

СТАНОВИЩЕ

от проф. дмн Евгения Стоименова

на дисертационен труд „**Метрични методи за анализиране и моделиране на наредени данни**” с автор Николай Иванчев Николов

за получаване на образователната и научна степен „доктор” в професионално направление: 4.5 Математика, научна специалност: „Теория на вероятностите и математическа статистика”.

Дисертацията е посветена на вероятностни модели за данни, съставени от наредби на N обекта. Задачите са актуални в последното десетилетие и намират приложения в различни области на природните и социалните науки в т.ч. маркетингови проучвания, провеждане на избори, класиране на участници в спортни събития и други. Изследваните модели се основават на статистическите свойства на метрики върху пермутации, които се използват като мерки за близостта между различни наредби на данни.

Дисертацията се състои от увод, пет глави и пет при и заключителни бележки, написана е на английски език с автореферат на български и на английски. Общият обем е 83 страници и включва библиография от 83 заглавия. Дисертацията се основава на 6 научни статии.

Дисертантът познава добре съвременната литература по изследваните проблеми. Цитирани са класически и съвременни резултати от областта на вероятностното моделиране на наредени данни. Избраната методика е подходяща за поставените цели.

В Първа глава се разглеждат основни статистически мерки за близост на наредени данни, породени от метрики върху пермутации. Основен акцент е метриката на Лии, за която са изследвани точното разпределение върху множество от пермутации, изведени са формули за средната стойност, дисперсията и е доказано асимптотичното разпределение при равновероятен избор на пермутации. Резултатите са нови за тази метрика и допълват изследванията върху мерки за близост на наредени данни, получени от други автори. Разгледани са примери за приложение тази и други метрики за анализ на данни. Свойствата на метриката на Лии се използват в моделите, построени в следващите глави.

Втора глава е посветена на два класа от параметрични модели за наредени данни, а именно клас от модели, базирани на разстояния, и маргинални модели. В модела на Малоус се предполага, че съществува латентна модална ранжировка и разпределението на наблюдаваните ранжировки се определя от близостта им до тази ранжировка. Моделът на Малоус

е изследван в случая, когато се използва разстоянието на Лии. Предложен е алгоритъм с постъпково оценяване и максимизиране на функцията на правдоподобие (ЕМ алгоритъм) за намиране неизвестните параметри в разглежданите модели. Предложен е ЕМ алгоритъм за намиране на неизвестните параметри на обобщените латентни модели на Малоус в случая когато се предполага, че данните са групирани в няколко клъстера със съответни модални ранжировки. Класът на маргиналните модели са развити с използване на разстоянието на Лии. Направени са сравнения на модели, конструирани чрез различни метрики върху пермутации.

В Трета глава се разглеждат задачи за клъстериране на наредени данни. Предложен е моделът на Малоус с метриката на Лии и е получено асимптотичното приближение на нормализиращата константа. Получен е резултат, чрез който значително се намалява времето за изчисляване на коефициента на съгъстеност при две групи. Демонстрирано в кои случаи новият модел има предимства пред други модели, базирани на разстояния.

В Четвърта глава се разглеждат модели на т.н. несъвършени наредби при балансирани извадки от наредени множества (ИНМ). Предложен е ЕМ алгоритъм за оценка на неизвестните параметри и доказана неговата сходимост към стационарна точка. Въведени са два нови статистически критерия за проверка на хипотезата за перфектна наредба. При алтернатива за неперфектна наредба се предлага модела на Малоус. Предложен е ЕМ алгоритъм с постъпково оценяване и масимизиране на функцията на правдоподобие за намиране неизвестните параметри в разглежданите модели за несъвършена наредба. Показана е сходимостта на алгоритъма към стационарна точка. Доказани са асимптотични теореми свързани с метриките на Спирман и разстоянието на Лий при пермутации с фиксиран елемент. Получени са асимптотични резултати за матриците от вероятностите за грешки при използването на модели с различни разстояния, включително разстоянието на Лии. Сравнена е мощността на предложените и съществуващите критерии. Предложените модели може да се използват за сравняване на два метода за нареждане на обекти, чрез което може да се увеличи ефективността на процедурата при ИНМ за бъдещи наблюдения.

В Пета глава се изследват непараметрични критерии за сравняване на разпределението на две извадки. Въведен е рангов критерий, базиран на разстоянието на Лии. Получени са точното разпределение на статистиката на критерия при нулевата хипотеза, средната стойност и дисперсията както и асимптотичното разпределение на статистиката. Сравнена е мощността на критерия с няколко други рангови критерии за разглежданата хипотеза. Дискутира се предимството на новия критерий при разпределения с тежка опашка.

Доказателствата на теоремите от глави 2-5 на дисертацията са предста-

вени в Приложение Е. **Научните приноси** са резюмирани от дисертанта в 5 точки. Съществената част от тях е лично дело на дисертанта. **Авторефератът** отразява коректно съдържанието на дисертацията и е във варианти на български и английски език.

Дисертацията е написана въз основа на 6 публикации. Две от тях са излезли от печат в реномирани списания с импакт фактор: *Metrika* и *AStA Advances in Statistical Analysis*, две са в поредицата *Studies in Computational Intelligence* на Springer, другите две са публикувани в сборници с доклади от международни конференции. Дисертантът е докладвал своите изследвания на 8 международни научни форуми и поне 5 национални семинари. Количеството и качеството на публикациите показват, че кандидатът съществено надвишава минималните изисквания за образователната и научна степен „доктор“.

Заклучение. Дисертационният труд напълно удовлетворява съвкупността от критерии и показатели за придобиването на научна степен „доктор“ съгласно ЗРАСРБ и съответните правилници за условията и реда за придобиване на научни степени в БАН. Ето защо убедено предлагам на Научното жури да присъди на неговия автор Николай Иванчев Николов научната степен „доктор“ в областта на висше образование 4. Природни науки, математика и информатика, професионално направление: 4.5 Математика, научна специалност: „Теория на вероятностите и математическа статистика“.

София,
25 март 2020 г.

Подпис:

Евгения Стоименова