

СТАНОВИЩЕ

от проф. д-мн Огнян Ив. Кунчев

по конкурс за заемане на академичната длъжност

"доцент" в професионално направление 4.5 Математика, научна специалност:

"Теория на вероятностите и математическа статистика"

за нуждите на Института по математика и информатика при БАН

Съгласно заповед номер 99/31.03.2015 г. на директора на ИМИ БАН и решение на научното жури с Протокол номер 1 от 10.04.2014 г. съм избран за член на жури по конкурс за доцент, обявен в Държавен вестник, бр. 8, от 30.01.2015 г. В съответствие с това е и настоящето становище.

Документи за участие в конкурса е подал д-р Деян Йорданов Палежев.

1. Представени материали и кратки биографични данни

За участие в конкурса кандидатът е представил 13 научни статии, публикувани в специализирани международни списания, като 12 са в списания с импакт фактор, а две са в списания с много висок импакт фактор, а именно, Science и Nature; три са в Proc. Natl. Acad. Sci. U. S. A., и една статия е в Serdica J. Computing; една статия е публикувана в Stat. Appl. Genet. Molec. Biol.; останалите статии са в специализирани научни списания от областта на генетиката, биологията и медицината. Само една от публикациите е самостоятелна.

Деян Палежев е роден през 1972 г. От 1990 до 1998 учи в ФМИ на Софийския Университет, като защитава дипломна работа в областта на математическия анализ. От 1998-2001 учи във втората степен на обучение по математика (graduate studies) в Бостънския университет, а през периода 2001-2006 получава магистърска и докторска степени в Департамента по статистика на Йейлския университет. В тези университети той има и разнообразна преподавателска и научна дейност, като някои от тях са в Департаментите по генетика и по Детски болести в Медицинския учебен център (Medical School) на Йейлския университет. В периода 2006-2011 той участва в пет научно-изследователски проекта в областта на приложение на статистически методи за анализ на данни в генетиката и биологията. В периода 2010-2011 е асистент в Софийския университет, а в периода 2012-2014 е асистент към ИМИ, БАН, като е участвал в съвместни магистърски програми с НБУ.

2. Научна и преподавателска дейност на кандидата

Деян Палежев има разнообразна научна и преподавателска дейност.

Той е защитил дисертация на тема „Тест за бимодалност в многомерни пространства“. Дейността на д-р Деян Палежев показва качества на изявен специалист в областта на

математическата статистика и нейните приложения, и по-специално в областта на статистическите и биоинформатични методи. Той е автор на 13 научни публикации, като резултатите им е докладвал на престижни научни конференции. Деян Палежев се е занимавал с преподавателска дейност в Департамента по статистика на университета в Йейл, където е чел курсове по случайни процеси, теория на вероятностите и теория на статистиката; преподавал е във ФМИ на СУ, 2012-2013 г., където е чел курс по статистика и емпирични данни, а също така и е преподавал в съвместната програма на ИМИ и НБУ през 2014 г., където е чел курс по статистически методи.

Деян Палежев е участвал в няколко научно-изследователски проекта, в областта на приложения на статистически методи в биологията, медицината и генетиката, в периода от 2006 до 2011 в университета в Йейл, а понастоящем участва в проект на ИМИ, финансиран от ФНИ и в двустранен проект с МАНИ. Деян Палежев е отличен с награди: John F. Enders Fellowship, Yale University (2005) и John Perry Miller Fund Award, Yale University (2004).

3. Анализ на научните постижения на д-р Палежев:

Научните и научно-приложни интереси на д-р Деян Палежев, съгласно представените публикации, са в научната специалност „Теория на вероятностите и математическа статистика“, и по-точно са в областта на Приложната статистика, като са посветени на приложения на статистиката в биологията, медицината и генетиката. В тях се разработват нови методи в областта на биостатистиката, биоинформатиката, статистическата генетика, изчислителната биология, многомерния анализ, класификации и клъстериране.

Трябва особено да се подчертае, че в настоящия момент научните интереси на д-р Палежев са концентрирани в областта на новите технологии за генетично секвениране - Next Generation Sequencing (NGS), и в свързаните с тях методи за анализ на данни, които са възникнали през последните десет години. Тези методи генерират огромно количество качествени данни от ДНК и РНК, в сравнение с предходните подходи. Наличието на тези Големи масиви от данни (Big Data), налага и изследването на нови математически, статистически и биоинформатични методи за техния анализ.

Работите представени в 13-те научни статии може да бъдат разделени на четири групи:

- 1) Приложение на статистически и биоинформатични методи към новите технологии за генетично секвениране на ДНК и РНК, а също и към предишното поколение технологии – ДНК и РНК микрочипове;
- 2) Анализ на данни в случаи, в които имаме извадка с размерност, много по-малка от размерността на пространството от данните; тези ситуации възникват много често при горе-изброените видове технологии.
- 3) Анализ на данни от индуцирани плурипотентни стволови клетки (iPS cells);
- 4) Анализ на популации от различни географски региони с различни фенотипни признаци.

Към първата област, свързана с NGS технологиите, които се появиха от средата на първото десетилетие на този век, се отнасят публикациите [1, 8, 5], като статия [1] е публикувана в Science през 2007 г., и е една от първите в света по тази тема; тази статия има и много висока цитируемост. В нея са идентифицирани 1300 структурни варианта чрез метод PEM (paired-end mapping), при което двата края на относително дълги ДНК сегменти се секвенират и се намират координатите им в референтния геном. Участието на д-р Палежев в тази публикация е съществено.

В статиите [13, 12, 7, 8] има нови статистически и биоинформатични методи за анализ на данни от секвениране на РНК. При анализ на такива данни се цели сравняване на генната експресия (gene expression) между две групи - здрави хора и пациенти. Проблемът е, че извадката е малка, с по-малко от 10 индивида във всяка група, а размерността на данните е няколко хиляди, като хората имат над 25000 гени. Възниква проблем, който се нарича „проклятие на размерността“ (curse of dimensionality), добре познат в областта на анализ на големи масиви от данни, в частност в Machine learning.

В статията [13] се оценява грешката от I род при множествени сравнения. Трябва да се отбележи, че тази публикация бе избрана за научно-приложно постижение на ИМИ за 2014 г.

В статията [12] се изследва разлика между генната експресия при пациенти от Синдрома на Турет и тази на контролна група.

В статията [7] са изследвани промените в генната експресия, свързани с възрастта, като е показано, че най-големите промени в генната експресия, свързани с възрастта, са в клетъчните сигнални системи и съответно в преноса на нервни импулси.

Статията [9] е основана на дисертацията на Деян Палежев, посветена на разработване на нов статистически критерий за идентифициране на клъстери в многомерно пространство, базиран на k-means метода за клъстеризиране. Тук проблемът е, че размерността на пространството на променливите е много по-голяма от броя на наблюденията. Резултатите могат да се приложат към биологични данни, свързани със секвениране.

В статиите [8, 6 и 5] е проведен анализ на данни от индуцирани плурипотентни стволови клетки (iPS cells), като статията [8] е публикувана в Nature през 2012 г., и е посветена на две модерни теми на научни изследвания (iPS клетки и секвениране).

Анализ на данни от микрочипове (microarrays) е проведен в статиите [2, 3 и 4]. Изследваните гени са свързани с растежа и миграцията на клетките, имунната система, стабилността на ДНК и др.

Интересно приложение има на резултатите от статиите [10,11], където са правени анализи на фенотипни признаци за географски региони. Използвайки клъстерен анализ е показано, че сортовете от западния регион са генетично по-близки помежду си, отколкото до тези от североизточния регион, а също така, че клъстерирането на сортовете от североизточния регион, по генетични признаци, до голяма степен съвпада с това по географски подрегиони.

4. Импакт на научните публикации на кандидата

Работите на Деян Палежев имат множество цитирания: забелязани са 722 цитирания в научни списания с импакт фактор, като статия [1] е цитирана поне 557, статия [8] е цитирана над 80 пъти. и статия [6] е цитирана над 50 пъти.

5. Оценка на личния принос на кандидата

Почти всички от представените публикации са в съавторство, което е характерно за областта на изследвания в биологията, медицината и генетиката, където експерименталната работа изисква големи колективи. В тези работи д-р Деян Палежев има равностойно участие, като в някои от работите той е единствен математик/статистик.

6. Лични впечатления

Имам отлични впечатления за Деян Палежев, като за по-млад колега и специалист по вероятности, статистика, приложна статистика, а също и като опитен специалист по компютърната имплементация на разработените методи към Високоразмерните данни, възникващи в медицината, генетиката и биологията. Имали сме нееднократно разговори относно новите методи в областите на съвместни интереси, като например Machine learning, Sparse clustering, и други.

Заклучение

Съгласно приведените от мен по-горе аргументи, считам, че д-р Деян Палежев е висококвалифициран специалист, който има доказани приноси в областта на вероятностите и статистиката, и техните приложения.

Считам, че демонстрираните от него компетентност, изследователски качества, стремежът му за прилагане на нови методи, а също и интегрирането му в българската и световната научна мрежа, и не на последно място широкият отзвук на неговите постижения, ми дават основание да препоръчам на уважаемото Научно жури да гласува положително за присъждането на д-р Деян Палежев на длъжността „доцент” по Професионално направление: 4.5. Математика, специалност „Теория на вероятностите и математическа статистика“.

София, 15 юни, 2015 г.

Подпис:

О. Кунчев