Рецензия

относно дисертационен труд за придобиването на научната степен “доктор ”

от доц. дн Младен Светославов Савов, катедра “Изследване на операциите, вероятности и статистика“, към ИМИ на БАН, член на журито съгласно Заповед No. 33/31.01.2020г. на Директора на ИМИ-БАН и председател и съставящ рецензия по решение на журито от 04.02.2020

заглавие дисертационен труд: “Метрични методи за анализиране и моделиране на наредени данни”
автор на дисертацията: Николай Иванчев Николов

област на висше образование: 4. Природни науки, математика и информатика
професионално направление: 4.5.Математика научна специалност : “Теория на вероятностите и математическа статистика”
научна организация: ИМИ-БАН

Рецензията е изготвена въз основа на Заповед No. 33/31.01.2020г. на Директора на ИМИ-БАН, издадена на основание чл. 4 от Закона за развитието на академичния състав в Република България (Обн. ДВ. бр.38 от 21 Май 2010г., изм. ДВ. бр.81 от 15 Октомври 2010г., изм. ДВ. бр.101 от 28 Декември 2010г., изм. ДВ. бр.68 от 2 Август 2013г., изм. и доп. ДВ. бр.30 от 3 Април 2018г., изм. ДВ. бр.17 от 26 Февруари 2019г.), решение на Научния съвет на ИМИ-БАН от 24.01.2020 с протокол номер 1. То е съобразено с изискванията на: Закона за развитието на академичния състав в Република България (ЗРАСРБ), Правилника за прилагане на ЗРАСРБ, Правилника на ИМИ за условията и реда за придобиване на научни степени и заемане на академични длъжности и указанията за изготвяне на рецензии и становища за придобиване на научна степен "доктор"към ИМИ-БАН.

1. Обща характеристика на научно-приложната дейност на кандидата.

Николай Николов защитава магистърската си дисертация по направление “Математическо моделиране в икономиката” през 2015г към ФМУ-СУ "Св. Климент Охридски". Същият е придобил бакалавърска степен по "Приложна математика"от ФМИ през 2013.

Н. Николов има 6 публикации в конференции и научни списания. Участвал е в 9 конференции като е изнесъл редица доклади по научната си тематика. Н. Николов е спечелил 2 индивидуални награди - Награда за най-млад учен на БАН „Иван Евстратиев Гешов", 2019, Първа награда на докторантския конкурс „Най-добра публикация" 2019, БАН, направление – „Информационни и комуникационни науки и технологии". Участва в научния колектив на 4 научни проекта, включително и 3 финансирани по Фонд Научни Изследвания.

Научните интереси на Н. Николов са в областта на вероятностите и статистиката с фокус върху статистика на наредени данни. Докторантът работи не само с научния си ръководител проф. Евгения Стоименова, но и с колеги от международната научна общност като проф. Анна Дембинска (Варшава, Полша), с която има един завършен ръкопис. Н. Николов участва активно в научната дейност на ИМИ-БАН и води упражнения за редица курсове във ФМИ-СУ.

2. Общо описание на дисертацията

Представеният дисертационен труд на тема: "Метрични методи за анализиране и моделиране на наредени данни" е развит на 75 стандартни машинописни страници. Използвани са 83 литературни източника, които включват и 6 статии с участието на Н.Николов (5 от тях в съавторство с научния ръководител проф. Е. Стоименова и 1 самостоятелна). Дисертацията е написана на английски език и е структурирана в увод, 5 глави, апендикс с доказателства и описание на основните научни приноси.

3. Актуалност на разработвания проблем

Статистиката на наредени данни е важна област/част от модерната наука и нейните приложения. Това се дължи на факта, че в множество конкретни изследвания предпочитанията не се характеризират числово, а чрез наредба и по конкретен критерий се рангират по важност. Също така те се появяват и в контекста на непараметричната статистика, дори когато са въвлечени дискретни и непрекъснати разпределения, например, при сравняването на две извадки. Методите в тази област се базират на конкретно разстояние върху пермутационната група, което се предполага, че отразява правилно дистанцията между всеки две пермутации. В този смисъл изборът на разстояние е важна задача и е обусловена от спецификите на проблема. Известни са редица разстояния като някои от най-известните са: разстояние на Кендал (Kendall's tau), разстояние на Улам, разстояние на Хаминг и прочее. Могат да се отбележат различни области, при които методи от наредените статистики, базирани на различни метрики, могат да бъдат от съществена полза: психология ([50,70]), клъстеризация ([38]), обработка на сигнали ([7]), процедури за ефикасно извличане на информация ([3,80,81,82,83]) и т.н. Интересно и сравнително неизследвано е разстоянието на Ли, което може да се мисли като разстоянието между точки върху кръг. То е и основен предмет на изследване в настоящата дисертация. Интересно е да се отбележи, че разстоянието на Ли е подходящо за ситуации, при които се предполага, че има антагоностични подгрупи от експерти и е показало робастни свойства ([7]). В този смисъл дисертацията, изследвайки различни методи/модели, базирани на разстоянието на Ли, има потенциално голяма приложна стойност. Това в комбинация с развитието на компютрите и науката за данни (data science) прави резултатите в дисертацията съществен принос в една бурно развиваща се област.

4. Характеристика на получените резултати и използваната методология

Ще разгледам главите поотделно. Разработките във всяка една от тях са директно свързани с разстоянието на Ли.

4.1. **Глава 1 - Уводни бележки.** В тази глава са дадени предварителни сведения за разстоянията върху пермутациите, като е направен и кратък обзор за важността на това понятие. Идеята е да се намери водеща пермутация и на базата на конкретни свойства и цели да се определи подходящо разстояние, което измерва отдалечеността на дадена пермутация от водещата такава. Уводният материал е представен добре, като е въведено и разстоянието на Ли. В частност то се оказва, че удовлетворява свойствата на метрика (неравенство на триъгълника) и е дясно инвариантно ($d_L(\pi,\sigma) = d_L(\pi\circ\tau,\sigma\circ\tau)$). Използвайки класическа теорема на Хьофдинг, са намерени очакването и дисперсията и след центриране и нормиране с тях е доказана асимптотичната нормалност на метриката на Ли ($D_L(\pi)$), индуцирана от равномерното разпределение на π в пермутационната група.

4.2. **Глава 2 - Вероятностни модели за наредени данни.** В Глава 2 са изследвани теоретично и сравнени симулационно вероятностни модели за рангови данни. Моделите са най-общо два вида. Първият е базиран на разстояние в пермутационната група $d(\cdot,\cdot)$ и се задава в следния вид

$$\mathbb{P}_{\theta,p,\pi_0}(\pi) = \sum_{j=1}^{K} p_j e^{\theta_j d(\pi,\pi_{0,j})-\psi_N(\theta_j)}, \pi \in \mathcal{S}_N, \theta \in \mathbb{R}^K, \tag{4.1}$$

където се предполага, че има K латентни групи, всяка от които има относителна тежест p_j и има параметър θ_j за $j \le K$. Разстоянието $d(\cdot,\cdot)$ може да бъде по принцип произволно, а $\psi_N(\theta)$ е нормиращата константа. Идеята е при дадени наблюдения (рангирани данни) да се изберат параматерите θ, p, π_0, така че моделът най-добре да ги обхваща в смисъла на максималното правдоподобие. Хубаво е да се отбележи, че ако $\theta \equiv 0$, то имаме равномерното разпределение. По-общо тези модели се конструират от равномерното разпределение с експоненциално изменение на вероятностите (exponential tilting). Когато $K = 1$ се наблюдава най-класическият модел, въведен от Малоус, и е от вида

$$\mathbb{P}_{\theta,\pi_0}(\pi) = e^{\theta d(\pi,\pi_0) - \psi_N(\theta)}, \pi \in \mathcal{S}_N. \tag{4.2}$$

В Глава 2 е направен добър обзор на литературата и историческата последователност при въвеждането на различните модели. Прилагането на метода на максималното правдоподобие за последния модел е автоматично при условие, че се разполага с информация за $\psi_N(\theta)$. В случая на $d = d_L$ е предложено да се използва факта, че при големи N разстоянието на Ли, подходящо нормирано (с ν) и центрирано (с μ), се схожда към $Z \sim N(0,1)$ и следователно се очаква, че

$$e^{\psi_N(\theta)} = N!\mathbb{E}\left[e^{\theta D_L}\right] \sim N! e^{\theta\mu + \frac{\theta^2}{2}\nu^2}.$$

Емпирично това е частично потвърдено от симулации, но теоретично това вероятно не е вярно, защото не е приложена правилно сходимостта към нормално разпределение. Таблица 2.1 по-скоро потвърждава наблюдението ми. Оригиналните приноси на докторанта като минимум са тези емпирични резултати.

Вторият модел се нарича модел на маргиналните разпределения (Marginals model). Най-общо той има вида

$$\mathbb{P}_\lambda(\pi) = e^{\sum_{j=1}^N \sum_{i=1}^N \lambda_i^{(j)} 1_{\pi(i)=j} - \psi(\lambda)}, \pi \in \mathcal{S}_N, \lambda \in \mathbb{R}^{N^2}, \tag{4.3}$$

откъдето произхожда и името му (отделят се подмножествата $\{\pi(i) = j\} \subset \mathcal{S}_N$). Целта е да се разберат/приближат количествата

$$m_{ij} = \sum_{\{\pi(i)=j\}} \mathbb{P}_\lambda(\pi) \tag{4.4}$$

или вероятността за наблюдение на пермутация с условието $\{\pi(i) = j\}$. Въведена е и матрицата $M = (m_{ij})$, която в смисъла на матрица на прехода е двойно стохастична. Когато

$$\lambda_i^{(j)} = \theta c_N(i,j) = \theta \min\{|i-j|, N - |i-j|\},$$

се казва, че моделът на маргиналните разпределения е базиран на разстоянието на Ли. Тогава

$$m_{ij}(\theta, N) = \sum_{\{\pi(i)=j\}} \mathbb{P}_{\theta,e_N}(\pi),$$

в смисъла на (4.2). Частичен принос на докторанта е асимптотичният резултат в Теорема 2.1, който дава емпирични/симулационни основания да се предполага, че

$$\lim_{N\to\infty} m_{ij}(\theta, N)\frac{N}{e^{\mu(N)\theta+\frac{\theta^2\nu^2(N)}{2}}} = 1.$$

Твърдението на теоремата, касаещо тази сходимост, не е доказано правилно. Сходимостта към нормалност не е използвана коректно, защото са използвани неправилно, фактът че

$$m_{ij}(\theta, N) = \frac{1}{N}\frac{\tilde{g}_{N-1}(\theta)}{g_N(\theta)}$$

и някои свойства на $\tilde{g}, g$. Това не е възможно по следните причини. Функциите $g_N, \tilde{g}_N$ са функции на моментите на $D_L(\pi)$ и още една сходна случайна величина. Знае се, че $(D_L(\pi) - \mu(N))/\sigma(N)$ се схожда към нормално разпределение. Това обаче не води по принцип до факта, че

$$g_N(\theta) \sim e^{\mu(N)\theta+\theta^2\frac{\sigma^2(N)}{2}}.$$

Дори не е лесно да се види, че

$$\frac{\tilde{g}_{N-1}(\theta)}{g_N(\theta)} \sim e^{(\tilde{\mu}(N)-\mu(N))\theta+\theta^2\left(\frac{\tilde{\sigma}^2(N)}{2}-\frac{\sigma^2(N)}{2}\right)}.$$

Затова мога да приема резултата като емпиричен, но не и като математически.

В следващата част на главата е направен обзор на метода на максималното правдоподобие в контекста на разглежданите модели. Въведени са понятия като *total nonuniformity of data* и е показано как те могат да се използват, за да се оцени колко добре даден модел напасва данните. В подглава 2.4 е разгледан EM-алгоритъм за намиране на параметрите θ, p, π_0 в (4.1), което поради адитивната структура на (4.1) е невъзможно да бъде направено аналитично. Разисквани са основните проблеми и изчислителни трудности пред прилагането на този алгоритъм и поради това е предложен обобщен EM-алгоритъм, който преодолява част от споменатите трудности и за него е показано, че очакваното правдоподобие винаги е ненамаляващо. Доказателството на последния факт е коректно.

В оставащите части са направени различни емпирични изследвания, които целят да покажат кога един тип модел е по-добър от друг. Най-интересно е да се отбележи, което е и принос на докторанта, че моделите, базирани на разстоянието на Ли, са особено подходящи за случаи, при които имаме антагонистични подгрупи от експерти.

4.3. **Глава 3 - Клъстериране на наредени данни.** В тази част се разглежда проблемът за клъстериране на пермутационни извадки, което представлява търсене на скрити подгрупи от пермутации, които са на близко разстояние до даден център (пермутация

). Този тип задачи не е нов в литературата и в дисертацията е направен добър исторически преглед на проблематиката. Методите за клъстериране могат да се формулират най-общо, като оптимизационния проблем

$$(\hat{\sigma}_1, \dots, \hat{\sigma}_K) = argmin_{(\sigma_1, \dots, \sigma_K)} \frac{1}{n} \sum_{i=1}^{n} \min\{j \leq K : d\left(\pi^i, \sigma_j\right)\}, \tag{4.5}$$

където $(\pi_j)_{j=1}^n$ е извадката. Въвежда се коефициентът

$$T_K = 1 - \frac{\hat{C}_K}{C_K^0},$$

където

$$C_K^0 := \min\{(\sigma_1, \dots, \sigma_K) : \frac{1}{N!} \sum_{\pi \in \mathcal{S}_N} \min\{j \leq K : d\left(\pi^i, \sigma_j\right)\}\}$$

$$\hat{C}_K := \frac{1}{n} \sum_{i=1}^{n} \min\{j \leq K : d\left(\pi^i, \hat{\sigma}_j\right)\}.$$

В Глава 3 на дисертацията е разгледан случаят когато $d = d_L$. Първият резултат е относно приближението на универсалната константа C_K^0. Това е сериозен теоретичен проблем особено при по-големи N и K, понеже цялата изчислимост на T_K зависи от това доколко е на разположение C_K^0. Основният оригинален резултат е Следствие 3.1, в което когато $K = 2$ и N четно е доказано, че задача (4.5) има решение, зададено от e_N, e_N^*, съответно - идентитета и неговата противоположна пермутация в смисъл на разстояние на Ли ($d_L(\pi, e_N) + d_L(\pi, e_N^*) = \frac{N^2}{2}$). Тогава

$$C_K^0 \sim \frac{N^2}{4} - \sqrt{\frac{N^4 + 8N^2}{24\pi(N-1)}}.$$

Доказателството е коректно направено и използва следните идеи: от дясната инвариантност може да се предположи, че едната пермутация е e_N; $d_L(\pi, e_N) \stackrel{d}{=} d_L(\pi, \sigma), \forall \sigma \in \mathcal{S}_N$; бидейки противоложна на e_N, пермутацията e_N^* е кандидат за минимум на

$$\min_{\sigma \in \mathcal{S}_N} \mathbb{E}\left[\min\{d_L(\pi, e_N), d_L(\pi, \sigma)\}\right] \stackrel{?}{=} \mathbb{E}\left[\min\{d_L(\pi, e_N), d_L(\pi, e_N^*)\}\right],$$

като задачата по-общо може да се сведе до

$$\min_{(X,Y)} \mathbb{E}\left[\min\{X, Y\}\right],$$

еквивалентна от своя страна на

$$\min_{(X,Y)} \mathbb{E}\left[|X - Y|\right],$$

където (X, Y) са всички възможни двойки (coupling) на X с условието $X \stackrel{d}{=} Y$. В този смисъл тази задача има връзка с други области на вероятностите и математиката.

Случаят $K = 2$ и N нечетно не е разгледан, защото поради загуба на симетрия не е ясно дали някоя от противоположните две пермутации ще минимизира (4.5). Вероятно това ще е предмет на бъдещи изследвания, макар че по-важната отворена задача изглежда да е случаят $K \geq 3$. В дисертацията е направен опит да се налучка решението, но естествените кандидати, базирани на структурата в случая $K = 2$, се оказват неподходящи.

В последната част на Глава 3 е направено симулационно изследване, което показва, че методът за клъстеризация, базиран на разстоянието на Ли, може да бъде подходящ за редуциране на броя на клъстерите (K).

4.4. **Глава 4 - Несъвършени наредби при извадки от наредени множества.** В тази част се разглеждат извадки от наредени множества (ИНМ) от балансиран тип. Най-общо имаме непрекъснати случайни величини, чиито наблюдения се рангират експертно (измерванията са скъпи и/или сложни). Така от извадката на първите k се избира най-малкият по ранг, от вторите k- вторият най-малък по ранг, докато се получат $\{X_{[1]}, \ldots, X_{[k]}\}$ или една извадка от наредени множества. Тези величини се измерват количествено, като дори при съвършенно експертно нареждане е възможно от закона за шанса, $X_{[j]}$ да не са монотонно нарастващи (еднакви стойности се случват с нулева вероятност). Възможно е да се направят произволен брой такива извадки. В литературата е пресметната вероятността конкретна наблюдавана наредба да произхожда от съвършенна наредба, уравнение (4.1) от дисертацията. Също така се знае при конкретен модел за грешка, т.е. при зададени вероятности елемент с ранг i да е поставен на $j-$та позиция (p_{ij}), какви са съответните вероятности да се наблюдава конкретна наредба. Дисертацията прави и обзор на статистиките, които се използват, за да се тества дали е налична съвършенна наредба или не. Реално тези статистики са базирани на някои от основните разстояния върху пермутационната група (Кендал, Спирман, Чебишев и Ли) и са приложими за произволен брой извадки от наредени множества. Една от трудностите при прилагането на тези статистики е фактът, че априори нямаме знания за p_{ij}. За да разреши този проблем, докторантът разглежда/въвежда маргиналният модел от Глава 2 с цел да моделира p_{ij}, като използва параметър $\theta \leq 0$, за да оптимизира избора на конкретен алтернативен модел. Така се получава

$$p_{ij} = q(i, j, \theta, k) = \sum_{\pi \in \mathcal{S}_k, \pi(i)=j} \mathbb{P}\left(\pi \middle| \theta\right),$$

съответната матрица $Q(\theta, k) = (q(., ., \theta, k))$ и се специфицира параметрично семейство, което описва модел с несъвършенна наредба.

В следващата част е предложен ЕМ-алгоритъм за максимално правдоподобна оценка на θ, като е дискутирана сходимостта към максимално правдоподобната оценка. Въпреки това остава проблемът за изчисляването на компонентите на $Q(\theta, k)$, което от първи принципи ще изисква сумирането по $\mathcal{S}_k$, което е много сложна изчислителна задача. За решаването на този проблем в дисертацията е предложено приближение на $q(i, j, \theta, k)$ в случаите, когато прилежащото разстояние е: на Ли (Теорема 2.1), на правилото на Спирман (Spearman's footrule) (Теорема 4.1) и на ро на Спирман (Spearman's ρ) (Теорема 4.2). В същността си $q(i, j, \theta, k)$ се приближат с количества от вида $e^{\theta\mu+\frac{\theta^2\nu^2}{2}}/k$, които идват от нормално приближение. Доказателствата са в апендикса, като предварителните леми, които осигуряват това приближение, са коректно направени и са базирани на основната теорема на Хьофдинг. За съжаление обаче доказателствата не са верни в цялост и страдат от същия недостатък като това в Теорема 2.1. Емпирично, приближенията изглежда, че работят добре, но теоретично не са доказани вярно, за повече информация за същността на проблем виж коментара към Теорема 2.1. Приемам тези резултати за емпирични.

В част 4.6 е направено изследване на мощността на различните статистики. Понеже съвършенната наредба не зависи от алтернативния модел, се пресмятат критичните нива на статистиките и се използват на последващ етап да се оцени тяхната мощност при условие, че алтернативният модел е този, споменат в горния параграф. Оказва се, че като цяло тестът, базиран на ро на Спирман, е най-мощният. В част 4.7 е разгледан пример от практиката, като е показано, че за него не може да се отхвърли хипотезата за съвършенна наредба. Въпреки това са пресметнати мощностите при алтернативния модел, базиран на петте основни теста. Разискани са предимствата от използването на този алтернативен модел, като убедително е изтъкнато, че оценената $\hat{\theta}$ може да служи за оценка на експертното рангиране.

4.5. **Глава 5 - Разстояние на Ли в рангови критерии за две извадки.** Ранговите статистики се използват в редица приложения, включително и в тестването на хипотези за равенство на две разпределение. В тази глава е разгледан случаят, когато при две независими извадки $(X_1, X_2, \ldots, X_m)$ и $(Y_1, Y_2, \ldots, Y_n)$, всяка една от които се състои от независими еднакво разпределени случайни величини с функции на разпределение F и G, се тества хипотезата $H_0 : F = G$ срещу $H_1 : F \geq G$ и $F > G$ за някое x. Унифициран подход, базиран на наредените наблюдения и дадено разстояние върху групата на пермутациите, за тестването на тази хипотеза, е предложен от Критчлоу. Той е показал, че редица добре известни тестове като на Уилкоксон, Колмогоров-Смирнво, Ман-Уитни и други отговорят на различни изходни разстояния. В тази глава е изследван тестът на Критчлоу, но с разстоянието на Ли. За целта е необходимо да се пресметнат редица количества, свързани със съответната статистика.

Общата идея на Критчоу е да нареди заедно $(X_1, X_2, \dots, X_m)$ и $(Y_1, Y_2, \dots, Y_n)$, като алтернативната хипотеза ще съответства на сценарии, при които $(X_1, X_2, \dots, X_m)$ заемат основно ранговете между 1 и m. В този смисъл и предвид, че пермутации от $\mathcal{S}_m \times \mathcal{S}_n \subseteq \mathcal{S}_{m+n}$ не променят броя Y_j, които попадат сред първите m компонента, то подходяща статистика е $d_L([\alpha], \mathcal{S}_m \times \mathcal{S}_n)$, където $[\alpha] := \{\alpha \circ \pi : \pi \in \mathcal{S}_m \times \mathcal{S}_n\}$. $d_L([\alpha], \mathcal{S}_m \times \mathcal{S}_n) = 0$, тогава и само тогава когато $\alpha \in \mathcal{S}_m \times \mathcal{S}_n$. В глава 5, използвайки дясната инвариантност и други съображения, е показано как да се изчисли $d_L([\alpha], \mathcal{S}_m \times \mathcal{S}_n)$, в смисъл да се намери пермутация σ, такава че

$$2L_{m,n} := d_L([\alpha], \mathcal{S}_m \times \mathcal{S}_n) = d_L(\alpha, \sigma).$$

В Твърдение 5.1 е изведено съвместеното разпределение на статистиката $L_{m,n}$ и $H_{m,n}$, където последната брои колко елемента от Y са попаднали сред първите m наредени наблюдения (всички $m+n$ се нареждат заедно). В същността си доказателството е кратко, но хитро, защото успява правилно да постави проблема, т.е. да изброи сътовените случаи, при които $\{L_{m,n} = l, H_{m,n} = k\}$. За съжаление в първоначалния си вид, сметките за изброяването са необозирими и затова в Твърдение 5.2 са изведени рекурентни формули за количвества от вида $N(m, n, k, l)$, т.е. броя случаи, когато $\{L_{m,n} = l, H_{m,n} = k\}$. Случаите $m + n$ четно и нечетно водят до различни отговори понеже при първия сценарий е възможно $|i - j| = N - |i - j| = N/2$. Доказателството е коректно, но липсват достатъчни обяснителни бележки при изложението му. Това е обща забележка за дисертацията. Твърдение 5.3 превежда резултата на Твърдение 5.2 на езика на вероятностите и е намерено разпределението на $L_{m,n}$. След тези подготвителни резултати е доказана Теорема 5.1, която съдържа в явен вид средното и дисперсията на $L_{m,n}$ и е показана асимптотичната нормалност на тази статистика. Доказателството използва рекурентните връзки от Твърдение 5.3 и методът на функцията на моментите. То е вярно и изисква прецизното доказателство на факта, че

$$\phi_{m,n}\left(\frac{t}{\sigma_{m,n}}\right) e^{-\mu_{m,n}\frac{t}{\sigma_{m,n}}} = e^{\frac{t^2}{d2}}\left(1 + O\left(\frac{t}{\sqrt{\min(m,n)}}\right)\right),$$

където $\phi_{m,n}$ е функцията на моментите на $L_{m,n}$.

В част 5.4 е направено сравнение на мощността на $L_{m,n}$ с класическите методи на Уилкоксон, Колмогоров-Смирнов и $t-$теста. Последният има голяма мощност извън случая на тежки опашки, но не е робастен. Другите методи се отличават с робастност и добра мощност, но и са независими от вида на разпределението. Асимптотичната нормалност на $L_{m,n}$ позволява изчислимост и показва добри резултати при тежки опашки.

5. Оценка приносите

Приносите на докторанта са правилно отразени в дисертацията и автореферата. Оригиналността на резултатите не буди съмнение, както и факта, че докторантът е положил необходимите усилия за тяхното постигане.

6. Преценка на публикациите

Публикациите към дисертацията са шест на брой, като две от тях са с импакт фактор ([59] и [61]), две са с SJR ([60] и [62]) и две са нереферирани. Пет от шест статии са в съавторство с научния ръководител проф. Евгения Стоименова и една е самостоятелна. Няма да разглеждам реда на авторство, тъй като той е в азбучен ред. Убеден съм, че докторантът има сериозен принос във всяка една от тях.

7. Оценка на автореферата

Авторефератът правилно отразява основните положения и научните приноси на дисертационния труд. Авторската справка правилно отразява приносите.

8. Критични бележки

Основната забележка към дисертацията са доказателствата на Теорема 2.1, Теорема 4.1 и Теорема 4.2. В частите, в които се твърди конкретно асимптотично поведение на $m_{ij}(\theta, N), q(i, j, \theta, k)$, тези теореми не са правилно доказани. Това е съществен недостатък на дисертацията, който е компенсиран от количеството други резултати и особено тези в Глава 5. Доколкото ми е известно Теорема 2.1 е публикувана в статия [59] и при невъзможност да се поправи, следва да се прибегне до "erratum".

Цялостен недостатък на дисертацията е нейният твърде сбит характер. Много от доказателствата са във вид, който може да е твърде кратък дори за рецензирано списание. Това прави четенето трудно. Но това не отнема от достойнствата на труда.

9. Заключение

Представеният дисертационен труд напълно отговаря на изискванията на закона за развитие на академичния състав в Република България (ЗРАСРБ) и правилника за неговото прилагане. Представените към дисертацията трудове са достатъчни по качество и количество, като няма съмнение относно значимостта на приносите в тях. Това ми дава основание да препоръчам на научното жури да гласува положително за присъждане на маг. мат. Николай Иванчев Николов образователната и научна степен “Доктор” в област на висшето образование 4, Природни науки, математика и информатика, професионално направление 4.5 Математика, докторска програма „Теория на вероятностите и статистика".

. . . .
доц дн М. Савов

гр. София
18.03.2020