

第八屆碩士生 統計研究成果研討會



2026/
5/30 星期六

會議地點

國立高雄大學圖書資訊大樓

主辦單位

國立高雄大學統計學研究所

協辦單位

中央研究院統計科學研究所

中華機率統計學會

國科會自然處補助科學推展中心數學組

國科會自然處統計學門

國立高雄大學巨量資料研究中心

第八屆碩士生 統計研究成果研討會

議程手冊

主辦單位：國立高雄大學統計學研究所

協辦單位：中央研究院統計科學研究所

中華機率統計學會

國科會自然處補助科學推展中心數學組

國科會自然處統計學門

國立高雄大學巨量資料研究中心

會議日期：2026 年 5 月 30 日 星期六

會議地點：國立高雄大學圖書資訊大樓

目 錄

致謝	1
大會簡介	2
交通資訊	4
會場資訊	5
議程	7
大會專題演講	8
海報分組摘要	9
A 組	9
B 組	16
C 組	24
D 組	31
E 組	40
F 組	48
G 組	57
H 組	65

致謝

本屆研討會得以順利籌辦，首先誠摯感謝各協辦與支持單位的大力協助與經費支持，包括中央研究院統計科學研究所、中華機率統計學會、國科會自然處補助科學推展中心數學組、國科會自然處統計學門，以及國立高雄大學巨量資料研究中心。因有各單位的支持與投入，方能使本次研討會順利推展，提供全國統計相關領域研究生良好的學術交流平台。

此外，特別感謝國立陽明交通大學統計學研究所盧鴻興特聘教授擔任本屆大會主講人。在 AI 與智慧醫療快速發展的時代，如何透過智慧醫療提升生活品質與守護健康，已成為重要議題。盧教授長期投入醫療統計、智慧醫療與跨領域研究，現亦擔任高雄醫學大學附設中和紀念醫院副院長，對智慧醫療之發展具有深厚實務經驗與前瞻視野。從籌備初期與盧教授聯繫的過程中，便深刻感受到其對統計推廣與人才培育的熱忱，不僅迅速回覆相關事宜，更以「我們一起努力推展統計應用」的理念欣然應允擔任主講人，令人十分感佩。

本屆研討會亦邀請來自不同統計領域之 16 位專家學者擔任海報論文評審委員。感謝各位評審委員於百忙之中撥冗參與，提供專業且寶貴的建議，協助學生精進研究內容，並提升本次研討會之學術品質。

最後，感謝國立高雄大學統計學研究所全體教師、助教與學生工作團隊於籌備期間的辛勞與付出，從活動規畫、場地安排與行政支援，皆投注大量心力，使本次活動能圓滿完成。同時，也感謝所有與會師長、同學與支持本次研討會的朋友們。期盼透過本次活動，能促進統計領域之交流與合作，並為臺灣統計教育與研究發展持續注入新的能量。

國立高雄大學統計學研究所
第八屆碩士生統計研究成果研討會
召集人 郭錕霖 敬上
2026 年 5 月

大會簡介

「碩士生統計研究成果研討會」由國科會自然司統計學門發起，旨在提升臺灣高等教育統計學生之專業能力，特別著重於培育碩士生具備良好的口語表達與研究成果展示能力。期望藉由本研討會促進統計研究生之間的觀摩學習與學術交流，增進新知，並進一步激發持續進修與跨校研究合作之契機。本研討會每年約有一百至兩百位師生報名參與，並有五十篇左右論文海報參與競賽，研究主題涵蓋「生物統計」、「工業統計」、「財務統計」、「生物資訊」、「機器學習」及「多元應用」等跨領域方向，充分展現統計方法於不同領域中的廣泛應用與發展潛力。針對海報競賽表現優異之學生，大會亦將擇優推薦至當年度南區統計研討會進行口頭論文發表，提供研究生更進一步之學術交流平台。

歷屆回顧

第一屆 由淡江大學數學學系於 2018 年 6 月主辦，邀請產業界資深研發人士進行專題演講，介紹統計於高科技產業中的重要角色與未來發展方向，使與會師生對統計在業界的實務應用有更深入的認識。

第二屆 由國立清華大學統計學研究所於 2019 年 6 月主辦，以「資料可視化」為主題，邀請馬匡六教授擔任大會主講。透過資料視覺化觀點，引導學生從不同角度理解資料與資訊傳遞的重要性。

第三屆 由逢甲大學統計學系於 2020 年 5 月主辦，邀請國內統計博士班高年級優秀學生進行研究分享，包括董弘平與楊郁成，藉由同儕經驗交流，鼓勵碩士生投入更深入之研究工作。

第四屆 由國立中山大學應用數學系於 2022 年 5 月接辦。該屆邀請黃彥棕研究員擔任主講者，分享統計於跨領域研究中的應用與發展，讓與會學生更深入理解跨領域學習的重要性與未來優勢。

第五屆 以「跨領域學習」為主題，由國立中央大學統計研究所主辦，邀請楊宏基先生、林祐賢董事長，以及許根寧博士擔任講者。透過產業與學術經驗分享，使學生更加了解資料科學與統計專業於業界之應用方向、人才需求與未來發展。

第六屆 由國立臺灣大學統計碩士學位學程舉辦，邀請張源俊研究員擔任大會主講，引領研究生從實際問題出發，思考研究問題之形成與發展；同時亦邀請畢業校友分享業界經驗，協助學生了解未來產業趨勢與應用需求。

第七屆 由逢甲大學統計學系舉辦，邀請黃昭惠與王俞凱分享博士研究歷程與成果，另邀請林文明博士分享產業實務與業界經驗，讓學生對學術研究與職涯發展有更多元的認識。

第八屆 本屆研討會由國立高雄大學統計學研究所主辦，並邀請國立陽明交通大學統計學研究所盧鴻興特聘教授擔任大會主講人。盧教授將以 AI 與智慧醫療為主題，介紹人工智慧於智慧醫療中的核心概念與實際應用，並探討如何從醫療數據中萃取資訊、支援臨床決策。期望透過本次演講，引導與會學生了解統計、資料科學與人工智慧於智慧醫療中的關鍵角色與未來發展方向。

歷屆主辦單位及大會專題演講

屆次	主辦單位	大會專題演講主講者	服務單位
第一屆 (2018)	淡江大學 數學學系	產業界資深研發人士	
第二屆 (2019)	國立清華大學 統計學研究所	馬匡六 教授 (Kwan-Liu Ma)	University of California-Davis, Dept. of Computer Science
第三屆 (2020)	逢甲大學 統計學系	董弘平 博士生 楊郁成 博士生	國立清華大學統計學研究所 國立中興大學應用數學系博士班
第四屆 (2022)	國立中山大學 應用數學系	黃彥棕 研究員	中央研究院統計科學研究所
第五屆 (2023)	國立中央大學 統計研究所	楊宏基 先生 林祐賢 董事長 許根寧 博士	台灣積體電路製造股份有限公司(前) 新北大眾捷運股份有限公司 英國 Phaster 公司
第六屆 (2024)	國立臺灣大學 統計碩士學位學程	張源俊 研究員	中央研究院統計科學研究所
第七屆 (2025)	逢甲大學 統計學系	黃昭惠 博士候選人 王俞凱 博士候選人 林文明 博士	國立清華大學統計學研究所 國立中央大學統計研究所 智城資訊股份有限公司
第八屆 (2026)	國立高雄大學 統計學研究所	盧鴻興 特聘教授	國立陽明交通大學統計學研究所

會場資訊

本研討會於國立高雄大學圖書資訊大樓舉行。各活動地點分布如下：

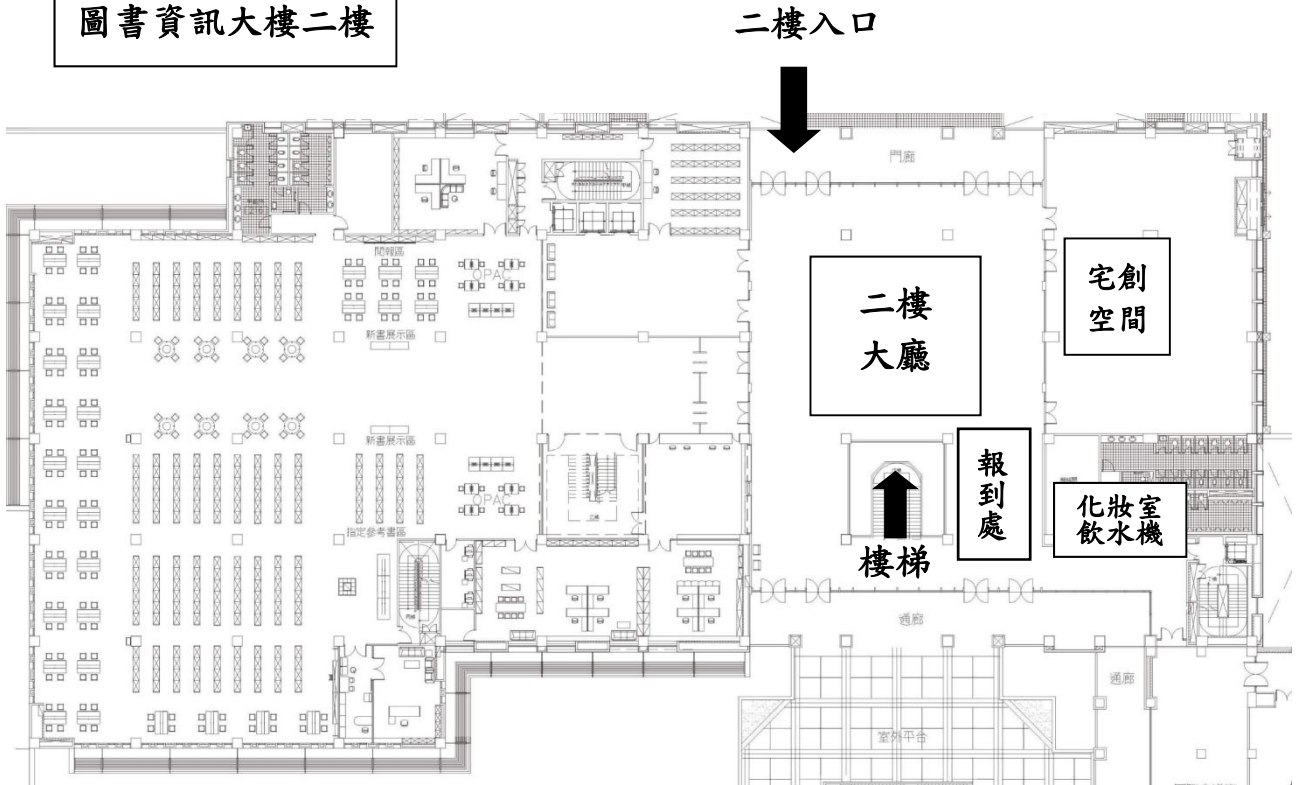
地點	用途
圖書資訊大樓 二樓大廳	報到、海報張貼、海報評審、茶會
圖書資訊大樓 二樓宅創空間	開幕、午餐、茶會
圖書資訊大樓 一樓多媒體教室	大會專題演講、綜合講評及頒獎

校園地圖

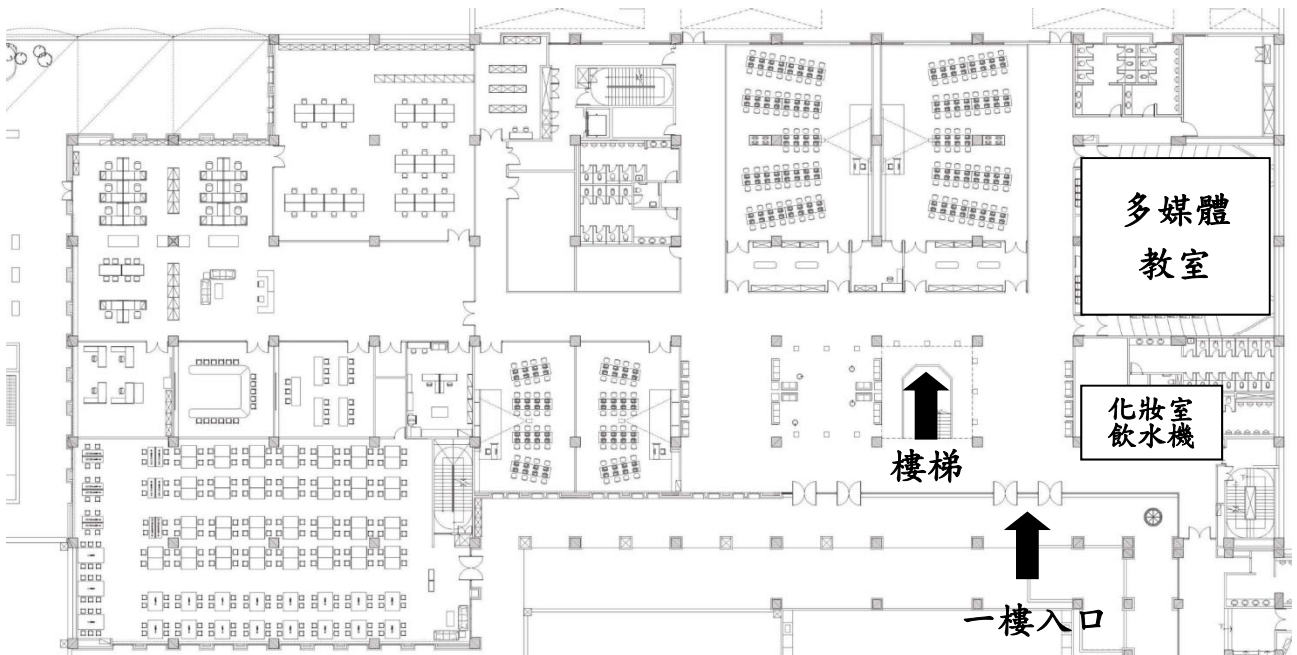


會場平面圖

圖書資訊大樓二樓



圖書資訊大樓一樓



議程

■ 地點：圖書資訊大樓

時間	活動	地點
09:00 – 09:40	報到 第一梯次海報張貼	二樓大廳
09:40 – 10:00	開幕	二樓宅創空間
10:00 – 11:40	第一梯次 (A、B、C、D 組) 評審	二樓大廳
11:40 – 13:00	午餐 第二梯次海報張貼	二樓宅創空間、一樓大廳 二樓大廳
13:00 – 14:40	第二梯次 (E、F、G、H 組) 評審	二樓大廳
14:40 – 15:00	茶會	二樓大廳、宅創空間
15:00 – 16:00	大會專題演講	一樓多媒體教室
16:00 – 16:10	休息	
16:10 – 17:00	綜合講評及頒獎	一樓多媒體教室

備註：第一梯次同學於上午 09:20~09:50，依大會指定之號碼張貼個人海報。

第二梯次同學於中午 12:20~12:50，依大會指定之號碼張貼個人海報。

大會專題演講

主講人

盧鴻興

高雄醫學大學人工智慧生醫研究院 院長
高雄醫學大學附設中和紀念醫院 副院長
高雄醫學大學學士後醫學系 特聘教授
國立陽明交通大學統計學研究所 特聘教授
國立陽明交通大學數據科學與工程研究所 合聘教授
國立陽明交通大學智能系統研究所 合聘教授
美國康乃爾大學統計與資料科學系 兼任教授



演講主題

從數據到決策：AI 驅動的智慧醫療

時間與地點

時間：15:00 – 16:00

地點：圖書資訊大樓一樓多媒體教室

主講人介紹

本屆榮幸邀請盧鴻興教授擔任本次研討會之大會主講。盧教授現為國立陽明交通大學統計學研究所特聘教授，並兼任高雄醫學大學附設中和紀念醫院副院長、及高雄醫學大學人工智慧生醫研究院院長，長期活躍於學術研究與臨床實務交會的第一線，是台灣推動生醫科學與精準醫療發展的重要代表人物。

盧教授受過完整而嚴謹的統計學訓練，並長期投入生物統計、醫學影像、人工智慧與醫療資料科學等跨領域研究，特別關注統計方法與人工智慧技術如何在真實醫療場域中發揮可解釋、可落地的決策價值。近年來，其研究與實務經驗更延伸至大型醫療資料分析、智慧醫療系統建構，以及生醫研究與臨床需求之間的整合。

在本次專題演講中，盧教授將以其橫跨統計理論、人工智慧與醫療管理的獨特視角，深入探討臺灣生醫科學在未來發展中的關鍵契機與挑戰，包括資料驅動醫療、跨領域人才培育，以及學研與醫療體系如何共同形塑具國際競爭力的生醫生態系，為與會者帶來前瞻且具啟發性的洞見。

演講摘要

介紹人工智慧在智慧醫療中的核心概念與應用，說明如何從醫療數據中萃取資訊並支援臨床決策。包含醫療 AI 發展與深度學習基礎、醫療影像與臨床資料分析、生成式 AI 與多模態 AI 在醫療中的應用，以及未來智慧醫療發展趨勢與挑戰，協助理解 AI 如何提升診斷效率、醫療品質與智慧醫療決策能力。

海報分組摘要

A 組 (10 : 00 – 11 : 40)

編號	報告人	題目
A1	蔡承濤	A Study on Change-point Detection for Multivariate Time Series
A2	劉沛禎	三區段門檻條件自迴歸變幅模型
A3	楊書暉	Modeling Vine Copulas Without Simplifying Assumption
A4	楊于韻	Precision-Constrained Optimal Test Planning in Degradation Analysis
A5	李語捷	結合中間資料與設限衰變值之維納過程壽命推估與加速模型分析
A6	吳彥蓁	考量退化異質性與有限資料之剩餘使用壽命推估研究

A Study on Change-point Detection for Multivariate Time Series

蔡承濤

國立中央大學統計研究所

摘要

This study proposes an innovative approach for identifying change-points in multivariate time series. The key idea is to transform the multivariate series into a collection of univariate series via multiple random linear projections. Change-point detection procedures are then applied to each projected series, yielding a set of candidate locations for structural breaks. Clustering techniques are subsequently employed to this set to simultaneously estimate the number and locations of change-points. Several simulation studies are conducted to evaluate the effectiveness of the proposed method. An empirical analysis based on stock market indices, government bond yields, and foreign exchange markets from eight countries over the period February 7, 2006, to December 30, 2024, is further presented to illustrate its practical performance. The numerical results demonstrate that the proposed method achieves improved detection accuracy with lower computational cost compared to existing approaches. The empirical findings also align with major global financial and economic events occurring during the study period.

三區段門檻條件自迴歸變幅模型

劉沛禎
東海大學統計學系

摘要

本研究提出三區段條件自迴歸變幅模型 (three-regime Threshold Autoregressive Range, TARR)，對 Chen 等人 (2008) 所建立之雙區段 TARR 模型的非線性結構上的延伸。相較於原始模型僅設定單一門檻值以區分兩種市場狀態，本研究引入兩個門檻值，將市場波動性劃分為三個區段，以更細緻地刻畫資產價格波動的非線性動態行為。同時發展貝氏方法，結合馬可夫鏈蒙地卡羅 (MCMC) 模擬方法，對模型中之未知參數進行有效估計，並完整處理門檻參數的不確定性。最後，藉由外匯市場的實證分析來比較不同模型的配適表現以及預測市場風險的能力。

關鍵字：TARR 模型、三區段、貝氏方法、MCMC

Modeling Vine Copulas Without Simplifying Assumption

楊書暉
國立中山大學應用數學系

摘要

We propose a novel framework for modeling complex dependence structures using non-simplified vine copulas. In conventional vine copula models, the simplifying assumption, that is, the conditional copula is independent of the conditional variable, is typically imposed. However, this approach may not adequately capture more sophisticated dependence patterns in the data. To address this issue, we apply the regression association measures to detect complex dependence structure in the pair-copula construction, allowing the vine structure to be estimated without the simplifying assumption. By employing a data-driven partitioning strategy, we further propose a flexible conditional copula model that allows the dependence structure to vary across the conditioning variable, thereby facilitating the assembly of a non-simplified vine structure. Simulation studies reveal improved performance in modeling complex dependence structures. Real data examples are analyzed for illustration.

Precision-Constrained Optimal Test Planning in Degradation Analysis

楊于頡
國立中央大學統計研究所

摘要

衰變試驗被廣泛應用於評估高可靠度產品的壽命資訊。在可容許的變異水準下，制定具成本效益的試驗計畫，對製造商而言能大幅提升經濟效益。本研究以常見隨機過程之衰變模型為基礎，在精確度限制下，理論推導最低成本試驗計畫中各決策變數包含樣本數、測量間隔時間及測量次數之存在且唯一之充分必要條件。此外，本研究進一步推導各決策變數對總成本之影響變化率，以排序其重要性，並分析實驗成本、模型參數及精確度對決策變數重要性排序之影響。最後，藉由實例呈現精確度限制之變動與總成本變化間的關係，為決策者制定試驗計畫提供具體參考。

結合中間資料與設限衰變值之 維納過程壽命推估與加速模型分析

李語捷

國立高雄大學統計學研究所

摘要

高可靠度產品的壽命試驗中，由於觀測時間受限，多數產品元件於試驗期間內尚未發生失效，使得僅依賴失效時間之方法難以有效推估壽命分布。當可觀測之品質特性（quality characteristics）隨時間呈現衰變，且其衰變行為與失效機制相關時，可透過衰變試驗取得額外可靠度資訊，以輔助壽命分布的推估。在衰變路徑符合維納衰變過程（Wiener degradation process）假設下，Tang 和 Su (2008) 提出利用中間資料（intermediate data），亦即衰變路徑通過預設非失效門檻（nonfailure threshold）之時間資訊（threshold crossing times），進行壽命分布估計。然而，在有限的試驗時間情況下，部分衰變路徑可能尚未通過非失效門檻，因而產生時間設限資料，使得僅依賴中間資料之推估方法受到限制。為解決同時包含中間資料與時間設限衰變資料之壽命分布參數估計問題，本研究結合 Lee 和 Tang (2007) 對時間設限維納衰變資料所提出之修正最大概似估計方法（modified maximum likelihood method），建構一套統一之概似推估架構，以同時整合門檻通過時間與設限時間下之衰變值資訊，並進一步擴展至多道非失效門檻之情境。此外，考量實務上常透過調整環境應力（如溫度）以加速產品衰變過程，本研究進一步將所提出之概似推估架構推廣至加速衰變模型（accelerated degradation model）。在維納衰變過程假設下，本研究建立模型參數與應力水準之關聯結構，使漂移與變異參數皆可隨應力條件進行調整，並據此整合不同應力水準下之衰變資料，以進行壽命分布之推估，並提供外推至正常使用條件之分析基礎。此外，透過實證分析顯示，在部分情境下，允許變異參數隨應力水準改變，或採用分層參數設定，能更充分反映衰變行為之異質性，並提升模型擬合與推估之表現。

考量退化異質性與有限資料之剩餘使用壽命推估研究

吳彥蓁

國立高雄大學統計學研究所

摘要

剩餘使用壽命 (Remaining Useful Life, RUL) 推估為預測與健康管理系統中的重要研究議題，主要目的在於根據設備運行資料掌握退化狀態，並預測其距離失效所剩餘之時間。然而，實務上完整運行至失效序列資料往往不易取得，且不同設備即使處於相似操作條件下，其退化趨勢方向與幅度仍可能存在差異。因此，本研究關注如何在僅使用少量運行資訊，且設備退化行為具有異質性的情境下，建構可描述健康狀態變化之模型，並進行剩餘使用壽命推估。本研究提出一套整合分群策略、健康指標建構與退化外推模型之 RUL 推估框架。在僅使用每台引擎運行至失效序列中的少量局部資料之條件下，首先透過分群策略刻畫不同引擎之退化型態，並針對各群分別建立高斯過程分類 (Gaussian Process Classification, GPC) 模型。各 GPC 模型以多維感測器資料為輸入，輸出對應之後驗機率，並將其定義為健康指標 (Health Index, HI)，以刻畫設備隨時間之健康狀態演變。於預測階段，測試引擎之感測器資料可透過已建立之 GPC 模型轉換為對應的 HI 序列，作為後續 RUL 推估之基礎。針對 HI 序列之退化外推，本研究採用兩種方法：其一為高斯過程迴歸 (Gaussian Process Regression, GPR)，透過非參數函數模型對 HI 進行連續化建模；其二為維納過程 (Wiener Process, WP)，將 HI 轉換為隨時間演進之隨機退化過程，並以首次達門檻時間推估壽命。實驗結果顯示，所提出之流程在資料受限情境下，能有效因應退化行為異質性，並展現較佳的 RUL 預測表現。

海報分組摘要

B 組 (10:00 – 11:40)

編號	報告人	題目
B1	賴盈宇	Flexible mixture modeling using the truncated skew-normal distribution: properties and applications
B2	黃柏凱	整合微觀模擬與機率敏感度分析之透析治療成本效益與預算影響評估：以臺灣為例
B3	游聲寬	雙有序中介下具中間干擾之路徑特定效應辨識方法
B4	陳智翔	基於條件分佈適合度檢定之吉氏取樣收斂診斷
B5	林一如	基於馬可夫鏈之 Copula 模型與 PELT 演算法對二項序列線下多重改變點之偵測
B6	李慧敏	Multivariate Slash Censored Regression Approach with Multiple Detection Limits

Flexible mixture modeling using the truncated skew-normal distribution: properties and applications

賴盈宇

國立中興大學統計學研究所

摘要

This paper introduces a finite mixture of truncated skew-normal model that provides a flexible and unified framework for analyzing heterogeneous, asymmetric, and truncated data. The proposed model can effectively accommodate both skewness and truncation, allowing for greater modeling flexibility in capturing intricate features often encountered in real-world applications. An expectation-conditional maximization (ECM) algorithm is developed to compute the maximum likelihood estimates of the model parameters. Standard errors are derived using a general information matrix based approach. The effectiveness and practical utility of the proposed methodology are assessed through simulation studies and applications to two real datasets.

整合微觀模擬與機率敏感度分析之透析治療成本效益與預算影響評估：以臺灣為例

黃柏凱
輔仁大學統計資訊學系

摘要

研究背景與目的

透析治療具有長期病程累積、高昂醫療支出與顯著個體族群異質性等特徵。傳統隊列平均模型（Cohort-average models）較難精確反映個體在不同路徑下的健康效益變動與財務負擔。本研究旨在建構個體層級微觀模擬（Individual-level Microsimulation）模型，結合機率敏感度分析，量化評估臺灣血液透析（HD）與腹膜透析（PD）之長期成本效益，並進一步推估其政策調整對年度醫保預算的短期影響。

方法

本研究建構以年為週期之離散時間模型，個體存活事件採參數存活模型描述；基準情境設定為指數分配，年度死亡機率由分齡生命餘命數據反推；敏感度分析則採用 Weibull 分配，以納入危險率隨時間遞增（IFR）之特徵。針對 30 至 79 歲各分齡組獨立進行 10,000 名虛擬個體模擬，並依據臺灣透析盛行人口結構進行權重彙整。模型外部驗證以《2024 臺灣腎病年報》之五年存活率為基準，結果顯示平均絕對百分比誤差（MAPE）為 4.99%，具備高度配適度。針對不確定性處理，執行 1,000 次機率敏感度分析（PSA），對醫療成本、效用值與分配參數進行蒙地卡羅抽樣，產出成本效益平面圖與接受曲線。

結果

於 3% 折現與年齡加權之基準分析中，HD 與 PD 之平均累積品質調整壽命（QALY）分別為 5.51 與 6.12，平均總成本分別為新臺幣 511 萬元與 420 萬元；PD 較 HD 增加 0.61 QALY 並節省約 91 萬元，增量成本效益比（ICER）為 -148.59 萬元/QALY（呈現 PD 優勢策略，PD dominant）。PSA 結果顯示，92.7% 之模擬點落於 PD 優勢

象限；於每 QALY 130 萬元之支付意願門檻下，PD 具備成本效益之機率為 99.1%。預算影響分析顯示，若政策介入使 PD 占比每年提升 0.5% 或 1.0%，於五年期內累積可分別節省健保支出約 6.15 億元與 12.30 億元。

結論與政策意涵

本研究證明個體層級微觀模擬結合機率敏感度分析，能在同一統計架構下處理個體異質性、長期外推累積與參數不確定性，同時為透析模態配置提供長期效率（CEA）與短期財政預估（BIA）之量化證據。結果顯示 PD 在臺灣醫療環境下具備顯著的經濟效益與預算節流潛力，可作為臨床實務與健保支付政策調整之參考依據。

關鍵字：個體層級微觀模擬、機率敏感度分析、成本效益分析、預算影響分析、透析治療

雙有序中介下具中間干擾之路徑特定效應辨識方法

游聲寬

國立陽明交通大學統計學研究所

摘要

在生物醫學、行為科學與健康研究中，暴露變項對結果變項的影響常經由多個具有時間順序的中介機制傳遞。例如，某個治療或危險因子可能先影響早期生理或行為反應，再進一步透過後續中介機制影響疾病結果。因果中介分析旨在拆解暴露變項對結果變項之總效應，以理解不同中介路徑所代表的因果機制；然而，當存在兩個具時間順序之中介變項時，較早的中介變項可能進一步影響後續中介與結果變項，形成中間干擾因子（intermediate confounder），使傳統自然效應框架下常用之 cross-world independence assumptions 無法成立，進而導致相關路徑特定效應不可辨識。

本研究針對雙有序中介且存在中間干擾因子之情境，提出兩組替代辨識策略，分別建立於無交互作用與無異質性假設之上，並據此推導路徑特定效應（path-specific effects）之辨識結果。在估計方面，本文採用 g-computation 建構估計方法，並透過模擬研究評估有限樣本表現。模擬結果顯示，當辨識假設成立時，所提方法相較於忽略中間干擾因子之 naive approach，可有效降低估計偏誤。本文提供一套適用於雙有序中介與中間干擾結構下之路徑特定效應辨識與估計方法。

基於條件分佈適合度檢定之吉氏取樣收斂診斷

陳智翔

國立高雄大學統計學研究所

摘要

吉氏取樣 (GS) 是一連串基於完全條件分佈 (full conditionals) 之迭代型取樣過程，在實務上如何判定迭代是否已足夠接近穩定分佈，仍為 MCMC 方法中的重要議題。傳統收斂診斷方法多半依賴單鏈之軌跡穩定性、自相關結構或鏈間變異比較，但較少直接利用條件分佈本身所隱含之結構資訊。本研究提出一種基於多鏈 GS 的條件分佈收斂診斷方法，在每一次迭代後，利用多條平行鏈所形成之橫截面樣本，檢定該樣本是否符合「其他」尚未更新之條件分佈。其核心概念在於：若目前樣本已不僅符合當前更新所使用之條件分佈，亦近似符合其他條件分佈，則表示 GS 已逐漸接近穩定狀態。因此，可透過連續迭代中的條件分佈適合度檢定結果，建立停止迭代之診斷準則。

基於馬可夫鏈之 Copula 模型與 PELT 演算法 對二項序列線下多重改變點之偵測

林一如

國立中央大學統計研究所

摘要

當資料序列的生成參數發生結構性異常時，我們稱之為改變點。改變點偵測的核心目的即是精準捕捉這些參數的不規則變化。改變點偵測可分為即時 (Online) 與線下 (Offline) 偵測，本篇論文主要聚焦於線下多重改變點偵測。實務上，由於序列型資料往往不滿足獨立性假設，我們採用一階馬可夫鏈模型來捕捉資料間的序列相關性。在聯合分配的建構上，本研究結合二項邊際分配，並探討三種具備不同尾部相依特性的 copula 函數：具左尾相關的 Clayton copula、具右尾相關的 Gumbel copula，以及同時具備雙尾相依特性的 BB1 copula。為了高效且準確地估計多個改變點的位置，我們將 Copula 模型估計出的概似函數結合訊息準則（如赤池訊息準則 AIC 或貝氏訊息準則 BIC）作為懲罰項，導入修剪精確線性時間 (Pruned Exact Linear Time, PELT) 演算法中，藉此大幅降低多變點搜尋的計算複雜度，並探討不同懲罰準則對變點偵測敏感度與模型選擇的影響。在實證分析中，本研究將此演算法應用於珠寶製造數據，旨在回溯並精確標定製造過程中不良率發生結構性改變的歷史節點，藉此協助檢視製程瑕疵，最終達到降低珠寶不良率之目的。

關鍵字：線下改變點偵測、copula 模型、馬可夫鏈、修剪精確線性時間演算法 (PELT)、訊息準則。

Multivariate Slash Censored Regression Approach with Multiple Detection Limits

李慧敏

國立成功大學統計學系

摘要

In many applications, the normality assumption of statistical modeling may not always be realistic due to heavy tails arising from atypical observations and skewness induced by detection limits (DLs) of equipment's quantitative arrays. As a flexible alternative by modeling multivariate data with leptokurtosis and censorship, this paper introduces the MSL with censoring (MSLC) model, in which more than one attribute may contain censored values, and establishes the multivariate slash censored regression (MSLCR) model where the random errors have the MSL distribution. To perform maximum likelihood (ML) estimation of model parameters, we develop expectation-maximization-type algorithms in which the E-steps rely on the first two moments of truncated MSL distributions. The asymptotic standard errors of the ML estimators of location parameters and regression coefficients are obtained by inverting the expected information matrix according to Louis' method. For illustrative purpose, our methodology is applied to three real datasets with genuinely censored observations. Simulation studies are undertaken to investigate the performance of the proposed estimation procedures.

海報分組摘要

C 組 (10 : 00 – 11 : 40)

編號	報告人	題目
C1	羅廷幃	Degradation Analysis of Multivariate Trend Process with Dependent Trend Function
C2	蕭鈺承	Standardized Minimax Design for Gamma Accelerated Degradation Tests
C3	蔡秉杰	基於時間序列模型之衰變分析
C4	黃暄桐	DuMouchel-Jones 修正 D-最適準則下之混合實驗近似設計
C5	胡中岩	Changepoint Detection via a BB1 Copula-Based PELT Algorithm with Mixture Normal Marginals
C6	施承峻	混合隨機效應 Tweedie 衰變模型

Degradation Analysis of Multivariate Trend Process with Dependent Trend Function

羅莚幃

國立成功大學統計學系

摘要

With the rapid advancement of technology, modern products exhibit high reliability and long lifetimes, making failure data difficult to obtain within limited testing periods. Degradation analysis provides an effective statistical method to complete the lifetime prediction by monitoring quality characteristics (QCs) over time. However, the degradation of industrial products is sometimes influenced by multiple QCs simultaneously, and these QCs may exhibit underlying dependence. Therefore, how to construct the degradation model based on the multiple QCs under considering their dependence is crucial for accurate lifetime assessment. This study proposes a multivariate trend process framework by incorporating dependent parameters into the trend functions to capture such correlations. The proposed framework allows different QCs to follow distinct stochastic trend processes (e.g., trend gamma process and trend inverse gaussian process) under a consistent modeling structure. Under the complex dependence structures, maximum likelihood estimation (MLE) is often computationally intractable due to the lack of closed-form joint likelihoods. To address this issue and the constraints of limited sample sizes typically encountered in high-reliability products, a Bayesian framework is adopted to incorporate prior knowledge and improve estimation stability. Markov chain Monte Carlo (MCMC) methods are used for posterior inference and reliability evaluation. Simulation studies and real-data analysis demonstrate the effectiveness of the proposed framework in reliability inference.

Standardized Minimax Design for Gamma Accelerated Degradation Tests

蕭鈺承

國立成功大學統計學系

摘要

For highly reliable products, observing failures under normal-use conditions within a limited testing period is often difficult. Accelerated degradation tests (ADTs) provide an efficient approach for collecting degradation measurements under elevated stress levels and inferring product lifetime. Optimal designs for gamma ADTs often depend on accurate specification of model parameters, which may be impractical when parameter uncertainty exists. To address this issue, this study investigates standardized minimax designs for gamma ADTs by incorporating a plausible range of parameter values and protecting the design against worst-case scenarios. Since the exact criterion is non-differentiable and involves a multi-layer nested optimization problem, a hybrid optimization strategy based on Particle Swarm Optimization (PSO) is used to search for the minimax design under the exact gamma process model. In addition, a saddlepoint approximation (SPA) is considered to simplify the Fisher information matrix and provide analytical insight into the design structure. This study explores standardized minimax design strategies for gamma ADTs from both exact model and approximation-based perspectives, aiming to develop robust and computationally efficient test plans under parameter uncertainty.

基於時間序列模型之衰變分析

蔡秉杰

國立成功大學統計學系

摘要

In this study, we consider two time series models to fit degradation data and predict the product's first hitting time (FHT) or remaining useful life (RUL). One is the integrated moving average with a drift term (IMAD) model. Alternatively, to capture the heterogeneity within the degradation paths, we treat the drift term as a random effect and then construct the integrated moving average with a random drift (IMADR) model. The likelihood functions of both proposed models are derived, and the corresponding maximum likelihood estimators are obtained. Based on the proposed models, the FHT distributions are analytically derived. Monte Carlo simulations confirm the validity of the proposed models. For real data applications, the fatigue crack growth data and NASA battery data are analyzed. First, we fit the IMADR model to perform one- and two-step-ahead predictions, as well as to estimate the corresponding FHT distribution. Next, consider the regeneration-jumps in the NASA dataset, we use the intervention analysis of time series model, and then complete the FHT and RUL distributions of battery.

DuMouchel-Jones 修正 D-最適準則下之混合實驗近似設計

黃暄楠

國立高雄大學統計學研究所

摘要

傳統 D-最適設計在模型假設正確時能有效提升參數估計精確度，然而當實際模型中存在未被納入模型的潛在重要項時，原先所建構的設計可能導致資訊矩陣行列式下降，進而降低參數估計的精確度。為降低設計對單一模型結構的敏感度，DuMouchel and Jones (1994) 提出修正 D-最適準則，透過在資訊矩陣中引入調整參數，以反映潛在模型項所帶來的不確定性。本研究即以該修正準則作為分析基礎，延續 DuMouchel 與 Jones (1994) 於混合實驗與 Scheffé 二次模型下之設計案例，進一步在近似設計理論架構中，系統性探討該準則所對應之最適設計結構。相較於既有文獻多於固定候選點集合中建構離散設計，本文改以連續混合設計空間為研究對象，並推導該修正準則下之等價定理，作為驗證修正 D-最適近似設計最適性的理論依據，並進一步於成分數 $q=3$ 與 $q=4$ 的情形下加以應用與驗證。研究結果顯示，在低維混合實驗與 Scheffé 二次模型假設下，修正 D-最適近似設計之支持點僅分佈於單純形之端點與邊界中心位置，且設計權重隨調整參數的變化呈現一致且可解釋的結構趨勢。此外，本研究進一步以數值方法延伸至成分數 $q=5$ 與 $q=6$ ，結果顯示在 $q=3$ 與 $q=4$ 所推導之設計結構轉換，由僅配置於頂點轉變為包含頂點與邊界中點之設計，在高維情形下仍然成立，且其對應臨界值隨成分數增加呈現規律性變化趨勢。本研究並以所推導之近似設計為基準，利用 D-efficiency 指標評估文獻中既有離散設計在相同準則與參數設定下之相對效率。於計算層面，本研究輔以粒子群最佳化演算法作為數值搜尋工具，並建置互動式 Shiny 介面，以協助使用者理解先驗設定對修正 D-最適設計結構之影響。

Changepoint Detection via a BB1 Copula-Based PELT Algorithm with Mixture Normal Marginals

胡中岩
國立中央大學統計研究所

摘要

Stochastic processes in financial and econometric time series frequently exhibit non-stationary behavior, characterized by abrupt shifts in both marginal distributions and serial dependence structures. Traditional linear models often fail to account for the complex dependencies including nonlinear association or tail structure observed during market extremes, such as joint booms and crashes, as well as the unobserved heterogeneity inherent in return distributions. We propose a robust framework for exact changepoint detection by integrating copula-based Markov chains with the 2-component Gaussian Mixture Model (GMM) to characterize marginal multimodality and skewness. The serial dependency is modeled using the Archimedean BB1 copula, which allows for the simultaneous and independent modeling of asymmetric upper and lower tail dependencies. To address the computational challenges associated with exact search in these complex structures, we implement the Pruned Exact Linear Time (PELT) algorithm. By integrating the joint transition density of the BB1-GMM framework into the PELT cost function, we demonstrate that exact segmentation of complex regimes can be achieved with linear computational complexity. The proposed methodology is validated through a rigorous simulation study and applied to historical financial data to identify structural transitions surrounding global economic crises.

混合隨機效應 Tweedie 衰變模型

施承峻

國立成功大學統計學系

摘要

對於高可靠度產品而言，若僅依賴傳統失效時間資料進行壽命分析，通常需耗費較長試驗時間，且可能面臨失效樣本不足的問題。相較之下，衰變分析可利用產品品質特徵隨時間變化之資訊，在尚未觀察到足夠失效時，提供較豐富的可靠度推論依據。由於衰變資料常同時存在個體間隨機差異與潛在群體間衰變速率差異，因此本研究目的在於兼顧這兩類異質性並基於更具彈性之分配建立衰變模型。本研究以 Tweedie 過程為基礎，提出混合隨機效應 Tweedie 過程 (mixture random effect Tweedie process, mRETP) 模型，用以分析具潛在分群結構之單調衰變資料。由於 Wiener 過程、Gamma 過程與 Inverse Gaussian 過程皆可視為 Tweedie 過程之特例，所提模型較傳統單一隨機過程模型更具一般性與彈性。本研究以隨機效應刻劃個體差異，並以混合模型描述群體差異，使模型能同時反映個體間與群體間之異質性。參數估計方面採用 EM 演算法，並進一步建立壽命分配，利用有母數 bootstrap 方法求得參數估計值與壽命分配累積分配函數之信賴區間。在實例分析中，本研究以裂縫成長資料進行模型比較，並將所提 mRETP 模型與混合隨機效應 Gamma 過程、混合隨機效應 Inverse Gaussian 過程，以及不含混合結構之模型進行配適比較。結果顯示，mRETP 模型具有最佳的整體表現，顯示其更能有效描述裂縫成長資料中同時存在的個體差異與群體差異。

海報分組摘要

D 組 (10 : 00 – 11 : 40)

編號	報告人	題目
D1	戴毓祥	Changepoint Detection for Zero-Inflated Poisson Processes via PELT
D2	黃芯鈺	相對機率密度比估計應用於高維資料下之離群值偵測—基於子空間搜尋方法
D3	黃以新	Statistical Inference for a Contaminated Zero-Inflated Compound Poisson–Gamma Process with Applications
D4	湯東翰	具動態零壹膨脹的離散整數型時間序列模型之貝氏推論
D5	陳志恆	Multiple Imputation Confidence Intervals for the Ratio of parameters with Missing Data: Binomial and Poisson Cases
D6	郭育維	Confidence Intervals for the Proportion Difference in Matched-Pair Binary Data with Missing Values
D7	張聖岳	Flexible Spatial Modeling with Variable Selection: A Deep Neural Network-Based Higher-Order Nonparametric Spatial Autoregressive Framework

Changepoint Detection for Zero-Inflated Poisson Processes via PELT

戴毓祥

國立中央大學統計研究所

摘要

This study employs the PELT method for detecting changepoints in time series data. We model the data using the Copula Markov Zero-Inflated Poisson (ZIP) framework, where parameters are estimated via Maximum Likelihood Estimation (MLE) using the Newton-Raphson method. To capture different dependence structures, we utilize both the Clayton Copula and the Joe Copula. Furthermore, we conduct extensive simulation studies to evaluate the performance of our approach. Specifically, we vary simulation settings, such as dependence strength and sample size, to assess their impact on the accuracy of both changepoint detection and parameter estimation.

相對機率密度比估計應用於高維資料下之離群值偵測— 基於子空間搜尋方法

黃芯鈺

國立成功大學統計學系

摘要

機率密度比 (probability density ratio) 可用來描述兩組資料分布之間的相對關係，降低分別估計兩個機率密度函數 (probability density function) 所造成的誤差，且能廣泛應用於離群值偵測 (outlier detection) 等問題。針對高維資料，本研究提出結合子空間搜尋與 α 相對機率密度比 (α -relative density ratio) 之估計方法，以同時達到降維與穩定估計。此方法透過子空間搜尋降低高維資料下之估計困難、減少不相關變數對估計結果的影響，並採用具上界 (upper bound) 性質的 α 相對機率密度比函數，避免分布差異較大時機率密度比估計值發散，進而提升估計的數值穩定性。最後，將所提之方法應用於離群值偵測。模擬研究結果顯示，本文方法在不同資料維度與離群樣本比例設定下，皆具有良好的辨識能力與估計穩定性。此外，本文以身體組成測量資料進行實證分析，驗證所提之方法的有效性與實務應用價值。

Statistical Inference for a Contaminated Zero-Inflated Compound Poisson–Gamma Process with Applications

黃以新
國立中央大學統計研究所

摘要

This study considers the estimation problem for a contaminated zero-inflated compound Poisson–Gamma (ZiCPG) process. The contaminated ZiCPG model provides a flexible framework for modeling the dynamics of stochastic processes with jumps by simultaneously incorporating zero inflation and random jump sizes into a Poisson process with noise contamination. Consequently, computing the maximum likelihood estimator is challenging. To address this issue, we propose an effective procedure based on jump-state identification and particle filtering to estimate latent jumps and mitigate the computational challenges arising from noise contamination. A prediction model is further developed based on the estimated jumps. Numerical results from various simulation scenarios demonstrate that the proposed estimation scheme achieves satisfactory performance. Moreover, when applied to an electricity consumption dataset, the proposed prediction model yields superior performance in forecasting potential abnormal electricity consumption compared with competing methods.

具動態零壹膨脹的離散整數型時間序列模型之貝氏推論

湯東翰

逢甲大學統計學系

摘要

隨著監測技術的普及，具備狀態受限 (Bounded) 與零壹膨脹 (Zero-and-One Inflation) 特性的計數型時間序列近年來成為研究重點。本研究旨在建構並改進具 Poisson 與負二項分配的受限零壹膨脹狀態的模型，透過負二項分配具備額外的分散參數之特性，強化對資料過度離散 (Over-dispersion) 狀態的捕捉能力。在模型架構方面，透過 INGARCH(1,1) 結構描述計數型資料的時間相關性，並創新引入羅吉斯函數 (Logistic Function) 動態捕捉膨脹參數的時間變動特性，以靈活描述計數型時間序列資料的變動。本研究在貝氏推論框架下使用馬可夫鏈蒙地卡羅法 (MCMC)，並結合 MH 演算法進行參數估計，同時監測參數估計的軌跡與自相關性，確保收斂效率與精確度。本研究最後以台灣空氣品質 AQI 等級日資料進行實證分析，並使用多項預測評估指標判斷預測效果，且結果顯示本研究提出的動態模型處理偏態與高度離散資料之優勢，可做為空氣品質預警監測的參考工具。

Multiple Imputation Confidence Intervals for the Ratio of parameters with Missing Data: Binomial and Poisson Cases

陳志恆
國立成功大學統計學系

摘要

Confidence interval estimation for the ratio of parameters is widely applied in clinical trials and epidemiological studies. While the MOVER-R (method of variance estimates recovery for the ratio) provides an effective framework for these estimations, its application is often limited by missing data issues commonly encountered in practice. Integrating multiple imputation with MOVER-R, this study develops a new procedure for constructing confidence intervals for the ratio of parameters under independent binomial and poisson distributions, specifically addressing data that are missing completely at random and missing at random. The simulation study shows that the coverage probabilities of the proposed intervals remain stable and close to the nominal level. Furthermore, this study summarizes optimal recommendations for the application of each interval across various data scenarios. Finally, the practical utility of the proposed intervals is validated through illustrative examples using real-world data.

Confidence Intervals for the Proportion Difference in Matched-Pair Binary Data with Missing Values

郭育維
國立成功大學統計學系

摘要

Matched-pair binary data are frequently encountered in clinical trials, where the difference in proportions is often the primary focus of interest. In practice, data missingness often restricts analysis. Within the framework of multiple imputation (MI), existing interval estimation methods for such missing data still suffer from insufficient coverage probabilities, a problem that is particularly pronounced when parameters approach their boundaries. To address these limitations, this study develops and compares several improved interval estimation procedures within the MI framework. Simulation results show that the new method proposed in this study effectively mitigates the issue of undercoverage, demonstrating the most stable overall performance and outperforming existing methods. Furthermore, while maintaining satisfactory coverage probabilities, the proposed procedure yields shorter average interval widths, thereby providing more precise interval estimates. Finally, the practical value of these methods is demonstrated through the analysis of real-world datasets.

Flexible Spatial Modeling with Variable Selection: A Deep Neural Network-Based Higher-Order Nonparametric Spatial Autoregressive Framework

張聖岳

國立成功大學統計學系

摘要

Deep neural networks have received increasing attention because of their strong predictive performance and flexibility in modeling complex nonlinear relationships. This thesis investigates the variable selection problem in a higher-order nonparametric spatial autoregressive model with nonparametric endogenous spatial effects. Unlike conventional spatial autoregressive models that rely primarily on prespecified spatial weight matrices, the proposed framework employs deep neural networks to learn multiple spatial weighting mechanisms from the data and to model nonlinear endogenous spatial effects and nonlinear main effects in a unified architecture.

The proposed model consists of a spatial branch and a main-effect branch. The spatial branch is constructed through multiple learned spatial weighting mechanisms and their corresponding nonparametric endogenous effect functions, allowing the model to accommodate complex and heterogeneous spatial dependence. The main-effect branch combines a linear skip component with a nonlinear neural network component. To achieve parameter learning and variable selection simultaneously, an ℓ_1 penalty and a hierarchical structured sparsity constraint are imposed on the main-effect branch, such that a variable excluded from the linear skip component is also removed from its associated nonlinear representation. A regularization-path training procedure with a hierarchical proximal update is further developed for model estimation and feature selection.

Simulation studies show that the proposed method can correctly identify important variables while maintaining competitive predictive performance under nonlinear and multiple spatial dependence structures. An empirical analysis using the California Housing Prices data further demonstrates that the proposed framework achieves improved predictive performance and identifies meaningful explanatory variables compared with a fixed-spatial-

weight benchmark. Overall, the proposed method integrates flexible spatial mechanism learning, nonparametric endogenous spatial effects, and interpretable variable selection within a unified deep learning framework.

Keywords: Variable selection; Higher-order nonparametric spatial autoregressive model; Nonparametric endogenous spatial effect; Deep neural networks; Spatial heterogeneity

海報分組摘要

E 組 (13:00 – 14:40)

編號	報告人	題目
E1	廖虹琳	GEPINet：結合生物通路資訊之基因－環境交互作用高維癌症存活分析深度學習架構
E2	葉宇倫	A Copula-Based Model for Residual Dependence between a Right-Censored Event Time and an Ordinal Variable
E3	李其郡	測量誤差下之多元羅吉斯迴歸影響點診斷
E4	余梓平	Weighted Cumulative Distribution Functions for Nonparametric Inference via Empirical Likelihood
E5	何嘉元	2019-2020 年台灣勞工的睡眠時長與性別、甲狀腺促進素之相關性研究
E6	仇偉恩	高維度癌症基因體資料中特徵篩選與存活模型評估之系統性架構

GEPINet：結合生物通路資訊之基因－環境交互作用高維 癌症存活分析深度學習架構

廖虹琳
國立中正大學數學系

摘要

動機 (Motivation)

準確的存活預測對於癌症預後評估與精準醫療具有重要意義。Cox 比例風險模型因具備良好的可解釋性，至今仍被廣泛應用於存活分析；然而，其在捕捉非線性效應以及處理高維基因體資料方面仍存在限制。近年來雖然深度存活模型在預測效能上有所提升，但多數方法忽略了基因－環境交互作用 (gene - environment interaction, GxE)，且缺乏具統計基礎之特徵篩選機制，進而影響模型之穩定性與生物學解釋能力。

研究結果 (Results)

本研究提出 GEPINet (Gene - Environment Pathway Interaction Network)，一種結合統計篩選、生物知識與深度神經網路之通路導向存活分析架構。GEPINet 首先利用逆機率設限加權 (inverse probability of censoring weighting, IPCW) 調整之 Kendall 偏相關係數，在高維資料情境下篩選與存活結果相關之基因－環境交互作用特徵。接著，將篩選後之基因映射至 Gene Ontology (GO) 生物通路，並透過資料驅動 (data-driven) 之 bottom-up 策略建構通路層級結構，以形成具有生物學一致性之特徵聚合表示。最後，將通路表示與臨床變數共同輸入至以 Cox 部分概似函數為最佳化目標之深度神經網路中進行建模。本研究於多個異質性癌症資料集上評估 GEPINet，包括肺腺癌、結直腸癌、食道癌以及頭頸部鱗狀細胞癌。實驗結果顯示，GEPINet 在 concordance index、風險分層能力以及模型穩定性等指標上，皆優於傳統 Cox 模型與現有深度存活分析方法。此外，所提出之架構在跨癌別分析中亦展現出良好的泛化能力，顯示其能有效捕捉不同癌症間共享之預後訊號。

結論 (Conclusion)

GEPINet 提供一套兼具統計嚴謹性、生物學解釋性與深度學習彈性之整合式存活分析架構。透過納入基於 IPCW 之基因-環境交互作用篩選，以及資料驅動之 bottom-up 生物通路建構策略，能有效提升模型之預測效能與穩健性。研究結果顯示，GEPINet 不僅能改善高維癌症存活預測表現，亦能提供具生物學意義之通路層級資訊，對癌症預後分析、精準醫療與臨床決策具有潛在應用價值。

A Copula-Based Model for Residual Dependence between a Right-Censored Event Time and an Ordinal Variable

葉宇倫
國立中山大學應用數學系

摘要

In many applications, a right-censored event time and an ordinal variable are commonly analyzed separately, or the ordinal variable is simply incorporated into the survival model only through regression effects. Such approaches do not directly account for residual dependence between the two variables, and ignoring this dependence may lead to biased results. In this work, we propose a copula-based joint model for a right-censored event time and an ordinal variable, combining a Cox proportional hazard model for the event time margin with a cumulative logit model for the ordinal margin. The dependence between the two margins is modeled through a parametric copula family. A Breslow-type estimator is used for the baseline hazard, and all finite-dimensional parameters are estimated jointly by maximum likelihood. The proposed model also yields time-varying class probability and prediction scores that can be used for cumulative/dynamic time-dependent receiver operating characteristic curve (ROC) analysis. The finite-sample performance of the proposed method is investigated through simulation studies. Real data examples are analyzed for illustrations.

測量誤差下之多元羅吉斯迴歸影響點診斷

李其郡

輔仁大學統計資訊學系

摘要

羅吉斯迴歸模型廣泛應用於流行病學與社會科學等領域，然而實務資料常面臨自變數含有測量誤差以及異常觀測值並存的挑戰。若忽略測量誤差，傳統分析將導致參數估計偏誤；而影響點則可能主導推論結果。儘管文獻上針對羅吉斯迴歸的測量誤差校正與影響點診斷已有個別探討，但結合兩者之研究仍相對匱乏。因此，本研究旨在建立含異質性測量誤差之羅吉斯迴歸模型，提出基於蒙地卡羅期望最大化 (MCEM) 演算法的參數估計與影響點診斷架構。在方法論上，本研究採用異質性測量誤差模型，利用 MCEM 演算法克服積分無閉式解之困難以進行參數估計。為解決運算效率問題，本研究應用一步近似法推導單點刪除與群組刪除後的參數變動近似解，並據此建構多種診斷指標，包括廣義 Cook's Distance (GD) 與修正 Cook's Distance (MD)、廣義 DFFITS (GDF) 與修正 DFFITS (MDF)、GCD.GSPR、mCD* 以及推廣之標準化偽判定係數 (standard pseudo R^2)。模擬研究結果顯示，相較於忽略誤差的 Naïve 方法，MCEM 估計法可降低參數估計偏誤；在影響點診斷效能方面，透過 $1 - AUC$ (AUC^*) 多數決勝率與平衡準確率 (BA) 之評估發現：在單點刪除架構下，MCEM 基礎之診斷指標整體上較 Naïve 具有較佳的影響點鑑別能力；常見表現較佳之指標包括 GCD.GSPR、GDF 與 MD。然而，在群組刪除架構下，整體結果多由 Naïve 取得較高 AUC^* 。進一步依汙染型態比較可知，MCEM 的優勢主要集中於測量誤差變異異常情境；無論採用單點刪除或群組刪除，MCEM 均明顯呈現更高的鑑別勝率，突顯其校正效益。

Weighted Cumulative Distribution Functions for Nonparametric Inference via Empirical Likelihood

余梓平
國立成功大學統計學系

摘要

This study develops an empirical likelihood-based nonparametric testing method for statistical functionals of an unknown distribution. The proposed approach constructs functional constraints using a weighted Gaussian kernel cumulative distribution function and incorporates them into the empirical likelihood ratio framework. Both the mean functional and the quantile functional are considered, with the median used as a representative quantile case. Simulation studies under Weibull, Gamma, and Lognormal distributions are conducted to compare the proposed method with an existing density-based empirical likelihood method. The results show that the proposed method maintains reasonable Type I error control as the sample size increases. It also achieves higher empirical power in the median functional case and comparable or slightly better power in the mean functional case, demonstrating its effectiveness for nonparametric inference on distributional functionals.

2019-2020 年台灣勞工的睡眠時長與性別、甲狀腺促進素之 相關性研究

何嘉元
輔仁大學統計資訊學系

摘要

以 2019 至 2020 年台灣勞工健檢資料，探討睡眠時長與甲狀腺促進素異常之關聯。利用勝算比比較各睡眠分組對 tsh 異常的關聯性。並納入生理變數（性別，年齡，BMI）交互觀察。

高維度癌症基因體資料中特徵篩選 與存活模型評估之系統性架構

仇偉恩

國立中正大學數學系

摘要

高維度基因體存活分析常受到 ($p \gg n$) 問題的挑戰，導致模型不穩定與過度配適。在此背景下，特徵篩選與模型選擇對於提升預測表現與模型穩健性具有關鍵作用。本研究旨在建立一套整合性分析架構，系統性評估高維度癌症基因體資料中特徵篩選方法與存活預測模型之組合效果。

本研究考量四種具代表性的特徵篩選方法，包括 Partial Likelihood (PL)、Feature Aberration at Survival Times (FAST)、Composite Robust Censored Distance Correlation Screening (CRCDCS) 以及 Network-adjusted IPCW Kendall's tau (NPN-MB)，分別涵蓋參數式、排序式、距離相關式與網路資訊調整式策略。上述方法進一步結合四種存活模型，包括 Minimax Concave Penalty (MCP) 懲罰 Cox 迴歸、CoxBoost、Oblique Random Survival Forest (ORSF) 以及 DeepSurv。透過多個 TCGA 癌症世代資料，本研究系統性比較不同「特徵篩選 × 模型」組合之表現。模型效能於測試資料上以 concordance index (C-index)、Integrated Brier Score (IBS)、deviance 及 log-rank test 進行評估，以衡量模型之判別能力、預測誤差與風險分層能力。

研究結果顯示，特徵篩選方法會顯著影響後續模型之預測表現，而不同存活模型對資料結構亦呈現不同程度的敏感性。特別是結合能考量變數相依性的特徵篩選方法與非線性存活模型之組合，在多數資料集中皆能取得較佳表現，顯示整合式建模策略之優勢。

本研究提供一套統一的分析架構，用以評估特徵篩選與存活模型之間的交互影響，並為高維度癌症存活分析中的方法選擇提供實務參考，進一步支持其於精準醫療領域之應用。

海報分組摘要

F 組 (13:00 – 14:40)

編號	報告人	題目
F1	謝佳穎	ESG 與信用評等對投資的影響與財務因子探討
F2	曾俊雄	加密貨幣市場風險結構與動態傳導機制： 異常值分解下的多層網絡解析
F3	陳衮綸	基於尾部風險管理的多因子策略實證分析： 以台股市場為例
F4	陳珈羽	基於集成學習之函數型時間序列預測方法
F5	林恩平	Joint Tail Risk Forecasting for High-Volatility Assets Based on Copula Semiparametric Framework
F6	吳家齊	基於技術分析指標信號之選股策略
F7	江芄幟	在貝氏框架下的零壹膨脹有序型時間數列 模型之股價報酬率區間建模應用

ESG 與信用評等對投資的影響與財務因子探討

謝佳穎

國立高雄大學統計學研究所

摘要

本研究旨在探討環境、社會與治理（ESG）評等和信用評等對於公司投資表現之影響。我們的核心問題聚焦在不同評等間與股票表現的關聯性、如何利用評等選取投資標的、並討論影響評等的相關財務因子等等。為解決上述問題我們分別使用決策樹分析與 Duncan 多重比較檢定，以各評等的夏普比率（Sharpe Ratio）為目標進行分群，並且搭配 Jonckheere-Terpstra 趨勢檢定排序績效，藉此篩選出最佳評等區間範圍，找出股票與評等間的關聯性；同時，我們想解決決策樹和 Duncan 多重檢定兩方法優劣性，本研究使用等權重（Equal Weight）與風險平價（Risk Parity）建立投資組合，採滾動式抽樣進行拔靴法（Bootstrapping）模擬，比較兩方法超額績效並確立「優良組」評等範圍。為了探討財務因子與評等之間的關聯性，我們運用順序式邏輯斯迴歸（Ordinal Logistic Regression）輔助分析，使用雙向逐步迴歸搭配 AIC 準則，其中研究納入獲利與營運能力、財務安全性與資金配置、市場評價與流動性和投資風險與波動控管等四大類財務指標，建構模型以識別顯著財務因子。本研究以 2020 至 2022 年間美國股票市場之 400 家上市公司為實證樣本。實證顯示，決策樹分析與 Duncan 檢定皆能有效識別顯著高績效資產，且投資組合的表現都明顯優於大盤與隨機抽取（naïve），績效最高超越對照組達 200%，成功將信用評等與 ESG 應用於投資組合；迴歸結果揭示，波動風險與平均交易量皆為兩種評等財務顯著變數。

最後我們探討僅透過評等挑選資產，或直接使用財務因子挑選資產，這兩者間對投資組合績效的異同性並總結原因。未來可以研究在分散投資時，如何有效結合信用評等和 ESG 評等，衡量風險以建構投資組合。

關鍵詞：ESG、信用評等、夏普比率、風險平價、等權重法則、投資組合、決策樹分析、Duncan 多重比較檢定、Jonckheere-Terpstra 趨勢檢定、拔靴法模擬、順序式邏輯斯迴歸

加密貨幣市場風險結構與動態傳導機制： 異常值分解下的多層網絡解析

曾俊雄
國立中山大學應用數學系

摘要

This paper develops a decomposition-based network framework to analyze persistent versus episodic contagion in cryptocurrency markets. Using high-frequency data for 16 major cryptocurrencies from 2019 to 2026, we apply a robust iterative procedure to decompose return and volatility series into regular components, episodic extreme shocks, and persistent level changes. This design allows for a clear distinction between stable market architecture and event-driven transmission channels. Our results reveal that cryptocurrency contagion is multi-layered: a stable backbone of persistent links represents the market's regular information-sharing structure, while major events activate episodic channels through temporary spillovers and synchronized shocks. Importantly, the assets central to jump episodes differ from those dominant in regular dependence networks. The framework provides a structured approach to distinguish the market's underlying architecture from temporary contagion paths, offering clearer insights into risk management during both normal and crisis times.

基於尾部風險管理的多因子策略實證分析： 以台股市場為例

陳衾綸

國立高雄大學統計學研究所

摘要

傳統投資績效常以風險價值 (VaR) 或預期損失 (ES) 作為評估指標，以衡量其風險部位。本研究希望結合這兩種尾部風險指標，探討其對投資組合績效的實證影響，並透過整合控管極端風險，使投資組合在各種景氣環境下，皆能維持預期內的回報與風險水準。為了具體檢驗此一構想，本文以台灣上市股票市場為例，採用歷史回測與滾動視窗分析，建構一套整合 VaR、ES 及 VIX 指數的複合投資策略。在研究設計中，VaR 與 ES 模型用以捕捉尾部風險，並引入 VIX 指數作為市場環境的動態濾網，在系統性風險上升時靈活調整投資組合曝險程度。實證結果顯示，結合 VaR 與 ES 的風險控管策略具備顯著的景氣循環適應力。在經濟擴張期，該策略能捕捉市場動能並創造超額報酬，例如一年期配置下風險濾網策略平均報酬達 28.05% (基準大盤為 9.11%)；在經濟收縮或震盪期，則有效抑制下檔風險，平均最大回撤穩定於 -10.6% (基準大盤約 -18.1%)，兩年長週期測試更將大盤 -48.48% 的最大回撤收斂至 -18.23%。本研究證實，此類納入尾部風險管理的複合策略，能改善傳統模型的限制，在多變的經濟環境下，提供具長期穩定風險調整後報酬的資產配置參考。

基於集成學習之函數型時間序列預測方法

陳珈羽

淡江大學統計與資料科學學系

摘要

隨著資料蒐集技術的進展，許多觀測資料可視為隨時間連續變化的曲線，此類資料稱為函數型時間序列（functional time series, FTS），常見於死亡率、電力負載、空氣品質與金融市場等應用領域。傳統 FTS 預測方法通常先透過函數型主成分分析（functional principal component analysis, FPCA）進行降維，再對主成分分數序列建立時間序列模型。然而，在資料具有結構變化、異質性或模型選擇不確定性的情況下，單一預測模型之表現可能受到限制。本研究以集成學習概念為基礎，探討函數型時間序列之預測方法。透過整合多個候選模型之預測結果，本研究期望降低單一模型選擇所造成的不確定性，並提升方法在不同資料情境下的適用性。為評估所提方法之預測表現，本研究將設計數值模擬，並比較其與現有方法在不同資料結構下的預測表現。研究結果可望為函數型時間序列預測提供一個具彈性且易於實作的分析架構。

Joint Tail Risk Forecasting for High-Volatility Assets Based on Copula Semiparametric Framework

林恩平
逢甲大學統計學系

摘要

This study proposes a copula-based bivariate ES-CAViaR framework to jointly forecast value-at-risk (VaR) and expected shortfall (ES), and to investigate inter-asset dependence in downside risk. For the marginal dynamics, we consider four ES-CAViaR-type specifications: the baseline ES-CAViaR model; ES-CAViaR-st, which redefines the exceedance term as a ratio; ES-CAViaR-v, which incorporates Rogers-Satchell volatility; and ES-CAViaR-R, which incorporates Garman-Klass volatility. To capture cross-asset dependence, we adopt the Clayton and the BB1 copulas based on pseudo-observations. The Clayton copula is well suited for modeling lower-tail dependence and assessing joint extreme downside risk, whereas the BB1 copula captures both lower- and upper-tail dependence, allowing us to examine asymmetric tail behavior. Pseudo-observations are used to map the return series into the copula domain. We estimate the unknown parameters using an adaptive Markov chain Monte Carlo (MCMC) scheme. In the burn-in phase, we employ a random-walk Metropolis algorithm, and then switch to an independent-kernel Metropolis-Hastings sampler for the subsequent draws. In the empirical study, we employ the copula-based bivariate ES-CAViaR framework to the cryptocurrency market using Bitcoin and Ethereum. We generate rolling-window forecasting, one-step-ahead out-of-sample forecasts at the 1% level over a five-year evaluation period. Forecast accuracy and coverage are examined via the violation rate (VRate), conditional violation probability, and backtest methods. Forecast performance is further evaluated using scoring criteria. The results indicate that the ES specification redefining exceedances as a ratio yields more robust joint VaR and ES forecasts, with the improvement most pronounced during periods of elevated downside risk. The Clayton copula captures lower-tail dependence, while the BB1 copula provides a more flexible structure, revealing distinct patterns in upper- and lower-tail dependence in the

empirical study. The conditional VaR violation probabilities summarize the co-occurrence of violations, with differences across conditional proportions primarily driven by marginal violation frequencies rather than directional effects.

基於技術分析指標信號之選股策略

吳家齊

國立高雄大學統計學研究所

摘要

本研究旨在探討技術分析指標信號對投資績效之影響。有別於傳統上信號的應用方式，我們進一步將信號區分為單日與多日；此外，考量到同一交易日中常有多檔資產同時發生黃金交叉，因此會依照不同信號強度的排序方式來建立不同的選股策略。首先，根據當日黃金交叉的股票中按照選股策略挑選資產，並且配合資產配置策略建構投資組合。其次，當持有資產發生死亡交叉時，本研究將後續交易方式會區分為遞補以及不遞補兩種模式。實證資料以台灣 2000 年至 2024 年間各產業隨機挑選一半每日收盤價資料，以模擬台灣市場，其中包含 31 種產業，共 479 家公司。投資期間設定半年、一年以及兩年，觀察各個策略在短、中、長期的投資表現，並且以夏普比值（Sharpe Ratio）和索提諾比值（Sortino Ratio）作為衡量指標，歸納出適合投資人的策略。研究結果發現，採用信號變化量極小值策略的投資表現普遍優於採用極大值之策略。其中，KD 指標在短、中、長期投資下，均為多日信號變化量極小值配合遞補以及等權重策略表現最好；MACD 指標在短、中期投資，則是以單日信號變化量極小值配合不遞補以及均值-變異數模型，而長期投資以單日信號變化量極小值配合均值-變異數模型，不論是在遞補或是不遞補的表現均不錯。

關鍵字：技術分析指標、投資組合、選股策略

在貝氏框架下的零壹膨脹有序型時間數列模型之 股價報酬率區間建模應用

江芄幟
逢甲大學統計學系

摘要

近年來隨著人工智慧科技產業的蓬勃發展，金融市場的交易也越顯熱絡，吸引許多的投資人及民眾紛紛投入大量資金到股票市場中，但同時，近年來在全球部分地區局勢動盪的干擾因素之下，也造成股票市場發生突發性的劇烈變化。因此，如何透過有效的風險評估工具來降低投資的損失，已是投資人長期關注的議題。在金融實務與風險管理中，投資人為了能快速掌握金融市場的變化，以做為先期策略的制定，往往關注的重點並非精確數值的預測，而是能有效且快速掌握趨勢的風險區間預測。本研究擬將連續的股價報酬轉換為離散且具有順序性的報酬區間整數型變數，同時考慮此整數型變數的區間變化幅度，以及正報酬或負報酬之方向性建模。由於整數型時間序列模型近年來廣泛應用於統計、經濟與財務領域，其中又以 INGARCH 模型 (Integer-Valued Generalized Autoregressive Conditional Heteroskedasticity Model) 最廣為使用，本研究將應用此 INGARCH 模型在區間幅度變化的建模上，並且針對正、負報酬的方向變動機率，考慮邏輯斯函數的建模。然而，考量在每日的金融交易實務中，幅度變數的變化通常為小幅波動或接近零報酬的變化，具有高度的集中特性，因此，本研究在幅度變數的分配假設中納入零壹膨脹 (zero-one inflation) 的結構，並且考慮具備捕捉過度離散(over-dispersion)變化的 Beta Negative-Binomial 分配，來進行股價報酬區間變化的預測。在模型的參數估計中，本研究採用貝氏分析方法結合馬可夫鏈蒙地卡羅 (Markov chain Monte Carlo) 抽樣，以獲得模型的參數估計值，進而對股價報酬區間進行預測。本文透過模擬研究來驗證所提方法的有效性，最後，並且利用美國股票市場的個股股價變動做為實例，來觀察本研究所提方法的實證分析結果。

海報分組摘要

G 組 (13 : 00 – 14 : 40)

編號	報告人	題目
G1	趙睿群	未測量混淆下利用代理變數之隨機介入效果的識別與估計
G2	黃纓婷	A Semi-Reference Bayesian Approach to Cell-Type Annotation and Marker Discovery in Single-Cell RNA-Seq
G3	黃子維	有序 Probit 迴歸模型下之貝氏稀疏群組選擇方法
G4	修子恆	Hierarchical Bayesian Estimation of Fatigue Limit Models from Small Samples
G5	朱宸儀	Proximal Causal Inference for a Path-Specific Effect in the Presence of a Recanting Witness
G6	王俊傑	Maximum Likelihood and Bayesian Inference for Finite Mixtures of Contaminated Normal Linear Mixed-effects Models

未測量混淆下利用代理變數之隨機介入效果的識別與估計

趙睿群

國立清華大學統計與數據科學研究所

摘要

因果中介分析常面臨兩項挑戰：其一，自然直接效應與自然間接效應的識別依賴跨世界獨立性假設，而此假設無法透過實證驗證，也無法由任何研究設計保證；其二，未測量混雜因素的存在會破壞標準識別條件。基於此，本文以隨機介入效應作為目標參數，在存在未測量混雜因素且暴露變數為二元的情況下，進行效應的識別與估計。相較於自然效應，隨機介入效應作為替代性效應參數，不直接依賴自然效應所需的跨世界獨立性假設。本文透過近端因果推論框架，引入一對代理變數，並利用橋接函數建立識別結果；在此基礎上，提出於半參數模型下對應的有效影響函數。估計方面，除了根據有效影響函數的多重穩健估計量外，進一步發展目標最大概似估計量，其透過目標化更新步驟，構造出相對應的代入式估計量；此外，並引入深度學習方法估計橋函數，以處理橋接函數求解所涉及的逆問題。

A Semi-Reference Bayesian Approach to Cell-Type Annotation and Marker Discovery in Single-Cell RNA-Seq

黃纓婷

國立成功大學統計學系

摘要

Single-cell RNA sequencing technology has become a vital tool for understanding cellular diversity and discovering new types of cells in complex biological systems. A fundamental step in this process is cell type annotation. This procedure involves identifying the identities of cells in a new dataset, known as the query dataset, by comparing them to a reference that provides known cell type information. However, a major challenge arises when the query dataset contains cells that are not represented in the reference. Traditional methods are often limited by forced classification, where they must assign a known label to every cell, leading to wrong annotations. Even when newer tools attempt to address this problem, they often rely on simple filtering methods based on low confidence scores. Because these methods lack a precise way to categorize cells that not present in the reference, it becomes difficult to distinguish between meaningful biological groups and low-confidence cells. To address these limitations, we distinguish between two types of cells that are not present in the reference. First, unknown cell types represent novel biological populations that are not yet recorded in the reference. Second, unassigned cells are defined as low-confidence cells that cannot be reliably assigned to either known reference cell types or unknown cell types. This lack of confidence often stems from poor experimental or the presence of undefined subtypes. Following this logic, we propose BaySCA, a novel semi-reference Bayesian approach that integrates hierarchical modeling with marker gene discovery. We develop a Markov chain Monte Carlo procedure via Gibbs sampling to simultaneously estimate cell-type-specific effects and identify marker genes for both known reference cell types and unknown cell types. A key feature of BaySCA is its ability to identify multiple unknown cell types simultaneously. By treating unknown populations as distinct groups rather than a single category, BaySCA enables a clearer identification of multiple new cell types and their specific

marker genes. Furthermore, BaySCA incorporates a posterior-based rejection threshold to effectively distinguish and filter out unassigned cells from structured cell types. To accommodate broader scenarios, we also extend this framework into a reference-free clustering model. Through simulation studies and real data analysis, we compare BaySCA with several benchmarking methods, including SingleR, scmap-cell, CaSTLe, SCINA, and scSorter. Our results demonstrate that BaySCA outperforms existing tools, particularly in the robust detection and classification of unknown cell types.

有序 Probit 迴歸模型下之貝氏稀疏群組選擇方法

黃子維

國立清華大學統計與數據科學研究所

摘要

針對反應變數為次序變數 (ordinal variable) 之情境，本研究提出一套應用於有序 Probit 迴歸模型之貝氏稀疏群組 (sparse group) 選擇方法。傳統的變數選擇方法多侷限於評估單一變數是否應納入模型；然而，在實務資料中，解釋變數往往具備自然的群組結構或存在高度相關性。倘若僅針對個別變數進行篩選，極易導致選擇結果欠缺穩定性與模型解釋力。基於上述考量，本研究於有序 Probit 迴歸模型之架構中，導入群組與變數雙層級之指標變數，藉此分別鑑別具備解釋力之變數群組，以及該群組內真正具關鍵影響力之核心變數。本方法透過馬可夫鏈蒙地卡羅方法對指標變數之後驗樣本進行抽樣，並進一步估算其後驗納入機率，以執行精確之群組與變數選擇。模擬實驗結果證實，本研究所提之方法於多樣的相關結構、訊號強度及高維度資料情境下，皆能有效識別關鍵群組與重要變數，同時維持優異之分類與預測效能。最後，本研究亦透過實際資料分析，進一步實證該方法之實務應用價值與整體有效性。

Hierarchical Bayesian Estimation of Fatigue Limit Models from Small Samples

修子恆
國立成功大學統計學系

摘要

In the reliability assessment of engineering materials, accurate fatigue life prediction and S-N curve establishment are essential for structural integrity. However, high-cycle fatigue testing is often limited by high costs and long durations, resulting in small sample sizes. Although Hong (2025) proposed the M2 and M3 frameworks using the Limited Failure Population (LFP) concept to handle censored data, traditional Maximum Likelihood Estimation (MLE) still suffers from systematic underestimation of the scale parameter in small-sample scenarios.

This study proposes a hierarchical Bayesian model (HBM) with a data-driven physical consistency prior centered on the fatigue limit. By treating the fatigue limit as the primary parameter and making the intercept and slope conditionally dependent on it, the model stabilizes parameter exploration. The prior automatically adjusts the distributional shape of the fatigue limit according to observed data, effectively controlling its proximity to the minimum observed failure stress and ensuring physical plausibility.

Empirical results on real metallic material fatigue datasets show that the proposed model effectively corrects the scale parameter underestimation bias of MLE. The resulting S-N curve confidence bands comprehensively account for hyperparameter uncertainty, providing more conservative and physically meaningful estimates. Overall, this study demonstrates a robust Bayesian framework that enhances the reliability of fatigue life prediction under limited data conditions.

Keywords: Hierarchical Bayesian Model (HBM), Fatigue Limit, Small-Sample Analysis, Physical Consistency Prior, S-N Curve.

Proximal Causal Inference for a Path-Specific Effect in the Presence of a Recanting Witness

朱宸儀

國立清華大學統計與數據科學研究所

摘要

A common challenge in causal mediation analysis is the presence of unmeasured confounding, which renders standard identification strategies invalid. A further complication arises when the treatment affects an intermediate variable, known as a recanting witness, that subsequently confounds the relationship between the mediator and the outcome, precluding identification of natural direct/indirect effects under conventional no-unmeasured-confounding assumptions. In this work, we consider identification and estimation of the path-specific effect of a binary treatment on an outcome through a mediator, in settings where both unmeasured confounding and a recanting witness are present. Building on the proximal causal inference framework, we establish four nonparametric identification results for the target estimand by leveraging a pair of proxy variables to recover information about the unmeasured confounder. We then derive the efficient influence function under a semiparametric model that imposes no restrictions on the observed data distribution beyond the existence of the relevant bridge functions, and propose a multiply robust estimator based on the efficient influence function that remains consistent whenever any one of four combinations of working models is correctly specified.

Maximum Likelihood and Bayesian Inference for Finite Mixtures of Contaminated Normal Linear Mixed-effects Models

王俊傑

國立成功大學統計學系

摘要

Finite mixtures of linear mixed-effects (FM-LME) models have become a commonly used tool for clustering longitudinal data with heterogeneity. However, the conventional assumption of normality in the FM-LME model may be sensitive to mild outliers or contaminated points. This thesis extends the FM-LME model by considering the multivariate contaminated normal (MCN) distributions for component random effects and within-subjects errors, referred to as the finite mixtures of contaminated normal linear mixed-effects (FM-CNLME) model henceforth. Comparing with the multivariate normal distribution, the MCN distribution has two extra parameters: one for controlling the proportion of mild outliers and the other for specifying the degree of contamination, improving the robustness of the estimation results. Regarding the inference of parameters as well as latent data in the model, we develop two approaches: maximum likelihood (ML) estimation and Bayesian inference. For carrying out ML estimation, an efficient alternating expectation conditional maximization (AECM) algorithm is provided. From a fully Bayesian viewpoint, Markov Chain Monte Carlo (MCMC) sampling procedure, which combines the Gibbs sampler and Metropolis-Hastings (M-H) algorithm, is developed to perform posterior inference. Numerical results show that, when dealing with complex longitudinal data that have grouped structures and contaminated observations, the FM-CNLME model can effectively identify hidden groups, achieves better fitness of data, and offer more accurate estimation of parameters as well as fitted responses compared with the existing model. Our proposed model and methods are applied to real-data examples to demonstrate their practical values and efficiency.

海報分組摘要

H 組 (13 : 00 – 14 : 40)

編號	報告人	題目
H1	黃天宇	兩層式拒絕分類架構於帕金森氏症診斷之研究
H2	陸曉芬	大型語言模型與協同過濾在電影推薦之應用研究
H3	郭軒丞	應用深度適應性袋形距離於深度局部中心分群之框架研究
H4	郭芳樞	半母數迴歸樹之變數選取方法
H5	林掄元	DeepOGS：結合重疊群組篩選與深度學習之穩健且具可解釋性的結直腸癌 CMS 分型方法
H6	林柏安	支持向量機之決策邊界的統計推論
H7	王思雯	類別型誤分類之校正方法比較

兩層式拒絕分類架構於帕金森氏症診斷之研究

黃天宇

國立高雄大學統計學研究所

摘要

本研究提出一個具有拒絕機制的兩層分類架構，其中第一層分類器使用決策樹，針對決策樹無法正確分類的樣本，進行下一步使用第二層分類器（如 SVM）配合特徵工程去雜訊、特徵篩選簡化模型、依賴型拒絕器避免錯誤分類之風險。實驗結果顯示，經過特徵工程與拒絕機制整合後，模型可在維持整體辨識能力的同時，提高對不確定樣本的保守性，因而更適合應用於需要兼顧準確性與安全性的醫學輔助診斷情境。

大型語言模型與協同過濾在電影推薦之應用研究

陸曉芬

國立中山大學應用數學系

摘要

近年來，推薦系統逐漸由傳統協同過濾走向結合深度學習與生成式模型之混合式架構。面對影音平台中大量且快速變動的內容，僅依賴使用者互動紀錄雖能掌握偏好趨勢，卻難以充分利用電影標題、描述與語意脈絡等文字資訊；若直接以大型語言模型進行推薦，又容易忽略序列行為中蘊含的時間依賴與個人化訊號。因此，本研究以 Amazon Movies & TV 資料集為例，提出一個結合協同過濾、序列建模與大型語言模型的兩階段推薦流程。本文在方法上的核心貢獻，集中於協同 - 文字對齊模組的損失設計。具體而言，本研究以既有的兩階段推薦架構為基礎，先使用 Self-Attentive Sequential Recommendation (SASRec) 提供穩定的序列偏好表示，再於協同 - 文字對齊模組中結合 pairwise Bayesian Personalized Ranking (BPR) 與對稱式 Information Noise-Contrastive Estimation (InfoNCE)，使協同過濾物品表示與 Sentence-BERT (SBERT) 文字嵌入在對齊時，同時保留推薦所需的排序性與跨模態對應所需的判別性；語言模型注入模組則再將對齊後的表示注入大型語言模型，透過提示模版生成最終推薦結果。實驗結果顯示，本研究所提出的 Stage 1 損失設計能有效支撐後續生成式推薦：採用 OPT-2.7B 作為語言模型骨幹，並對 100 組獨立抽樣（每組 10,000 位使用者）進行評估，Hit@1 平均達 0.6214 ± 0.0044 。相較於原始 A-LLMRec 以 OPT-6.7B 所得之 0.6237，兩者絕對差距僅為 0.0023（約 0.37%），而本研究所使用之語言模型參數量（2.7B）僅為原始方法（6.7B）之約 40%，顯示跨模態對齊設計能在更輕量的模型規模下維持相近的推薦效能。整體而言，本研究驗證了以「BPR + 對稱式 InfoNCE」重新設計 Stage 1 對齊損失的可行性，並顯示精確的跨模態對齊能在較小模型規模下維持相近的推薦效能，為後續建構可解釋且適於實際部署之生成式推薦系統提供基礎。

應用深度適應性袋形距離於深度局部中心分群之框架研究

郭軒丞

國立臺灣大學統計碩士學位學程

摘要

在群集分析中，對具備非凸幾何、嚴重密度異質性與非對稱結構的資料進行無監督劃分，始終是一項根本性的挑戰。本研究擴展深度局部中心分群（DLCC）演算法，以處理其在資料分派上的限制。為緩解由嚴重密度異質性引發的結構失衡，本方法基於鄰域重疊準則保留稀疏中心，並限制過渡群集的初始擴張，從而防止稀疏觀測值遭系統性誤派。針對未分派的觀測值，本程序相對於結構核心，評估連續的深度適應性袋形距離。此幾何適應性距離能有效捕捉局部非等向性特徵與非對稱結構。經數值實驗證實，本框架能維持非凸幾何中的結構連通性，並防止在嚴重密度異質性與結構非對稱下發生系統性分類錯誤。

半母數迴歸樹之變數選取方法

郭芳樞

國立中正大學數學系

摘要

本研究提出一種適用於群集資料之迴歸樹 (Clustered Response Generalized Regression Tree, CGRT) 方法。在群集資料中，同一群組內的觀測值常具有相關性。為了處理這種群內相關結構，我們採用 Liang and Zeger (1986) 提出的廣義估計方程式 (Generalized estimating equations, GEE) 作為樹模型節點分割與預測模型建構之基礎。在樹的建立上，我們參考 Loh (2002) 所提出的 GUIDE (Generalized, Unbiased Interaction Detection and Estimation) 的方法，分開選取切割變數及切割點。切割變數的選擇方式則是透過重抽樣 (permutation) 方式計算經驗 p 值 (empirical p-value)，再根據經驗 p 值的大小來決定切割變數。模擬研究結果顯示，CGRT 在切割變數的選取上不具有選取偏差且對切割變數亦具備不錯的辨識能力，在切割點選取上也具有良好的準確性。另一方面，進一步比較預測能力後發現，在真實資料的平均數結構並非廣義線性模型時，CGRT 在預測能力的表現會優於非樹狀的估計模型，包括廣義線性混和效果模型 (Generalized Linear mixed-effects model, GLME) 及 GEE 模型。此外，我們也將 CGRT 與 Fokkema et al. (2018) 提出的廣義線性混合效應模型之樹狀方法 (Generalized linear mixed-effects model trees, GLMM trees) 進行比較，模擬結果顯示 CGRT 相較於 GLMM trees 更具有強韌特性。最後，我們將所提出的迴歸樹方法應用到一筆高中生薪資的實際資料上。

關鍵字：群集資料、迴歸樹、廣義估計方程式、選取偏差、重抽樣

DeepOGS：結合重疊群組篩選與深度學習之 穩健且具可解釋性的結直腸癌 CMS 分型方法

林掄元
國立中正大學數學系

摘要

背景 (Background)

結直腸癌 (Colorectal Cancer, CRC) 具有高度的分子異質性，這使得精確分型與個人化治療面臨重大挑戰。共識分子亞型 (Consensus Molecular Subtypes, CMS) 框架將 CRC 分為四種標準化亞型 (CMS1 - CMS4)；然而，其應用仍受到若干限制，包括單一資料集中的預測穩定性不足、跨資料集泛化能力有限，以及對高維度基因表現資料中非線性交互作用建模不足。此外，現有方法往往需要在生物學可解釋性與預測效能之間進行取捨。

結果 (Results)

本研究提出一種多階段學習框架 DeepOGS，整合重疊群組篩選 (Overlapping Group Screening, OGS)、路徑導向特徵選擇，以及深度神經網路，用於結直腸癌 CMS 分型。DeepOGS 將多分類問題拆解為兩兩分類任務，以提升特徵辨識能力，同時在 OGS 架構下建模基因主效應 (main effects) 與基因間交互作用 (gene - gene interactions)。此外，本研究納入 Gene Ontology (GO) 路徑資訊，以進行具生物學意義的特徵選擇。在四個獨立的結直腸癌轉錄體資料集 (TCGA COAD/READ、GSE39582、GSE14333 與 GSE13294) 中，DeepOGS 在大多數情況下皆達到最佳或具競爭力的表現，分類準確率普遍超過 85%。在跨資料集分析中，該模型仍維持穩定效能，顯示其具有良好的泛化能力。進一步分析顯示，DeepOGS 所學得的潛在表徵 (latent representations) 與 CMS 類別具有良好一致性，且各類別之間具有明顯區隔。此外，模型亦能穩定辨識與神經訊號傳遞、離子運輸、血管新生，以及細胞外基質重塑等相關的重要生物機制。

結論 (Conclusion)

DeepOGS 結合統計式特徵篩選、生物路徑資訊與深度非線性建模，在提升結直腸癌 CMS 分型準確度的同時，也兼顧模型穩定性與生物學可解釋性。研究結果顯示，將結構化統計方法與深度學習結合，能有效解決高維度生醫資料分析中的挑戰，並具有延伸應用於其他癌症亞型分析的潛力。

支持向量機之決策邊界的統計推論

林柏安

國立高雄大學統計學研究所

摘要

支持向量機 (SVM) 的研究多聚焦於分類表現與參數調整，較少討論決策邊界位置本身的隨機性。本文從一維度出發，探討線性 SVM 決策邊界位置的分佈特性，並進行統計推論。本研究採用順序統計量，分別在硬間隔與軟間隔的情況下進行探討，包含理論推導、模擬驗證到實務單次資料判讀的分析流程，有助於理解 SVM 決策邊界的位置變動性與結果穩定性。

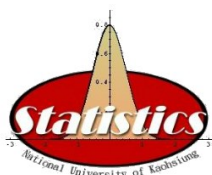
類別型誤分類之校正方法比較

王思雯

輔仁大學統計資訊學系

摘要

測量誤差存在於含有測量行為的資料中，若未妥善處理，可能將導致迴歸係數偏誤，進而影響模型的解釋力或預測能力。本研究透過模擬設計，有邏輯斯迴歸模型為基礎，針對類別型變數誤分類情境，設定不同誤分類結構，比較不同校正方法之估計偏誤、均方根誤差以及覆蓋率的差異，再進一步透過實際資料分析，以說明各方法在實務中的應用情形。



國立高雄大學統計學研究所

Institute of Statistics, National University of Kaohsiung

本所簡介

本所為配合國內統計研究與人才培育之需求而設立，主要設置目的如下：

● 提昇我國科技與經濟之發展

塑造我國成為具有高度競爭力的發展環境，並在生物科技、電子資訊等高科技產業，能夠跟上世界的研究水準，以及強化我國的經濟體質。從資料的收集、分析，到相關成效的評估，都需要統計人員的參與協助。將統計的理論與方法，適當地應用在各個領域中，對於我國未來科技水準的提昇與經濟的持續發展，均能產生極大的助益。

● 培養優秀統計人才

統計方法被廣泛使用，使得專業統計人才愈顯不足。良好的數據品質與正確的分析研判，均須有受過嚴格統計訓練的人員參與才能完成。本所致力培養具備良好溝通能力、易於與其他領域人員合作之統計人才，期能成為國內產業界之研發骨幹。

● 支援校內其他系所

統計學已普遍應用於各領域。本所的設置，除可提供本校其他系所教師研究上所需之統計諮詢，並使學生在大學時期，即能培養清晰的機率統計概念，瞭解隨機現象對日常生活的影響，並具備應用統計方法解決實際問題的基本能力。

● 教育目標

除基本統計理論外，本所依學生志趣、國家產業發展方向及教師專長，在工業統計、財金統計、網絡研究，與生醫科學等方面，開設適當的課程。訓練學生多元發展，並強化學生善用統計工具與解決問題的能力，為本所教師在教學及研究上的共同努力目標。此外，本所搭配「巨量資料研究中心」與「人工智慧與資料科學學程」，亦開設相關領域之課程，培養數據科學專業人才。

本所資訊

地址：81148 高雄市楠梓區高雄大學路 700 號

電話：07-5919362

Email：stat@nuk.edu.tw

統計所網頁





中華機率統計學會

The Chinese Institute of Probability and Statistics

地 址：300044 新竹市光復路二段 101 號

國立清華大學第三綜合大樓 8F

e-mail：cips@stat.nthu.edu.tw

電 話：(03)571-5131 ext 33173

學會網址



歡迎您的加入



地址：237303 新北市三峽區大學路 151 號

國立台北大學大數據與智慧城市研究中心

e-mail：jcsa39@mail.ntpu.edu.tw

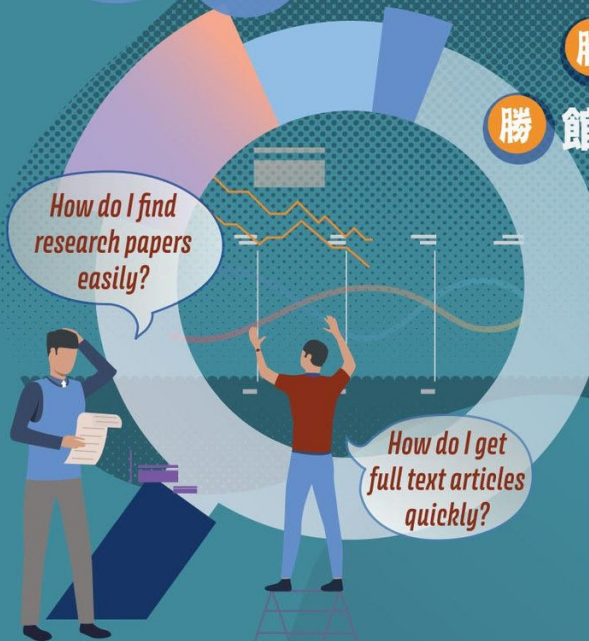
電話：(02)8674-1111 ext 66966

歡迎投稿

投稿方式：

請將論文 email 至 jcsa39@mail.ntpu.edu.tw，信件中註明論文作者及服務單位等相關資訊。投稿論文皆須經評審，評審過程為雙盲 (double-blind) 制。故在投稿的文稿中必須移除作者姓名及所有能提供作者身份的線索，包括未發表論文之引用及研究贊助來源等。此外，被接受的論文由中國統計學社在取得作者的非專屬授權後，進行出版。

統計文獻館際合作



- 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

DDS

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



統計科學研究所
Institute of Statistical Science

中央研究院

<https://www.stat.sinica.edu.tw/>

2026
新科院士演講暨統計高峰會議
Academia Sinica Statistical Summit

July 3^{FRI} 2026 Lecture Hall 1F 潤學廳
環境變遷研究大樓
Environmental Changes Research Building

Distinguished Lectures

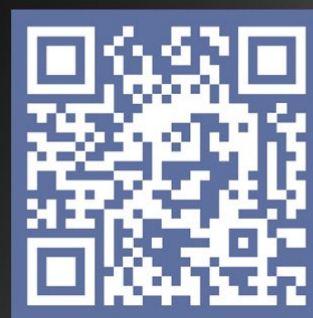
Tailen Hsing University of Michigan, U.S.A.		Featured Talk by Newly Elected Academician
Jianqing Fan Princeton University		Kung-Yee Liang Feng Chia University
Yu Shyr Vanderbilt University		Wah Chiu Stanford University



新科院士演講暨統計
高峰會議

[https://www3.stat.sinica.edu.tw/
2026newacademician/](https://www3.stat.sinica.edu.tw/2026newacademician/)

STATISTICAL SCIENCE CAMP
統計科學營
數據探索 無所不在
 2026.8.25(二)-8.26(三)
 人文社會科學館.國際會議廳.
 中央研究院 統計科學研究所主辦



統計科學營

<https://www3.stat.sinica.edu.tw/camp2026/>

STATISTICAL SCHOOL 2026
統計研習營
 問題導向的統計思維與方法：理論、應用及挑戰
 2026.7.14(二)-7.24(五)
 中央研究院 人文社會科學館 3F 第二會議室
 申請資格：對統計科學有興趣之國內外大學部在學學生 (含應屆畢業生)
 申請辦法：線上申請，即日起至2026年5月27日(三)止
 WWW3.STAT.SINICA.EDU.TW/SCHOOL2026
 中央研究院 統計科學研究所主辦



統計研習營

<https://www3.stat.sinica.edu.tw/school2026/>

中央研究院

2026 資料價創與挑戰工作坊

Workshop on Data Value Creation and Challenges

提供國內跨領域人員
「資料的價值創造與研究挑戰」交流與討論的平台

2026.7.17 FRI. 人文社會科學館 3F 國際會議廳

報名開放中

主辦單位	協辦單位
中央研究院 統計科學研究所	資料科學 統計合作社
	中央研究院人社中心 調查研究專題中心
	臺灣實證資料 推動與加值計畫



資料價創與挑戰工作坊
<https://www3.stat.sinica.edu.tw/dvcc2026/>

中央研究院統計科學研究所圖書館 DDS

統計文獻館際合作

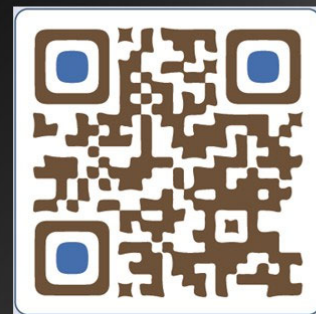
How do I find research papers easily?

How do I get full text articles quickly?

- 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 由國科會自然科學及永續研究圖書服務計畫-數學領域支應



申請館際合作
<https://sites.stat.sinica.edu.tw/library/dds/>



國立高雄大學 | 統計學研究所

811高雄市楠梓區高雄大學路700號

700, Kaohsiung University Rd., Nanzih District,

Kaohsiung 811, Taiwan, R.O.C.

TEL: 07-5919362

