-control study, 6,444 pancreatic cancer cases during 2015-2017 were included, and 12,888 controls were selected by matching seven demographic characteristics. Concentrations of annual, 5-year, and 10-year averages of air pollutants prior to cancer occurrence were estimated for residential locations. Conditional logistic regressions were implemented to estimate odds ratios between air pollutants and pancreatic cancer risk. Population attributable risk (PAR) was calculated to assess the impact of various risk factors. Mediation analyses were conducted to investigate the mechanism of pancreatic carcinogenesis.

Result

Our ecological study revealed a strong positive correlation (0.51; 95% confidence interval [CI]: 0.17, 0.74) between the ozone (O_3) concentration and the incidence of pancreatic cancer in European countries. Similarly, O_3 exposure was significantly

associated with pancreatic cancer risk in the Taiwanese population, with odds ratios of 1.22 (95% CI: 1.15, 1.31), per interquartile range increase and a conspicuous dose-response relationship. Moreover, O_3 exposure had a much higher PAR (12.34%) than pancreatic diseases (1.73%) and diabetes (0.60%). Pancreatic diseases and diabetes were significant mediators between O_3 exposure and pancreatic cancer with proportions of mediation being 1.19% and 0.02%, respectively.

Conclusion

We conclude that O_3 may increase the risk of pancreatic cancer.

Keywords: air pollutants; ecological study; matched case-control study; ozone; pancreatic cancer;

分組議程 SIII 時間表/摘要

分組議程 SIII : 2023 年 6 月 30 日 (星期五) 11:00 - 12:20

SIII-1	**可靠度衰變資料分析**	**Chair：李宜真**	**理 A310**	**p.179**

王義富 - 國立成功大學統計學系

Gamma Process Model for the Voltage Decay Data in Lithium-Ion Batteries

董弘平 - 國立陽明交通大學工業工程與管理學系

Optimizing Two-Variable Gamma Accelerated Degradation Tests with a Semi-Analytical Approach

鄭雅珊 - 國立清華大學統計所

Optimum Test Planning for Heterogeneous Inverse Gaussian Processes

SIII-2	**Detection of implicit patterns in data**	**Chair：曾聖澧**	**理 B201**	**p.183**

黃信誠 - 中央研究院統計科學研究所

Fast Spatial Prediction for Nonstationary Processes with a Divide-and-Conquer Strategy

黃世豪 - 國立中央大學數學系

Inference for the dimension of a regression relationship using pseudo-covariates

鍾思齊 - 國立中山大學應用數學系

A novel framework of 3D orientation estimation for cryo-EM analysis

SIII-3	**實驗設計**	**Chair：陳瑞彬**	**理 A212**	**p.187**

陳瑞彬 - 國立成功大學統計系

Indicator-based Bayesian Variable Selection for Gaussian Process Models in Computer Experiments

蔡風順 - 中央研院統計科學研究所

Row-column designs for symmetric parallel line assays with block size odd.

蔡碧紋 - 國立臺灣師範大學數學系

Design of Order of Addition Experiment

SIII-4	**空間統計與資訊**	**Chair：詹大千**	**理 D241**	**p.191**

唐嘉宏 - 中央研究院人文社會科學研究中心

An improved geographical and temporal weighted regression analysis based on spatiotemporal weights with application to acute hemorrhagic conjunctivitis in China

陳建州 - 輔仁大學統計資訊學系

Spatial Analysis on Supply and Demand of Adult Surgical Masks in Taipei Metropolitan Areas in the Early Phase of the COVID-19 Pandemic

林柏丞 - 國立臺北大學不動產與城鄉環境學系

應用時空梯度演算法分析電信信令人口統計資料之人流移動特徵

分組議程 SIII : 2023 年 6 月 30 日 (星期五) 11:00 - 12:20

SIII-5 **International session (CIPS&KSS)** **Chair：Cathy W. S. Chen** **理 A209** **p.196**

Cathy W. S. Chen - 逢甲大學統計學系

Bayesian Modeling of Spatial Integer-Valued Time Series

Yondai Kim - 首爾大學統計學系

Bayesian learning for deep neural network

Keun Baik Lee - 成均館大學統計學系

Analysis of Long Series of Multivariate Longitudinal Data using Multivariate t-Linear Models

SIII-6 **International session** **Chair:Henry Horng-Shing Lu** **理 A211** **p.200**

Shaw-Hwa Lo - 哥倫比亞大學統計學系

A Novel Approach to Adopt Explainable Artificial Intelligence in X-ray Image Classification

Inchi Hu - 喬治梅森大學

From I-score, a model-free feature selection statistic, to explainable AI

Jacky Chung-Hao Wu - 國立陽明交通大學統計研究所

Test-Fairness Deep Learning with Influence Score

SIII-7 **工業統計** **Chair：蘇南誠** **理 A318** **p.204**

許湘伶 - 國立高雄大學統計學研究所

Multi-output Gaussian Process based on Neural Kernel Learning and its Prediction Applications

李名鏞 - 靜宜大學資料科學暨大數據分析與應用學系

ON THE OPTIMALITY OF STEP-STRESS ACCELERATED DEGRADATION TESTS FOR WIENER DEGRADATION PROCESSES WITH TIME-CENSORING

黃偉恆 - 逢甲大學統計學系

Multiple Exponentially Weighted Moving Average Control Schemes for Monitoring And Diagnostics of Correlated Quality Variables of Different Types with Individual Observations

SIII-8 **財務計量** **Chair：黃士峰** **理 A314** **p.209**

邱海唐 - 國立中正大學數學系

Model selection for high-dimensional linear regression with highly correlated variables

陳亭甫 - 國立中央大學數學系

狀態相關財務模型之參數估計

翁新傑 - 國立臺北大學統計系

Mixed Frequency Analysis: Consistent and Asymptotic Normality of Semiparametric MIDAS Regression Model

分組議程 SIII : 2023 年 6 月 30 日 (星期五) 11:00 - 12:20

SIII-9	**象徵性資料分析與應用**	**Chair：吳漢銘**	**理 A316**	**p.213**

吳漢銘 - 國立政治大學統計學系

Exploratory Symbolic Data Analysis

高君豪 - 淡江大學統計學系

時序地理資料基於排序法之地圖上色機制與矩陣視覺化分析

林良靖 - 國立成功大學統計學系

Monitoring Photochemical Pollutants based on Symbolic Interval-Valued Data Analysis

SIII-10	**應用統計**	**Chair：鍾冬川**	**理 A210**	**p.217**

陳秀雯 - 中原大學工業與系統工程學系

Properties of Doubly Truncated Mean Past to Failure Function and its Application

劉峰旗 - 逢甲大學統計學系

Integer-valued transfer function models for counts that show zero inflation

郭珈妤 - 靜宜大學資科系

TRAVELING WAVES FOR A THREE-SPECIES COMPETITION SYSTEM WITH TWO WEAK ABORIGINAL COMPETITORS

分組議程 SIII : 2023 年 6 月 30 日 (星期五) 11:00 - 12:20

SIII-11 Contributed Session - E **Chair：鄭順林** **理 A106** **p.221**

鄭順林 - 國立成功大學統計學系暨數據科學研究所

汰役電池篩選機器學習機制開發

吉怡宗 - 國立高雄大學統計學研究所

基於深度學習框架的剩餘使用壽命預測

林楷崙 - 國立臺北大學統計學系

以 GPU 加速小樣本多變量常態分配檢定之工具

吳紫豪 - 輔仁大學統計資訊學系應用統計所

基於模擬方法概化結合 K-Means 與分散/聚集式快速過濾大量資料建構預測排名之應用

劉家銘 - 國立臺灣大學統計碩士學位學程

A novel GNN-based Network Intrusion Detection using Ensemble Framework and E-SageGAT Architecture

SIII-12 Contributed Session - F **Chair：林聖軒** **理 A109** **p.227**

柯鋅瑀 - 國立中山大學應用數學系

透過藥物的分子作用機制篩選高惡化漿液型卵巢癌的治療藥物

張韶恩 - 國立成功大學統計學研究所

對 OTU 資料分類和對病患分群之研究

陳昱豪 - 國立成功大學統計所

結合連續型馬可夫鏈建構混合復發模型

謝綸豪 - 國立臺北大學統計學系

探討空間統計擷取惡性黑色素瘤影像特徵的適切性

高興翰 - 國立臺北大學統計學系

運用統計方法擷取二頭肌周圍積水超音波影像特徵

SIII-13 Contributed Session - G **Chair：鄭宗琳** **理 A307** **p.233**

賴暐婷 - 國立中央大學統計研究所

A Modified VAR-GARCH Model for Asynchronous Multivariate Financial Time Series via Variational Bayesian Inference

古裕彥 - 國立臺灣大學會計學系

The Impact of Inflation on Housing Prices in Taiwan: Evidences from Quantile Regression Method

Giap Van Su - 國立中山大學應用數學系

Spectral properties of Laplacian matrix of Stochastic block models

Session III-1

June 30, 11:00 – 12:20

理工一館 A310

◆Organizer/Chair :

李宜真 – 國立成功大學統計學系

◆Speaker:

1. 王義富 – 國立成功大學統計學系
2. 董弘平 – 國立陽明交通大學工業工程與管理學系
3. 鄭雅珊 – 國立清華大學統計所

Gamma Process Model for the Voltage Decay Data in Lithium-Ion Batteries

Yi-Fu Wang

Department of Statistics, National Cheng Kung University

Abstract

Currently, the rechargeable lithium-ion battery has been widely used in various electronic products and equipment. Therefore, it is very important to predict the end of performance (EOP) for lithium-ion battery. In the battery discharging experiments, the voltage decay data is easy to obtain and the discharge voltage curves are stable, so our purpose is to establish a statistical model based on discharge voltage decay data to improve the accurate of EOP prediction. In this article, motivated by NASA battery voltage decay data, we propose a gamma process model incorporated with the biexponential model to provide the EOP prediction. Finally, the case study and simulation study indicate that an accurate EOP prediction of lithium-ion battery can be obtained by employing the proposed methodology. (This is joint work with En Lin)

Keywords: Gamma process; Lithium-ion battery; End of performance; Voltage.

Optimizing Two-Variable Gamma Accelerated Degradation Tests with a Semi-Analytical Approach

董弘平
國立陽明交通大學 工業工程與管理學系

Abstract

Gamma accelerated degradation tests are widely used to assess timely lifetime information of highly reliable products when the degradation path of quality characteristic of products follows a monotonic process. In this talk, a semi-analytical approach is proposed to determine the optimal designs for two-variable gamma accelerated degradation tests under three criteria: D-optimality, A-optimality and V-optimality. We first use general equivalence theorem to prove that the optimal approximate designs only allocate test units at the four vertices of a rectangular design region, and the corresponding optimal proportion of total number of measurements at each stress level is derived. Next, we introduce a cost constraint function to further determine optimal integer designs. More specifically, a numerical approach is used to resolve the number of test units and number of measurements at each stress level.

Optimum Test Planning for Heterogeneous Inverse Gaussian Processes

Chien-Yu Peng[1], Hideki Nagatsuka[2], Ya-Shan Cheng[3]

[1] Institute of Statistical Science, Academia Sinica

[2] Department of Data Science for Business Innovation, Chuo University

[3] Institute of Statistics, National Tsing Hua University

Abstract

The heterogeneous inverse Gaussian (IG) process is one of the most popular and most considered degradation models for highly reliable products. One difficulty with heterogeneous IG processes is the lack of analytic expressions for the Fisher information matrix (FIM). Thus, it is a challenge to find an optimum test plan using any information-based criteria with decision variables such as the termination time, the number of measurements and sample size. In this article, the FIM of an IG process with random slopes can be derived explicitly in an algebraic expression to reduce uncertainty caused by the numerical approximation. The *D*- and *V*-optimum test plans with/without a cost constraint can be obtained by using a profile optimum plan. Sensitivity analysis is studied to elucidate how optimum planning is influenced by the experimental costs and planning values of the model parameters. The theoretical results are illustrated by numerical simulation and case studies.

Keywords: Conjugate distribution, Destructive degradation, Orthogonality, Random effects, Repeated measures

Session III-2

June 30, 11:00 – 12:20

理工一館 B201

◆Organizer/Chair :

曾聖澧 – 國立中山大學應用數學系

◆Speaker:

1. 黃信誠 – 中央研究院統計科學研究所
2. 黃世豪 – 國立中央大學數學系
3. 鍾思齊 – 國立中山大學應用數學系

Fast Spatial Prediction for Nonstationary Processes with a Divide-and-Conquer Strategy

Yung-Huei Chiou[1], Chun-Shu Chen[1], Hsin-Cheng Huang[2]

[1] Institute of Statistics, National Central University

[2] Institute of Statistical Science, Academia Sinica

Abstract

Spatial data over a large domain generally shows nonstationary features. However, appropriately specifying a nonstationary covariance function is challenging, and the computation of the corresponding inverse matrix in kriging is intractable when the dataset is massive. In this research, we utilize a methodology based on a linear combination of stationary processes with spatially varying weights for modeling nonstationary processes. Instead of considering piecewise stationary processes, our spatial model allows the spatial covariance function to change smoothly or sharply across regions governed by a tuning parameter. Importantly, it reduces to a global stationary process when all stationary components share a common spatial covariance structure. A maximum composite likelihood method is developed for efficient estimation. A novel doubled kriging method is designed for fast spatial prediction based on a divide-and-conquer strategy with almost no loss of efficiency. The proposed method is expected to be highly flexible and computationally efficient. Numerical results show that it is effective in nonstationary spatial covariance function estimation and spatial prediction.

Keywords: Composite likelihood; cross-covariance; kriging; multivariate Matern covariance function.

Inference for the dimension of a regression relationship using pseudo-covariates

Shih-Hao Huang

Department of Mathematics, National Central University

Abstract

In data analysis using dimension reduction methods, the main goal is to summarize how the response is related to the covariates through a few linear combinations. One key issue is to determine the minimal number of such linear combinations, which is the dimension of the sufficient dimension reduction (SDR) subspace. In this work, we propose an easily-applied approach to conduct inference for the dimension of the SDR subspace, based on augmentation of the covariate set with pseudo-covariates. Applying the partitioning principal to the possible dimensions, we use sequential testing to select the dimensionality, by comparing the strength of the signal arising from the actual covariates to that appearing to arise from the pseudo-covariates. Under a *uniform direction* condition, our test statistic asymptotically follows a beta distribution, and the family-wise type I error rate of our sequential testing is rigorously controlled. Simulation studies and an analysis of newborn anthropometric data demonstrate the robustness of the proposed approach, and indicate that the power is comparable to or greater than the alternatives. (Work done jointly with Kerby Shedden and Hsin-wen Chang).

Keywords: inference for dimension, nonparametric regression, sequential testing, sufficient dimension reduction, variable augmentation

A novel framework of 3D orientation estimation for cryo-EM analysis

鍾思齊

國立中山大學應用數學系

Abstract

Cryo-electron microscopy (cryo-EM) shows great potential for determining protein 3D structures. However, the current workflow is slowed down by the complicated process of estimating 3D orientations from 2D projection images, especially as contaminated or low-quality images are challenging to identify in the dataset. Recent deep learning-based approaches aim to accelerate the process by employing amortized inference, eliminating the need for parameter estimation for each image. Despite this, these methods overlook outlier presence and network generalization capabilities.

In this presentation, I will introduce a preliminary framework designed to recover orientations directly from the acquired 2D projections in an end-to-end manner. Our approach uses a 10-dimensional feature vector to represent an orientation, followed by a Quadratically-Constrained Quadratic Program to obtain the orientation prediction in unit quaternion format with uncertainty statistics. Furthermore, contrastive learning is employed to improve our model's generalization ability by incorporating pairwise distances between images. Lastly, I will discuss ongoing research directions, such as integrating domain knowledge into the framework and applying the method to real cryo-EM datasets.

Keywords: Cryogenic electron microscopy, amortized inference, pose estimation, contrastive learning

Session III-3

June 30, 11:00 – 12:20

理工一館 A212

◆Organizer/Chair :

陳瑞彬 – 國立成功大學統計系

◆Speaker:

1. 陳瑞彬 – 國立成功大學統計系
2. 蔡風順 – 中央研究院統計科學研究所
3. 蔡碧紋 – 國立臺灣師範大學數學系

Indicator-based Bayesian Variable Selection for Gaussian Process Models in Computer Experiments

Ray-Bing Chen

Department of Statistics & Institute of Data Science, National Cheng Kung University

Abstract

Gaussian process (GP) models are commonly used in the analysis of computer experiments. Variable selection in GP models is of significant scientific interest but existing solutions remain unsatisfactory. For each variable in a GP model, there are two potential effects with different implications: one is on the mean function, and the other is on the covariance function. However, most of the existing research on variable selection for GP models has focused only on one of the effects. To tackle this problem, we propose an indicator-based Bayesian variable selection procedure to take into account the effects from both the mean and covariance functions. A variable is defined to be inactive if both effects are not significant, and an indicator is used to represent the variable being active or not. For active variables, the proposed method adopts different prior assumptions to capture the two effects. The performance of the proposed method is evaluated by both simulations and real applications in computer experiments.

Row-column designs for symmetric parallel line assays with block size odd.

Feng-Shun Chai

The Institute of Statistical Science, Academia Sinica

Abstract

The row-column designs for bioassays which can estimate three specific contrasts of primary interest with full efficiency are called L-designs. The necessary condition for the existence of the row-column L-design are the block sizes even. For odd block sizes, we provide here a class of highly efficient designs, called nearly row-column L-designs. The magic rectangles and nearly magic rectangles play an important role in constructing the nearly row-column L-designs. One challenge is to show the connectedness of the constructed nearly row-column L-designs.

Design of Order of Addition Experiment

Pi-Wen Tsai
National Taiwan Normal University

Abstact

The purpose of order-of-addition experiments is to identify the best order of adding *m* different components in a system. It is often unaffordable to test the all possible *m*! orders, and the design problem arises to choose a subset of orders for comparison. The popular way of analysis order-of-addition experiments is based on pair-wise ordering (PWO) factors and efficient fractional PWO designs are available. In this talk, we propose an alternative model considering nearest neighbor effects and the construction of new class of optimal order-or-addition experiments is studied.

Session III-4

June 30, 11:00 – 12:20

理工一館 D241

◆Organizer/Chair :

詹大千 – 中央研究院人文社會科學研究中心

◆Speaker:

1. 唐嘉宏 – 中央研究院人文社會科學研究中心
2. 陳建州 – 輔仁大學統計資訊學系
3. 林柏丞 – 國立臺北大學不動產與城鄉環境學系

An improved geographical and temporal weighted regression analysis based on spatiotemporal weights with application to acute hemorrhagic conjunctivitis in China

Jia-Hong Tang[1], Ta-Chien Chan[1]

[1] Research Center for Humanities and Social Sciences, Academia Sinica

Abstract

Most epidemic models wish to quantify the impact of this space-time heterogeneity on epidemiologic dynamics. Geographical and Temporal Weighted Regression (GTWR) has emerged as an effective extension of Geographically Weighted Regression (GWR) for modeling nonstationary spatiotemporal processes such as those that commonly arise from epidemiological study. In GTWR, any spatiotemporal nonstationary in the relationship of interest is considered through a local estimation of model coefficients via a spatiotemporal weight matrix. The core issue of GTWR is a mechanism for weighting the effects of both temporal and spatial variation. The spatiotemporal weights used in GTWR are defined as multiplicative combination of temporal and spatial kernels, such as Gaussian kernels or bi-square kernels, which may not be plausible in some situations such as the transmissibility of acute infectious diseases. The multiplication formulation makes the spatiotemporal kernel function decay faster so that spatiotemporal weights may be underestimated. To address this issue, we propose an improved GTWR equipped with a new spatiotemporal kernel function to account for both spatial and temporal relationships in spatiotemporal analysis of acute infectious diseases. A case study of acute hemorrhagic conjunctivitis cases in China from 2010 to 2016 was used to assess the model performance. Results show that the proposed spatiotemporal kernel function significantly improves the quality of fit and accuracy.

Keywords: GTWR, Epidemiology, Spatiotemporal weights, Kernel function.

Spatial Analysis on Supply and Demand of Adult Surgical Masks in Taipei Metropolitan Areas in the Early Phase of the COVID-19 Pandemic

Chien-Chou Chen[1], Guo-Jun Lo[1], Ta-Chien Chan[2,3]

[1] Department of Statistics and Information Science, Fu Jen Catholic University[2]

Center for Humanities and Social Sciences, Academia Sinica

[3] Institute of Public Health, School of Medicine, National Yang Ming Chiao Tung University

Abstract

This study aimed to assess the gap between the supply and demand of adult surgical masks under limited resources. Owing to the implementation of the real-name mask rationing system, the historical inventory data of aggregated mask consumption in a pharmacy during the early period of the COVID-19 outbreak (April and May 2020) in Taiwan were analyzed for supply-side analysis. We applied the Voronoi diagram and areal interpolation methods to delineate the average supply of customer counts from a pharmacy to a village (administrative level). On the other hand, the expected number of demand counts was estimated from the population data. The relative risk (RR) of supply, which is the average number of adults served per day divided by the expected number in a village, was modeled under a Bayesian hierarchical framework, including Poisson, negative binomial, Poisson spatial, and negative binomial spatial models. We observed that the number of pharmacies in a village is associated with an increasing supply, whereas the median annual per capita income of the village has an inverse relationship. Regarding land use percentages, percentages of the residential and the mixed areas in a village are negatively associated, while the school area percentage is positively associated with the supply in the Poisson spatial model. The corresponding uncertainty measurement: villages where the probability exceeds the risk of undersupply, that is, Pr (RR < 1), were also identified. The findings of the study may help health authorities to evaluate the spatial allocation of anti-epidemic resources, such as masks and rapid test kits, in small areas while identifying priority

areas with the suspicion of undersupply in the beginning stages of outbreaks.

Keywords: Bayesian hierarchical modeling, Voronoi diagram, small area estimation, supply and demand, surgical mask

應用時空梯度演算法分析電信信令人口統計資料之人流移動特徵

林柏丞[1]，詹大千[2]

[1]國立臺北大學不動產與城鄉環境學系

[2]中央研究院人文社會科學研究中心地理資訊科學研究專題中心

摘要

人群移動的探索是疫情控制、都市規劃、交通運輸管理以及災害防救管理的重要參考依據。在資料隱私保護下，行動網路數據利用規則時空立方體單元進行資料的加總以進行資料去識別化，人群的移動特徵因而隱藏在其中，本研究應用三維時空梯度演算法，分析時空加總的電信信令人口統計資料，萃取其中之人流移動特徵並探索城市的脈動。國外的研究常常使用的是個體的軌跡數據，雖然可以很精準的看到人流的變化，但是此類資料的機敏性太高，難以釋出給一般研究人員使用，因此，本研究試圖使用 10 分鐘間隔、250 公尺空間解析度的電信服務網格加總數據進行空間視覺化的分析，主要是透過三維時空梯度演算法的開發與模擬實驗，確認使用此種演算法可以在兼顧隱私的考量下，直覺地呈現網格人流資料的流向與流速，並可與其他的空間數據進行結合，另藉由人流屬性資料的分組，例如年齡或性別分組，可以對不同行為人口進行分層分析。期望透過此方法的導入，未來研究人員或政府單位可以對都市規劃、交通規劃與社會經濟的發展有更動態且精細的了解。

關鍵詞：電信信令資料、人流、時空梯度

Session III-5

June 30, 11:00 – 12:20

理工一館 A209

◆Organizer/Chair :

Cathy W. S. Chen – 逢甲大學統計學系

◆Speaker:

1. Cathy W. S. Chen – 逢甲大學統計學系
2. Yondai Kim – 首爾大學統計學系
3. Keun Baik Lee – 成均館大學統計學系

Bayesian Modeling of Spatial Integer-Valued Time Series

Cathy W.S. Chen(陳婉淑)[1], Chun-Shu Chen(陳春樹)[2], & Mo-Hua Hsiung(熊墨樺)[1]

[1]Department of Statistics, Feng Chia University

[2]Graduate Institute of Statistics, National Central University

Abstract

Many diseases spread through contact transmission, from person to person, either by direct or indirect contact. Understanding these transmission patterns is crucial for effective disease control and prevention. To this end, incorporating spatial-temporal patterns into the multivariate integer-valued GARCH models with generalized Poisson distribution or zero-inflated generalized Poisson is proposed. In this setup, the effect of each neighboring location on the target series is examined, eliminating the need to pre-assign a spatial weight matrix. The generalized Poisson distribution emphasizes over-dispersion cases, while the zero-inflated generalized Poisson distribution further accounts for large proportions of zero counts. Newly designed models are utilized to investigate time series of counts in epidemiology, making inferences, forecasting, and model selection within a Bayesian framework via Markov chain Monte Carlo (MCMC) algorithms. As an illustration, design simulation studies and multivariate weekly dengue cases are examined for the performance of the Bayesian methods. The results prove that the proposed models are successfully fitted in terms of predictions.

Keywords: Markov chain Monte Carlo method, Time series of counts, Generalized Poisson distribution, Spatial INGARCH model, Dengue, Euclidean distance.

Bayesian learning for deep neural network

Yongdai Kim

Department of Statistics, Seoul National University

Abstract

Bayesian approaches for learning deep neural networks (BNN) have been received much attention and successfully applied to various applications. Particularly, BNNs have the merit of having better generalization ability as well as better uncertainty quantification. For the success of BNN, search an appropriate architecture of the neural networks is an important task, and various algorithms to find good sparse neural networks have been proposed. In this paper, we propose a new node-sparse BNN model which has good theoretical properties and is computationally feasible.

We prove that the posterior concentration rate to the true model is near minimax optimal and adaptive to the smoothness of the true model. In particular the adaptiveness is the first of its kind for node-sparse BNNs. In addition, we develop a novel MCMC algorithm which makes the Bayesian inference of the node-sparse BNN model feasible in practice.

Analysis of Long Series of Multivariate Longitudinal Data using Multivariate t-Linear Models

Keunbaik Lee

Department of Statistics, Sungkyunkwan University

Abstract

To analyze long series of multivariate longitudinal data, multivariate linear models typically assume a multivariate normal error distribution with a covariance matrix of an autoregressive moving-average (ARMA) structure. When there are outliers or when the data have heavy tails in their distribution, the assumption of multivariate normality could be too strong in practice. Estimates of mean parameters may be sensitive in this situation. To solve this problem, we propose multivariate t linear models (MTLM) with an ARMA structured covariance matrix. On the other hand, modeling the covariance matrix for multivariate longitudinal data is difficult because the covariance matrix is high dimensional and must be positive-definite. We use modified ARMA Cholesky decomposition and hypersphere decomposition to overcome these problems. The parameter estimation for the models is more flexible than MTLM with an AR structured covariance matrix. Several simulation studies make clear the performance of the proposed models, and data from non-alcoholic fatty liver disease are also provided to demonstrate the methodology.

Session III-6

June 30, 11:00 – 12:20

理工一館 A211

◆Organizer/Chair :

Henry Horng – Shing Lu – 國立陽明交通大學統計研究所

◆Speaker:

1. Shaw-Hwa Lo – 哥倫比亞大學統計學系
2. Inchi Hu – 喬治梅森大學
3. Jacky Chung-Hao Wu – 國立陽明交通大學統計研究所

A Novel Approach to Adopt Explainable Artificial Intelligence in X-ray Image Classification

Shaw-Hwa Lo

Department of Statistics, Columbia University

Abstract

Robust “Blackbox” algorithms such as Convolutional Neural Networks (CNNs) are known for making high prediction performance. However, the ability to explain and interpret these algorithms still require innovation in the understanding of influential and, more importantly, explainable features that directly or indirectly impact the performance of predictivity. In view of the above needs, this study proposes an interaction- based methodology – Influence Score (I-score) – to screen out the noisy and non-informative variables in the images hence it nourishes an environment with explainable and interpretable features that are directly associated to feature predictivity. We apply the proposed method on a real-world application in Pneumonia Chest X-ray Image data set and produced state- of-the-art results. We demonstrate how to apply the proposed approach for more general big data problems by improving the explain ability and interpretability without sacrificing the prediction performance. The contribution of this paper opens a novel angle that moves the community closer to the future pipelines of XAI problems.

From I-score, a model-free feature selection statistic, to explainable AI

Inchi Hu

George Mason University

Abstract

In this talk, we study some further properties of I-score, introduced in Chernoff, Lo and Zheng (2009), and explore their implications to explainable AI. The I-score properties studied include its relationship with AUC, predictivity of a subset of variables, and why I-score can increase prediction accuracy without sacrificing explainability etc. Examples are highlighted that incorporating I-Score in convolutional neural network (CNN) and recurrent neural network (RNN) can improve both prediction accuracy and explainability. We compare I-score with other existing model-free variable selection methods. If time allows, we will also discuss some properties of a backward dropping algorithm - a greedy algorithm to identify a subset of variables that achieves maximum I-score.

Test-Fairness Deep Learning with Influence Score

Jacky Chung-Hao Wu

Institute of Statistics, National Yang Ming Chiao Tung University

Abstract

The prevalence of artificial intelligence has led to many significant problems that required new perspectives. Among them, the hidden bias in AI systems is a serious problem that should be carefully addressed. The source of bias usually comes from data collection focusing on heterogeneous and imbalanced groups. The data with bias will often generate the model with inhomogeneous discrimination that has diverse prediction performances for different groups. Therefore, it is crucial to eliminate the phenomenon of inhomogeneous discrimination and improve the fairness. To mitigate the inhomogeneous discrimination, we propose a new feature selection method based on deep learning and statistics to enhance the model fairness and preserve the prediction performance simultaneously. Integrating with deep learning models, we adopt the influence score (I-score), which develops the statistical methodology that can detect the interaction patterns between multiple features in the proposed method. The features related to the bias information will be detected and excluded in the fair model. We call this method the fair I-score method. The fair I-score method will explore the features unassociated with the discriminatory factors so that the resulting prediction performance is homogeneous for distinct groups. We conduct the empirical study on skin lesion datasets and show that the fair I-score method can produce a model that can correctly classify the types of skin lesions by eliminating the bias information inherent in the diverse groups. This report is based on the joint work with Professor Shaw-Hwa Lo in Columbia University, Professor Inchi Hu in George Mason University, Professor Henry Horng-Shing Lu and the related members in National Yang Ming Chiao Tung University.

Keywords: deep learning; fairness; influence score.

Session III-7

June 30, 11:00 – 12:20

理工一館 A318

◆Organizer/Chair :

蘇南誠 – 國立臺北大學統計系

◆Speaker:

1. 許湘伶 – 國立高雄大學統計學研究所
2. 李名鏞 – 靜宜大學資料科學暨大數據分析與應用學系
3. 黃偉恆 – 逢甲大學統計學系

Multi-output Gaussian Process based on Neural Kernel Learning and its Prediction Applications

Hsiang-Ling Hsu(許湘伶)* and Fu-Hsuan Zhang(張馥瑋)

Institute of Statistics, National University of Kaohsiung

Abstract

The flexibility and uncertainty measures inherent in predictions make the Gaussian process regression models widely applicable to various fields. The basic concepts of multivariate normal distribution, kernels, non-parametric models and conditional probability are essential to build a Gaussian process. However, Gaussian process regression is based on the similarity of the sample function space and independently builds each predictive model but ignores the correlation between time points. In this work, we utilize a multi-output Gaussian process to capture the correlation between time points and may simultaneously preserve the characteristics of the traditional Gaussian process. On the other hand, the prediction results of the Gaussian process are influenced by the structure assumptions of covariance function, so we introduce neural kernel network to automatically acquire the combination of kernel functions instead of manually setting specific kernel functions. In brief, this work proposes a multi-output Gaussian process regression model that attempts to combine different kernel functions to acquire the complicated correlation patterns between data for predictions. We analyze the electric load data of Kyushu, Japan and the apparent temperature of Lingya, Kaohsiung City based on the proposed method with two kinds of evaluation indexes, the root mean square error (RMSE) and mean absolute percentage error (MAPE), to measure the prediction performances. Compared with persistence, seasonal autoregressive integrated moving average (SARIMA) and traditional Gaussian process regression (GPR), the experimental results show that the proposed method possesses low MAPE and RMSE in forecasting the electricity load and apparent temperature in the year of 2020, which displays a good trade-off and flexibility between accuracy and interpretation.

Keywords: Kernel function, Model selection, Multi-output Gaussian process, Neural network, Time series.

ON THE OPTIMALITY OF STEP-STRESS ACCELERATED DEGRADATION TESTS FOR WIENER DEGRADATION PROCESSES WITH TIME-CENSORING

Ming-Yung Lee

Department of Data Science and Big Data Analytics, Providence University

ABSTRACT

To design an optimal degradation test for reliability, existing researches normally pre-assume the form of the test plan (e.g., parallel-stress, step-stress, or ramp test plan) and then determine the values of the plan design parameters to optimize certain objective functions, without considering whether the plan so obtained remains optimal within a broader class of possible test plans. In this paper we prove that, for any given test plan (optimal or not), there exists an equivalent or nearly equivalent step-stress plan having the same objective function value and satisfying or nearly satisfying certain constraints. Our result provides a formal justification for considering only easily-implemented step-stress degradation tests. In this paper we assume the product's degradation follows a Wiener process with time-varying parameters, and intermediate degradation data at certain pre-fixed times during the test are available. Failed test units are removed and the test is time-censored. Various statistical objective functions, such as the estimated asymptotic variances of the maximum likelihood estimators of Mean-Time-To-Failure and *p*th quantile of the lifetime distribution, are considered with a budget constraint. Cost-Performance Index (CPI) is used to evaluate the efficiency of acceleration and an ADT is considered to be a better experiment if its CPI is higher. A similar result is given for the "dual" problem, where the above-mentioned statistical objective function and the budget constraint are switched. In addition, we show that, under some mild conditions, the proposed step-stress test plans can reduce to simpler (e.g., 2-step) test plans. Finally, we use light emitting diode (LED) testing as an example to illustrate our results.

Keywords: Optimal degradation test, step-stress function, lifetime percentiles, asymptotic variance, expected cost, Cost-Performance Index (CPI)

Multiple Exponentially Weighted Moving Average Control Schemes for Monitoring and Diagnostics of Correlated Quality Variables of Different Types with Individual Observations

Wei-Heng Huang(黃偉恆)

Department of Statistics, Feng Chia University (逢甲大學統計系)

Abstract

As more data, both in quantity and variety, are being collected for quality control applications in the era of big data, new challenges inevitably arise, especially for multivariate statistical process control (MSPC). One such challenge, which has received very little attention, is multivariate control chart (MCC) for monitoring correlated quality variables of different types. A recent study by Huang et al. (2022), which develops a Shewhart type MCC, is the first attempt at tackling such a challenge. The proposed chart of Huang et al. (2022), which utilizes the step-down multiple testing procedure of Holm (1979), not only is capable of monitoring correlated variables of different types, but also able to provide instantaneous diagnostics of which parameters are out of control when the chart signals. In this study, we adapt and extend the methodology of Huang et al. (2022) to develop an exponentially weighted moving average (EWMA) chart specifically for the case when the sample size is one. The proposed chart is shown to be effective in detecting parameter changes as well as diagnosing which parameters are out of control when the chart signals.

Keywords: Diagnostics, Exponentially weighted moving average, Holm's step-down multiple testing procedure, Multivariate control chart, Phase-II monitoring.

Session III-8

June 30, 11:00 – 12:20

理工一館 A314

◆Organizer/Chair :

黃士峰 – 國立高雄大學統計學研究所

◆Speaker:

1. 邱海唐 – 國立中正大學數學系
2. 陳亭甫 – 國立中央大學數學系
3. 翁新傑 – 國立臺北大學統計系

Model selection for high-dimensional linear regression with highly correlated variables

邱海唐*、陳亮潔

國立中正大學數學系

Abstact

This work aims at selecting relevant variables in high-dimensional linear regression models with highly correlated variables. To achieve this goal, we propose an approach relying on the orthogonal greedy algorithm, the high-dimensional information criterion, and the group selection concept. Simulation studies are conducted to evaluate the finite sample performances of the proposed approach.

Keywords: Group selection, High-dimensional information criterion, Orthogonal greedyalgorithm

狀態相關財務模型之參數估計

陳亭甫

國立中央大學數學系

摘要

本研究嘗試在貝氏估計架構下，結合 copula 函數與粒子濾波法來估計狀態相關的狀態空間模型參數。粒子濾波法的應用相當廣泛，在金融領域中可用來刻劃市場的狀態變數，並進一步搭配市場價格動態過程，估計財務模型的參數。市場價格模型可由多種不同的隨機過程組成，而隨著模型考量的風險因子越多，狀態變數的維度、模型的參數個數也隨之增加。狀態變數之間的相關性結構會使參數之間的相關係數各數大幅增加，因此本研究結合 copula 函數與 particle filter 方法，使用 copula 函數來描述變數之間的相關結構，發展多維度且具有狀態相關之下的 copula-particle filter 結構，再透過最大概似期望演算法(Expectation-Maximization algorithm)來估計模型參數。除了透過隨機模擬資料驗證 copula-particle filter 的穩健性，也能透過市場資料實際應用此估計方法，藉由篩選後的粒子觀察隱藏變數所代表的市場風險特性。

關鍵詞：Copula、粒子濾波法、EM 演算法

Mixed Frequency Analysis: Consistent and Asymptotic Normality of Semiparametric MIDAS Regression Model

Wen-Jen Tsay[1], Hsin-Chieh Wong[2]

[1]Institute of Economics, Academia Sinica

[2] Department of Statistics & Fintech and Green Finance Center, National Taipei University

Abstract

Thanks to big data innovation, macroeconomic forecasting with regard to daily decision-making, which aims at estimating data sampled at different frequencies, has been widely appreciated. An innovative mixed frequency analysis, based on data sampled at different frequencies, called MIxed DAta Sampling, or MIDAS, regression models, was proposed previously to construct a new time series for matching the length of the lower frequency dependent variable. However, the advantages of regression models with mixed sampling frequencies are relatively far from clear. We derive the asymptotic properties of an ordinary least squares (OLS)-based method for conducting mixed data-sampling analysis, interestingly leading to the consistent property and asymptotic normality of the long run impact. To do so, we use a Vandermonde matrix to approximate the unknown weighting function (or unweighted scheme) for the MIDAS regression model, enabling us to easily compute and achieve dimension reduction in practice.

Keywords: Semiparametric regression, MIDAS model, Vandermonde matrix, consistent, asymptotic normality.

Session III-9

June 30, 11:00 – 12:20

理工一館 A316

◆Organizer/Chair :

吳漢銘 – 國立政治大學統計學系

◆Speaker:

1. 吳漢銘 – 國立政治大學統計學系
2. 高君豪 – 淡江大學統計學系
3. 林良靖 – 國立成功大學統計學系

Exploratory Symbolic Data Analysis

Bo-Syue Jiang[1], Po-Wei Chen[2], Han-Ming Wu[2]

[1]Department of Statistics, National Taipei University

[2]Department of Statistics, National Chengchi University

Abstract

Exploratory Data Analysis (EDA) serves as a preliminary yet essential tool for summarizing the main characteristics of a dataset before appropriate statistical modeling can be applied. Quite often, EDA employs traditional graphical techniques such as boxplots, histograms, and scatterplots, and is equipped with various dimension reduction methods and computer-aided interactive functionalities. Over the years, data collected has become increasingly large and complex. Data descriptions have moved beyond single-value representations, encompassing intervals, histograms, and distributions. These are examples of the so-called symbolic data. In response to this development, we have created two R packages: **dataSDA** and **ggESDA**. The **dataSDA** package is designed to collect a diverse range of symbolic data and offers a comprehensive set of functions that facilitate the conversion of traditional data into the symbolic data format. These datasets can serve as benchmarks for evaluating symbolic data analysis methods. In addition, the package implements various R functions for computing symbolic descriptive statistics. The **ggESDA** package extends ggplot2 to offer a variety of plots specifically designed for exploratory symbolic data analysis. In this talk, we will discuss how **ggESDA** is implemented. We will demonstrate its utility through the analysis of two real symbolic datasets found in **dataSDA**.

Keywords: Exploratory data analysis, Interval-valued data, R packages, Statistical graphics

時序地理資料基於排序法之地圖上色機制與矩陣視覺化分析

高君豪

淡江大學統計學系

摘要

隨著資料時代的來臨，利用 IoT 監測設備不僅會產生大量的時間序列監測資料，也因為設備放置的位置，資料同時包含地理資訊。因此，時序地理資料(Geospatial-Temporal Data)的計算技術和統計分析方法成為目前重要的研究趨勢。然而探索式資料分析 (Exploratory Data Analysis, EDA)，為現今資料科學經常使用之地理相關資料呈現方式。當資料結合了地理資訊和時序資料，EDA 中的地理資料視覺化方法便需要能在地圖上呈現完整時間序列資料的資訊，而不只是單一時間點或綜整後的單一數據。然而，利用各時間點的資料分別繪出地圖影像，或者使用動畫或影片播放所有地圖影像，對資料科學家於探索其資料樣態而言並不容易。本研究針對時序地理資料之地圖視覺化提出一新的呈現方式。利用大家熟知的排序(Seriation)演算法，結合 HSV 的色彩空間，開發出新的地圖上色模式。透過單一顏色呈現各個地理樣本於時間變數中之相關性，樣本顏色的相似程度呈現時間維度資料之綜觀表徵。以此方法除了傳統的連續與類別資料外，更可以解決象徵性時序資料於地圖視覺化的困難。另外也透過矩陣視覺化(Matrix Visualization)的技術呈現原始資料矩陣的微觀表徵。我們也重新設計了針對時序地理資料的矩陣視覺化上色方式。讓地圖與矩陣圖使用相同上色機制，使用者可以在最直觀的情況下，透過綜觀與微觀的角度深入分析與探索時序地理資料。

關鍵詞：地圖視覺化、矩陣視覺化、時序地理資料、象徵性資料分析、探索式資料分析、排序

Monitoring Photochemical Pollutants based on Symbolic Interval-Valued Data Analysis

Liang-Ching Lin*[1], Meihui Guo[2], Sangyeol Lee[3]

[1] Department of Statistics and Institute of Data Science, National Cheng Kung University.

[2]Department of Applied Mathematics, National Sun Yat-sen University

[3]Department of Statistics, Seoul National University

Abstract

This study considers monitoring photochemical pollutants for anomaly detection based on symbolic interval-valued data analysis. For this task, we construct control charts based on the principal component scores of symbolic interval-valued data. Herein, the symbolic interval-valued data are assumed to follow a normal distribution, and an approximate expectation formula of order statistics from the normal distribution is used in the univariate case to estimate the mean and variance via the method of moments. In addition, we consider the bivariate case wherein we use the maximum likelihood estimator calculated from the likelihood function derived under a bivariate copula. We also establish the procedures for the statistical control chart based on the univariate and bivariate interval-valued variables, and the procedures are potentially extendable to higher dimensional cases. Monte Carlo simulations and real data analysis using photochemical pollutants confirm the validity of the proposed method. The results particularly show the superiority over the conventional method that uses the averages to identify the date on which the abnormal maximum occurred.

Keywords: Control chart, Monitoring photochemical pollutants, Symbolic principal component analysis, Symbolic data analysis

Session III-10

June 30, 11:00 – 12:20

理工一館 A210

◆Organizer/Chair :

鍾冬川 – 逢甲大學統計學系

◆Speaker:

1. 陳秀雯 – 中原大學工業與系統工程學系
2. 劉峰旗 – 逢甲大學統計學系
3. 郭珈妤 – 靜宜大學資料科學暨大數據分析與應用學系

Properties of Doubly Truncated Mean Past to Failure Function and its Application

Hsiu-Wen Chen[1], Dong Shang Chang[2]

[1]Department of Industrial and Systems Engineering, Chung Yuan Christian University

[2] Department of Business Administration, National Central University

Abstract

The properties of the doubly truncated mean past to failure function for continuous distributions have not been discovered. We show that the function goes to 0 as time increases toward the upper bound of lifetime and its failure rate function diverges to infinity over the same range. The decreasing property of the function is further investigated and its failure rate function is increasing with time over the same range. The width of both confidence bands converges at nearly the end of life. These properties give reasonable measures of product reliability for real life applications. We present one numerical example of the lognormal distribution to illustrate the usefulness of these functions.

Keywords: Mean residual life, Failure rate, Doubly truncated random variable

Integer-valued transfer function models for counts that show zero inflation

Feng-Chi Liu (劉峰旗)

Department of Statistics, Feng Chia University

Abstract

This study introduces a class of transfer function models that accommodates exogenous variables with a delay parameter, which is in the response of the system to delay. We propose zero-inflated integer-valued transfer function models with generalized Poisson or negative binomial distribution that can describe overdispersion, a large proportion of zeroes, and relation with exogenous variables. We employ Bayesian adaptive Markov chain Monte Carlo (MCMC) methods for inference and model selection. We conduct a simulation study to evaluate the effectiveness of the proposed methods and apply them to real examples. (This is a joint work with Cathy W.S. Chen and Aljo Clair Pingal)

Keywords: Integer-valued time series; Markov chain Monte Carlo method; Generalized Poisson; Negative binomial; Zero counts.

TRAVELING WAVES FOR A THREE-SPECIES COMPETITION SYSTEM WITH TWO WEAK ABORIGINAL COMPETITORS

郭珈妤

靜宜大學 資料科學暨大數據分析與應用學系

Abstract

In this paper, we study the traveling wave solutions for a three species competition system with two weak aboriginal competitors and one strong alien competitor. We are concerned with the existence of traveling waves such that these two co-existence aboriginal competitors are wiped out by the invading alien strong competitor. First, we derive the existence of wave profiles based on an application of Schauder's fixed point theorem with the help of constructing suitable generalized upper-lower solutions to capture the unstable wave tail limit. Then a new method for deriving the stable wave tail limit is introduced. Finally, the minimal invading speed is characterized.

Session III-11

June 30, 11:00 – 12:20

理工一館 A106

◆Chair:：

鄭順林 – 國立成功大學統計學系暨數據科學研究所

◆Speaker:

1. 鄭順林 – 國立成功大學統計學系暨數據科學研究所
2. 吉怡宗 – 國立高雄大學統計學研究所
3. 林楷崙 – 國立臺北大學統計學系
4. 吳紫豪 – 輔仁大學統計資訊學系應用統計所
5. 劉家銘 – 國立臺灣大學統計碩士學位學程

汰役電池篩選機器學習機制開發

鄭順林*、黃子靜、林俊源、徐淳姿、曾以諾
國立成功大學統計系暨數據科學研究所
黃建中
工業技術研究院綠能與環境研究所

摘要

電動汽車的數量急速成長，使得汰役電池將大量產生。車用動力電池一般使用約 5~6 年，電池健康狀態就會衰退到 80%以下，也就是動力電池在多次循環使用後，電池的最大容量只剩新品時的 80%，一般認為此時的電池狀態不再適合應用於動力載具上，必須從載具上卸除汰換，此時汰換下來動力電池可以稱為「汰役電池」。從電動汽車退役後，鋰電池通常含有 70%~80% 的額定容量。若能經過有效的分組後，這些汰役電池仍然可以用於能量或功率要求相對較低的應用，例如住宅和商業建築的儲能系統以及低效能的電動汽車。然而，在這種再次使用之前，最重要的問題是如何有效地將大量的汰役電池分類，篩選出可用電池再次服役。本研究計畫以汰役動力電池轉用儲能應用為主軸，根據短和長時間測試數據建立篩選的指標，汰役電池模組需經過篩選利用才能確保性能與安全性。本研究導入機器學習中決策樹分類模型，同時考量開路電壓、放電容量、內部阻抗這三種電池性能，並量測電池膨脹量作為新增參數，以及放電循環的量測參數，建立基於多參數分類之汰役電池篩選機器學習機制。

關鍵詞：汰役電池、決策樹、分類篩選、機器學習

基於深度學習框架的剩餘使用壽命預測

吉怡宗
國立高雄大學統計學研究所

摘要

系統的剩餘使用壽命(Remaining Useful Life, RUL)估計是工業應用中的主要預測活動之一。在文獻中提到 RUL 預測的研究方向大致有兩種。第一種是使用深度學習的模型估計系統的剩餘使用壽命；另一種是文獻所提出的研究方法，即為結合深度學習工具和基於健康指標(Health Index, HI)相似性的曲線匹配技術(簡稱 HI 曲線匹配法)，估計系統的剩餘使用壽命。因此我們根據這兩種研究的方向提出三個方法估計 RUL，第一個為建立卷積長短期記憶網路(Convolutional LSTM Network, ConvLSTM)的模型估計 RUL。第二個為修改的 HI 曲線匹配法，主要導入 CNN 建構 HI 及變更相似性的計算方式。最後，將這兩種方法的個別預測結果進行簡單平均以預測 RUL。在 NASA C-MAPSS 飛機渦輪引擎數據集的應用結果表明，所提出的三種方案在準確度(Accuracy, AC)、分數(Score)和均方根誤差(Root Mean Square Error, RMSE)等三種預測評估指標函數的表現上都比原始文獻的方法擁有更佳的預測性能，且以簡單平均後的平均預測表現為最優。

關鍵字：卷積長短期記憶網路、相似性計算、深度學習、剩餘使用壽命

以 GPU 加速小樣本多變量常態分配檢定之工具

林楷崙 [1]

國立臺北大學統計學系碩士班

[1]kevin940087@gmail.com

摘要

許多統計模型會假設某些變數服從多變量常態分配，使用模型之前必須要檢查該假設是否成立。在現實中，一般多變量常態檢定操作是基於檢定統計量在大樣本下會趨近某一種分配，以該理論上的分配計算出的 p-value 或 critical value 判斷是否拒絕虛無假設。但是若於小樣本或統計檢定量沒有漸進分布之時，需改以模擬出的 critical value 判斷是否要接受多變量常態分配之假設。

本文以蒙地卡羅模擬多變量常態檢定的檢定統計量的經驗分配，其中檢定包含：Mardia 偏態與峰態檢定、Henze and Zirkler 檢定與 *Wmin(5)*檢定，取得設定顯著水準下的 critical value。比較以 GPU 或以 CPU 計算的所需消耗的時間差距，得出 GPU 在多數的樣本數與分配維度下較 CPU 計算來的有效率。並以 Python 製作多變量常態檢定的應用程式作為以 GPU 加速與檢定的互動介面。

關鍵詞：多變量常態分配，蒙地卡羅模擬，Python，CuPy，GUI

基於模擬方法概化結合K-Means與分散/聚集式快速過濾大量資料建構預測排名之應用

杜逸寧

輔仁大學統計資訊學系

吳紫豪

輔仁大學統計資訊學系應用統計研究所

摘要

杜逸寧(2019)針對資料挑選過程提出 K-Means 結合 Scatter and Gather 形成的大量資料快速篩選方式。方法為將資料初始分群後接著數次挑選，再將挑選後的資料合併後再次進行分群與挑選，執行多次 Scatter 與多次 Gather 後產生結果。本研究試圖將其方法概化，假設有一個篩選資料並且排名的過程，初始有 a 個品項與 b 個使用者，每個使用者皆對所有品項有著喜好排序。在各品項中會含有 c 個關於該品項內容的資料，資料含有名目尺度與連續型尺度是否會影響 K-means 與 Scatter and Gather 分群方式。首先，本研究使用斯皮爾曼等級相關係數用以檢視本研究之預測排名是否與資料之排名有相同的趨勢，實驗過後斯皮爾曼等級相關係數落在 0.3917，表示在資料筆數多的情況下，本研究依然能僅用幾次的迭代次數，得到中度相關的標準。第二，在尺度選擇上，實驗後發現相關係數不會有顯著差異，得到無論是何種尺度的資料都不會影響結果的結論。最後，本研究使用平均絕對百分比誤差計算樣本估計母體的情況，發現樣本為母體的十分之一即可接近母體實驗結果，平均絕對百分比誤差為 0.02。綜上所述，本研究藉由設計不同情境的模擬過程，使方法能概化在不同資料中。

關鍵詞：K-Means 分群法、分散/聚集式、斯皮爾曼等級相關係數

A novel GNN-based Network Intrusion Detection using Ensemble Framework and E-SageGAT Architecture

Chia-Ming Liu, Chen-An Tsai

Master Program in Statistics, National Taiwan University

Abstract

This work aims to addresses the challenging tasks for robust Network Intrusion Detection Systems (NIDS) capable of tackling increasingly sophisticated and complex cyberattacks. Although the conventional machine learning methods have been widely applied to cyberattack detections and show promising results, little attention has been given to the inherent topological structure of network flow data. Therefore, we present a novel ensemble framework for Graph Neural Networks (GNN) and a revolutionary model architecture, E-SageGAT, which applies self-attention mechanisms at the edge level of graphs. These innovative contributions empower the exploration of the structural information of networks and traffic correlations, thereby enhancing the accuracy of detecting malicious activities. When applied to three benchmark datasets, our approach outperforms existing methods (E-GraphSAGE/E-ResGAT/XGBoost) with a 1-2% improvement in weighted F1-score, while the ensemble framework further improves performance by 2-3%. By retaining the original topological structure and transforming network packet flow tabular data into multi-directed graphs, our proposed methods accurately reveal the latent network flow structures and set the stage for significant progress in the field of network intrusion detection.

Keywords: Graph Neural Networks, Ensemble learning, Network Intrusion Detection, Cyber Security

Session III-12

June 30, 11:00 – 12:20

理工一館 A109

◆Chair：

林聖軒 – 國立陽明交通大學統計學研究所

◆Speaker:

1. 柯鋅瑀 – 國立中山大學應用數學系
2. 張韶恩 – 國立成功大學統計學系
3. 陳昱豪 – 國立成功大學統計學系
4. 謝綸豪 – 國立臺北大學統計學系
5. 高興翰 – 國立臺北大學統計學系

透過藥物的分子作用機制篩選高惡化漿液型卵巢癌的治療藥物

柯鋅瑀

國立中山大學應用數學系

摘要

卵巢癌是女性生殖系統中常見的惡性腫瘤之一，由於早期症狀不明顯，許多患者在發現時已進展到晚期，使得治療更具挑戰性。目前治療方式主要為使用鉑金類藥物進行化療，但許多患者仍會出現抗藥性或復發，因此找到合適的藥物是現今重要的課題。

本篇研究以卵巢癌中最常見的亞型——高惡化漿液型卵巢癌(High-Grade Serous Ovarian Cancer)為主要研究對象。先前的研究透過 HDMAC (High-Dimensional Analysis of Molecular Alterations in Cancer) 流程及臨床數據，篩選出影響病人存活風險且與抗藥性相關的基因：ASAP3 及 NEBL。本篇研究以 ASAP3 及 NEBL 作為目標基因探討其與藥物的關係。然而，由於 ASAP3 及 NEBL 的基因效應分數不夠低，以這兩個基因作為標的基因進行篩藥對組織與細胞的影響不夠明顯。因此本研究透過分析基因表現量、基因效應等數據，找到 CDCA5、ANKRD65 這兩個基因分別作為 ASAP3 及 NEBL 的伴隨基因，並一同進行藥物篩選，期望藉此找到適合的藥物來治療高惡化漿液型卵巢癌患者。

本研究接續使用兩個線性混合效應模型，分別探討藥物效應和基因效應以及藥物效應和基因表達量之間的關係，以進一步篩選出更合適的藥物。未來，還需要進行臨床實驗，以確保所篩選出的藥物可以真正幫助到患者，並且不會對患者造成不良的副作用或是風險。

關鍵詞：高惡化漿液型卵巢癌、基因表現量、基因效應、藥物效應、分子作用機制、篩選藥物

對 OTU 資料分類和對病患分群之研究

張韶恩、馬瀰嘉
國立成功大學統計學研究所

摘要

分類操作單元 (Operational Taxonomic Unit; OTU)的作用是對群體進行分類。通過對生物的基因序列進行標記,將相似的基因序列片段分為一組,視為一個OTU。在基因定序技術愈發成熟的現代,OTU的運用也越來越廣,許多領域的研究都使用OTU來做分析。根據研究,直腸癌、口腔癌與人類乳突瘤病毒感染皆與人體微生物有關,透過對微生物進行分類,找出哪些病患有上述病症相關的OTU亦是現今醫學非常重要的課題。由於基因資料具有離散性,本研究假設OTU資料呈康威-麥克斯韋-卜瓦松分配 (Conway–Maxwell–Poisson; CMP),透過比較OTU資料偏態係數的方法先將OTU分類,接著利用不同的分群參數使資料的驗後機率密度函數最大化,來達到對OTU資料做分類和對病患分群之雙分類效果。

接著,我們分別使用模擬數據和實例以評估對病患分群的表現。在統計模擬中,我們生成二維和三維的資料,來觀察不同維度下對於此方法所得出的錯誤分類率有何差異。Holmes et al. (2012)假設微生物群落服從狄利克雷多項式分配(Dirichlet-multinomial distribution),在實例中,我們使用 Holmes et al. (2012)文獻提供的資料來比較,以了解我們修改過分配後的方法在對OTU資料做分類和對病患分群的雙分類表現。

關鍵詞:康威-麥克斯韋-卜瓦松分配、EM 演算法、資料分類

結合連續型馬可夫鏈建構混合復發模型

陳昱豪、馬瀰嘉

國立成功大學統計學研究所

摘要

隨著醫學技術的不斷進步，許多疾病已經可以被控制，使得患者的狀況得到改善，在足夠長的時間內沒有發現死亡事件。但是，部份疾病可能隨著時間的推進反覆復發，因此單純的混合治癒模型並不能全面描述此種疾病的狀態。本研究使用連續型馬可夫鏈建立隨時間改變的復發機率，進一步結合未復發群體和復發群體的存活函數，採用以韋伯分配為基礎風險函數的 Cox 比例風險模型，透過概似函數和演算法，分別找出影響復發機率和存活時間的解釋變數，並估計其相對應的係數。接著，我們使用模擬數據和實例，以評估係數估計的好壞。在統計模擬中，我們生成復發資料，來計算多次復發結合 Cox 比例風險模型和韋伯分配、及多次復發結合受風險因子影響之轉移機率矩陣，並計算患者的存活機率。在實例中，我們採用 R 語言內建膀胱癌多腫瘤復發資料進行實例分析，透過係數的估計和檢定，找出對復發機率和存活時間具有影響的解釋變數，並利用該模型計算復發機率和平均存活時間。

關鍵詞：混合治癒模型、復發事件、連續型馬可夫鏈、Cox 比例風險模型、韋伯分配

探討空間統計擷取惡性黑色素瘤影像特徵的適切性

謝綸豪、黃怡婷

國立臺北大學統計學系

摘要

病灶需要仰賴專業人員來判定，但人員判讀的正確性會與經驗有關。近年來，由於資訊設備的改良，讓醫院有能力收集許多醫學影像，用來訓練深度學習模型，試圖找出可以客觀協助判讀的工具。但影像需要先進行判讀，才可以用來訓練模型，再者，不是所有病灶影像都可以很容易取得，然而，深度學習模型需要大量影像來擷取特徵，影像資料過少，勢必影像後續模型的預測能力。Tzeng 等人 (2021) 以截斷 Karhunen-Loéve 展式來建構醫學影像的空間相關，再使用特徵分解方式讓每個特徵函數代表某個特別影像的結構，依此當成影像特徵。本研究將 Tseng 等人的方法套用在黑色素瘤判讀，但影像為公開資料，本研究先將體毛覆蓋病灶的影像使用 CV2 套件以及黑帽算法來去除毛髮，可以更聚焦於病灶部分，再刪除雜訊過高的影像資料，經由人工裁切、對齊等前處理，最後再擷取特徵，透過決策樹、隨機森林、支援向量機等方法進行分類，以正確率、F1 分數來評估。結果顯示分類正確率平均將近八成，在專業人員判讀惡性黑色素瘤時，提供準確的客觀觀點予以參考。

關鍵詞：空間統計、黑色素瘤、分類器、特徵向量

運用統計方法擷取二頭肌周圍積水超音波影像特徵

高興翰、黃怡婷
國立臺北大學統計學系

摘要

現今影像辨識幾乎都是透過深度學習，隨機設定權重擷取大量特徵值，經由類神經網路訓練與分類。該模型需要大量的影像資料，才可以有很高精確度，再者，因特徵擷取方式架構在深度學習的演算法，較無法從模型的架構解釋訓練的模型，且若影像辨識結果錯誤，也很難理解模型誤判的原因。
利用長庚醫院所提供二頭肌周圍積水超音波 (Bicipital peritendinous effusion)，本研究使用兩種統計方法來進行特徵提取。第一種方法是利用克里金法 (Kriging) 計算出的預測方差，以此作為影像的第一類特徵變數。另一種方法則是將影像轉換成灰階共生矩陣，並利用灰階共生矩陣計算摘要統計量，作為第二類影像特徵變數。由受試者積水判定發炎的嚴重度，將前兩種特徵變數合併以隨機森林、支持向量機、KNN 等方法將受試者分類，以準確度、F1 分數等評估特徵變數分類的表現。結果顯示在準確度以及 F1 分數上最高有達到接近 6 成的分類效果。

關鍵詞：影像分類、克里金法、灰階共生矩陣、二頭肌周圍積水超音波

Session III-13

June 30, 11:00 – 12:20

理工一館 A307

◆Chair：

鄭宗琳 – 國立彰化師範大學數學系

◆Speaker:

1. 賴暐婷 – 國立中央大學統計研究所
2. 古裕彥 – 國立臺灣大學會計學系
3. Giap Van Su – 國立中山大學應用數學系

A Modified VAR-GARCH Model for Asynchronous Multivariate Financial Time Series via Variational Bayesian Inference

Wei-Ting Lai[1,2], Ray-Bing Chen[2] , Shih-Feng Huang[1]

[1]Graduate Institute of Statistics, National Central University

[2] Department of Statistics, National Cheng-Kung University

Abstract

This study proposes a modified VAR-GARCH model, called M-VAR-GARCH, for modeling asynchronous multivariate financial time series with GARCH effects and simultaneously accommodating the latest market information. A variational Bayesian (VB) procedure is developed to infer the M-VAR-GARCH model for structure selection and parameter estimation. We conduct extensive simulations and empirical studies to evaluate the fitting and forecasting performances of the M-VAR-GARCH model. The simulation results reveal that the proposed VB procedure produces satisfactory selection performances. In addition, our empirical studies find that the latest market information in Asia can provide helpful information to predict market trends in Europe and South Africa, especially when momentous events occur.

Keywords: Asynchronous time series, GARCH, variational Bayesian inference, vector autoregressive model

The Impact of Inflation on Housing Prices in Taiwan: Evidences from Quantile Regression Method

Yu-Yen Ku[1], Yi-Hung Kung[2], Chien Chia-Chien[3], Wang Tzu-Ling[1]

[1]Department of Accounting, National Taiwan University

[2]Department of Statistics and Information Science, Fu Jen Catholic University

[3]Department of Financial Engineering and Actuarial Mathematics, Soochow University

Abstract

We empirically investigate the impact of inflation on housing prices using a quantile regression method in the Taiwanese quarter data over the 2012Q3-2022Q4 period.
Our results suggest that the nonlinear relationship between inflation and housing prices is positively significant across the mostly quantiles in the effect of general inflation on housing prices distribution over time.
Interestingly, the effect of inflation on housing prices is stronger in the lower quantile of housing prices distribution than that in the higher quantile of housing prices distribution. Our results are robust after using alternative inflation or housing prices index.

Keywords: inflation, housing prices, quantile regression

Spectral properties of Laplacian matrix of Stochastic block models

Giap Van Su[1,2], Mei-Hui Guo[1], May-Ru Chen[1]

[1] Dept. of Applied Math., National Sun Yat-Sen University, Kaohsiung, Taiwan

[2] Department of Mathematics, Thai Nguyen University of Education, Thai Nguyen, Vietnam

Abstract

The stochastic block model, abbreviated to SBM, is an extension of the Erdös-Rényi graph. It is a generative model for random graphs, which generates graph communities modeled by being connected with particular edge densities. The SBM play a crucial role in numerous field of science, including clustering problems, machine learning, and complex network science, where it serves as a robust tool for recovering community structure in graph data. In this study, we consider the general SBM with K communities of different size. By employing the Stieltjes transform and random matrix theory, we derive an explicit formula for the limiting empirical spectral density function for the normalized Laplacian matrix of general SBM under mild conditions, as the matrix size tends to infinity. We also analyze an upper bound of the spectral norm and the properties of eigenvalues of such random matrices. Furthermore, we perform numerical simulations to validate our theoretical results.

Key words: limiting spectral density function, Normalized Laplacian matrix, Stochastic block model.

分組議程 SIV 時間表/摘要

分組議程 SIV : 2023 年 6 月 30 日 (星期五) 13:30 - 14:50

SIV-1 **計量經濟** **Chair：黃彥棕** **理 A310** **p.241**

陳樂昱 - 中央研究院經濟研究所

Robust Testing of the Matrix Semidefiniteness Hypothesis: A Weighted Eigenvalue Sum Approach

顏佑銘 - 國立政治大學國際經營與貿易學系

Robust Estimation of the Quantile Mediation Treatment Effect with Machine Learning

劉祝安 - 中央研究院經濟研究所

Model Averaging Prediction for Possibly Nonstationary Autoregressions

SIV-2 **計量經濟學理論與應用** **Chair：冼芻蕘** **理 B201** **p.245**

陳由常 - 國立臺灣大學經濟學系

Two-Step Empirical Bayes Methods

莊雅婷 - 中央研究院經濟研究所

Who Are the Low-carbon Captains? Energy Consumption and Bargaining Power in University Dorms

李金全 - 中央研究院經濟研究所

碳關稅之引力：台灣是否應課徵碳稅以應對歐盟碳邊境調整機制(CBAM)?

冼芻蕘 - 國立清華大學經濟系

Re-balancing hedge position with statistics of hedge ratios : concepts and applications

SIV-3 **空間統計方法及應用** **Chair：陳春樹** **理 A212** **P.250**

余清祥 - 國立政治大學統計學系

空間異質性檢定的比較與應用 A Study of Spatial Inhomogeneity and its Application

詹大千 - 中央研究院人文社會科學研究中心

Dynamic modelling for noise mapping in urban areas

沈宗荏 - 國立中興大學統計學研究所

Measuring the level of aggregation in the spatial distribution of multiple species using line transect data

SIV-4 **生物統計** **Chair：王秀瑛** **理 D241** **P.254**

王秀瑛 - 國立陽明交通大學統計學研究所

The distance distribution of human microRNAs

林惠文 - 東吳大學數學系

Integrating longitudinal and cross-sectional data to adjust for missing confounding factors

施嘉翰 - 中央研究院統計科學研究所

Estimation of correlation ratios using Gaussian process regression models

分組議程 SIV : 2023 年 6 月 30 日 (星期五) 13:30 - 14:50

SIV-5	**應用機率**	**Chair：陳美如**	**理 A209**	**p.258**

李志煌 - 國立臺灣大學數學系

PushASEP model on a periodic ring

林偉傑 - 國立臺灣大學數學系

Universality of superconcentration in the Sherrington-Kirkpartrick model

Kyung-Youn Kim - 國立中興大學應用數學系

Dirichlet heat kernel estimates for aniostropicMarkov processes

SIV-6	**高維度資料分析**	**Chair：林建同**	**理 A211**	**p.262**

陳素雲 - 中央研究院統計科學研究所

Robust Self-tuning Semi-parametric PCA for Contaminated Elliptical Distribution

陳定立 - 中央研究院統計科學研究所

Variable selection via information gain

黃名鈺 - 中央研究院統計科學研究所

Dimension Reduction With Incomplete Data

SIV-7	**統計方法**	**Chair：李燊銘**	**理 A318**	**p.266**

謝淑惠 - 中央研究院人文社會科學研究中心調查專題研究中心

Examining the validity of stochastic lie detector model in estimating prevalence of sensitive attributes

盧馬汀 - 國立成功大學統計學系

A Kernel-assisted Multiple Imputation Estimation of a Zero-Inflated Model When Some Covariates Are Partially Observed

李燊銘 - 逢甲大學統計學系

Estimation of zero-inflated proportional odds regression with missing covariates

分組議程 SIV : 2023 年 6 月 30 日 (星期五) 13:30 - 14:50

SIV-8 **Contributed Session - H** **Chair：孫立憲** **理 A316** **p.270**

楊竣皓 - 國立中山大學應用數學所

探討自監督學習方法於 Cryo-EM 影像去噪的潛力

錢映伶 - 國立中山大學應用數學系

透過加速 γ-EM 演算法提高 Cryo-EM 2D 分類效能

陳虹君 - 國立高雄大學統計學研究所

Dynamic asset allocation based on machine learning

羅育聖 - 國立高雄大學統計學研究所

Boosted Hybrid Method for Non-stationary Data Analysis

陳諾霖 - 國立臺北大學統計學系

編碼方式對生成對抗網路生成音樂旋律之比較與分析

劉銘軒 - 國立臺北大學統計學系

編碼方式對長短期記憶類神經網路生成音樂旋律之比較與分析

SIV-9 **Contributed Session - I** **Chair：張志浩** **理 A210** **p.277**

龔一鴻 - 輔仁大學統計資訊學系

A Tree-based Model for Rank Data with Applications in European Values Studies

林宣廷 - 國立陽明交通大學統計學研究所

利用捲積自編碼器實現保有空間結構的稀疏主成分分析

蔡羽涵 - 國立中山大學應用數學系

基於自動編碼器和高斯混合模型比較不同深度學習超參數對 Cryo-EM 3D 構象分析中的效果

游宗諺 - 輔仁大學統計資訊學系應用統計所

基於模擬方法於概化區間排名演算法之應用

李翊瑄 - 國立中山大學應用數學系

服飾電商的二階段推薦系統

邱崴泰 - 國立臺北大學統計學系

探討執行長面部因子評分與資本結構調整能力之關聯

分組議程 SIV : 2023 年 6 月 30 日 (星期五) 13:30 - 14:50

SIV-10 **Contributed Session - J** **Chair：林聖軒** **理 A106** **p.284**

李易儒 - 中央研究院統計科學研究所

Cracking the Parkinson's Code: Using Chaotic-based Deep Learning and Large-Scale Brain Dynamics for Disease Identification

江皓瑋 - 國立中山大學應用數學系

3D 卷積神經網路在大腦核磁共振影像於阿茲海默症分類之應用

林珈卉 - 國立陽明交通大學統計學研究所

Progression Detection of Alzheimer's Disease for Taiwanese Patients via MRI and Hierarchical Classification with Deep Learning

解振成 - 國立陽明交通大學統計學研究所

Deep-learning-based Anomaly Detection and Localization of EEGs in Intensive Care Units

石亞喬 - 國立臺北大學統計學系

非參數回歸、非線性回歸和時間序列模型在台股指數預測表現之比較

呂亮葳 - 國立臺北大學統計學研究所

半遞迴無母數 Hammerstein-ARX 模型於網路聲量之應用

SIV-11 **Contributed Session - K** **Chair：施銘杰** **理 A109** **p.292**

賴恩語 - 中央研究院統計科學研究所

Illustrating the pleiotropic spectra of cancers using the Taiwan Precision Medicine Initiative

黃怡甄 - 國立清華大學計算與建模科學研究所

A time dependent approach to detect the change points of Basic Reproduction Number for COVID-19

林奕勳 - 國立陽明交通大學統計學研究所

Develop a Graphical User Interface System for LVH Disease Classification Using MRI SAX View Cine Images

謝譯瑭 - 國立中山大學應用數學系

An automatic detection model for cardiovascular disease based on time-frequency domain features of PCG signals.

陳冠憲 - 國立陽明交通大學統計學研究所

The deep learning technique in prediction of extra touch-up ablation and outcomes in cryoballoon ablation for atrial fibrillation

黃乾哲 - 國立陽明交通大學數據科學與工程研究所

Predicting Multi-Vessel Diseases by Polar Maps of SPECT MPI and Deep Learning

Session IV-1

June 30, 13:30 – 14:50

理工一館 A310

◆Organizer/Chair：

黃彥棕 – 中央研究院統計科學研究所

◆Speaker:

1. 陳樂昱 – 中央研究院經濟研究所
2. 顏佑銘 – 國立政治大學國際經營與貿易學系
3. 劉祝安 – 中央研究院經濟研究所

Robust Testing of the Matrix Semidefiniteness Hypothesis: A Weighted Eigenvalue Sum Approach

Le-Yu Chen[1], Jerzy Szroeter[2]

[1]Institute of Economics, Academia Sinica

[2] Department of Economics, University College London

Abstract

The hypothesis that a given parameter matrix is positive (or negative) semidefinite occurs widely in econometric models of optimizing agents. However, testing such a hypothesis remains a challenging problem when matrix eigenvalues have unknown multiplicity. In this paper, we develop a robust test method based on a sum of estimated eigenvalues weighted by data-dependent factors which compensate for ignorance of true multiplicities. To support the theory behind this, we derive new non-asymptotic bounds on errors incurred when approximating eigenvalue sum perturbations by linear functions of the corresponding matrix perturbations. We show that the test has correct asymptotic size in the uniform sense over the null-restricted data distribution space allowing for arbitrary eigenvalue multiplicity. We establish test consistency and obtain local power. The test is operationally simple having fixed critical value, in fact standard normal. We illustrate finite-sample performance of our proposed test via Monte Carlo simulations.

Keywords: Test, Eigenvalue, Matrix Perturbation, Multiplicity, Uniform Inference

Robust Estimation of the Quantile Mediation Treatment Effect with Machine Learning

Martin Hube[1], Yu-Chin Hsu[2], Yu-Min Yen[3]

[1]Department of Economics, University of Fribourg

[2]Institute of Economics, Academia Sinica

[3]Department of International Business, National Chengchi University

Abstract

In this paper, we study estimation of the quantile mediation treatment effect by using a double/debias estimator for estimating the cumulative distribution function (c.d.f.) of the potential outcome. The proposed estimator is based on an efficient score function for the c.d.f. of the potential outcome, and is robust to misspecification of nuisance parameters. We estimate the nuisance parameters of the proposed estimator with a machine learning method and use cross-fitting to reduce estimation bias from overfitting and/or regularization of the machine learner. A multiplier bootstrap procedure is then used to conduct statistical inferences. Relevant uniform consistency of the proposed estimator and validity of the multiplier bootstrap procedure are established. We illustrate performance of the proposed estimator by conducting a simulation.

Keywords: Causal inference, efficient score, mediation analysis, quantile treatment effect, semiparametric efficiency

Model Averaging Prediction for Possibly Nonstationary Autoregressions

Tzu-Chi Lin[1] and Chu-An Liu[2]

[1]Federal Reserve Bank of Philadelphia

[2] Institute of Economics, Academia Sinica

Abstract

As an alternative to model selection (MS), this paper considers the model averaging (MA) for integrated autoregressive processes of infinite order. We derive an uniformly asymptotic expression for the mean squared prediction error (MSPE) of the averaging prediction with fixed weights and then propose a Mallows-type criterion to select the data-driven weights that minimize the MSPE asymptotically. We show that the proposed MA estimator and its variants, Shibata and Akaike MA estimators, are asymptotically optimal in the sense of achieving the lowest possible MSPE. We further demonstrate that MA can provide significant MSPE reduction over MS when the model misspecification bias is algebraic decay. These theoretical findings are supported by Monte Carlo simulations and real data analysis.

Keywords: Asymptotic improvability, Asymptotic optimality, Integrated autoregressive processes, Model averaging

Session IV-2

June 30, 13:30 – 14:50

理工一館 B201

◆Organizer/Chair：

冼芻蕘 – 國立清華大學經濟系

◆Speaker:

1. 陳由常 – 國立台灣大學
2. 莊雅婷 – 中央研究院
3. 李金全 – 中央研究院經濟研究所
4. 冼芻蕘 – 國立清華大學經濟系

Two-Step Empirical Bayes Methods

陳由常

國立臺灣大學經濟學系

Abstract

Empirical Bayes (EB) methods are widely employed in the estimation of fixed effects. Prominent examples in economics include teacher value-added measures (VAM) and neighborhood effects. In these applications, the estimated effects are often further used as inputs to subsequent statistical analysis. This paper studies the validity of such procedures, in which case when EB estimates are used as explanatory variables in a regression is of particular interest. Our result shows that, although the ordinary least square (OLS) estimator is consistent in linear models, the standard errors could be substantially downward biased if one does not correct for generated regressors. For example, the standard errors can be 40 percent larger after the correction in a typical teacher VAM application. For non-linear regressions, we find that the two-step procedure is harder to justify, as the resulting estimator is generally inconsistent without imposing further distributional assumptions.

Who Are the Low-carbon Captains? Energy Consumption and Bargaining Power in University Dorms

莊雅婷
中央研究院經濟所

Abstract

This study investigates how the interaction between personality traits and bargaining power influences room electricity usage in university dormitories. First, we explore the association between electricity usage and five personality traits—independent thinking, social justice, self- discipline, interaction, and innovation to identify traits that predict low electricity consumption. We create a jointly defined traits variable using principle component analysis. Then, we exam- ine whether residents with such traits in the meantime with the highest bargaining power affect their room's electricity consumption. Using family income and physical strength to approximate bargaining power, we find that, in general, residents with high bargaining power seem to domi- nate the electricity usage decision-making in a room. Our study provides empirical evidence to better understand how personal traits correlate with actual electricity consumption. We also further advance the literature by including bargaining power dynamics to explain energy usage behaviors in a group setting.

碳關稅之引力：
台灣是否應課徵碳稅以應對歐盟碳邊境調整機制（CBAM）？

李金全，楊宗翰

中央研究院經濟研究所

摘要

歐盟計畫今年（2023）十月開始實施碳邊境調整機制（CBAM）的過渡期。歐盟希望透過CBAM的施行，防止他們的產業外流到對環境標準較低的國家，以抑止全球的碳排放總量。大部分研究使用可計算一般均衡模型。但在此研究中，作者將建立一個多國、多產業的結構性引力模型，以用來衡量碳稅、碳關稅以及 CBAM 對於各國，特別是台灣的實質所得、福利等經濟衝擊以及碳排放量的影響。此模型將各國的經濟體透過國際貿易，以及全球供應鏈連結起來，並同時將碳排放內生化。在此研究，作者預計運用世界投入產出表資料庫（World Input-Output Database）2016 來校準模型的所有參數，並努力做到以下幾點：(1) 在當前實際的碳定價的情況下，量化碳洩漏（carbon leakage）的程度。(2) 使用結構性引力模型來評量歐盟 CBAM 對於實質所得、社會福利以及碳排放變化的影響，(3) 並計算台灣向國內能源密集產業可能需要課徵的最佳碳稅稅率，以因應歐盟 CBAM，以及 (4) 探討當未來 CBAM 關稅稅率提升或台灣的主要貿易夥伴決定實施自己的 CBAM 時，這種最優的國內碳稅稅率可能的變化。

Re-balancing hedge position with statistics of hedge ratios: concepts and applications

Chor-yiu (CY) SIN

College of Technology Management, National Tsing Hua University

Abstract

In a recent article (2022, *The Journal of Derivatives*), Fabozzi and Fabozzi conclude "..... there is strong evidence that these models (advanced econometric models) do not improve hedge efficiency significantly, if at all". As a matter of fact, dynamic hedging attempts to strike the balance between hedging effectiveness and transactions costs. In this paper, using the Garch asymptotic theories developed by Ling and McAleer (2003, *Econometric Theory*), Francq and Zakoïan (2012, *Econometric Theory*), and Zhang, Sin and Ling (2015, *Stochastic Processes and their Applications*), we derive the asymptotic properties of the hedge ratio. As a result, we construct a natural and simple statistic of re-balancing, namely, the (asymptotic) standard deviation of hedge ratio. We apply our method to a number of paired variables such as WTI Crude Oil Futures and Spot Price. Empirical results are compared with those obtained in Tsuji (2018, *Economic Modelling*), Choudhry, Hasan and Zhang (2019, *International Journal of Banking, Accounting and Finance*), and Wang, Lin, Lin and Lai (2020, *Journal of Risk*).

Keywords: Dynamic hedging; Garch; model complexity; statistics of hedge ratios; transaction costs

Session IV-3

June 30, 13:30 – 14:50

理工一館 A212

◆Organizer/Chair：

陳春樹 – 國立中央大學統計研究所

◆Speaker:

1. 余清祥 – 國立政治大學統計學系
2. 詹大千 – 中央研究院人文社會科學研究中心
3. 沈宗荏 – 國立中興大學統計學研究所

空間異質性檢定的比較與應用

A Study of Spatial Inhomogeneity and its Application

余清祥 (Jack C. Yue)

國立政治大學統計

摘要

臺灣土地面積雖然不大，各地居民的壽命差異卻相當大，民國 110 年臺北市、臺東縣男性平均壽命高達九歲，隨著公共政策、社會資源的調整與挹注，預期國人健康及壽命的地區不平等將逐漸縮小。空間異質性可用於檢視健康與壽命的地區差異，而空間異值指的即是空間上的差異，例如特定的高死亡率地區、低生育率地區，相關檢定大致可分為三大類：總體檢定、區域檢定、焦點檢定，總體檢定可用於檢定全區域資料是否為空間同質，區域檢定多用於偵測高風險地區（或稱為群聚），焦點檢定則用於確認特定地區周圍是否有較高發生率。本文先使用電腦模擬比較三種異質性檢定中各一個具代表性的方法：Moran's I（總體檢定）、SaTScan（區域檢定）及 Tango score test（焦點檢定），比較資料具有空間自相關及群聚時的偵測效果，以提供實務分析的參考。其中模擬的實驗地區為二度空間，大小為 5x5、7x7、9x9、…、21x21 的格子點，檢測各方法在空間同質、空間自相關、群聚的效果。研究發現三種方法在空間同質性的結果大致相同，Moran's I 在空間自相關時最敏感，而群聚存在時則以 SaTScan 效果最佳。本文套用上述研究結果至縣市、鄉鎮市區層級全死因及前三大主要死因（惡性腫瘤、心臟疾病、肺炎），藉由探索性資料分析、空間分析等方法，包括空間異質性的統計檢定，探討我國國民死亡率的地區差異與時間變化趨勢。分析發現主要死因的死亡率具有空間異質性，惡性腫瘤及心臟疾病熱區大多落在東南部山區，地區間的差異未隨時間有明顯改變；肺炎死亡率在西半部有逐年上升的趨勢，推測與都市化的空氣品質惡化有關。

關鍵詞：空間異質性、空間自相關、平均餘命、主要死因、探索性資料分析

Dynamic modelling for noise mapping in urban areas

Jia-Hong Tang[1], Bo-Cheng Lin[2], Jing-Shiang Hwang[3], Ling-Jyh Chen[4], Bing-Sheng Wu[5], Hong-Lian Jian[1], Yu-Ting Lee[1], Ta-Chien Chan[1]*

[1]Research Center for Humanities and Social Sciences, Academia Sinica

[2]Department of Real Estate and Built Environment, National Taipei University

[3]Institute of Statistical Science, Academia Sinica

[4]Institute of Information Science, Academia Sinica

[5]Department of Geography, National Taiwan Normal University

Abstract

Environmental noise has been a major environmental nuisance in metropolitan cities. To achieve the goal of sustainable community, noise reduction is an important approach. Without systematic noise mapping, the spatio-temporal distribution of noise variations is hard to capture. This study proposes a new methodology framework to combine statistical models and acoustic propagation for dynamic updates of 2D and 3D traffic noise maps by using a limited number of noise sensors in Taipei City based on multisource data including noise monitoring, vehicle detectors, meteorological data, road characteristics, and socio-demographic data. The hourly mean difference between the predicted and measured noise level is within the range of −6.25 dBA to −4.46 dBA in the 2D noise model. For the 3D noise model, the hourly mean prediction error is within the range of 0.02 dBA to 1.93 dBA. Based on the WHO benchmark for excessive road traffic noise, we found at least 30% of inhabitants in Taipei City are exposed to levels exceeding 53 dBA Lden, and >25% are exposed to noise levels exceeding 45 dBA Lnight. The noise maps not only can help identify vulnerable communities to adopt proper approaches for noise reduction but also can remind the residents to take action to improve their quality of life.

Keywords: Traffic noise, Noise mapping, Noise monitoring, Noise modeling, Noise prediction

Measuring the level of aggregation in the spatial distribution of multiple species using line transect data

沈宗荏

國立中興大學統計學研究所

Abstract

It is common knowledge that some specific species in an ecological community present aggregate distributions, but this does not necessarily imply that the community as a whole presents an aggregate distribution. Using the conspecific-encounter index derived from the Markov non-independent sampling model, this talk will introduce a legible definition of community-level distributional aggregation as an interspersed or cluster-like distribution of different species. In practical applications, by utilizing the conspecific-encounter index that accounts for the non-independent sampling of consecutive individuals along line transects, the result reveals that tree assemblages in tropical forest ecosystems can present a strong signal of extensive distributional interspersion. By contrast, for the amphibian assemblages, the conspecific-encounter index was consistently high, implying that amphibian communities tend to be highly aggregate in space.

Session IV-4

June 30, 13:30 – 14:50

理工一館 D241

◆Organizer :

程毅豪 – 中央研究院統計科學研究所

◆Chair：

王秀瑛 – 國立陽明交通大學統計學研究所

◆Speaker:

1. 王秀瑛 – 國立陽明交通大學統計學研究所
2. 林惠文 – 東吳大學數學系
3. 施嘉翰 – 中央研究院統計科學研究所

The distance distribution of human microRNAs

Hsiuying Wang

Institute of Statistics, National Yang Ming Chiao Tung University

Abstract

microRNAs (miRNAs) are small single-stranded non-coding RNAs of approximately 22 nucleotides that play an important role in cell differentiation, development, regulation of cell cycle, and apoptosis. miRNAs target many disease-related genes so that they can be used as biomarkers for many diseases. The relationships between different diseases were investigated and discussed based on their miRNA biomarkers in literature as well as the relationships between vaccines and diseases. So far, there are around 2000 miRNAs that have been discovered in humans. In this study, we calculate the pairwise distances of these human miRNAs and then fit statistical models for these distances. Then we use these distance distributions to discuss the disease associations in some examples.

Integrating longitudinal and cross-sectional data to adjust for missing confounding factors

Hui-Wen Lin[1], Yi-Hau Chen[,2]

[1] Department of Mathematics, Soochow University

[2] Institute of Statistical Science, Academia Sinica

Abstract

This study aims to address the challenges associated with analyzing longitudinal data in medical research by proposing a two-stage research approach that combines cross-sectional and longitudinal data. By collecting a large amount of cross-sectional data in Stage 1 and selecting a smaller sample for long-term monitoring in Stage 2, the proposed approach reduces costs while improving model estimation performance and better representing the population. Specifically, the study focuses on the integration of logistic regression models for cross-sectional data with generalized estimating equations (GEEs) for longitudinal data. This two-stage method not only corrects estimation biases in confounding factors but also provides a practical solution for effectively integrating longitudinal and cross-sectional data in medical research.

Keywords: integrating longitudinal and cross-sectional data, generalized estimating equations (GEEs), two-stage method.

Estimation of correlation ratios using Gaussian process regression models

Jia-Han Shih (施嘉翰)[1], Yi-Hau Chen (程毅豪)[1]

[1] Institute of Statistical Science, Academia Sinica

Abstract

Recently, measures of regression association received a lot of attention. In this talk, we consider the so-called correlation ratio of a dependent univariate variable Y on an independent multivariate variable $\mathbf{X}$, i.e., the proportion of variation in Y that can be explained by $\mathbf{X}$. It is regarded as a regression association measure. The existing nonparametric estimation method based on the concept of the nearest neighbor (Azadkia and Chatterjee, Ann. Stat. 49(6), 3070-3102, 2021) may have large estimation bias if the dimension of $\mathbf{X}$ is large. To resolve this issue, we propose to estimate the correlation ratio by using Gaussian process regression models. In addition, we also consider some extensions such as the conditional correlation ratio and the partial correlation ratio. Both these two extensions allow a multivariate control variable $\mathbf{Z}$. Simulation studies are conducted to examine the performance of the proposed methods. We conclude this talk with some discussions.

Keywords: Covariance, Directional association, Explained variance, Functional association, Pearson correlation coefficient

Session IV-5

June 30, 13:30 – 14:50

理工一館 A209

◆Organizer/Chair :

陳美如 – 國立中山大學應用數學系

◆Speaker:

1. 李志煌 – 國立臺灣大學數學系
2. 林偉傑 – 國立臺灣大學數學系
3. Kyung-Youn Kim – 國立中興大學應用數學系

PushASEP model on a periodic ring

Jhih-Huang Li[1], Axel Saenz[2]

[1]National Taiwan University, Taiwan

[2]Oregon State University, United States

Abstract

In this ongoing joint work with Axel Saenz (Oregon), we are interested in the following interacting particle system. On a periodic ring of size L, there are N particles jumping according to some specific rules. At rate $p \geq 0$, each particle tries to jump to the right by 1 and the jump happens only if the target site is not occupied (called TASEP dynamics). At rate $q \geq 0$, each particle jumps to the left by 1, by propagating its jump if the target site is occupied (called Push dynamics). We want to understand the behavior of such a system at the relaxation time scale for large L and N with the ratio N/L fixed.

Keywords: statistical mechanics, interacting particle systems

Universality of superconcentration in the Sherrington-Kirkpartrick model

Wei-Kuo Chen[1], Wai-Kit Lam[2]

[1] University of Minnesota

[2] National Taiwan University

Abstract

We study the universality of superconcentration for the free energy in the Sherrington-Kirkpatrick (SK) model. It was shown by Chatterjee that when the system consists of N spins and Gaussian disorders, the variance of this quantity is superconcentrated by establishing an upper bound of order N/logN, in contrast to the O(N) bound obtained from the Gaussian-Poincaré inequality. In this talk, I will discuss a joint work with Wei-Kuo Chen, showing that superconcentration indeed holds for any choice of centered disorders with finite third moment, where the upper bound is expressed in terms of an auxiliary nondecreasing function f that arises in the representation of the disorder as f(g) for g standard normal. Under an additional regularity assumption on f, we further show that the variance is of order at most N/logN.

Keywords: Sherrington-Kirkpatrick model, Universality, Superconcentration

Dirichlet heat kernel estimates for aniostropicMarkov processes

Kyung-Youn Kim
Applied Mathematics, National Chung Hsing University

Abstract

Corresponding to the non-local operator, there exists a discontinuous Markov process with this operator as infinitesimal generator. Also the heat kernel of the non-operator is the transition density of the Markov process. In this talk, we discuss the heat kernel bounds for the anisotropic discontinuous Markov process. Let L_i, $i = 1, \ldots, d$, be identical and independent 1–dimensional symmetric Lévy processes whose characteristic functions satisfy the weakly scaling condition. Define a Markov process $M := (M_1, \ldots, M_d)$ whose jumping kernel is comparable to that of $L := (L_1, \ldots, L_d)$. Then M is a pure jump process that jumps parallel to the coordinate axes. We discuss the sharp two-sided heat kernel estimates on $\mathbb{R}^d$ and $C^{1,1}$–open set $D \subset \mathbb{R}^d$. This is the joint work with Lidan Wang.

Session IV-6

June 30, 13:30 – 14:50

理工一館 A211

◆Organizer :

呂恆輝 – 東海大學統計學系

◆Chair :

林建同 – 逢甲大學

◆Speaker:

1. 陳素雲 – 中央研究院統計科學研究所
2. 陳定立 – 中央研究院統計科學研究所
3. 黃名鈺 – 中央研究院統計科學研究所

Robust Self-tuning Semi-parametric PCA for Contaminated Elliptical Distribution

Hung Hung (洪弘)[1], Su-Yun Huang (陳素雲)[2], Shinto Eguchi (江口真透)[3]

[1]National Taiwan University

[2]Academia Sinica

[3]Institute of Statistical Mathematics

Abstract

Principal component analysis (PCA) is one of the most popular dimension reduction methods. The usual PCA is known to be sensitive to the presence of outliers, and thus many robust PCA methods have been developed. Among them, the Tyler's M-estimator is shown to be the most robust scatter estimator under the elliptical distribution. However, when the underlying distribution is contaminated and deviates from ellipticity, Tyler's M-estimator might not work well. In this article, we apply the semiparametric theory to propose a robust semiparametric PCA, which is shown to be a re-weighted Tyler's M-estimator. The merits of our proposal are twofold. First, it is robust to both heavy-tailed elliptical outliers and nonelliptical outliers. Second, it pairs well with a data-driven tuning procedure, which is based on active ratio and can adapt to different degrees of data outlyingness. Simulation studies and a data analysis demonstrate the superiority of our method.

Keywords: active ratio, elliptical distributions, influence function, PCA, robustness, semiparametric theory, Tyler's M estimator.

Variable selection via information gain

Ting-Li Chen

Institute of Statistical Science, Academia Sinica

Abstract

Variable selection is a critical step in building statistical models, as it helps to identify the most important predictors for explaining the response variable. In this talk, we will focus on an information gain-based variable selection technique that can effectively handle non-linear relationships between the response and the covariates. We will introduce the information gain criterion, which compares the entropy of the response variable to the conditional entropy of the response variable given an explanatory variable, to identify the variables that best explain the response. Finally, we will demonstrate the strength of this technique through examples, and compare it with other variable selection methods.

Keywords: variable selection, entropy, criterion

Dimension Reduction With Incomplete Data

Ming-Yueh Huang
Institute of Statistical Science Academia Sinica

Abstract

The prevalence of multi- or high-dimensional data is a significant trend in modern research, often a consequence of sophisticated data collection techniques. However, the gathered data may not be comprehensive, reflecting the full scope of the targeted population. This incompleteness could result from restrictive data collection policies or inherent characteristics of the population itself. In this talk, I will highlight the primary challenge of performing dimension reduction on such incomplete high-dimensional data. To mitigate this issue, we propose the utilization of an imputation method, serving as a stand-in for missing or incomplete responses. I will demonstrate this technique with two distinct types of data: observational data for causal inference and right-censored survival data.

Session IV-7

June 30, 13:30 – 14:50

理工一館 A318

◆Organizer/ Chair :

李燊銘 – 逢甲大學統計學系

◆Speaker:

1. 謝淑惠 – 中央研究院人文社會科學研究中心調查專題研究中心
2. 盧馬汀 – 國立成功大學統計學系
3. 李燊銘 – 逢甲大學統計學系

Examining the validity of stochastic lie detector model in estimating prevalence of sensitive attributes

Shu-Hui Hsieh

Center for Survey Research, Research Center for Humanities and Social Sciences,

Academia Sinica

Abstract

The traditional direct questioning methods result often in biased data due to nonresponse or social desirability bias when sensitive topics are included in the survey. The randomized response technique (RRT), an indirect questioning techniques (IQTs), aims to control for this bias by protecting the respondent's privacy and increasing cooperation. Unfortunately, some participants choose not to follow the instructions and provide untruthful responses. This study examines the validity of the stochastic lie detector (SLD) model in estimating the prevalence of a sensitive attribute. The results indicate that SLD produces higher prevalence estimates but with a lower proportion of carriers responding truthfully than the evidence supporting its validity. Moreover, if noncarriers are attracted to this answer to avoid suspicion, it violates the model assumptions and leads to distortion of the prevalence estimates. These findings highlight the importance of examining the validity of IQTs and the need for proper survey design to minimizes bias in estimating prevalence of sensitive attributes.

A Kernel-assisted Multiple Imputation Estimation of a Zero-Inflated Model When Some Covariates Are Partially Observed

Martin Lukusa
Department of Statistics and Institute of Data Science, NCKU

Abstract

Count data with excess of zeros are common in many areas of applied statistics research such as Transportation, Epidemiology, Ecology, Manufacturing, etc. These zeros arise from the events of interest which never occurred and which did not occur. One popular approach to model the excess of zeros problem is to use a zero-inflated model which mingles two distributions including the count distribution and the zero mass distribution models. In practice, the count and the zero mass parts are modeled as linear predictors. We consider that some covariates used in the linear predictors are subject to missingness. Under the missing data at random (MAR) process, the naive estimation method is known to be biased estimator. Hence, we propose two semipara- metric, kernel-assisted imputation, estimators that mimic Lee et al. estimators'(2020). The proposed estimators are unbiased with an asymptotic normal distribution. A simulation study is used to investigate the finite sample performance of the proposed estimators. Additionally, we exploit an example of traffic data. The overall results is satisfactory.

Keywords: Zero-inflated data, Missing at Random, Semiparametric Estimation, Nonpara- metric imputation, Kernel-assisted, Traffic data.

Estimation of zero-inflated proportional odds regression with missing covariates

Shen-Ming Lee[1], Chin-Shang Li[2]

[1]Department of Statistics, Feng Chia University

[2]School of Nursing, The State University of New York University at Buffalo

Abstract

A zero-inflated proportional odds (ZIPO) regression model is proposed as an alternative model for analysis of ordinal categorical response data with more zeros than accounted for by a proportional odds regression model. The inverse probability weighting method and nonparametric multiple imputation (MI) method are suggested to address the issue of estimation of the parameters in the ZIPO regression model when covariates are missing at random. Variance estimation is performed by using Rubin-type and two proposed bootstrap methods. Simulations are conducted to investigate the finite-sample performances of the Rubin-type method and two proposed bootstrap methods. Simulation results indicate that the performances of the two proposed bootstrap methods are comparable and generally superior to the Rubin-type method in terms of the agreement between standard deviation and asymptotic standard error. To demonstrate the practical use of the proposed estimation methods, we utilized a survey data set of violations of traffic rules of motorcyclist respondents in Taiwan.

Keywords: Bootstrap; estimating function; inverse probability weighting; missing at random; multiple imputation

Session IV-8

June 30, 13:30 – 14:50

理工一館 A316

◆Chair：

孫立憲 – 國立中央大學統計研究所

◆Speaker:

1. 楊竣皓 – 國立中山大學應用數學所
2. 錢映伶 – 國立中山大學應用數學系
3. 陳虹君 – 國立高雄大學統計學研究所
4. 羅育聖 – 國立高雄大學統計學研究所
5. 陳諾霖 – 國立臺北大學統計學系
6. 劉銘軒 – 國立臺北大學統計學系

探討自監督學習方法於 Cryo-EM 影像去噪的潛力

楊竣皓

國立中山大學應用數學所

摘要

本文旨在探討利用深度學習自監督的方式應用於 Cryo-EM 70S 核糖體的合成影像進行去噪，對比於傳統的小波轉換、總變差及 BM3D 的去噪方法，是否有比較好，測試過程中主要使用盲點網路法去替換掉原有的像素點，產生可行的自監督的去噪損失函數，一共採用了四個神經網路的方法：Noise2Void 在只有吵雜的影像下進行去噪；Self2Self 利用單張的吵雜影像即可進行去噪任務；Noise2Same 考量到盲點網路的使用對於 J−不變性的嚴格程度不同，所影響的去噪效；Neighbor2Neighbor 利用產生兩個吵雜的子圖，用到更多的吵雜影像資訊對於去噪的幫助。並探討這些方法的超參數對於去噪效果的影響。

關鍵詞：U-Net、影像去噪、深度學習、自監督學習

透過加速γ-EM 演算法提高 Cryo-EM 2D 分類效能

錢映伶

國立中山大學應用數學系

摘要

本篇論文探討了如何透過加速γ-EM 演算法來提高低溫電子顯微鏡(cryo-EM)影像分類效能。由於低溫電子顯微鏡影像通常具有低訊噪比，需要進行影像對齊和分類來消除噪聲，因此相關演算法非常重要。γ-EM 為此領域新提出有潛在的演算法，相較於傳統 cryo-EM 分類法的優勢在於對離群值較不敏感和能自動決定潛在的群集數量，因此相當適合拿來做探索式分析。然而其運算時間冗長且未測試於不同的 cryo-EM 資料集，因此實際應用上仍有其限制。本論文提出了透過平行化參數估計、fast subset 演算法來大幅度縮短運行時間。另外更進而用 phase-flip 與 HDBSCAN 演算法來改善演算法前後處理步驟進而提升其分類效果。模擬研究結果顯示，本研究所提出的改良後γ-EM 在不同 cryo-EM 資料集表現良好。同時，我們也研究了提出各種方法對γ-EM 的影響，以及與現有的分類演算法 ML2D 進行比較以評估其實際應用效果。未來可進一步研究如何在實際的資料集中獲得準確的分類結果，並探索其在其他領域的應用潛力。

關鍵詞：影像分群、γ-最大化期望值演算法、平行化參數估計、fast subset 演算法、phase-flip、HDBSCAN

Dynamic asset allocation based on machine learning

Hong-Chun Chen(陳虹君)

Institute of Statistics, National University of Kaohsiung

Abstract

This study proposes a dynamic asset allocation strategy based on machine learning. The first part of the strategy uses daily stock prices and some popular variables to cluster the trading days by Single-cell Interpretation via Multi-kernel LeaRning. The associated operation direction, optimal take-profit threshold, and optimal stop-loss point for each cluster are calculated accordingly. The second part confirms the operation position according to several technical indicators and the operation direction suggested in the first part. The third part predicts operation positions for new data by combining multiple machine learning methods. We employ the Largan stock price data from March 2012 to February 2022 from the TEJPro database to investigate the investment performance of the proposed method. Numerical results reveal that the proposed trading strategy it has a more stable investment performance than competitors.

Keywords:optimal take-profit threshold, optimal stop-loss point, trading strategy

Boosted Hybrid Method for Non-stationary Data Analysis

羅育聖

國立高雄大學統計學研究所

摘要

針對非平穩數據的分析，文獻上已提出多種對線性模型殘差進行非線性建模之技巧來兼顧模型解釋力與擬合度。這類方法多將數據透過時間序列模型預測平穩分量後，將剩餘殘差以類神經網絡等機器學習技法進行擬合。本研究基於此類混合模型，進一步引入增強(boosting)學習與指數移動平均(EWMA)的方法，來增進模型預測效果。在實證分析的部分，我們將強化混合(Boosting hybrid, BH)的方法應用於文獻研究中三筆真實數據之分析，並與前述文獻之混合模型方法進行比較。數據分析結果顯示，增強混合方法可以在預測上得到較為平穩的表現。

關鍵詞：Boosting Hybrid、EWMA、非平穩數據

編碼方式對生成對抗網路生成音樂旋律之比較與分析

陳諾霖

國立臺北大學統計學系

摘要

本研究使用了兩種不同風格的樂曲作為音樂旋律生成的數據集。其中一個數據集是麻省理工學院所開發的 Music21 套件中的巴哈聖詠（Bach Chorales）；另一個數據集則是來自公開資源網站 KernScores 的德國民謠。在進行音樂編碼方面，我們使用了兩種不同的編碼方式，而兩種風格的樂曲皆使用了這兩種不同的音樂編碼方式進行編碼。在進行訓練方面，我們選擇了深度學習（Deep Learning）中的生成式對抗網路（Generative Adversarial Network, GAN）。此模型能在訓練過程中不斷地進行生成和判斷的交互，從而逐漸提高生成的品質。為了評估生成結果的表現，本研究使用了多種可解釋的客觀評估指標，這些指標包括平均音高間隔、音符長度直方圖、音高集合轉移矩陣、音高範圍等。

關鍵詞：音樂生成、深度學習、生成對抗網路、編碼

編碼方式對長短期記憶類神經網路生成音樂旋律之比較與分析

Music Melody Generation using Long Short-Term Memory Neural Network with Different Encoding Schemes – A Comparative Analysis

劉銘軒

國立臺北大學統計學系

摘要

LSTM (Long Short-Term Memory)模型是一種循環神經網路(Recurrent Neural Network)的變型，常用於處理序列資料。我們的研究在於探討 LSTM 於生成音樂旋律上之應用。在資料編碼方式上，目前最常使用的是獨熱編碼(one-hot encoding)，但它忽略了音高順序的資訊，而此資訊對應到的是旋律起伏走勢的重要性質。我們嘗試使用數值編碼保留此資訊，希望更能捕捉音樂集中旋律的特性。我們將兩種編碼方式的模型實作於德國民謠音樂集以及巴哈詠唱音樂集，分別生成音樂旋律。我們進一步使用多項客觀指標，評估與探討兩種編碼方式所生成之音樂旋律的表現。

關鍵詞：LSTM、音樂生成、編碼

Session IV-9

June 30, 13:30 – 14:50

理工一館 A210

◆Chair：

張志浩 – 國立高雄大學統計學研究所

◆Speaker:

1. 龔一鴻 – 輔仁大學統計資訊學系
2. 林宣廷 – 國立陽明交通大學統計學研究所
3. 蔡羽涵 – 國立中山大學應用數學系
4. 游宗諺 – 輔仁大學統計資訊學系應用統計所
5. 李翊瑄 – 國立中山大學應用數學系
6. 邱崴泰 – 國立臺北大學統計學系

A Tree-based Model for Rank Data with Applications in European Values Studies

Yi-Hung Kung[1], Yu-Yen Ku[2], Shih Yu-Shan[3], Chien Chia-Chien[4], Wang Tzu-Ling[2]

[1] Department of Statistics and Information Science, Fu Jen Catholic University

[2] Department of Accounting, National Taiwan University

[3] Department of Mathematics, National Chung Cheng University

[4] Department of Financial Engineering and Actuarial Mathematics, Soochow University

Abstract

A decision tree method for analyzing complete or partial rank data is developed in this article. It utilizes Pearson's chi-squared tests of independence to guide split variables and points selection. A Bonferroni type stopping rule is applied to declare a node terminal. Through simulations, we find that this method is relatively unbiased and more powerful in selecting the informative split variables. A real data study, the European Values Study, further demonstrates the effectiveness of our method.

Keywords: classification and regression tree, distance-based model, independence test, selection bias

利用捲積自編碼器實現保有空間結構的稀疏主成分分析

林宣廷

國立陽明交通大學統計學研究所

摘要

本研究提出一種結合捲積神經網路(Convolutional Neural Network)和自編碼器(Autoencoder)的架構，以實現保有空間結構的稀疏主成分分析。我們利用自編碼器找出最能解釋空間變異性的主成分，同時透過捲積神經網路捕捉區域性的空間特徵。此外，我們在損失函數中加入稀疏懲罰項以促進主成分的解釋性，透過模擬實驗，我們呈現了所提方法比傳統主成分分析和稀疏主成分分析等常見方法的優越之處。

關鍵詞：類神經網路、自編碼器、主成分分析、空間主成分分析、稀疏主成分分析

基於自動編碼器和高斯混合模型比較不同深度學習超參數對 Cryo-EM 3D 構象分析中的效果

蔡羽涵

國立中山大學應用數學系

摘要

Cryo-EM 技術可協助研究人員觀察蛋白質等生物分子的結構和其運作機制，而基於神經網路的三維高斯混合模型是近年來能自動解析蛋白質複合物中的結構異質性的重要工具之一，然而在過去文獻當中對於此演算法中各項參數對最後效能的影響卻未有量化的探討，因此很難去進一步改良其架構或探索其使用上的限制性。本研究旨在利用 70S 核糖體的合成數據、CryoDRGN 核糖體的合成數據以及真實數據 EMPIAR-10076 來量化各參數影響，我們將比較及調整深度學習中的超參數以及最後分群是否使用 UMAP 降維方法對 Cryo-EM 三維構象分析效果的影響，並以 purity 和 impurity 的平均值作為評估指標。其中，超參數以高斯混合模型的位置、幅度和標準差，以及神經網路層數、學習率、是否使用卷積網絡、池化層、激活函數和初始化函數逐步進行調整。在實驗的後期，為了達到更好的成果，引入了變分自動編碼器的架構、Exponential Decay、One Cycle AdamW 的學習率排程方法，以及q-高斯分佈。最終，綜合最佳參數及架構的組合，我們發現基於調整後自動編碼器的三維q-高斯分佈加上 One Cycle AdamW 的分群效果表現最優異，能夠讓合成數據和真實數據依據各自的類別清楚的分開來。我們相信改良後的架構更能適用於廣泛的資料集，進而幫助後續的構象分析研究。

關鍵詞：Cryo-EM 3D 構象分析、自動編碼器、三維高斯混合模型、深度學習、超參數搜尋

基於模擬方法於概化區間排名演算法之應用

杜逸寧
輔仁大學統計資訊學系
游宗諺
輔仁大學統計資訊學系應用統計所

摘要

林勝義(2015)針對評估學生成績表現的問題而提出一種區間排名演算法，該演算法延伸資料探勘方法 K 最近鄰的概念，將相似的資料形成多個區間，並將區間進行排名。本研究認為在資料具有隨時間變化的特性下，若資料變化符合該跳脫區間條件，將進行區間排名的調整。假設在一場比賽中擁有 m 位選手，每位選手擁有 n 次測驗紀錄，接著找出每位選手長期實力鄰近的 k 位選手形成各自的實力區間並定義區間內排名，在區間中的前段與後段選手將透過區間排名演算法判斷是否需要調整，調整後會建議選手是否需要重測以獲得更佳的紀錄，並檢驗調整後的選手是否滿意建議決策。另外本研究想探討區間內 k 位選手的數量是否會影響排名調整的結果。經過多次實驗模擬後，本研究發現當 k 值介於 m 的 3%至 7%之間能達到較好的結果，並在 5%時效果最佳，平均滿意程度為 64.79%。因此，本研究欲透過資料模擬的方式進行實驗設計，經過大量不同情境的模擬過程，使結果更能反應出資料是否符合長期的趨勢。

關鍵詞：區間排名演算法、K 最近鄰、資料模擬、實驗設計

服飾電商的二階段推薦系統

Two-stage recommendation system for fashion E-Commerce System

李翊瑄、郭美惠

國立中山大學應用數學系

摘要

在推薦系統優化問題中，協同過濾(Collaborative Filtering)是一種常見的方法，透過分析用戶的行為和偏好，推薦用戶感興趣的產品或內容。推薦系統的協同過濾方法使用矩陣分解(Matrix Factorization)將稀疏的用戶-產品評分矩陣分解成兩個低維度的矩陣，預測用戶對產品的喜好分數，並進行排序推薦給用戶。然而，在實務上的電商系統中，用戶及產品的資料量巨大，僅使用矩陣分解進行推薦將使得產品的排序不夠精準。學習排序是源自於資料檢索領域的機器學習演算法，考慮產品及用戶的特徵(features)，建立用戶對產品偏好的模型，以增進產品排序的精準度。在本研究我們將協同過濾和學習排序演算法結合為兩階段推薦系統，對每個用戶可能感興趣的產品進行評分的預測，從而進行最終排序。我們使用服飾電商系統的用戶交易資料進行實證分析，結果顯示採用協同過濾和學習排序的機器學習模型，能夠有效提高推薦系統的平均精度均值（MAP@k）。

關鍵詞：推薦系統、協同過濾、矩陣分解、學習排序、電商系統

探討執行長面部因子評分與資本結構調整能力之關聯

邱崴泰

國立臺北大學統計學系

摘要

一些新興的財務與會計文獻將公司執行長的面部結構與經濟成果及公司績效聯繫起來，而本論文則是研究面部特徵與資本結構調整之間的關聯，探討國外高階經理人的能力是否與他們的面部特徵相關，是否面部特徵分數較高的經理人能更有效率的改善公司的資本結構。

本文使用 Compustat Excucomp 數據集中 2012 年到 2021 年一千五百大公司財務報表的數據，為 Flannery and Rangan（2006)論文中最適資本結構的部分調整模型去計算公司的目標槓桿，再來將數據集中的人臉照片套入機器學習模型計算出三個不一樣的面部因子評分，即吸引力、可信度和支配力，並探討高面部因子評分(更有個性)的執行長是否有更大的能力調整公司槓桿。

再來本文還觀察高評分的執行長在公司不同槓桿下以及不同公司成長機會下，可能進行的財務行為，是偏好舉債或是增資。

關鍵詞：資本結構調整、面部因子評分、公司成長機會

Session IV-10

June 30, 13:30 – 14:50

理工一館 A106

◆Chair：

林聖軒 – 國立陽明交通大學統計學研究所

◆Speaker:

1. 李易儒 – 中央研究院統計科學研究所
2. 江皓瑋 – 國立中山大學應用數學系
3. 林珈卉 – 國立陽明交通大學統計學研究所
4. 解振成 – 國立陽明交通大學統計學研究所
5. 石亞喬 – 國立臺北大學統計學系
6. 呂亮葳 – 國立臺北大學統計學研究所

Cracking the Parkinson's Code: Using Chaotic-based Deep Learning and Large-Scale Brain Dynamics for Disease Identification

Yi-Ju Lee[1,2]

[1] Institute of Statistical Science, Academia Sinica

[2] Smart Healthcare Project, Academia Sinica

摘要

With the rise of complexity science and advanced analytics, extracting crucial health-related information from high-dimensional data has become more efficient. The power law scaling in physics has been validated for understanding the complexity of brain signals at temporal scales, which represents their dynamical properties. Deep learning models have shown potential in medical image analysis for detection and classification. In particular, the recent chaotic-based models like ChaosNet simulate the dynamic system of human neuronal firing patterns, representing the closest resemblance to human behavior.

This research integrates and evaluates the cross-scale property of power law in fMRI and ChaosNet model to identify Parkinson's Disease (PD) from healthy adults. Using the Parkinson's Progression Markers Initiative (PPMI) cohort data, we extracted power-law scaling of brain activity from functional images of 100 age and sex-matched PD patients and 100 cognitively normal elderly subjects. These images were transformed into a heatmap for model training, using ChaosNet and DenseNet 121 as model backbones. The models were trained using Python 3.8 for 200 epochs with 1 Nvidia DGX A100 (40G) GPU.

Our results show that the complexity-transformed image data with ChaosNet has significantly decreased training time from 29.1 hours to 0.82 hours, with similar classification results. The best-performing model achieved an average testing accuracy of 89.3 and identified key brain regions related to PD, such as both

Putamen, dorsolateral superior frontal gyrus, and cuneus under Bonferroni correction.

This approach identifies abnormal brain activities in PD patients by using signal complexity as features in chaotic-based models. It requires skills and knowledge in statistical science, computer science, and neurosciences. The altered pathological hemodynamics in PD indicate that the loss of brain signal complexity may play a role in precise clinical diagnosis. The approach denotes the properties of chaotic neuron-induced complex behavior, which is more efficient in comparison with other models. Future work will focus on applying complexity index in cooperating segmentation models to identify the key brain regions for disease classification.

Keywords: Signal Complexity, Chaotic-based Model, Parkinson’s Disease, fMRI

3D 卷積神經網路在大腦核磁共振影像於阿茲海默症分類之應用

江皓瑋
國立中山大學應用數學系

摘要

阿茲海默症 (Alzheimer's disease) 是一種大腦疾病，俗稱老年癡呆，該疾病會導致記憶，思考和行為問題，是一種不正常的老化現象。隨著人口高齡化，患有阿茲海默症的人口逐年上升，台灣失智協會推估民國 120 年失智人口近 46 萬人，屆時每 100 位台灣人有 2 位失智者。

在醫學影像中，結構性核磁共振影像可以很精細的提供腦部各區域的狀況，能呈現具體的多方向切面影像。在現代臨床檢查中，也能夠直接觀察全腦構造，又可偵測特定腦區的形態，是阿茲海默症的檢測基石，故本研究希望建立核磁共振影像的二分法（正常與異常）和三分法（正常、輕度認知障礙與異常）模型，輔助醫師診斷預測是否患有阿茲海默症。

結構性核磁共振影像為 3D 影像，由許多 2D 影像堆疊而成，資料前處理方面對每位受試者做顱骨剝離和配準等影像處理。在資料收集的部分，使用前處理過後的部分 2D 切片做為資料集，可以節省記憶體以及加快模型運算，再對資料集切分為訓練集和測試集，然後使用自建的 3D 卷積神經網路和基於 3D ResNet 架構的神經網路模型進行訓練，訓練時進行資料增強以增加模型泛化的能力。最終以 softmax 函數計算每個類別的機率，再以準確率評估模型的優劣，並使用 Gradient-weighted Class Activation Mapping (Grad-CAM) 熱點圖的方式呈現 3D 卷積神經網路看重患者腦部的區域。

關鍵詞：阿茲海默症、結構性核磁共振影像、3D 卷積神經網路、3D ResNet

Progression Detection of Alzheimer's Disease for Taiwanese Patients via MRI and Hierarchical Classification with Deep Learning

林珈卉
國立陽明交通大學統計學研究所

摘要

Alzheimer's disease (AD) is the most common form disease of dementia, it is not a normal phenomenon of aging, but rather an irreversible and progressive neurodegenerative brain disease. Currently, there is no effective treatment for AD. Therefore, early prediction of the disease is necessary to enable real-time treatment and postpone the progression of AD. Most of the published AD-related papers are derived from the publicly available ADNI database, but most of the data sources are from Western patients. These known development results may not be suitable for Taiwanese patients. Therefore, we hope to develop an Alzheimer's disease progression detection method applicable to Taiwanese patients. We collect magnetic resonance imaging (MRI) scans and clinical data of Taiwanese patients from the Chang Gung Medical system. Taiwanese patients are classified into three categories: cognitively normal (CN), mild cognitive impairment (MCI), and AD. In our research, we adopt a deep learning-based approach and construct a hierarchical classification consisting of multiple binary classification models to classify Taiwanese patients. The hierarchical classification model using MRI data achieves a macro sensitivity of 0.801. In addition, the predictive performance of the physician-assisted clinical diagnosis model further improves the model, achieving a macro sensitivity of 0.874 in model predictions. These research results can assist in making an early diagnosis of diseases in Taiwanese patients, and further help patients receive effective treatment and relieve the progression of AD. This is the joint work with Dr. Jacky Chung-Hao Wu and Prof.
Henry Horng-Shing Lu in National Yang Ming Chiao Tung University, Dr. Min-Yu Lan and Dr. Chiung-Chih Chang in Chang Gung Memorial Hospital.

Keywords: Alzheimer’s disease, convolutional neural network, deep learning, MRI

Deep-learning-based Anomaly Detection and Localization of EEGs in Intensive Care Units

解振成

國立陽明交通大學統計學研究所

摘要

The Intensive Care Unit (ICU) is a specialized hospital unit that cares for critically ill patients who require intensive medical care. The medical staff in the ICU uses a variety of instruments and equipment to continuously monitor patients' vital signs. Patients are admitted to the ICU for a variety of reasons, and analyzing large amounts of monitoring data is a significant challenge, resulting in a heavy workload for medical staff. To alleviate this burden, we are developing an automated anormaly detection and localization model for monitoring data. Our study focuses on electroencephalography (EEG), a real-time monitoring method that records a patient's brainwaves to help diagnose brain-related diseases. Since EEG data is time-sensitive, we are building deep learning models to detect abnormal brainwaves, especially spike and sharp waves. We collect EEG data from 50 ICU patients, 16 of whom have abnormal brainwaves; the abnormal brainwaves annotation includes the occurrence time and electrode position. To obtain a comprehensive understanding of brain activity, we construct a window-based detection and localization model for detecting sets of abnormal brainwaves. Our window-based model enables rapid detection of the presence and localization of abnormal EEGs over short periods of time. To establish the localization model, we use the multi-label method to define the problem and design a post-processing to recover the actual position. The multi-label results in a macro balanced accuracy of 90%, and the post-processing balanced accuracy exceeds 80%. This research has the potential to reduce the workload of ICU and assist medical staff in identifying the position of abnormal brainwaves in patients. This is the joint work with Dr. Jacky Chung-Hao Wu and Prof. Henry Horng-Shing Lu in National Yang Ming Chiao Tung University, Dr. Chieh-Liang Wu and the related researchers in Taichung Veterans General Hospital.

Keywords: EEG Spike Detection, EEG Spike Localization, CNN

非參數回歸、非線性回歸和時間序列模型在台股指數預測表現之比較

石亞喬

國立臺北大學統計學系

摘要

本研究旨在比較非參數回歸模型、非線性回歸模型與時間序列模型在模擬與預測台股指數上的效果。我們使用了 2020 年 1 月到 2023 年 3 月的台股指數數據進行了模擬和預測實驗，對比了多元 GARCH 時間序列模型、非線性回歸模型和 Thin-Plate Spline 非參數回歸模型的準確性。結果顯示，thin plate spline 模型在股票價格模擬和預測方面有著較好的表現，因此能夠為股票市場的投資決策提供更為準確的參考依據。

關鍵詞：GARCH、Thin-Plate Spline

半遞迴無母數 Hammerstein-ARX 模型於網路聲量之應用

呂亮葳、鍾麗英
國立臺北大學統計學研究所

摘要

GME 事件後大眾開始重視網路聲量對股市所帶來的影響，而這些在網路論壇上受到高度關注之股票被命名為迷因股，許多學者也對此進行一連串的研究分析。本研究運用半遞迴無母數 Hammerstein-ARX 模型，並將網路聲量套入此模型中進而去預測股票報酬率，其方法中的非線性子系統之半遞迴無母數估計量滿足漸近不偏性和一致性且模型在實務應用上也取得了不錯的預測效果，且半遞迴的估計方式可節省數據儲存空間。本研究也會將 XGBoost 與 HARX 模型一同納入研究中進行預測與評比，期望找出最佳模型。研究結果顯示使用半遞迴無母數 Hammerstein-ARX 模型且選擇高斯核函數可以得到更準確的預測結果。此外，在大部分的迷因股中，此方法相較於 XGBoost 和 HARX 模型更能夠精確預測報酬率，可提供後續學者在相關領域研究上的參考。

關鍵詞：網路聲量、Hammerstein 系統、核函數估計

Session IV-11

June 30, 13:30 – 14:50

理工一館 A109

◆Chair：

施銘杰 – 國立東華大學應用數學系

◆Speaker:

1. 賴恩語 – 中央研究院統計科學研究所
2. 黃怡甄 – 國立清華大學計算與建模科學研究所
3. 林奕勳 – 國立陽明交通大學統計學研究所
4. 謝譯瑭 – 國立中山大學應用數學系
5. 陳冠憲 – 國立陽明交通大學統計學研究所
6. 黃乾哲 – 國立陽明交通大學數據科學與工程研究所

Illustrating the pleiotropic spectra of cancers using the Taiwan Precision Medicine Initiative

En-Yu Lai[1] and Yen-Tsung Huang[1]
[1]Institute of Statistical Science, Academia Sinica

Abstract

Cancer, as a complex disease, involves intricate mechanisms of genetic polymorphism and following regulations. Some variants play pivotal roles in almost all cancer types since they regulate cell growth and division, but others may be specific to certain cancer types. Understanding the roles of cancer-driver genes in the regulation network that leads to cancer susceptibility can be a great help in establishing cancer etiology.

In this paper, we investigate all cancer types provided by the Taiwan Precision Medicine Initiative (TPMI). We first performed the GWAS analyses to identify the variants associated with certain cancer. Then we applied the composite test of inverse-variance weighting normal-product (CINP) to select the pleiotropic genes that regulate more than one cancer. The proposed gene-level pleiotropy test considers two cancers as the target of the evaluated gene, so we finally applied CINP on 153 combinations among 18 cancers. We further clustered the cancer pairs using the co-regulating pleiotropic genes and analyzed the hierarchical structure of the network. Based on the pleiotropic hierarchy, we also provide functional annotation of pleiotropy levels for each gene. The annotation aims to assist researchers in better understanding the underlying causes of cancer and developing new approaches for prevention and treatment.

Keywords: Pleiotropic gene, Polygenic traits, Composite null, CINP, Cancer etiology

A time dependent approach to detect the change points of Basic Reproduction Number for COVID-19

Yi-Zhen Huang

National Tsing Hua University, Institute of Computational and Modeling Science

Abstract

The coronavirus pandemic (COVID-19) has rapidly evolved into a global catastrophe from late 2019. Every time we think the disease is about to die down, there are always new variant appear, causing the confirmed cases to reach record highs. Such as Alpha (B.1.1.7) in late 2020, Delta (B.1.617.2) in mid-2021, from late 2021 continues to the present, the original BA.1 variant, and several sub-variants of Omicron: BA.2, BA.3, BA.4, BA.5, BQ.1 and BQ.1.1.

We propose a simple method, using the susceptible-infectious-removed (SIR) model to estimate the parameters by minimizing the sum of the squared distances between the actual observations and the simulated outputs. This allows us to calculate the parameters of the SIR model. We then use the change point detection on the basic reproduction number (R_0) and other parameters to identify the points of drastic change with the aim of observing the process of the basic reproduction number variation.

We have applied our method to analyze the emergence time of variant strains in several countries, and gained insights into the outbreak of variant strains in these countries.

In the future, our method can be applied to the emergence of new diseases by detecting change points through the number of confirmed and removed cases, thereby detecting the presence of new variant strains or community outbreaks at an early stage.

Keywords: basic reproduction number, compartmental models, time-dependent SIR model, the Hotelling T_2 control Chart.

Develop a Graphical User Interface System for LVH Disease Classification Using MRI SAX View Cine Images

林奕勳

國立陽明交通大學統計學研究所

Abstract

We developed a deep learning model for diagnosing left ventricular hypertrophy (LVH), hypertrophic cardiomyopathy (HCM), or Fabry disease using cine images from a short-axis (SAX) view. Our model achieved promising results, with an F1 score of 0.846, an accuracy of 0.909 on the Taiwan Veterans General Hospital (TVGH) dataset, and an AUC of 0.914. In addition, We also conducted a single-blind study and external testing using data from Taichung Veterans General Hospital (TCVGH) showed an F1 score of 0.727, an accuracy of 0.806, and an AUC of 0.918. Furthermore, To further refine our model, we collected additional data and developed five classification models, which achieved high accuracy and macro F1 scores, with an F1 score of 0.867 and an accuracy of 0.853 on the TVGH dataset. Finally, as our ultimate goal is to assist doctors in diagnosing patients' diseases, we built a graphical user interface (GUI) for doctors. The system displays the probability predictions of various disease, which can provide doctors with a reference due to possible patient comorbidities. The probability model can also save the predicted results and be integrated into clinical outpatient systems in the future, We hope that our research can contribute to the ongoing efforts to improve the accuracy and reliability of medical diagnoses, ultimately leading to better patient outcomes.

This is the joint work with Mr. Wei-Wen Chen and Prof. Henry Horng-Shing Lu in National Yang Ming Chiao Tung University, Dr. Ling Kuo and the related researchers in Taipei Veterans General Hospital.

Keyword：Left Ventricular Hypertrophy, deep learning, MRI SAX View Cine Images

An automatic detection model for cardiovascular disease based on time-frequency domain features of PCG signals.

Yi-Tang Hsieh[1], Mei-Hui Guo[1], Ming-Chun Yang[2], Yu-Jung Huang[3]

[1]Department of Applied Mathematics National Sun Yat-sen University

[2]Department of Pediatrics, E-DA Hospital, I-Shou University

[3]Department of Electrical Engineering I-Shou University

Abstract

Cardiovascular disease is one of the leading causes of death today. Therefore, early detection and treatment are essential to improve patient survival rates. The stethoscope is a common diagnostic tool that converts heart sounds collected through the stethoscope into a phonocardiogram (PCG). However, even experienced doctors can sometimes make misdiagnoses. Therefore, in the past few decades, it is a popular research topic to develop automatic heart sound classification models by analyzing phonocardiograms and combining machine learning methods. Although some studies have shown good results, yet there are still shortcomings, such as most studies only consider clean data and conduct research under ideal conditions, which doesn't correspond to the clinical situation. In this study, we propose a new method that uses time-frequency domain analysis to extract multiple features and then combines a stacked recurrent neural network model and machine learning methods to improve the accuracy and practicality of heart sound recognition.

We apply the proposed methods to 838 clinical data; the results show that a deep learning model combine with XGBoost obtains the best classification results, achieving an accuracy of 85.26% (sensitivity: 79.22%, specificity: 87.93%) on the test set. Compared with the results of experienced cardiologist, achieving an accuracy of 81.04% (sensitivity: 68.98%, specificity: 98.65%), this shows that our proposed method has higher accuracy and sensitivity. We apply the proposed approach to the famous 2016 PhysioNet/CinC Challenge public dataset to further validate the model's reliability. Deep learning combined with XGBoost achieves an accuracy rate of 95% (sensitivity: 91.81%, specificity: 98.88%).

Keywords: Constant Q Transform, Deep Learning, Deep Scattering, Heart sound classification, PCG

The deep learning technique in prediction of extra touch-up ablation and outcomes in cryoballoon ablation for atrial fibrillation

陳冠憲

國立陽明交通大學統計學研究所

Abstract

Atrial fibrillation is the most common type of arrhythmia in clinical practice and has a high recurrence rate. The current treatment is catheter ablation, which can significantly reduce the recurrence rate compared to drug therapy. However, it can cause some damage to heart tissue and is not suitable for children or certain patients. Cryoballoon ablation is a newer treatment for atrial fibrillation that takes less time and causes less damage to heart tissue than catheter ablation. This makes it a better option for children with smaller hearts. However, a significant limitation of this technique is that there is still a certain rate of recurrence in patients who undergo this procedure, and those who have a recurrence need to undergo another catheter ablation (extra touch up). At present, it is still not possible to predict which patients will experience recurrence after cryoballoon ablation before the surgery. However, some studies have suggested that this may be associated with the size of the patient's pulmonary veins and the angle of their connection to the left atrium. Therefore, we want to use deep learning techniques to establish several classification models with the goal of identifying features that are imperceptible to the human eye. We built a point cloud model and a MeshCNN model using data provided by Taipei Veterans General Hospital, and achieved AUC values of 0.6855 and 0.6089, respectively, on the test dataset. The dataset consisted of 3D meshes of the hearts of 190 patients who underwent cryoballoon ablation. These 3D meshes were constructed from 2D CT images. We have built a foundation in this area and will continue to optimize these models. Our goal is to develop a deep learning model that is reliable enough to assist doctors in determining whether a patient is suitable for cryoballoon ablation before surgery.

This is the joint work with Mr. Wei-Wen Chen and Prof. Henry Horng-Shing Lu in

National Yang Ming Chiao Tung University, Dr. Chih-Min Liu and the related medical experts in Taipei Veterans General Hospital.

Keyword：Cryoballoon ablation, Deep learning, Extra touch-up ablation, Pointnet, MeshCNN

Predicting Multi-Vessel Diseases by Polar Maps of SPECT MPI and Deep Learning

Chien-Che Huang[1], Jui-Jen Chen[2], Chang, Yen-Hsiang[2], Henry Horng-Shing Lu[1,3]

[1] Institute of Data Science and Engineering, National Yang Ming Chiao Tung University

[2] Department of Nuclear Medicine, Kaohsiung Chang Gung Memorial Hospital

[3] Institute of Statistics, National Yang Ming Chiao Tung University

Abstract

This study aimed to explore the feasibility of using deep learning methods for coronary artery disease (CAD) assessment. Specifically, we focused on assessing the severity of CAD based on SPECT MPI using Efficient-Net and Deep SMOTE methods and classifying CAD impairment into Multi-Vessel Disease (MVD) and Single-Vessel Disease (SVD). Our study utilized a dataset of 158 patients, consisting of 105 MVD and 53 SVD patients We use the images of SPECT MPI from Department of Nuclear Medicine, Kaohsiung Chang Gung Memorial Hospital (KCGMH) between 2019 and 2021.

The effectiveness of our method was validated through experiments, and our model achieved an accuracy of 84.38% and an area under the curve (AUC) of 0.9221. This research provides a new and effective approach to the early prediction and diagnosis of MVD, which has the potential to improve patient outcomes and reduce healthcare costs. Our findings demonstrate the potential value of using deep learning methods for CAD prediction and diagnosis and suggest that deep SMOTE is an effective technique for augmenting sparse medical image data.

Keywords: Deep Learning (DL), coronary artery disease (CAD), SPECT MPI, Multi-Vessel Disease (MVD)

分組議程 SV 時間表/摘要

分組議程 SV : 2023 年 6 月 30 日 (星期五) 15:00 - 16:20

SV-1	**空間統計**	**Chair：張雅梅**	**理 A310**	**p.305**

林培生 - 國家衛生研究院

Disease Mapping for Correlated Zero-inflated Data

楊洪鼎 - 國立高雄大學統計學研究所

Variation-based model average prediction for spatial data with nonstationary correlation structure

張雅梅 - 淡江大學統計學系

利用相似度指標於有無地圖上偵測物種關聯性

SV-2	**High-dimensional data analysis**	**Chair：黃灝勻**	**理 B201**	**p.309**

彭柏翔 - 國立清華大學統計學研究所

Feature Selection for High-Dimensional Heteroscedastic Regression and Its Applications

黃學涵

Sparse Matrix Estimation Based on Greedy Algorithms and Information Criteria

賴驥緯 - 國立清華大學統計學研究所

A Basis Function Approach for Finding Sparse Dominant Patterns in Climate Data

SV-3	**工業統計及應用**	**Chair：許湘伶**	**理 A212**	**p.313**

張明中 - 中央研究院統計科學研究所

Supervised Stratified Subsampling for Big Data Prediction

李宜真 - 國立成功大學統計學系

Statistical analysis for fatigue tests

林長鋆 - 國立中興大學統計學研究所

Generalized Bayesian D-optimal supersaturated multistratum designs

SV-4	**財務金融**	**Chair：蕭維政**	**理 D241**	**p.317**

鄭宏文 - 東吳大學財務工程與精算數學系

透過機器學習探討波動度差與股票報酬之預測

陳虹吟 - 淡江大學會計學系

Non-Parametric Inference on Risk Measures for Integrated Returns

匡顯吉 - 國立政治大學金融學系

Good Jump and Bad Jump Variance Risk Premium: Evidence from S&P500 Returns and Options

分組議程 SV : 2023 年 6 月 30 日 (星期五) 15:00 - 16:20

SV-5	**Causal Inference**	**Chair：林聖軒**	**理 A209**	**p.321**

楊子霆 - 中央研究院經濟研究所

因果推論在政策分析上的應用

林聖軒 - 國立陽明交通大學統計學研究所

From linear structural equation modeling to generalized multiple mediation formula

溫啟仲 - 淡江大學數學學系

Is“the seven year itch” real? ("七年之癢"是真的嗎?)

SV-6	**財務工程**	**Chair：孫立憲**	**理 A211**	**p.325**

王俞凱 - 國立中央大學統計研究所

Quantile three-factor model with heteroskedasticity, skewness, and leptokurtosis

吳牧恩 - 國立臺北科技大學資訊與財金管理系

PASS: Portfolio Analysis of Selecting Strategies on Quantitative Trading via Multi-Objective Genetic Algorithm

孫立憲 - 國立中央大學統計研究所

Online Changepoint Detection under Copula-Based Markov Chain Models for Gaussian Sequential Data

SV-7	**臨床試驗之統計方法與應用**	**Chair：林資荃**	**理 A318**	**p.330**

林資荃 - 財團法人醫藥品查驗中心

Statistical considerations in clinical trials

陳瑱芳 - 財團法人醫藥品查驗中心

調整設計應用於臨床試驗之統計考量

林奕廷 - 艾昆緯股份有限公司

資料分析在臨床試驗之實例分享

黃智揚 - 艾昆緯股份有限公司

Statistical methods and applications in clinical trials: Multiple Endpoints

SV-8	**統計方法**	**Chair：張志浩**	**理 A314**	**p.335**

顏佐榕 - 中央研究院統計科學研究所

Link Prediction via Exploring Common Neighborhoods

朱基祥 - 東海大學統計學系

Bayesian structure selection approaches for multiple binary responses via multi-task learning

張志浩 - 國立高雄大學統計學研究所

Estimation of Threshold Boundary Regression Models

分組議程 SV : 2023 年 6 月 30 日 (星期五) 15:00 - 16:20

SV-9	Contributed Session - L	Chair：曾玉玲	理 A316	p.339

鄭竣元 - 中原大學

Friedman 檢定的新近似法

林建璋 - 國立陽明交通大學管理科學系

雙組等效檢定的檢討

王律涵 - 國立臺北大學統計學系

雙偏斜常態之 ROC 曲線的統計性質探討

陳柏豪 - 國立臺北大學統計學系

雙偏斜常態母體之類別資料的 AUC 估計探討

林君儴 - 國立東華大學應用數學系

Steam 遊戲視覺化

林俊源 - 國立成功大學統計學系

RUL Prediction of Lithium-ion Batteries under Ambient Temperature Based on Piecewise Linear Random Coefficient Degradation Models

SV-10	Contributed Session - M	Chair：陳忠和	理 A210	p.346

陳忠和 - 南臺科技大學工業管理與資訊系

Economic Manufacturing Quantity Model with Quality Investment

范姜翔星 - 國立臺北大學統計學系

隱含波動率應用於台指選擇權配對交易之實證研究

劉佳穎 - 國立臺北大學統計學系

臺灣股市熊市之預測

洪煒傑 - 國立高雄大學統計學研究所

正交化資產對風險平價投資組合的影響與應用

李曉妮 - 國立高雄大學統計學研究所

重要風險因子及標的資產的選取及其投資組合的表現

温筑涵 - 國立臺北大學統計學系

社群媒體活動及網路聲量對股票價格與波動的影響

分組議程 SV : 2023 年 6 月 30 日 (星期五) 15:00 - 16:20

SV-11 **Contributed Session - N** **Chair：曹振海** **理 A106** **p.353**

呂一昕 - 國立東華大學應用數學系

爵士音樂家資料視覺化

陳君毅 - 國立臺北大學統計學系

卷積多實例學習方法應用於影像特徵提取

蔡明憲 - 國立臺北大學統計學系

基於深度學習的車牌辨識系統的對抗性攻擊:以生成對抗網絡為例

管彥鳴 - 國立陽明交通大學多媒體工程研究所

Detecting the Local Advanced Rectal Cancer from CT Images by Deep Learning Methods

何郁涵 - 國立高雄大學統計學研究所

High-dimensional clustering via features segmentation and fusion

何婉瑜 - 國立高雄大學統計學研究所

Dimension Reduction for High-Dimensional Regression by Iteratively Subgrouping Selection Procedure

SV-12 **Contributed Session - O** **Chair：蘇南誠** **理 A109** **p.360**

廖靖芸 - 國立政治大學統計學系

由《聯合報》及《人民日報》報導風格看兩岸差異

陳慧霜 - 國立政治大學統計學系

影像分析與深偽影片的偵測 A Statistical Approach of Deepfake Video Detection

施承翰 - 國立東華大學應用數學系

BGG 桌遊圖

黃文顥 - 國立清華大學

使用隨機配置方法處理大數據中不完整地址資料 - 以新北市為例

薛皓澤 - 國立高雄大學統計學研究所

電子零件參數辨識系統

黃凱琪 - 國立臺北大學統計學系

運用電信資料探討臺鐵車站人流量

Session V-1

June 30, 15:00 – 16:20

理工一館 A310

◆Organizer/Chair：

張雅梅 – 淡江大學統計學系

◆Speaker:

1. 林培生 – 國家衛生研究院
2. 楊洪鼎 – 國立高雄大學統計學研究所
3. 張雅梅 – 淡江大學統計學系

Disease Mapping for Correlated Zero-inflated Data

林培生

Institute of Population Health Sciences, National Health Research Institutes

Department of Mathematics, National Chung Cheng University

Abstract

Disease mapping is a research field to estimate spatial pattern of disease risks so that areas with elevated risk levels can be identified. The motivation of this paper is from a study of dengue fever infection, which causes seasonal epidemics in almost every summer in Taiwan. For analysis of zero-inflated data with spatial correlation and covariates, current methods would either cause a computational burden or miss associations between zero and non-zero responses. In this paper, we develop estimating equations for a mixture regression model that accommodates spatial dependence and zero inflation for study of disease propagation. Asymptotic properties for the proposed estimates are established. A simulation study is conducted to evaluate performance of the mixture estimating equations; and a dengue dataset from southern Taiwan is used to illustrate the proposed method.

Keywords: disease mapping model; estimating equations; mixture density function; spatial data; zero inflation.

Variation-based model average prediction for spatial data with nonstationary correlation structure

Chih-Hao Chang, Hsiang-Ling Hsu, Ying-Pin Zhuang, and Hong-Ding Yang*

Institute of Statistics, National University of Kaohsiung

Abstract

When analyzing spatially dependent data with noise, predicting the spatial variables of interest is essential. Many spatial prediction methods have been developed, each with its applicable occasion. Therefore, under the unknown mechanism of generating spatial variables of interest, there will be uncertainty in the choice of prediction by a single method. On the other hand, the experimenter will conduct data analysis under the assumption of stationary spatial correlation. However, the spatial correlation of data often does not satisfy stationary, and inferences under inappropriate assumptions may lead to unreasonable conclusions. Therefore, we propose three model average methods for spatial predicting based on the thin-plate smoothing spline and the spatially deformed Kriging predictions. For each model average prediction, the corresponding weight of each candidate prediction will be determined by the prediction variance meter of the candidate prediction methods and has the characteristics of fast computation and local adjustment of coordinates for prediction. After deciding on the final prediction method based on the simulation results, we analyzed Taiwan Agricultural Research Institute's heavy metal content in the soil. We estimated the concentration of Cadmium (Cd) at each monitoring station. Anticipation can improve forecast performance by avoiding uncertainty caused by choosing a forecasting method.

Keywords: Spatial prediction, Non-stationary covariogram, Prediction uncertainty, Prediction variation, Model averaging.

利用相似度指標於有無地圖上偵測物種關聯性

張雅梅、謝佩穎

淡江大學統計學系

摘要

本篇研究主要利用相似度指標於有無地圖上偵測物種關聯性，我們使用了Sorensen、BAM (best attainable match)、Morisita 與 Horn 指標，前兩種指標可直接應用於有無地圖上，後兩種指標則利用核密度估計法進行調整，使其適用於有無地圖上。於模擬研究中，在獨立的假設下，隨機生成兩個物種，建立各個指標的經驗分配，透過雙尾檢定去判別物種關聯性，並觀察各個指標的型一誤差及檢定力。從模擬的結果得知，出現率較低的資料中，指標較無法正確判別其關聯性；有用核密度估計修正後的指標容易把兩個具有關聯性的物種，判斷為不具關聯性。最後將本篇所提及的相似度應用於百靈鳥及甲蟎資料上，指標大多可以發現資料間是具有關聯性的。

關鍵詞：Sorensen、BAM、Morisita、Horn、核密度估計

Session V-2

June 30, 15:00 – 16:20

理工一館 B201

◆Organizer/Chair：

黃灝勻 – 國立清華大學統計學研究所

◆Speaker:

1. 彭柏翔 – 國立清華大學統計學研究所
2. 黃學涵 – 國立清華大學
3. 賴驥緯 – 國立清華大學統計學研究所

Feature Selection for High-Dimensional Heteroscedastic Regression and Its Applications

Po-Hsiang Peng[1], Hai-Tang Chiou[2*], Hsueh-Han Huang[3], and Ching-Kang Ing[1]

[1]Institute of Statistics, National Tsing Hua University

[2]Department of Mathematics, National Chung Cheng University

[3]Institute of Statistical Science, Academia Sinica

Abstract

We consider feature selection for high-dimensional linear heteroscedastic models. Inspired by the connection between the linear heteroscedastic function and the interaction model, we design a two-stage algorithm to choose the relevant features in the functional form of heteroscedasticity is linear or multiplicative, we provide a data-driven method to select between the two alternatives. We prove selection consistency of the proposed method and illustrate its performance via numerical simulations. We further apply the method to identify defective tools in the semiconductor manufacturing process.

Keywords: High-dimensional interaction model, Linear heteroscedasticity, Multiplicative heteroscedasticity

Sparse Matrix Estimation Based on Greedy Algorithms and Information Criteria

Hsueh-Han Huang, Ching-Kang Ing[1], Ruey S. Tsay[2]

[1] National Tsing Hua University

[2]University of Chicago

Abstract

We consider the problem of estimating the covariance matrix of serially correlated vectors whose dimension is allowed to be much larger than the sample size. We propose using the orthogonal greedy algorithm (OGA) together with a high-dimensional Akaike's information criterion (HDAIC) to estimate the matrix, and show that the proposed estimate is rate optimal under a sparsity condition more flexible than those in the existing literature. When the covariance matrix is bandable, we introduce a banding/tapering estimate whose parameters are chosen by a novel information criterion. The rate optimality of the latter estimate is also established.

Keywords: Banding/Tapering estimates, High-dimensional sparse covariance matrices, High-dimensional Akaike's information criterion, Orthogonal greedy algorithms, Optimal rates

A Basis Function Approach for Finding Sparse Dominant Patterns in Climate Data

Chi-Wei Lai[1], Hsin-Cheng Huang[2]

[1]Institute of Statistics, National Tsing Hua University

[2]Institute of Statistical Science, Academia Sinica

Abstract

Empirical orthogonal function (EOF) analysis is commonly used to identify dominant modes of variability in climate data. However, interpreting EOFs can be challenging as they may not always be physically meaningful. In this research, we propose a novel approach that uses basis functions to represent the spatial processes and the corresponding spatial covariance function. By searching for sparse patterns in the eigenspace formed by linear combinations of the first few EOFs, our method provides a more interpretable approach for studying localized interactions in climate data. We achieve sparsity by adapting a sparse principal component analysis (PCA) approach incorporating basis functions and utilizing rotation of the first few eigenfunctions, where sparsity is achieved by using an $\ell 1$-norm penalty. Our approach offers a flexible framework for analyzing high-dimensional spatial data that can be used in various scientific applications.

Session V-3

June 30, 15:00 – 16:20

理工一館 A212

◆Organizer/Chair：

許湘伶 – 國立高雄大學統計學研究所

◆Speaker:

1. 張明中 – 中央研究院統計科學研究所
2. 李宜真 – 國立成功大學統計學系
3. 林長鋆 – 國立中興大學統計學研究所

Supervised Stratified Subsampling for Big Data Prediction

Ming-Chung Chang

Institute of Statistical Science, Academia Sinica

Abstract

Extraordinary amounts of data are generated across disciplines owing to advanced technology. Such data richness, however, may yield difficulties in statistical model fitting and prediction either in terms of time cost or numerical stability. On the other side, stratified sampling has been used to control the homogeneity of data in the literature. In this talk, I will introduce a new subsampling approach, referred to as supervised stratified subsampling, which integrates nonparametric regression and stratified sampling for big data prediction. Theoretical and numerical results are provided to show the benefits of the proposed method.

Keywords: Large-scale data, Partitioning estimate, Bagging

Statistical analysis for fatigue tests

李宜真

國立成功大學統計學系

ABSTRACT

To ensure the product can last long enough in the field, the accelerated life test (ALT) is a useful experiment to obtain the material lifetime information and then make prediction at the normal use condition for the field performance. In the literature, the relationship between stress and lifetime is usually set to be a linear model. However, due to the physical-based information, some applications show that the stress-life relationship is sometimes set to be a nonlinear function, even a partial differential equation without closed form. Thus, to construct a procedure to analyze the data from an ALT with the physical-based stress-life model and probability models is a challenge research problem. This study uses the metal fatigue test as the motivating example with a physical-based model, Manson-Coffin-Basquin (MCB) curve. Assume the fatigue lifetime follows a distribution of the log-location-scale family, we construct an efficient estimation method for the fatigue data with the MCB relationship. We also compare the proposed estimation method with existing estimation method on performance of estimation.

Keywords: Fatigue test, Manson-Coffin-Basquin curve.

Generalized Bayesian D-optimal supersaturated multistratum designs

Chang-Yun Lin

Department of Applied Mathematics and Institute of Statistics, National Chung Hsing University

ABSTRACT

Supersaturated designs are useful in the initial stage of experiments to identify important factors from many of interest with a small number of runs. Traditional supersaturated designs were mainly constructed for completely randomized experiments, which have single-stratum structures. They cannot be used for experiments that have multistratum structures, such as the split-plot, strip-plot, and staggered-level experiments. How to construct supersaturated multistratum designs for complex experiments has gained much attention recently. In this paper, we consider the situation in which the experimenters have prior knowledge of which factors are more likely to be important (called the primary factors) than the others (called the potential factors). By taking primary and potential factors into account, we propose an approach using the generalized Bayesian D (GBD) criterion to con- struct a new class of supersaturated multistratum designs. The GBD-optimal supersaturated multistratum designs provide guidelines on how to assign factors to the designs, which enhances efficiency on identifying active factors. A case study shows that the proposed supersaturated design (32 runs with 19 factors) is as effective as the full 26 factorial design (64 runs with 6 factors) to identify important factors in a battery cell experiment.

Keywords: potential terms; primary terms; split-plot; staggered-level; strip-plot

Session V-4

June 30, 15:00 – 16:20

理工一館 D241

◆Organizer/Chair：

蕭維政 – 東吳大學財務工程與精算數學系

◆Speaker:

1. 鄭宏文 – 東吳大學財務工程與精算數學系
2. 陳虹吟 – 淡江大學會計學系
3. 匡顯吉 – 國立政治大學金融學系

透過機器學習探討波動度差與股票報酬之預測

陳韻安、鄭宏文*

東吳大學財務工程與精算數學系

摘要

本研究探討傳統的線性迴歸與機器學習模型在股票報酬的預測上，誰的表現較好。首先驗證波動度變量的可預測性。透過投資組合分析及 Fama-MacBeth 迴歸分析發現實質波動度、隱含波動度（買權、賣權）及歷史-隱含波動差都與預期報酬存在顯著負相關，同時預期報酬與買-賣權隱含波動度差之間存在顯著正相關。包含這些波動度變量對機器學習技術進行實證分析也分析波動度預測因子對預期報酬的重要性，並追踪波動度預測因子和預期報酬之間的邊際關係。最後，根據每種方法的預測報酬構建投資組合，其中以迴歸樹及類神經網絡模型表現最佳。研究結果顯示：機器學習比傳統的預測技術更好地描述預期報酬的行為，並支持機器學習在創新金融科技中日益增長的角色。

關鍵詞：實質波動度、隱含波動度、波動度差、股票預測、機器學習。

Non-Parametric Inference on Risk Measures for Integrated Returns

Henghsiu Tsai[1], Hwai-Chung Ho[1] and Hung-Yin Chen[2]

[1] Institute of Statistical Science, Academia Sinica

[2] Department of Accounting, Chung Yuan Christian University

Abstract

When evaluating the market risk of long-horizon equity returns, it is always difficult to provide a statistically sound solution due to the limitation of the sample size. To solve the problem for the value-at-risk (VaR) and the conditional tail expectation (CTE), Ho et al. (2016, 2018) introduce a general multivariate stochastic volatility return model from which asymptotic formulas for the VaR and the CTE are derived for integrated returns with the length of integration increasing to infinity. Based on the formulas, simple non-parametric estimators for the two popular risk measures of the long-horizon returns are constructed. The estimates are easy to implement and shown to be consistent and asymptotically normal. In this chapter, we further address the issue of testing the equality of the CTEs of integrated returns. Extensive finite-sample analysis and real data analysis are conducted to demonstrate the efficiency of the test statistics we propose.

Keywords: Conditional tail expectation, Equality of tail risks, Inference, Integrated process, Quantile, Stochastic volatility model, Test, Value at risk

Good Jump and Bad Jump Variance Risk Premium: Evidence from S&P500 Returns and Options

Hsing-Hua Chang[1], Hung-Wen Cheng[2], Hsiang-Ting Huang[1], Xian-Ji Kuang*[1], and Shih-Kuei Lin[1]

[1]Department of Money and Banking, National Chengchi University

[2]Department of Financial Engineering and Actuarial Mathematics, Soochow University

Abstract

In the field of asset pricing, how to measure the relationship between the expected return and volatility of an asset is a crucial topic. In this paper, we extend the asym- metric double exponential distribution jump model under the affine GARCH model proposed by Yang (2018), referencing the return dynamics process designed by Kilic and Shaliastovich (2019). We attempt to propose a model under the affine GARCH model that characterizes good and bad jumps using two exponential distributions, and derive the closed-form solution for option pricing under this setting. In the empirical study, through model estimation, we can calculate the variance risk premiums from good and bad jumps. Finally, we verify the relationship of these two variance risk premiums to returns through cross-sectional regression and time series analysis.

Keywords: Variance risk premium, Good jump and bad jump, Option pricing, Cross- sectional regression, Time series analysis

Session V-5

June 30, 15:00 – 16:20

理工一館 A209

◆Organizer/Chair：

林聖軒 – 國立陽明交通大學統計學研究所

◆Speaker:

1. 楊子霆 – 中央研究院經濟研究所
2. 林聖軒 – 國立陽明交通大學統計學研究所
3. 溫啟仲 – 淡江大學數學學系

因果推論在政策分析上的應用

楊子霆

中央研究院經濟研究所/國立政治大學應用經濟學程

摘要

在這場演講中，我將利用幾個自己正在進行的研究，簡介因果推論方法在經濟學與政策分析上的應用。

關鍵詞：因果推論

From linear structural equation modeling to generalized multiple mediation formula

林聖軒

國立陽明交通大學統計學研究所

Abstract

Causal mediation analysis is advantageous for mechanism investigation. In settings with multiple causally ordered mediators, path-specific effects (PSEs) have been introduced to specify the effects of certain combinations of mediators. However, most PSEs are unidentifiable. Interventional analogue of PSE (iPSE) is adapted to address the non-identifiability problem. Moreover, previous studies only focused on cases with two or three mediators due to the complexity of the mediation formula in large number of mediators. In this study, we provide a generalized definition of traditional PSEs and iPSEs with a recursive formula, along with the required assumptions for nonparametric identification. This work has three major contributions: First, we developed a general approach (that includes notation, definitions, and estimation methods) for causal mediation analysis with an arbitrary number of multiple ordered mediators and with time-varying confounders. Second, we demonstrate identified formula of iPSE is a general form of previous mediation analysis. It is reduced to linear structural equation model under linear or log-linear model, to causal mediation formula when only one mediator. Third, a flexible algorithm built based on g-computation algorithm is proposed along with a user-friendly software online. This approach is applied to a Taiwanese cohort study for exploring the mechanism by which hepatitis C virus infection affects mortality through hepatitis B virus infection, abnormal liver function, and hepatocellular carcinoma. All methods and software proposed in this study contribute to comprehensively decompose a causal effect confirmed by data science and help disentangling causal mechanisms when multiple ordered mediators exist, which make the natural pathways complicated.

Keywords: causal inference, mediation analysis, multiple mediation, SEM, counterfactual, interventional approach.

Is“the seven year itch”real? ("七年之癢"是真的嗎?)

溫啟仲*[1], 程毅豪 [2]

[1]淡江大學數學學系

[2]中央研究院統計科學研究所

Abstract

The randomized response techniques (RRTs), including the related-question RRT of Warner (1965) and the unrelated-question RRT of Greenberg et al. (1969), have been utilized to reduce under-reporting of sensitive characteristics in survey studies by enhancing privacy protection. Currently, the RRT is mainly applied for prevalence estimation of some sensitive event. This work extends the application of the RRT to the analysis of time-to-event outcome. We apply the proposed method to the Taiwan extramarital sex data surveyed by the RRT to make statistical inferences on the time to the incidence of extramarital sex since marriage. In final, we discuss the authenticity of “the seven year itch” based on the Taiwan survey data.

Keywords: Randomized response technique, Semiparametric maximum likelihood estimation, Sensitive issue, Semiparametric transformation model.

Session V-6

June 30, 15:00 – 16:20

理工一館 A211

◆Organizer/Chair：

孫立憲 – 國立中央大學統計研究所

◆Speaker:

1. 王俞凱 – 國立中央大學統計研究所
2. 吳牧恩 – 國立臺北科技大學資訊與財金管理系
3. 孫立憲 – 國立中央大學統計研究所

Quantile three-factor model with heteroskedasticity, skewness, and leptokurtosis

Kai Y.K. Wang [1*], Cathy W.S. Chen [1], Mike K.P. So [2]

[1] Department of Statistics, Feng Chia University

[2] Department of Information Systems, Business Statistics and Operations Management, The Hong Kong University of Science and Technology

Abstract

The Fama-French three-factor model advances the capital asset pricing model by expanding size risk and value risk factors to market risk factors. A quantile Fama-French three-factor model with GARCH-type dynamics, leptokurtosis, and skewness via asymmetric Student t errors is proposed to overcome the limitations of the existing ones. One can investigate how the daily volatility and market risk factors act under different market conditions represented by quantile levels via the proposed model. Bayesian adaptive Markov chain Monte Carlo methods are used to estimate model parameters in the proposed model over various quantile levels. This study assesses the Bayesian inference performance via simulation studies in which the designated models are misspecified and considers some daily stock returns from NASDAQ to help further select the best model via the posterior odds ratio. It is clear that the various market conditions and the GARCH effect should be incorporated into the model. Findings show that the estimation of the size factor turns insignificant for lower quantiles - i.e., when the market is in a panic, investors ignore the size effect of a company's assets.

Keywords: Markov chain Monte Carlo methods, GARCH, Posterior odds ratio, Quantile regression, Asymmetric Laplace distribution, Asymmetric student t distribution

PASS: Portfolio Analysis of Selecting Strategies on Quantitative Trading via Multi-Objective Genetic Algorithm

Mu-En Wu[1], Da-Wei Chiang[2]

[1]Department of Information and finance management, National Taipei University of Technology

[2] Department of Information and finance management, National Taipei University of Technology

Abstract

Quantitative trading has become increasingly popular in recent years with the advancement of information technology. Numerous researchers have utilized machine learning or traditional mathematical formulas to construct indicators and signals for trading strategies. While the performance of a strategy can be evaluated through backtesting using historical data, ensuring that the strategy is funded adequately or maintaining a stable and profitable strategy to survive independently is critical. Risk management is a critical issue in achieving this goal, and combining multiple strategies to maintain a stable profit is one suitable method. Thus, developing an efficient search algorithm to combine strategies into a portfolio is another critical issue, and evolutionary computation, particularly multiobjective evolutionary algorithms (MOEAs), has high-quality solutions for multiobjective search problems. This paper integrates the need to maintain stable and profitable trading strategies with the development of a search algorithm to create a portfolio with two main contributions. The first contribution is a measurement indicator called "Drawdown Duration" (DDD), which reports the stability of strategy combinations. The second contribution is the Portfolio Analysis of Selecting Strategies (PASS) model's proposal to construct strategy combinations to maintain stable and profitable trading strategies. The results of our experiments indicate that the PASS model reduces DDD risk by 73.9\% and increases profit by 4.1\% compared to other MOEAs. Lastly, we applied the PASS model to construct a profitable strategy in other trading markets, including NYMEX, WTI, and Min YM. Overall, our study can

support investors in developing profitable trading strategies and managing risk in highly-leveraged trades.

Keywords: Quantitative trading, Efficient search algorithm, Multi-objective optimization, Drawdown.

Online Changepoint Detection under Copula-Based Markov Chain Models for Gaussian Sequential Data

Li-Hsien Sun[1], Dong-Hwa Kuo[1], Ming-Hua Hsieh[2]

[1]Graduate Institute of Statistics, National Central University

[2] Department of Risk Management and Insurance, National Chengchi University

Abstract

Online changepoint detection is a procedure to identify whether a sequential data structure changes over time. In practice, the dependence structure is an important issue for time series analysis. To achieve flexibility for the dependence structure, we propose a copula-based Markov chain model based on the Clayton copula and the marginal distribution being a Gaussian distribution. In order to relax the conjugate structure based on the independence structure, we consider the Markov chain Monte Carlo (MCMC) method to detect the changepoints online. The simulation results indicate that the proposed model outperforms the competitors in the precision and mean absolute error (MAE) in various scenarios. For empirical studies, we consider the daily log returns of the S\&P 500 Index, the Nikkei 225 Index, and the FTSE 100 Index to identify the changepoints in the period of the financial crisis in 2008 and the COVID-19 pandemic in 2020. The corresponding results reveal that the proposed model is able to capture the structure change in real application.

Keywords: Online changepoint detection, copula, Bayesian inference, Markov model

Session V-7

June 30, 15:00 – 16:20

理工一館 A318

◆Organizer/Chair：

林資荃 – 財團法人醫藥品查驗中心

◆Speaker:

1. 林資荃 – 財團法人醫藥品查驗中心
2. 陳瑱芳 – 財團法人醫藥品查驗中心
3. 林奕廷 – 艾昆緯股份有限公司
4. 黃智揚 – 艾昆緯股份有限公司

Statistical considerations in clinical trials

Tzy-Chy Lin

Division of New Drugs, Center for Drug Evaluation

Abstract

Clinical trials are mainly divided into three phases: Phase I, Phase II and Phase III. Each Phase has its main purpose. For example, the primary objectives of a phase I trial are to provide information on the safety, tolerability, pharmacokinetics (PK) and pharmacodynamics (PD) of an investigational medicinal product (IMP). The primary objective of a phase II trial is to determine the dose(s) and regimen for Phase III trials. Trials in Phase II are typically conducted in a group of patients who are selected by relatively narrow criteria, leading to a relatively homogeneous population and are closely monitored. Trials in Phase III are designed to confirm the efficacy and safety for the intended indication and recipient population. In this talk, we provide the statistical considerations in Phase I, Phase II and Phase III trials. Finally, some examples are illustrated to understand the trial design in different phases.

Keywords: Clinical trials, Phase I trials, Phase II trials, Phase III trials

調整設計應用於臨床試驗之統計考量

陳瑱芳

財團法人醫藥品查驗中心

摘要

隨機臨床試驗是臨床研究的標準，然而隨機臨床試驗往往需要大量的樣本數和較長的研究時間，缺乏評估重要次群眾療效的檢定力以及成本考量等缺點。為彌補傳統試驗設計的不足，近年來靈活的調整設計（adaptive design）快速發展，它允許試驗執行過程中，根據試驗中受試者累積的資料，前瞻地規劃去修改試驗設計，提升新藥試驗成功的機率。美國食品藥品管理局（US FDA）於 2010 年發佈第一版調整設計指引草案後，於 2018 年 9 月再次發佈了第二版調整設計指引草案「Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry」。 本主題將針對調整性臨床試驗，介紹試驗設計、執行與分析的關鍵性原則及統計考量。

關鍵詞：臨床試驗、調整設計、統計考量

資料分析在臨床試驗之實例分享

林奕廷

艾昆緯股份有限公司生物統計分析

摘要

新型冠狀病毒（簡稱新冠病毒，SARS-CoV-2）自 2019 年 11 月從中國武漢爆發以降，疫苗與藥物的開發逐漸受到一般大眾的注視，亦是國內外高度關注的議題。其中臨床試驗係此等疫苗與藥物供大眾使用前不可或缺的步驟，為相關療效與安全性建立實證基礎。本次將根據在臨床試驗受託研究機構（Contract research organization，簡稱 CRO）之實務經驗，分享臨床試驗從原始數據之收集到圖表之產出等過程的資料流，以及如何運用臨床數據交換標準聯盟（Clinical Data Interchange Standards Consortium，簡稱 CDISC）之標準，來進行臨床試驗資料之標準化，讓法規單位利於進行數據審閱。本次分享將不會揭露任何真實數據。

關鍵詞：臨床試驗、受託研究機構、臨床數據交換標準聯盟

Statistical methods and applications in clinical trials: Multiple Endpoints

Johnson Huang[1]

[1]IQVIA

Abstract

Clinical trials are increasingly emphasizing multiple endpoint outcomes. Compared to assessing a single endpoint, the use of multiple endpoint outcomes can more comprehensively reflect treatment effects, reducing research costs and time for patients and investigators. Multiple endpoint outcomes can include various indicators, ranging from disease symptoms to survival rates, which can be used in combination to reflect treatment effects on different aspects. When designing trials with multiple endpoint outcomes, careful consideration is needed regarding endpoint interaction, appropriate statistical methods, and suitable sample sizes and experimental designs.

To help control the issue of multiplicity brought by multiple endpoint outcomes in clinical trials, the FDA has released the "Guidance for Clinical Trials with Multiple Endpoints". This guidance provides systematic recommendations and guidance on how to select, evaluate, and analyze multiple endpoints.

This section aims to share with you the statistical analysis methods that can be used to control Type I errors in clinical trials and the strategies that can be used to establish endpoint indicators during the planning and trial design stages.

Keywords: Type I errors, multiple endpoints, multiple comparisons

Session V-8

June 30, 15:00 – 16:20

理工一館 A314

◆Organizer/Chair：

張志浩 – 國立高雄大學統計學研究所

◆Speaker:

1. 顏佐榕 – 中央研究院統計科學研究所
2. 朱基祥 – 東海大學統計學系
3. 張志浩 – 國立高雄大學統計學研究所

Link Prediction via Exploring Common Neighborhoods

Tso-Jung Yen [1], Wei-Chung Liu[1]

[1]Institute of Statistical Science, Academia Sinica

Abstract

Social network analysis aims to establish properties of a network by exploring link structure of the network. However, due to concerns such as confidential and privacy, a social network may not provide full information on its links. As some of the links are missing, it is difficult to establish the network's properties by exploring its link structure. In this paper we propose a method for recovering such missing links. We pay attention on a situation in which some nodes have fully-observed links. The method relies on exploiting the network of these “anchor” nodes to recover missing links of nodes that have neighborhoods overlapping with the anchor nodes. It uses a graph neural network to extract information from these neighborhoods, and then applies the information to a regression model for missing link recovery. We demonstrated this method on real-world social network data. The results show that this method can achieve better performance than traditional methods that are solely based on node attributes for missing link recovery.

Keywords: Graph neural networks, Link prediction, Missing data, Social network analysis

Bayesian structure selection approaches for multiple binary responses via multi-task learning

Chi-Hsiang Chu[1], Ray-Bing Chen[2]

[1] Department of Statistics, Tunghai University

[2] Department of Statistics, National Cheng Kung University

Abstract

In this work, we concentrate on the Bayesian structure selection problems for the categorical response. We focus on solving the selection problem for multiple binary responses and use the probit model for each response. Here, we consider the group structure with sparsity property on the rows of coefficient matrix where each row corresponds to one variable. Then we identify the relevant variables for the responses and the selection problems can be treated as the multi-task learning problem. The effectiveness of our proposed method will be demonstrated through simulation studies.

Keywords: Multi-task learning, component-wise algorithm, group structure, Markov chain Monte Carlo

Estimation of Threshold Boundary Regression Models

Chih-Hao Chang [1], Takeshi Emura[2], Shih-Feng Huang[3]

[1]National University of Kaohsiung

[2]Kurume University

[3]National University of Kaohsiung

Abstract

This talk considers the threshold boundary regression (TBR) model for sample splitting. The TBR model accommodates covariates in both the regression and threshold functions. The threshold function is allowed to be a nonlinear function of multiple covariates, constituting a hyperplane to describe data dynamics in two different states. We propose TBR-WSVM, a two-stage method that incorporates the weighted support vector machine (WSVM) and least-squares (LS) methods to estimate the TBR model. Under regularity conditions, we evaluate the consistency of the TBR-WSVM estimators with their optimal convergence rates. We conduct several simulation experiments to investigate the finite sample performance of the TBR-WSVM estimator. Compared with two recently proposed methods, TBR-WSVM enjoys three advantages: (i) threshold parameters need not be prefixed with nonzero values, (ii) threshold parameter ranges need not be specified, and (iii) the threshold boundary can be non-linearly estimated. Finally, we apply the TBR model to a real data analysis.

Keywords: consistency, iterative estimators, least-squares estimation, support vector machine, threshold estimation.

Session V-9

June 30, 15:00 – 16:20

理工一館 A316

◆Chair：

曾玉玲 – 國立東華大學應用數學系

◆Speaker:

1. 鄭竣元 – 中原大學
2. 林建璋 – 國立陽明交通大學管理科學系
3. 王律涵 – 國立臺北大學統計學系
4. 陳柏豪 – 國立臺北大學統計學系
5. 林君儀 – 國立東華大學應用數學系
6. 林俊源 – 國立成功大學統計學系

Friedman 檢定的新近似法

鄭竣元*

中原大學教務處

鄭秀麗

中原大學應用數學系

謝國文

國立陽明交通大學管理科學系

摘要

在單一觀測值的雙向實驗設計中，假設處理效應和區集效應之間沒有交互作用時，Friedman 檢定法是對應於雙向變異數分析的無母數統計方法。然而 Friedman 統計量的近似卡方分布較保守，因此本研究提出將 Friedman 統計量轉換成一種新的 F 近似法，以期能夠控制顯著水準。研究進而模擬方式比較新的 F 近似法和現有卡方近似法在檢定力和樣本數估算的表現。

關鍵詞：雙向變異數分析、檢定力、樣本數

雙組等效檢定的檢討

林建璋*、謝國文
國立陽明交通大學管理科學系

摘要

本研究針對 Anderson and Hauck (1983)的雙組等效檢定進行三項討論。鑒於文獻中並未公平評估，本文重新在一致的顯著水準下，進行該檢定與雙邊單尾方法的檢定力比較。另為改善其拒絕域無上限的缺陷，提供限縮拒絕域範圍的簡單法則，促進檢測的合理性。此外，考量不同樣本配置和成本問題，提供最佳樣本數規畫建議，藉以拓展其實用價值。

關鍵詞：雙組等效檢定、檢定力、樣本數

雙偏斜常態之 ROC 曲線的統計性質探討

王律涵

國立臺北大學統計學系

摘要

ROC 曲線，包含其 AUC 指標，是最常被用來做評估新的醫療技術或藥物療效的統計工具。目前最常被使用的相關分布特性是基於常態母體的假設下或大樣本的情況下所推論；然而很多時候資料並非具有常態分布，甚至會有因取樣不易而導致樣本數不多的情形產生。本研究將在偏斜常態母體分布的設定下探討 ROC 曲線的性質，包含其 AUC 的性質等，並進一步考慮其相關的檢定情況。

關鍵詞：ROC 曲線、ROC 曲線下的面積、偏斜常態分佈、診斷測試

雙偏斜常態母體之類別資料的 AUC 估計探討

陳柏豪、蘇南誠
國立臺北大學統計學系

摘要

診斷試劑因用於衡量患者的健康狀況，故要求準確度來給出有用的判斷資訊是相當重要的。ROC 曲線下的面積 AUC 是常用來衡量診斷試劑準確性的指標之一。傳統上，雙常態母體假設下的 AUC 具完整的統計推論，而當資料為有序類別型時，過去文獻則主要以三種統計方法：參數型的最大概似估計方法、非參數型的 Wilcoxon-Mann-Whitney 方法、結合前面兩者的混合型方法來估計 AUC。我們觀察到文獻上的實務資料常不對稱，本研究將探討在雙偏斜常態母體假設下，重新探討這三種方法的影響。首先我們會利用拔靴法(Bootstrap)來探討雙偏斜常態 AUC 的情況，檢定該診斷試劑是否具有高準確性，以及給出 AUC 的信賴區間，接著我們也將使用蒙地卡羅法來探討不同的母體分配、不同的兩母體差距、不同的類別群組數等情況下，比較這三種 AUC 估計方法的成效。

關鍵詞：拔靴法(Bootstrap)、有序類別型資料、參數型估計、非參數型估計、偏斜常態

Steam 遊戲視覺化

Visualization of Steam Games

林君儀

國立東華大學應用數學系統計研究所

指導教授：曹振海 博士

摘要

遊戲數量眾多，屬性、標籤、遊戲種類、熱門程度等變數多而混雜。從資料處理上來說，是屬於高維度而稀疏的資料。視覺化遊戲圖能一次性呈現數千款遊戲，並包含遊戲名稱、遊戲群落、熱門程度等資訊，同時以遊戲距離反映近似程度。以此為基礎可以幫助使用者尋找遊戲並探索其關聯，如同樣為動作遊戲，包含解謎要素與否，這樣的差異如何表現在遊戲距離上。借助視覺化的功能，可以觀察到遊戲的不同標籤或屬性對遊戲相似度的影響，並直觀以距離體現。遊戲圖中以圖標尺寸反映遊戲熱門程度，藉此得以快速找到自己熟悉的遊戲，並以此為基礎找到相近遊戲，此外也能用以發掘新遊戲。相比起 Steam 的搜尋列表，遊戲圖能夠以清楚的圖標反映熱門程度，並藉由瀏覽遊戲整體建立對多樣遊戲的認識。同時以聚落的形式表現遊戲，形成的遊戲群不只利於瀏覽，也使挑選近似遊戲更加快速。使用者藉由視覺化的遊戲圖，除了能夠全面地看到遊戲的整體資訊，還能根據偏好遊戲所在的聚落，進行拉近聚焦檢視以尋找可能的目標遊戲。如以遊戲《人中之龍》為例，顯示相關遊戲如《Azur lane crosswave》、《Evoland2》等。

關鍵詞：Visualization、Latent Dirichlet Allocation、Steam tags

RUL Prediction of Lithium-ion Batteries under Ambient Temperature Based on Piecewise Linear Random Coefficient Degradation Models

Chun-Yuan Lin

Department of Statistics, National Cheng Kung University

Abstract

In recent years, the performance of lithium-ion batteries has become one of the main issues due to the fast growing of electric vehicles. Accurate prediction of the remaining useful life of these batteries is essential for equipment prognostics and health management. In our experiment, three factors that affect the degradation of lithium-ion battery discharge capacity were considered, namely, the number of continuous charge-discharge cycles, the number of rest days, and the ambient temperature. The most valuable aspects of the experiment involved the battery data collection under fluctuating temperatures and considered the capacity regeneration phenomenon, which mimicked the real-world usage of lithium-ion batteries in electric vehicles. We proposed piecewise linear random coefficient degradation models to capture the capacity degradation paths given specified continuous charge-discharge cycles and rest days under fluctuated ambient temperatures. The likelihood function of our proposed model contains multiple integrals, which can result in computational challenges. We calculated the approximated integrals using the Monte Carlo method. We provided the point estimates, interval estimates, and distribution estimates of remaining useful life.

Keywords: Lithium-Ion Batteries, Remaining Useful Life, Capacity Regeneration Phenomena, Piecewise Linear Random Coefficient Model, Ambient Temperature

Session V-10

June 30, 15:00 – 16:20

理工一館 A210

◆Chair：

陳忠和 – 南臺科技大學工業管理與資訊系

◆Speaker:

1. 陳忠和 – 南臺科技大學工業管理與資訊系
2. 范姜翔星 – 國立臺北大學統計學系
3. 劉佳穎 – 國立臺北大學統計學系
4. 洪煒傑 – 國立高雄大學統計學研究所
5. 李曉妮 – 國立高雄大學統計學研究所
6. 温筑涵 – 國立臺北大學統計學系

Economic Manufacturing Quantity Model with Quality Investment

Chung-Ho Chen

Department of Industrial Management and Information

Southern Taiwan University of Science and Technology

Abstract

In this work, the author proposes the modified economic manufacturing quantity model by considering sampling inspection plan application for determining the optimal production quantity, process mean, and quality investment. Assume that the quality characteristic of product is normally distributed with unknown process mean and known process standard deviation. The quality investment function is the declining exponential reduction function of process variability. Taguchi's asymmetric quadratic quality loss function is adopted for measuring the product quality.

Keywords: Economic manufacturing quantity (EMQ), Process mean, Process standard deviation, Quality investment, Taguchi's asymmetric quadratic quality loss function.

隱含波動率應用於台指選擇權配對交易之實證研究

范姜翔星
國立臺北大學統計研究所

摘要

台灣期貨交易所(TAIFEX)的台指選擇權(TXO)除了是台灣市場上最活躍的選擇權產品之一外，在亞洲市場上的交易量也是名列前茅，故台指選擇權的交易策略也相對多樣，包括日內交易、波動率交易、時間價值衰減交易等等，吸引了許多投資者和交易者的關注參與。

本文主要探討配對交易在台指選擇權市場是否具有獲利空間，希望根據選擇權的隱含波動率可以反映市場對未來資產波動預期的特性，來判斷當前市場價格對該履約價商品是高估還是低估。如果我們發現一個選擇權的隱含波動率相對較高，而另一個選擇權的隱含波動率相對較低，這可能暗示市場對他們未來價格波動性的預期存在差異，基於這種觀察，我們可以考慮採取配對交易的策略，預期從價格回歸或從波動率收斂中獲取利潤。

本文利用thin-plate spline方法建構隱含波動率微笑曲線，將收集到的資料分成五分鐘一組來建構曲線，以此與接下來五分鐘市場上成交的選擇權的隱含波動率做比較，以其差別設定門檻來找尋高估與低估的選擇權來進行配對交易。

關鍵詞：選擇權隱含波動率、配對交易、thin-plate spline

臺灣股市熊市之預測

劉佳穎

國立臺北大學統計學系

摘要

對於投資人而言，股票市場熊市的發生通常伴隨著巨大的風險，若能預先觀測到市場趨勢和風險，進而預測股票市場的熊市發生，投資人不僅可以減少損失也可以做出更明智的投資決策。 本研究目的為預測臺灣股票市場熊市的發生，研究方法採用非參數法之 Bry-Boschan 法則定義臺股熊市，並建立 Probit 模型來探討總體經濟、財務變數及投資人情緒等因素對未來半年內臺灣股票市場的影響。研究結果顯示投資人情緒變數中，消費者信心指數、消費者投資股票信心指數、賣買權比、VIX 波動率對半年內臺股熊市有顯著的預測能力；總體經濟變數中，期間利差、通貨膨脹率、台幣兌美元匯率、工業生產指數、外資買賣超金額、M1B 對半年內臺股熊市有顯著的預測能力；財務變數中股市本益比、股市報酬變動率、股市殖利率、證券劃撥存款對半年內臺股熊市有顯著的預測能力；國際變數中美國半導體指數、美國那斯達克指數對半年內臺股熊市有顯著的預測能力。

關鍵詞：熊市、Prorbit 模型

正交化資產對風險平價投資組合的影響與應用

洪煒傑
國立高雄大學統計學研究所

摘要

本研究的目的在於觀察並比較不同正交化資產所建構的投資組合及其表現。本研究參考 Kamauchi & Yokouchi（2021）的文獻，透過主成分分析（Principal Component Analysis）以及經由資本資產定價模型（CAPM）排序後資產的 Gram-Schmidt 正交化兩種方法挑選資產，並使用風險平價投資組合（Risk Parity Portfolio）找出適當的權重，比較其投資報酬與風險等。在實證研究當中，本研究針對 2001 年至 2020 年，一共 20 年間，在台灣股市大盤裡上市的 300 檔股票標的，並分別考慮投資期為半年、一年與兩年的投資組合進行比對。根據比對的結果發現，Gram-Schmidt 正交化投資組合的各項結果，包含平均報酬、風險與 Sharpe Ratio 等，相較於主成分投資組合的各項結果均更為理想。另外本研究也針對投資標的之間的相關性進行分析，比較主成分投資組合及 Gram-Schmidt 正交化投資組合在不同相關性資產之間的差異。

關鍵詞：風險平價、主成分分析、Gram-Schmidt 正交化、相關性

重要風險因子及標的資產的選取及其投資組合的表現

李曉妮

國立高雄大學統計學研究所

摘要

本研究的目的在於探討只用財務模型和進一步使用統計模型所建構出的投資組合之表現。Fama-French (1992, 1993, 1995)透過一系列文獻發現三個重要的風險因子(RMRF, SMB, HML)。除了這三個風險因子，Bali and Cakici (2004)還發現另外三個重要的風險因子(HVARL, HTVL, HILLIQL)。本研究根據上述組成一個三因子模型和兩個五因子模型，由這三個財務模型分別去挑選高超額報酬的投資標的，並組成投資組合，比較其風險、報酬和 Sharpe Ratio 等。進一步以統計選模方法，從前述的財務模型中選取較影響報酬的風險因子，一是不同時期是否有不同的風險因子、一是個別公司採用不同的風險因子，一樣挑選出高超額報酬的投資標的組成投資組合，並與前述財務模型組成的投資組合、大盤和 naive 投資組合做比較。在實證研究中，本研究依照大盤比例隨機選取台灣 300 家上市公司股票，使用 2006 年初到 2020 年底的股價進行分析，統計模型投資組合表現在平均報酬和 Sharpe ratio 上都比財務模型投資組合表現高 20%以上。長期來看，統計模型投資組合表現也比大盤好。此外，本研究使用 LASSO 挑選資產來模擬大盤，並進一步用統計選模方法去修正，是否能得到一個能模擬大盤又比大盤更好的投資組合。

關鍵詞：套利定價理論、Fama-French 3 因子、投資組合、逐步回歸

社群媒體活動及網路聲量對股票價格與波動的影響

溫筑涵、鍾麗英

國立臺北大學統計學系

摘要

現今社會中網路的普及造就了社群媒體的發達，而人們使用社群及網路搜尋所留下的紀錄會影響周遭其他人的想法，本研究討論 Reddit 的 WallStreetBets 論壇每日提及次數、Twitter 的每日追蹤人數和 Google Trend 的關鍵字搜尋這三個變數是否和股票收盤價與報酬有關係。資料包含五支大型科技股和兩支迷因股在 2018 年 8 月至 2022 年 12 月的股價整理成縱向資料的形式，使用 panel data 迴歸分析得到 WallStreetBets 提及次數增加和股價和報酬有正影響；Twitter 追蹤人數和 Google 的搜尋趨勢增加和股價與報酬有負影響。此外，利用不對稱 EGARCH 模型探討三個變數和報酬的關聯性與是否存在波動不對稱的效果，研究結果顯示 WallStreetBets 提及次數和 Google 的搜尋趨勢對報酬波動影響較大；Twitter 追蹤人數對報酬波動影響較小。整體而言，兩種不同方法都顯示社群媒體活動和網路聲量是會影響股價報酬。

關鍵詞：panel data、EGARCH、社群媒體、網路聲量

Session V-11

June 30, 15:00 – 16:20

理工一館 A106

◆Chair：

曹振海 – 國立東華大學應用數學系

◆Speaker:

1. 呂一昕 – 國立東華大學應用數學系
2. 陳君毅 – 國立臺北大學統計學系
3. 蔡明憲 – 國立臺北大學統計學系
4. 管彥鳴 – 國立陽明交通大學多媒體工程研究所
5. 何郁涵 – 國立高雄大學統計學研究所
6. 何婉瑜 – 國立高雄大學統計學研究所

爵士音樂家資料視覺化

呂一昕

國立東華大學應用數學系

摘要

我們使用爵士音樂家的資料建立一些視覺化圖形, 這些圖能顯示音樂家在樂器、曲風和活躍年代上的相似程度。針對爵士音樂家的資料視覺化, 選擇以 Wynton Marsalis 和 Roy Hargrove 為首的 229 位音樂家, 並以他們的樂器、曲風和活躍年代做為資料矩陣的變數。我們將這些變數分別處理、運算得出各自的相似度矩陣, 再將這些矩陣合併, 用 PCA 和 t-SNE 降至二維。最後把二維的資料點繪製在平面上得到目標圖形, 這些圖形比現有的 Linked Jazz 能顯示出更多的資訊。當我們用變數上色時, 發現大部分的音樂家都不只演奏單一樂器或曲風。這兩項為多值變數（multi-value variable), 它導致上色的困難。我們讓三原色分別代表三種類型、接著用音樂家在三種類型中的比例進行調色, 解決這個問題。

關鍵詞：爵士音樂家、資料視覺化、多值變數

卷積多實例學習方法應用於影像特徵提取

陳君毅、張升懋

國立臺北大學統計學系

摘要

本論文結合卷積神經網路(Convolutional Neural Network) 的資料擷取方式與多實例邏輯斯迴歸(Multiple-instance Logistic Regression; MILR) 之模型來進行圖像辨識，以下稱本模型為卷積多實例邏輯斯迴歸(CMILR)。 此方法延伸自多實例學習(Multiple-instance learning)的構想，利用包(bag) 與實例(instance) 之間的關係進行建模與預測，有別於卷積神經網路 ，本模型藉由邏輯斯迴歸分類器的特性可尋找出圖像中的特徵並預測其特徵位置。綜合上述特性， CMILR 具有卷積神經網路的高準確率與多實例邏輯斯迴歸中線性模型之可預測性。經模擬實驗後，發現 CMILR 可以精準找出特徵的形狀與正確的預測特徵位置，並且對比卷積神經網路與正規化邏輯斯迴歸(Penalized Logistic Regression)皆具有較有高的準確率。承接模擬資料的結論，並應用於手寫數字辨識與 COVID-19 肺部 CT 掃描影像。 CMILR 在各式資料的圖像中有找出相對應的特徵，在圖像中實際特徵辨識上是佔有優勢；另一方面，在邏輯斯迴歸的建模方式下，其參數值具有解釋上的意義，且 CMILR 使用的參數數量遠少於卷積神經網路。

關鍵詞：Convolutional Neural Network、Multiple-instance Logistic Regression、LASSO

基於深度學習的車牌辨識系統的對抗性攻擊：以生成對抗網絡為例

蔡明憲

國立臺北大學統計學系

摘要

車牌識別系統(LPRS)是一種常見的技術，特別是使用深度學習神經網路的 LPRS 更是具備高效和準確性而廣泛應用於日常生活中，像是智能停車場、高速公路 ETC 等。然而，許多文獻表明深度學習模型相當容易受到對抗性攻擊(Adversarial Attack)。其中一種攻擊方式為對抗性樣本(Adversarial Examples)是透過加入微量的雜訊(Perturbation)在圖片中，使其作為模型的輸入，導致模型產生誤判，從而可能導致車牌識別結果出現錯誤。

在先前對抗性攻擊的研究中，大多假設攻擊者已經駭入系統，並從系統內部發起攻擊，此攻擊方式即為白盒攻擊(White box attack)，但是這種攻擊方法不太符合實際情況，因為大多數深度學習模型都是在私有環境中部署和運行的，攻擊者很難直接訪問這些模型。在本研究中，提出一個基於生成對抗網路(Generative Adversarial Network，GAN)技術來生成對抗性示例的方法，攻擊者無需事先了解該模型的架構和參數。具體而言，透過訓練一個生成器和一個判別器，我們可以產生與原始圖像非常相似但能夠誤導深度學習模型的對抗性示例，且有別於常見的 FGSM、BIM 等攻擊方式，能夠更有效的誤導 LPRS。

關鍵詞：車牌識別、對抗性攻擊、生成對抗網路

Detecting the Local Advanced Rectal Cancer from CT Images by Deep Learning Methods

Yen-Ming Kuan[1], Chun-Yu Lin[1,2], Jacky Chung-Hao Wu[2], Henry Horng-Shing Lu[1,3]

[1] Institute of Multimedia Engineering, National Yang Ming Chiao Tung University

[2] Taichung Veterans General Hospital

[3] Institute of Statistics, National Yang Ming Chiao Tung University

Abstract

Rectal cancer is a highly prevalent disease that requires a complex multimodal treatment plan. For patients with locally advanced rectal cancer, neoadjuvant chemoradiotherapy is recommended to increase the circumferential resection margin (CRM) and reduce the risk of local recurrence (LR). Early detection of locally advanced disease is critical for initiating treatment, but it often relies on the physician's experience to develop a treatment plan at the time of the initial workup.

To address this challenge, we developed a deep learning artificial intelligence model to help identify patients with locally advanced rectal cancer based on their CT images. Specifically, we used EfficientNet-B0 as our pre-trained model and adapted the classifier to our problem. We also performed image preprocessing based on the characteristics of the CT images. Our study found that our model achieved an AUC of 0.86 for image classification and an AUC of 0.88 for patient classification using the image classification results. These results suggest that our deep learning model has the potential to assist physicians in early detection and treatment planning for patients with locally advanced rectal cancer. This is the joint work with Dr. Jacky Chung-Hao Wu and Prof. Henry Horng-Shing Lu in National Yang Ming Chiao Tung University, Dr. Chun-Yu Lin and the related researchers in Taichung Veterans General Hospital

Keywords: LARC, CT, CNN, deep learning

High-dimensional clustering via features segmentation and fusion

Yu-Han He(何郁涵)

國立高雄大學統計學研究所

摘要

This paper proposes a clustering method based on feature segmentation and fusion, denoted by SFC. We divide the features into several groups and apply the clustering method with the features in each feature group separately for the subjects. A fusion method is used to fuse the clustering results. Under mild conditions, we prove that SFC has the advantage of improving the clustering effect compared to using all features for clustering at one time. In addition, the SFC can employ parallel computing when conducting clustering in each feature group to save computational time. We design several simulation scenarios and apply the SFC to four scRNA-seq gene datasets to investigate the performance of the proposed method. The numerical results reveal that the SFC obtains better clustering performances than competitors.

Keywords：feature segmentation, fusion, clustering

Dimension Reduction for High-Dimensional Regression by Iteratively Subgrouping Selection Procedure

何婉瑜

國立高雄大學統計學研究所

摘要

本研究針對高維度資料提出一個有效降維的三階段變數選取演算法。於第一階段中，採用迭代方式將變數隨機分群用以配適迴歸模型計算並保留顯著 p 值，直至總保留變數個數小於一半樣本數。於第二階段中，對保留變數在不同樣本數下，配適多次迴歸模型並測量 p 值之平穩表現，而後進一步保留較高平穩 p 值對應之變數。最後於第三步階段中，透過傳統變數選取方法進行最終變數選取。於模擬試驗中，我們應用此三階段變數選取演算法於不同相關結構之模擬數據集，模擬結果顯示相較於傳統高維度選取方法，當資料的稀疏性不明顯時，我們的方法在選取重要變數上有較好的表現。

關鍵詞：高維度、p 值、迴歸

Session V-12

June 30, 15:00 – 16:20

理工一館 A109

◆Chair：

蘇南誠 – 國立臺北大學統計學系

◆Speaker:

1. 廖靖芸 – 國立政治大學統計學系
2. 陳慧霜 – 國立政治大學統計學系
3. 施承翰 – 國立東華大學應用數學系
4. 黃文顥 – 國立清華大學
5. 薛皓澤 – 國立高雄大學統計學研究所
6. 黃凱琪 – 國立臺北大學統計學系

由《聯合報》及《人民日報》報導風格看兩岸差異

廖靖芸*、余清祥

國立政治大學統計學系

摘要

俗話說:「一方水土、養一方人」，由於環境制度、生活方式、觀念思想等之差異，即便同文同種的兩地，其居民的人文素質及文化特徵可能截然不同。中國與臺灣同屬於華夏民族，擁有相似的語言文化及家庭制度，但 1950 年代至今兩岸採用不同政治體制，加上外來文化及民族融合等因素，臺灣及中國的風俗民情之歧異性隨時間而愈發明顯。本研究以中國與臺灣兩地報紙的文字報導為研究對象，作為比較兩者差異的依據，透過文字採礦等方法分析寫作風格，找出兩岸在用字遣詞及思想觀念有哪些明顯不同。其中，中國部分選擇 1946 年~2021 年《人民日報》頭版報導，《人民日報》屬於中國共產黨機關報，紀錄中華人民共和國建國至今發生的重要新聞；臺灣部分選擇 1960 年~2021 年《聯合報》社論，《聯合報》屬於臺灣三大報之一，其歷史最為悠久。此次研究採用探索性資料分析（Exploratory Data Analysis），將單字及雙字詞視為生物物種（Species），引進生物多樣性（Species Diversity）及棲息地（Habitat）等概念，探索用字風格及關鍵字詞的關連及聚落。此外，我們也考量兩岸報紙的標點符號、虛詞（Function Words）、句子結構，藉由 Entropy、Simpson Index、TTR（相異字比例；Type-Token Ratio）等豐富度及不均度指標，萃取兩種報紙的重要特徵。

關鍵詞：文字探勘、風格變化、生物多樣性、關鍵詞、關聯分析

影像分析與深偽影片的偵測

A Statistical Approach of Deepfake Video Detection

陳慧霜[1]　　余清祥[2]

Huei-Shuang Chen　　Jack C. Yue

國立政治大學統計學系

摘要

大數據發展也面臨一刀兩刃的窘境，在提升民眾的生活便利及健康之餘，同時會帶來負面的影響，深偽影片 Deepfake 就是知名範例。藉由深度學習等技術偽造照片及影像，顛覆以往「眼見為憑」、「有圖有真相」的認知，眼睛容易受到偽造影片的愚弄，讓我們愈來愈難判斷訊息的真偽。我們認為偽造影片經過人工修改，由 Deepfake 產出之圖像會有過度平滑的傾向，套用一階差分之類的分析方式，或許可有效測量出圖像的平滑程度，作為判斷影像真偽的依據。本文使用基本的光三原色（紅綠藍，Red、Green、Blue），再加入兩種常見色彩空間：HSV（色相、飽和度、明度；Hue、Saturation、Value）、YCrCb（流明、紅色色度、藍色色度；luminance、Red difference、Blue difference）作為圖像資料的變數。另外，除了一階差分，為了更有效判別色彩空間變數的平滑程度，我們也考慮圖像切割（Segmentation），以及搭配 Kolmogorov-Smirnov 檢定來設定較為合宜平滑程度之門檻值，最後結合統計、機器學習模型衡量真偽偵測的結果分類。藉由交叉分析的驗證，以深偽影片常見的兩個資料庫為實驗對象，分析發現加入 HSV、YCrCb 可以增加二元分類問題之準確度，且在圖像分割為 16 ×16 時（過多分割未必較佳），偵測準確度超過 90%，優於過去深偽影片的相關研究。

關鍵詞：深偽影片 、影像辨識、多媒體鑑識、面部紋理、分割

[1] 陳慧霜，國立政治大學統計學系碩士班研究生，email：110354002@nccu.edu.tw

[2] 余清祥，國立政治大學統計所教授，email：csyue@nccu.edu.tw

BGG 桌遊圖

施承翰

國立東華大學應用數學系

摘要

我們以 BGG 桌遊資料爲原始素材，將遊戲主題、機制等單/多值類別變數進行 Jaccardization 等資訊萃取，進而以 PCA 以及 t-SNE 建構桌遊地圖，具體化桌遊之間的異同。以花磚物語（Azul）和貓與花毯（Calico）為例，相較於 BGG 網站的呈現，桌遊圖可以更容易地處看出兩者的相似性，以及周邊的類似桌遊。分析結果與介面整理以 Shiny R 呈現，以供研究者和桌遊愛好者方便使用，自主探索桌遊世界。

關鍵字: 視覺化、桌遊圖

使用隨機配置方法處理大數據中不完整地址資料 – 以新北市為例

黃文顥

國立清華大學

摘要

新北市房屋地址受到地名路名更新、填寫錯誤等原因，有 10.5%的房屋地址不完整，造成無法透過地址資訊找到其對應的所在座標，稱其為房屋座標缺失。我們嘗試五種不同的配置方法，透過模擬並搭配 Accuracy、Total distance 和 MISE 三個指標評估後，我們發現將不完整的地址分配至所在區域內的隨機座標，取得最佳的結果。另外，我們在五種配置方法中也嘗試了唯一一種連續型的方法，即 Kernel density estimation(Kde)作為填補的方法之一，但因為地理分佈通常並不連續，故我們猜測這將造成 Kde 在處理地理資料上的效果不佳。因此，我們針對 Kde 在不連續的情況探討 Bias 的變化，讓讀者了解在不連續與連續間 Bias 的差異。

關鍵詞：房屋座標缺失、隨機配置方法、Kernel density estimation

電子零件參數辨識系統

薛皓澤

國立高雄大學統計學研究所

摘要

本研究提出一自不同電子零件封裝之規格書中，自動辨識出零件製造商，且進而辨識出繪製電子零件電路板所需重要參數之系統，以降低人工辨識所產生的誤植率並提升繪製效率。所提出之參數辨識系統分為三個步驟，首先透過深度學習與接收者操作特徵（Receiver Operating Characteristic，ROC）曲線建立零件製造商標誌辨識模型，以自動自規格書中辨識出零件製造商。其次針對各零件製造商，透過建立電路板重要繪圖參數與零件視圖之質化特徵（如：視圖類別、擷取之數值在視圖的方向、參數代號...等）的關聯矩陣，同時針對各個重要繪圖參數之量化特徵分別建立其邊際分佈，再結合所建立之關聯矩陣與重要繪圖參數之邊際分佈建構一「擷取數值 vs. 繪圖參數」之分數矩陣。最後應用任務指派演算法與該分數矩陣，建議最佳之擷取數值與繪圖參數的辨識結果。實證研究方面，我們將所提出之方法應用於24種不同的電子零件封裝，皆能達到94%以上的電子零件製造商辨識準確率（accuracy）與89%以上重要繪圖參數辨識召回率（recall）。

關鍵詞：自動化、電子零件、任務指派演算法

運用電信資料探討臺鐵車站人流量

黃凱琪、黃怡婷

國立臺北大學統計學系

摘要

臺灣鐵路配合地方發展建設車站，建構出完整鐵道運輸系統，已是許多旅客出遊的首選。每個車站因區位會有不同的人流樣態。臺鐵可經由購買車票、電子票卡等資料總計每個車站的人流量，但周遭的商家與其他車流監控的單位就沒有相關資訊。隨著資訊科技的發展，手機普及化產生大量信令資料，經去識別化後，這些資料可以推估地區的人群流動活動情形。

本研究使用內政部統計處「109 年電信信令資料整合委外服務案」兩家電信業者提供用戶取樣資料，涵蓋範圍為 2020 年 11 月，篩選落在車站的資料點，總共 240 個車站，並將資料整合成時流量，以樣條曲線（Spline）和長短期記憶（Long Short-Term Memory, LSTM）模型進行配適，切割前 21 天的資料做為訓練資料集、後 9 天做為測試資料集，以測試資料的均方誤差（mean-square error, MSE）評估模型配適效果，比較何者有較佳的預測解釋能力。經過研究比較後發現，臺灣車站的人流資料，以長短期記憶模型來配適能得到較佳的效果。

關鍵詞：信令資料、LSTM、Spline、人流分析、臺鐵車站

第32屆南區統計研討會暨2023年中華機率統計學會年會及學術研討會與會人員名單

編號	服務單位、就讀學校	系所、部門	姓名	職稱
1	Feng Chia University	Department of Statistics	Cathy W. S. Chen	教授
2	National Taiwan University	Department of Economics	Hon Ho Kwok	助理教授
3	Seoul National University	Department of Statistics	Yondai Kim	教授
4	Seoul National University,	Department of Statistics	Chae Young Lim	教授
5	Sungkyunkwan University	Department of Statistics	Keun Baik Lee	教授
6	University of California, Los		Weng Kee Wong	教授
7	University of Cincinnati	Department of Environmental and Public	Wei-Wen Hsu	副教授
8	University of Michigan	Department of Statistics	邢泰侖	教授
9	工業技術研究院	綠能與環境研究所	黃建中	專案經理
10	中央研究院	人文社會科學研究中心	詹大千	研究員
11	中央研究院	人文社會科學研究中心 調查專題研究中心	謝淑惠	副研究員
12	中央研究院	人文社會科學研究中心	唐嘉宏	博士後研究
13	中央研究院	統計科學研究所	黃信誠	特聘研究員
14	中央研究院	統計科學研究所	黃景祥	特聘研究員
15	中央研究院	統計科學研究所	潘建興	研究員
16	中央研究院	統計科學研究所	陳君厚	研究員
17	中央研究院	統計科學研究所	黃彥棕	研究員
18	中央研究院	統計科學研究所	張源俊	研究員
19	中央研究院	統計科學研究所	陳素雲	研究員
20	中央研究院	統計科學研究所	蔡風順	研究員
21	中央研究院	統計科學研究所	楊欣洲	研究員
22	中央研究院	統計科學研究所	陳定立	副研究員
23	中央研究院	統計科學研究所	張明中	助理研究員
24	中央研究院	統計科學研究所	張馨文	助理研究員
25	中央研究院	統計科學研究所	梁埈豪	助理研究員
26	中央研究院	統計科學研究所	黃名鈜	助理研究員
27	中央研究院	統計科學研究所	余日彰	博士後研究
28	中央研究院	統計科學研究所	李易儒	博士後研究
29	中央研究院	統計科學研究所	施嘉翰	博士後研究
30	中央研究院	統計科學研究所	張漢利 Hendri Sutrisno	博士後研究
31	中央研究院	統計科學研究所	符慧敏	博士後研究
32	中央研究院	統計科學研究所	陳彥賓	博士後研究
33	中央研究院	統計科學研究所	陳淑君	博士後研究
34	中央研究院	統計科學研究所	賴恩語	博士後研究
35	中央研究院	統計科學研究所	黃意婷	研究助理
36	中央研究院	統計科學研究所	熊嘉妮	研究助理
37	中央研究院	統計科學研究所	顏佐榕	副研究員
38	中央研究院	統計科學研究所	蘇家瑩	博士生
39	中央研究院	歐美研究所	何之行	副研究員
40	中央研究院	經濟研究所	陳樂昱	研究員
41	中央研究院	經濟研究所	楊子霆	副研究員
42	中央研究院	經濟研究所	劉祝安	副研究員
43	中央研究院	經濟研究所	莊雅婷	助理研究員
44	中央研究院	經濟研究所	李金全	博士後研究
45	中央研究院		陳柏均	學生
46	中央研究院		陳學蒲	研究助理
47	中原大學	工業與系統工程學系	陳秀雯	助理教授

編號	服務單位、就讀學校	系所、部門	姓名	職稱
48	中原大學		鄭竣元	專案助理
49	中華經濟研究院	綠色經濟研究中心	張軒瑜	助理研究員
50	中華經濟研究院		陳馨蕙	助理研究員
51	元智大學	管院財金組	鄭應謎	
52	內政部統計處		賴威宇	科長
53	台中榮民總醫院	醫學研究部	蕭自宏	研究員
54	永豐金證券股份有限公司		呂正傑	資深經理
55	永豐金證券股份有限公司		鄭維	資深副理
56	永豐金證券股份有限公司		林煒紘	高級副理
57	永豐金證券股份有限公司		江承翰	經理
58	永豐金證券股份有限公司		柯季綸	襄理
59	全球人壽		謝明華	獨立董事
60	艾昆緯股份有限公司		林奕廷	生物統計分析資深專員
61	艾昆緯股份有限公司		黃智揚	Biostatistician
62	杜克大學	統計所	劉妍君	學生
63	亞馬遜網路服務公司		柯家琪	公部門業務經理
64	東吳大學	數學系	林惠文	教授
65	東吳大學	財務工程與精算數學系	鄭宏文	副教授
66	東吳大學	財務工程與精算數學系	蕭維政	助理教授
67	東吳大學	財務工程與精算數學系	簡嘉蒨	學生
68	東海大學	統計學系	呂恆輝	教授
69	東海大學	統計學系	張玉媚	教授
70	東海大學	統計學系	朱基祥	助理教授
71	東海大學	應用數學系	袁子倫	助理教授
72	東海大學	應用數學系	吳鳴迅	學生
73	南臺科技大學	工業管理與資訊系	陳忠和	教授
74	屏東科技大學	工業管理系	洪宗乾	教授
75	美國 George Mason University		胡膺期	教授
76	美國辛辛那提大學		許緯文	副教授
77	美國哥倫比亞大學	統計系	羅小華	教授
78	財團法人國家衛生研究院	群體健康科學研究所	林培生	研究員
79	財團法人國家衛生研究院		徐靖然	碩士研究生
80	財團法人國家衛生研究院		謝沅祐	學生
81	財團法人電信技術中心		彭俊能	副理
82	財團法人醫藥品查驗中心		林資荃	資深統計審查員
83	財團法人醫藥品查驗中心		陳瑱芳	統計審查員
84	高雄醫學大學	基礎科學教育中心	簡莉珠	副教授
85	啟碁科技		李丕強	資訊長
86	國立中山大學	應用數學系	羅夢娜	榮譽退休教
87	國立中山大學	應用數學系	張中	教授
88	國立中山大學	應用數學系	郭美惠	教授
89	國立中山大學	應用數學系	陳美如	教授
90	國立中山大學	應用數學系	曾聖澧	助理教授
91	國立中山大學	應用數學系	鍾思齊	助理教授
92	國立中山大學	應用數學系	林佩宜	研究助理
93	國立中山大學		Giap Van Su	博士生
94	國立中山大學	應用數學研究所	唐聖育	研究生
95	國立中山大學		林亮言	碩士生
96	國立中山大學	應用數學系	林登右	學生

編號	服務單位、就讀學校	系所、部門	姓名	職稱
97	國立中山大學	應用數學系	江皓瑋	學生
98	國立中山大學	應用數學系	李昱圻	學生
99	國立中山大學	應用數學系	李祐瑄	學生
100	國立中山大學	應用數學系	周柏呈	學生
101	國立中山大學	應用數學系	林暐庭	學生
102	國立中山大學	應用數學系	洪卉芬	學生
103	國立中山大學	應用數學系	孫瑞鴻	學生
104	國立中山大學	應用數學系	張維展	學生
105	國立中山大學	應用數學系	陳維佐	學生
106	國立中山大學	應用數學系	魏郁真	學生
107	國立中山大學	應用數學系	羅祖佑	學生
108	國立中山大學		江宇鴻	學生
109	國立中山大學		李翊瑄	學生
110	國立中山大學		黃聖勛	學生
111	國立中山大學		趙伯昇	學生
112	國立中山大學		錢映伶	學生
113	國立中山大學		沈怡姍	學生
114	國立中山大學		林良垵	學生
115	國立中山大學		柯鋅瑀	學生
116	國立中山大學		許育禎	學生
117	國立中山大學		郭郁琪	學生
118	國立中山大學		楊竣皓	學生
119	國立中山大學		蔡羽涵	學生
120	國立中山大學		鄭允礽	學生
121	國立中山大學		鄭喬文	學生
122	國立中山大學		謝宛臻	學生
123	國立中山大學		謝秉宏	學生
124	國立中山大學		李仁傑	學生
125	國立中山大學		謝譯瑭	學生
126	國立中央大學	土木工程學系	林遠見	副教授
127	國立中央大學	大氣科學系	柯緁盈	博士後研究
128	國立中央大學	大氣科學系	鄭芳怡	教授
129	國立中央大學	太空科學與工程學系	劉正彥	教授
130	國立中央大學	數學系	黃世豪	副教授
131	國立中央大學	數學系	陳亭甫	助理教授
132	國立中央大學	統計研究所	曾議寬	教授
133	國立中央大學	統計研究所	Shih-Feng Huang	教授
134	國立中央大學	統計研究所	陳春樹	教授
135	國立中央大學	統計研究所	鄒宗山	教授
136	國立中央大學	統計研究所	樊采虹	教授
137	國立中央大學	統計研究所	孫立憲	副教授
138	國立中央大學	統計研究所	賴暐婷	博士後研究
139	國立中央大學	統計研究所	王俞凱	博士生
140	國立中央大學	統計研究所	董奕賢	博士生
141	國立中央大學	統計研究所	賴龍斌	研究生
142	國立中央大學	統計研究所	陳稚蕙	碩士生
143	國立中央大學	統計研究所	蘇湘筠	碩士生
144	國立中央大學	統計研究所	江為民	學生
145	國立中央大學	統計研究所	洪誠彣	學生
146	國立中央大學	統計研究所	王彥霖	學生
147	國立中央大學	統計研究所	李宗祐	學生
148	國立中央大學	統計研究所	李雨澈	學生

編號	服務單位、就讀學校	系所、部門	姓名	職稱
149	國立中央大學	統計研究所	林易進	學生
150	國立中央大學	統計研究所	林詠筑	學生
151	國立中央大學	統計研究所	邱詠惠	學生
152	國立中央大學	統計研究所	馬繻嬪	學生
153	國立中央大學	統計研究所	許卜仁	學生
154	國立中央大學	統計研究所	陸彥廷	學生
155	國立中央大學	統計研究所	彭紫涵	學生
156	國立中央大學	統計研究所	黃柏惟	學生
157	國立中央大學	統計研究所	葉霈彤	學生
158	國立中央大學	統計研究所	劉懿萱	學生
159	國立中央大學	統計研究所	賴堉溱	學生
160	國立中央大學		JANNYENQ LIU	
161	國立中央大學	統計研究所	童義軒	學生
162	國立中正大學	統計科學研究所	魏子恒	研究生
163	國立中正大學	統計科學研究所	丁鈺展	學生
164	國立中正大學	統計科學研究所	吳佳萱	學生
165	國立中正大學	統計科學研究所	吳姿穎	學生
166	國立中正大學	統計科學研究所	林書岑	學生
167	國立中正大學	統計科學研究所	廖俊凱	學生
168	國立中正大學	資管系	許巍嚴	教授
169	國立中正大學	數學系	李宗翰	助理教授
170	國立中正大學	數學系	邱海唐	助理教授
171	國立中正大學	數學系	高正雄	退休副教授(兼任教師)
172	國立中正大學	數學系	謝進見	教授
173	國立中正大學		曾浩瑋	研究生
174	國立中正大學		卓辰樺	學生
175	國立中正大學		徐維廷	學生
176	國立中正大學		黃于安	
177	國立中興大學	統計學研究所	林長鋆	教授
178	國立中興大學	統計學研究所	林宗儀	教授
179	國立中興大學	統計學研究所	呂亞諭	碩士生
180	國立中興大學	統計學研究所	黃子曜	碩士生
181	國立中興大學	統計學研究所	廖昱凱	碩士生
182	國立中興大學	統計學研究所	鄭乃慈	碩士生
183	國立中興大學	統計學研究所	劉志軒	研究生
184	國立中興大學	統計學研究所	張家豪	學生
185	國立中興大學	統計學研究所	游昭庭	學生
186	國立中興大學	統計學研究所	賴韋智	學生
187	國立中興大學	循環經濟研究學院	徐佩君	助理教授
188	國立中興大學	應用數學系	沈宗荏	教授
189	國立中興大學	應用數學系	Kyung-Youn Kim	副教授
190	國立中興大學	應用數學系	金璟允	助理教授
191	國立中興大學	應用數學系	楊郁成	博士生
192	國立中興大學	應用數學系	莊立勝	學生
193	國立中興大學		洪子傑	碩士生
194	國立台灣大學	應用數學研究所	江金倉	教授
195	國立台灣大學	數學系	李志煌	助理教授
196	國立台灣大學	經濟學系	陳由常	助理教授
197	國立台灣大學	統計與數據科學研究所	陳裕庭	助理教授
198	國立台灣大學	會計學系	古裕彥	博士生
199	國立台灣大學	資訊工程學系	鄭詠堯	學生

編號	服務單位、就讀學校	系所、部門	姓名	職稱
200	國立台灣師範大學	數學系	蔡碧紋	教授
201	國立成功大學	統計所	茆玉麟	博士生
202	國立成功大學	統計所	楊宏基	博士生
203	國立成功大學	統計所	劉純帆	研究生
204	國立成功大學	統計所	何佩欣	研究生
205	國立成功大學	統計所	劉柏均	學生
206	國立成功大學	統計所	蘇靖雅	學生
207	國立成功大學	統計所	原靜旻	學生
208	國立成功大學	統計所	陳昱豪	學生
209	國立成功大學	統計所	陳晉群	學生
210	國立成功大學	統計所	鄭佳鈴	學生
211	國立成功大學	統計所	林虔毅	學生
212	國立成功大學	統計所	徐瑋柔	學生
213	國立成功大學	統計所	楊文博	學生
214	國立成功大學	統計所	鄔宜芳	學生
215	國立成功大學	統計學系	陳瑞彬	教授
216	國立成功大學	統計學系	王婉倫	教授
217	國立成功大學	統計學系	李政德	教授
218	國立成功大學	統計學系	蘇佩芳	教授
219	國立成功大學	統計學系	鄭順林	副教授
220	國立成功大學	統計學系	李宜真	副教授
221	國立成功大學	統計學系	李俊毅	副教授
222	國立成功大學	統計學系	林良靖	副教授
223	國立成功大學	統計學系	馬瀰嘉	副教授
224	國立成功大學	統計學系	王義富	助理教授
225	國立成功大學	統計學系	盧馬汀 (Martin Lukusa)	助理教授
226	國立成功大學	統計學系	戴安順	助理教授
227	國立成功大學	統計學系	王偉銘	博士後研究
228	國立成功大學	統計學系	尹健璋	學生
229	國立成功大學	統計學系	朱信榮	學生
230	國立成功大學	統計學系暨 數據科學研究所	吳岱凌	學生
231	國立成功大學	統資系應用統計所	黃自邇	研究生
232	國立成功大學	數據科學研究所	楊子萱	研究生
233	國立成功大學	數據科學研究所	廖浣婷	研究生
234	國立成功大學	數據科學研究所	王健安	碩士生
235	國立成功大學	數據科學研究所	馬欣蒂	學生
236	國立成功大學	數據科學研究所	吳啓榮	學生
237	國立成功大學	數據科學研究所	曾以諾	學生
238	國立成功大學	數據科學研究所	黃子軒	學生
239	國立成功大學	數據科學研究所	楊淳先	學生
240	國立成功大學	數據科學研究所	葉嘉浤	學生
241	國立成功大學		潘宥亦	博士生
242	國立成功大學		蔡茜婷	碩士生
243	國立成功大學		吳婕琪	學生
244	國立成功大學		林俊源	學生
245	國立成功大學		苗鈺	學生
246	國立成功大學		張韶恩	學生
247	國立成功大學		郭佳修	學生
248	國立成功大學	統計學系	李國榮	副教授
249	國立東華大學	資訊工程學系	林信鋒	副校長

編號	服務單位、就讀學校	系所、部門	姓名	職稱
250	國立東華大學	財務金融系	林金龍	榮譽教授
251	國立東華大學	應用數學系	王家禮	教授
252	國立東華大學	應用數學系	曹振海	教授
253	國立東華大學	應用數學系	黃延安	教授
254	國立東華大學	應用數學系	吳韋瑩	副教授
255	國立東華大學	應用數學系	曾玉玲	副教授
256	國立東華大學	應用數學系	官彥良	助理教授
257	國立東華大學	應用數學系	施銘杰	助理教授
258	國立東華大學	應用數學系	曾柏恩	碩士生
259	國立東華大學	應用數學系	石家睿	碩士生
260	國立東華大學	應用數學系	呂一昕	碩士生
261	國立東華大學	應用數學系	林君儴	碩士生
262	國立東華大學	應用數學系	林佳儀	碩士生
263	國立東華大學	應用數學系	林家弘	碩士生
264	國立東華大學	應用數學系	施承翰	碩士生
265	國立東華大學	應用數學系	張秝穎	碩士生
266	國立東華大學	應用數學系	黃紫瑄	碩士生
267	國立東華大學	應用數學系	劉哲瑋	碩士生
268	國立東華大學	應用數學系	戴傳軒	碩士生
269	國立東華大學	應用數學系	羅世昌	碩士生
270	國立東華大學	應用數學系	吳雨瑾	學生
271	國立東華大學	應用數學系	姜秉勳	學生
272	國立東華大學	應用數學系	蕭禾晴	學生
273	國立東華大學	應用數學系	鍾宜玲	學生
274	國立政治大學	金融學系	林士貴	教授
275	國立政治大學	金融學系	匡顯吉	博士生
276	國立政治大學	國貿系	顏佑銘	副教授
277	國立政治大學	統計學系	余清祥	教授
278	國立政治大學	統計學系	楊素芬	教授
279	國立政治大學	統計學系	鄭宗記	教授
280	國立政治大學	統計學系	吳漢銘	副教授
281	國立政治大學	統計學系	周珮婷	副教授
282	國立政治大學	統計學系	陳怡如	副教授
283	國立政治大學	統計學系	林政寬	學生
284	國立政治大學	統計學系	陳慧霜	學生
285	國立政治大學	統計學系	廖靖芸	學生
286	國立政治大學	統計學研究所	林正和	學生
287	國立政治大學	統計研究所	周鈺宸	學生
288	國立政治大學	應用數學系	黃啟瑞	講座教授
289	國立政治大學	應用數學系	陳隆奇	教授
290	國立政治大學	應用數學系	洪芷漪	助理教授
291	國立政治大學		華軒甫	博士生
292	國立政治大學		簡廷儒	研究生
293	國立政治大學		藍思皓	研究生
294	國立政治大學		吳宥群	學生
295	國立政治大學		林郁嘉	學生
296	國立政治大學		張君瑋	學生
297	國立政治大學		連啟雄	學生
298	國立政治大學		陳韋豫	學生
299	國立政治大學		陳慬瑜	學生
300	國立政治大學		黃政嘉	學生
301	國立政治大學		潘禹翔	學生

編號	服務單位、就讀學校	系所、部門	姓名	職稱
302	國立高雄大學	統計學研究所	俞淑惠	教授
303	國立高雄大學	統計學研究所	黃士峰	教授
304	國立高雄大學	統計學研究所	黃文璋	教授
305	國立高雄大學	統計學研究所	張志浩	副教授
306	國立高雄大學	統計學研究所	許湘伶	副教授
307	國立高雄大學	統計學研究所	郭鋸霖	副教授
308	國立高雄大學	統計學研究所	楊洪鼎	專案助理教
309	國立高雄大學	統計學研究所	洪浩鑫	碩士生
310	國立高雄大學	統計學研究所	薛皓澤	碩士生
311	國立高雄大學	統計學研究所	陳虹君	學生
312	國立高雄大學	統計學研究所	張果全	學生
313	國立高雄大學	統計學研究所	吉怡宗	學生
314	國立高雄大學	統計學研究所	何杏文	學生
315	國立高雄大學	統計學研究所	何郁涵	學生
316	國立高雄大學	統計學研究所	何婉瑜	學生
317	國立高雄大學	統計學研究所	李振愷	學生
318	國立高雄大學	統計學研究所	李曉妮	學生
319	國立高雄大學	統計學研究所	林大鈞	學生
320	國立高雄大學	統計學研究所	洪煒傑	學生
321	國立高雄大學	統計學研究所	紀宗原	學生
322	國立高雄大學	統計學研究所	徐玉玲	學生
323	國立高雄大學	統計學研究所	張雅慧	學生
324	國立高雄大學	統計學研究所	莊佑捷	學生
325	國立高雄大學	統計學研究所	許舒婷	學生
326	國立高雄大學	統計學研究所	郭冠麟	學生
327	國立高雄大學	統計學研究所	陳彥霖	學生
328	國立高雄大學	統計學研究所	劉印展	學生
329	國立高雄大學	統計學研究所	蔡帛諭	學生
330	國立高雄大學	統計學研究所	羅育聖	學生
331	國立高雄大學	應用數學所	鄭詠融	學生
332	國立高雄大學		陳彥佑	學生
333	國立高雄大學		陳昶旭	學生
334	國立高雄大學		蕭銘浚	學生
335	國立高雄大學	統計學研究所	鄭家妏	學生
336	國立清華大學	統計系	徐南蓉	教授
337	國立清華大學	統計學研究所	黃文瀚	教授
338	國立清華大學	統計學研究所	黃禮珊	教授
339	國立清華大學	統計學研究所	銀慶剛	教授
340	國立清華大學	統計學研究所	黃灝勻	博士後研究及研究助理
341	國立清華大學	統計學研究所	林子妤	專任助理
342	國立清華大學	統計學研究所	賴驥緯	博士生
343	國立清華大學	統計學研究所	陳煒傑	碩士生
344	國立清華大學	統計學研究所	沈來興	學生
345	國立清華大學	統計學研究所	林嘉韡	學生
346	國立清華大學	統計學研究所	康暐郡	學生
347	國立清華大學	統計學研究所	曾羿文	學生
348	國立清華大學	統計學研究所	鄭家豪	
349	國立清華大學	經濟學系	冼芻蕘	教授
350	國立清華大學	經濟學系	謝承佑	學生
351	國立清華大學	數理教育研究所	王姿陵	教授
352	國立清華大學	數學系	李卓彥	助理教授

編號	服務單位、就讀學校	系所、部門	姓名	職稱
353	國立清華大學	數學系	黃佳盈	碩士生
354	國立清華大學	數學系	潘星翔	研究生
355	國立清華大學		彭柏翔	博士生
356	國立清華大學		鄭雅珊	博士生
357	國立清華大學		黃怡甄	碩士生
358	國立清華大學		周秉儒	學生
359	國立清華大學		莊明儒	學生
360	國立清華大學		廖怡華	學生
361	國立清華大學		黃昭惠	學生
362	國立清華大學		黃文顥	學生
363	國立清華大學	統計學研究所	林佳霈	學生
364	國立清華大學		潘彥碩	
365	國立陽明交通大學	工業工程與管理學系	董弘平	助理教授
366	國立陽明交通大學	工程與計算生醫產業碩士專班	林承曄	碩士生
367	國立陽明交通大學	公共衛生研究所	楊惠沂	
368	國立陽明交通大學	統計學研究所	王秀瑛	教授
369	國立陽明交通大學	統計學研究所	黃冠華	教授
370	國立陽明交通大學	統計學研究所	盧鴻興	教授
371	國立陽明交通大學	統計學研究所	林聖軒	副教授
372	國立陽明交通大學	統計學研究所	吳崇豪	博士後研究
373	國立陽明交通大學	統計學研究所	黃乾哲	碩士生
374	國立陽明交通大學	統計學研究所	陳冠憲	碩士生
375	國立陽明交通大學	統計學研究所	林奕勳	學生
376	國立陽明交通大學	統計學研究所	林珈卉	學生
377	國立陽明交通大學	統計學研究所	孫長駿	學生
378	國立陽明交通大學	統計學研究所	陳芃辰	學生
379	國立陽明交通大學	統計學研究所	解振成	學生
380	國立陽明交通大學	統計學研究所	林宣廷	
381	國立陽明交通大學	管理科學系	林建璋	碩士生
382	國立陽明交通大學	數據科學與工程研究所	葉詠富	碩士生
383	國立陽明交通大學	應用數學系	千野　由喜	助理教授
384	國立陽明交通大學		陳彥霖	研究助理
385	國立陽明交通大學		褚育誠	博士生
386	國立陽明交通大學		管彥鳴	學生
387	國立彰化師範大學	統計資訊研究所	李宥辰	學生
388	國立彰化師範大學	數學系	鄭宗琳	教授
389	國立彰化師範大學	數學系	蕭守仁	教授
390	國立彰化師範大學	數學系	林子暘	學生
391	國立彰化師範大學	數學系	施承峻	學生
392	國立彰化師範大學	數學系	傅俊元	學生
393	國立彰化師範大學	數學系	韓明澄	學生
394	國立臺北大學	不動產與城鄉環境學系	林柏丞	助理教授
395	國立臺北大學	統計系	蘇南誠	教授
396	國立臺北大學	統計系	黃怡婷	教授
397	國立臺北大學	統計系	王鴻龍	副教授
398	國立臺北大學	統計系	張升懋	副教授
399	國立臺北大學	統計系	翁新傑	助理教授
400	國立臺北大學	統計系	石亞喬	學生
401	國立臺北大學	統計系	范姜翔星	學生
402	國立臺北大學	統計系	徐采邑	學生
403	國立臺北大學	統計系	王律涵	學生

編號	服務單位、就讀學校	系所、部門	姓名	職稱
404	國立臺北大學	統計系	王嘉顥	學生
405	國立臺北大學	統計系	江彥頡	學生
406	國立臺北大學	統計系	呂亮葳	學生
407	國立臺北大學	統計系	周琦雅	學生
408	國立臺北大學	統計系	林佳霓	學生
409	國立臺北大學	統計系	邱士軒	學生
410	國立臺北大學	統計系	邱崴泰	學生
411	國立臺北大學	統計系	高興翰	學生
412	國立臺北大學	統計系	張以謙	學生
413	國立臺北大學	統計系	張博智	學生
414	國立臺北大學	統計系	張智硯	學生
415	國立臺北大學	統計系	陳君毅	學生
416	國立臺北大學	統計系	陳柏豪	學生
417	國立臺北大學	統計系	陳琬淇	學生
418	國立臺北大學	統計系	陳諾霖	學生
419	國立臺北大學	統計系	温筑涵	學生
420	國立臺北大學	統計系	黃凱琪	學生
421	國立臺北大學	統計系	劉佳穎	學生
422	國立臺北大學	統計系	劉銘軒	學生
423	國立臺北大學	統計系	蔡明憲	學生
424	國立臺北大學	統計系	鄭亦甯	學生
425	國立臺北大學	統計系	謝綸豪	學生
426	國立臺北大學		嚴潔翎	學生
427	國立臺北大學		王薇淳	學生
428	國立臺北大學		林楷崙	學生
429	國立臺北大學		柯怡瑄	學生
430	國立臺北大學		張惠雯	學生
431	國立臺北大學		陳怡升	學生
432	國立臺北科技大學	資訊與財金管理系	吳牧恩	副教授
433	國立臺北科技大學	資訊與財金管理系	江大衛	學生
434	國立臺灣大學	統計碩士學位學程	劉家銘	碩士研究生
435	國立臺灣大學	氣候天氣災害研究中心	陳柏孚	助理研究員
436	國立臺灣大學	統計與數據科學研究所	江其祉	副教授
437	國立臺灣大學	統計與數據科學研究所	莊沅蓉	學生
438	國立臺灣大學	會計學系	王子凌	學生
439	國立臺灣大學	經濟學系	楊睿中	副教授
440	國立臺灣大學	經濟學系	郭漢豪	助理教授
441	國立臺灣大學	數學系	林偉傑	助理教授
442	國家太空中心		黃成勇	副研究員
443	國家太空中心	地面研發處	劉小菁	處長
444	國家科學及技術委員會	自然科學及 永續研究發展處	陳美慧	副研究員
445	國家科學及技術委員會	自然科學及 永續研究發展處	林蕙昕	專任助理
446	國家理論科學研究中心	數學組	上島芳倫	計畫博士後研究
447	國家理論科學研究中心	數學組	王靜雯	秘書
448	國家衛生研究院		張憶壽	名譽研究員
449	淡江大學	統計學系	李百靈	副教授
450	淡江大學	統計學系	張雅梅	副教授
451	淡江大學	統計學系	高君豪	助理教授
452	淡江大學	統計學系	邱俊捷	學生

編號	服務單位、就讀學校	系所、部門	姓名	職稱
453	淡江大學	統計學系	陳品樺	學生
454	淡江大學	統計學系	黃永駖	學生
455	淡江大學	統計學系	黃悅熏	學生
456	淡江大學	會計學系	陳虹吟	助理教授
457	淡江大學	數學學系	溫啟仲	教授
458	淡江大學	數學學系	黃逸輝	教授
459	淡江大學	數學學系	溫啟仲	教授
460	淡江大學	數學學系	周孟穎	助理教授
461	淡江大學	數學學系	吳林建宏	學生
462	淡江大學		吳芊穎	學生
463	淡江大學		趙芊茹	學生
464	逢甲大學	統計學系	陳婉淑	教授
465	逢甲大學	統計學系	黃偉恆	教授
466	逢甲大學	統計學系	李燊銘	教授
467	逢甲大學	統計學系	鍾冬川	教授
468	逢甲大學	統計學系	林建同	助理教授
469	逢甲大學	統計學系	劉峰旗	助理教授
470	逢甲大學	統計學系	林家鳴	研究助理
471	逢甲大學	統計學系	陳彥昌	研究助理
472	逢甲大學	統計學系	林旻萱	碩士生
473	逢甲大學	統計學系	吳岳昇	學生
474	逢甲大學	統計學系	邵維瑄	學生
475	逢甲大學	統計學系	張琬渝	學生
476	逢甲大學	統計學系	陳泓愷	學生
477	逢甲大學	統計學系	閔宥睿	學生
478	逢甲大學	統計學系	簡子芯	學生
479	逢甲大學	數據科學碩士學位學程	劉馨隃	學生
480	逢甲大學		張育瑋	研究生
481	逢甲大學		張家華	學生
482	逢甲大學		黃奕馨	學生
483	逢甲大學		趙泓諭	學生
484	逢甲大學		周延	學生
485	逢甲大學		陳俊硯	學生
486	逢甲大學		蔡鳳馨	學生
487	陸軍軍官學校	管理科學系	蘇志成	副教授
488	陸軍軍官學校	管理科學系	洪淑玲	副教授
489	朝陽科技大學	會計系	呂明哲	助理教授
490	慈濟大學	公共衛生學系	溫淑惠	教授
491	義守大學	醫學影像暨放射科學系	陳泰賓	教授
492	輔仁大學	統計資訊學系	蔡嘉仁	副教授
493	輔仁大學	統計資訊學系	陳建州	助理教授
494	輔仁大學	統計資訊學系	龔一鴻	助理教授
495	輔仁大學	統計資訊學系	曾毓豪	學生
496	輔仁大學	統計資訊學系應用統計所	王俞媜	學生
497	輔仁大學	統計資訊學系應用統計所	吳紫豪	學生
498	輔仁大學	統計資訊學系應用統計所	游宗諺	學生
499	輔仁大學		黃佳宣	學生
500	輔仁大學		周雨萱	學生
501	輔仁大學		楊倍欣	學生
502	靜宜大學	資料科學暨大數據分析與應用學系	李名鏞	副教授

編號	服務單位、就讀學校	系所、部門	姓名	職稱
503	靜宜大學	資料科學暨大數據分析與應用學系	郭珈妤	助理教授
504			劉妍君	學生
505			盧玉茹	學生
506			王彥樺	
507			黃學涵	
508			魏培峰	
509			辜凱琳	

中研 統計

sites.stat.sinica.edu.tw/library/dds

Document Delivery Services 大募集

中央研究院統計科學研究所圖書館

統計文獻館際合作

DDS

- 勝 免親赴本館
- 勝 館藏複印不麻煩
- 勝 期刊申請更便利
- 勝 個人讀者亦適用

服務範圍

中央研究院統計科學研究所 館藏

服務費用 免費

由國科會自然司科學及永續研究圖書服務計畫-數學領域支應

想成為合作單位

有意合作單位請來信提供
圖書館(室)聯繫窗口以及相關資訊
(承辦人姓名、電子郵件、聯繫電話)。

連絡信箱 stlib@stat.sinica.edu.tw

申請成為文獻讀者

非合作單位
有文獻需求者
請填寫
(個人)

我是合作單位或本所文獻讀者

文獻申請流程

申請者填寫申請表單

填入想要尋找的
文獻資料及
相關連結(如DOI)

館員處理作業

- 不成功 → 電子郵件通知申請者
- 成功 → 電子郵件通知申請者至貴系所圖書窗口領取紙本

結案

房租待繳中

直接搜尋也可以唷！

paytoearth_cathay

掃描 QR code
追蹤國泰人壽永續 IG

房租待繳中，關心氣候、培力與健康議題，
邀請大家追蹤 IG，一起用行動為
#永續地球 #永續社會 #永續自我
盡份心力吧！

國泰人壽

WNC
Wistron NeWeb Corp.

開啓通訊新契碁

啓碁科技為網路通訊解決方案領導廠商，總部位於台灣新竹科學園區，年營收達32億美元，在台灣設有研發基地，並於台、越、中、美、英、德、日等地設有業務或製造據點，全球員工超過13,000名。啓碁提供短/中/長全距離通訊解決方案，涵蓋5G邊緣端基礎設施、衛星寬頻、衛星廣播、企業網通、Wi-Fi設備、車用連網和智慧家庭，同時憑藉其在天線/RF設計和系統整合的專業能力，成為網通設備品牌商、電信巨擘、車廠和車用電子產品公司的重要供應商。

熱門職缺

- 品保工程師
- 硬體工程師
- 熱流工程師
- 產品工程師
- 資訊工程師
- 軟體工程師
- 天線工程師
- 工業工程師
- 資料分析工程師
- 自動化工程師
- 系統工程師
- RF工程師

WNC招募

WNC官網

Let Your Dreams Take Flight

永續共好
與愛同行

讓愛扎根土地 日漸茁壯
全球的成就 是全球人的功勞
每多一分榮耀 就多一分責任
全球人壽 與世界攜手共好

全球人壽與你一起共創未來

歡迎統計、精算、數學、風控、財務、
資訊等領域專業人才加入我們

更多招募資訊及
履歷投遞請關注

全球人壽保險股份有限公司 免付費服務及申訴專線：0800-000-662
110台北市信義區市民大道六段288號16樓 (112)全壽宣字第0515001號

最會追蹤的投資工具

指數投資證券 ETN

證券商發行的ETN（指數投資證券Exchange Traded Note）
追蹤指數報酬，指標價值與標的指數走勢一致，是投資理財的新選擇！

- **設有發行人資格及發行總額控管，強化保障**
- **指數標的多元，投資選擇性高**
- **上市交易，持有成本低，靈活透明**

TAIWAN STOCK EXCHANGE 臺灣證券交易所
流通證券・活絡經濟

地址：11049 台北市信義路五段7號9F
電話：(02)2792-8188
網址：www.twse.com.tw/zh/page/ETN/summary.html

臺灣期貨交易所
TAIWAN FUTURES EXCHANGE

活絡期貨交易　服務實質經濟

100 臺北市羅斯福路2段100號14樓
(02)2369-5678　www.taifex.com.tw

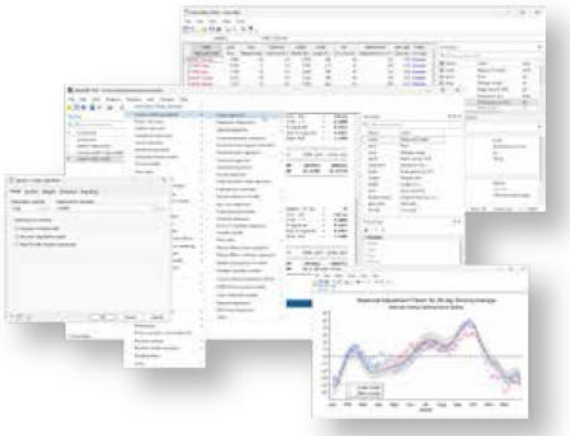
人工智慧
機器學習
資料科學

S F I
昊青股份有限公司
SCIENTIFIC FORMOSA, INC.

多元活動 ● 軟體代理 ● 專業諮詢 ● 線上購買

Stata 18

提供所有資料科學的需求 —
資料處理、視覺化、統計與自動化的報表



Stata 是一套具整合性的統計套裝軟體，不需要購買任何附加的模組，即可透過直覺簡易的指令操作與完整的下拉式選單，快速地完成資料管理、統計分析與圖形呈現，眾多領域皆廣泛應用。

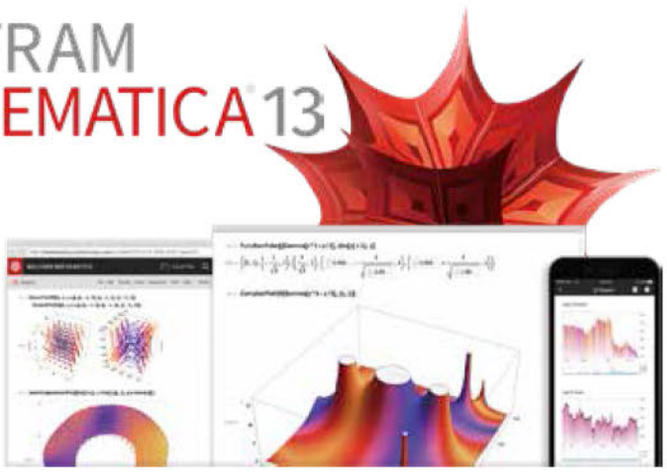
提供數百種統計工具，從基本標準方法，例如：基本的列表與概要的統計量、線性迴歸、ANOVA/MANOVA、ARIMA...等，到高階模型，例如：多階層模型、存活模型、SEM、貝氏分析...等。

WOLFRAM MATHEMATICA 13

高階數學及符號運算軟體

Mathematica 在30 年的歷史中擁有數百萬用戶，在科學計算領域確立了最先進的技術，為全球技術創新人員、教育領域、與各業人士提供最完善的計算環境。

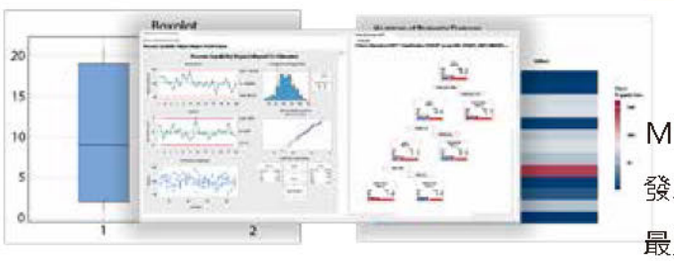
Mathematica不但容易上手，還集合了廣泛領域的知識，並擁有持續整合與擴展的系統，不僅用於計算，也能用於建模、資料視覺化、軟體開發、建立文件或是輸出成果，全方面的功能讓此軟體成為許多機構指定的科學運算工具。



圖：Mathematica可在桌面、雲端、行動裝置中使用

Minitab®

最適合用於統計教學資料分析的軟體



數據無所不在，您能否真正發揮出所擁有數據的價值？Minitab 統計軟體可以檢視當前及過往數據，找出趨勢並預測模式、發現變數之間隱藏的關係， 並建立非凡驚豔的視覺化效果，從容應對最嚴峻的挑戰和機遇。 讓強大的統計威力、領先業界的數據分析和動態視覺化為您所用，探索無限可能。

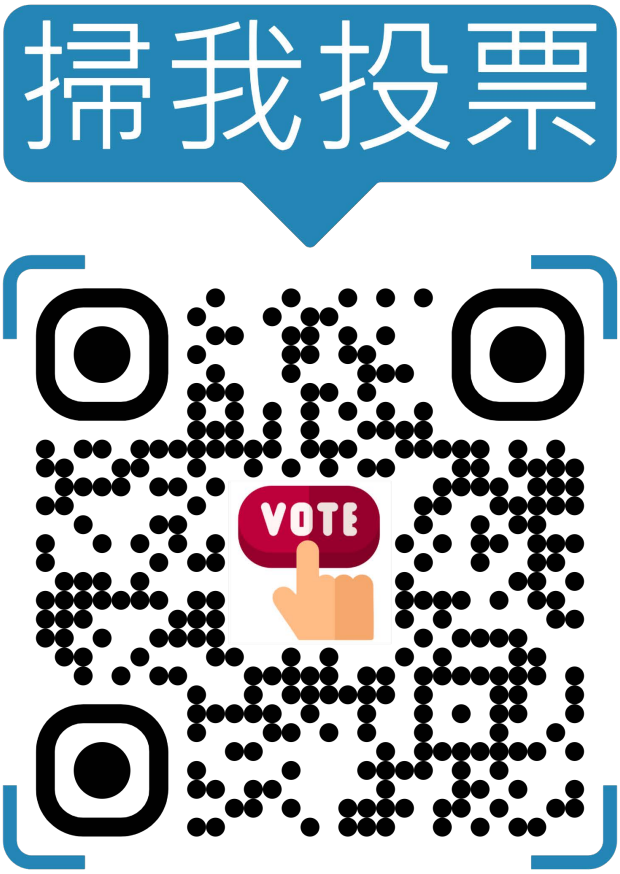
*掃QR CODE進官網
Fax: (02)2503-1680 Tel: (02)2505-0525
www.sciformosa.com.tw
104 台北市中山區復興北路354號11樓

Memo

Memo

Memo

Memo


掃我投票
VOTE