

РЕЦЕНЗИЯ

на дисертационен труд

Статистически методи за оценяване и анализ на многотипови разклоняващи се процеси

за присъждане на образователната и научна степен „Доктор“,
Област 4. Природни науки, математика и информатика,
Професионално направление 4.5. Математика, Научна специалност
„Теория на вероятностите и Математическа статистика“ (01-01-13)

Автор: **Ана Иванова Станева**,
Докторант към катедра „Вероятности, операционни изследвания и
статистика“, Факултет по математика и информатика,
СУ „Св. Климент Охридски“

Научен ръководител: Доцент д-р Весела Кирилова Стоименова, ФМИ- СУ

Рецензент: Професор дмн Николай М. Янев, ИМИ-БАН

1. Основни сведения.

Представеният дисертационен труд е на български език и съдържа 188 страници текст, от които 10 стр. са библиография с 114 заглавия. Освен това е представен Автореферат (на български език) с 39 стр. текст, от които 5 стр. са библиография със същите 114 заглавия като в дисертацията.

Към документацията е приложена дискета, която съдържа дисертацията, автореферата, автобиография, публикациите и Програми_R.

Авторката **Ана Иванова Станева** е била докторант към катедра „Вероятности, операционни изследвания и статистика“, Факултет по математика и информатика, СУ „Св. Климент Охридски“ през периода февруари 2013 - февруари 2017. През 1981-1986 г. следва във ФМИ – СУ и завършва като магистър по математика, а през 1991-1993 г. получава и клас квалификация като учител по математика. Има завършен 5-семестриален бакалавърски курс по макроикономика в УНСС (1997-1999). От 1. 03.2016 г. е асистент в МГУ- София, през периода 2014-2016 г е асистент в ТУ-София, през 2012-2014 г е асистент във ВСУ „Любен

Каравелов“. Била е асистент в Лесотехническият университет (1993-2000) и във ВВИСУ „Любен Каравелов“ (1996-1998). В СМГ „Паисий Хилендарски“ е била учител по програмиране през 1999-2005, 2008-2012, 2016-... . Заемала е длъжност проектант и програмист в министерството на транспорта (1986-1991). През 2005-2008 е работила като програмист в редица престижни фирми.

Журито е назначено с заповед (РД 38-164/06.03.2018) на Ректора на СУ „Св. Климент Охридски“. Предзащитата на дисертационния труд е проведена във ФМИ на СУ на 29.01.2018 г. Документацията и всички процедури по защитата са в съответствие с изисванията на закона и съответните правилници.

2. Структура, анализ, научни и научно-приложни приноси на дисертационния труд.

Най-общо казано, дисертацията разглежда въпроси, свързани със статистика на разклоняващи се стохастични процеси с краен брой типове частици. Често в литературата тези процеси се наричат многотипови или многомерни процеси на Биенайме-Галтон-Уотсън, които са с дискретно време, което се интерпретира като номер на съответното поколение. В случая състоянията на процеса са d -мерни целочислени вектори, които представляват броя на частиците от съответните типове в съответното поколение.

Дисертационният труд има следната структура: Увод (12 стр.), четири основни глави и заключение, както и четири приложения (А – Г). Ще се спрем подробно на всяка една от тези части със съответния структурен, научен и научно-приложен анализ.

Всъщност дисертацията започва с подробен списък на фигурите (около 25), списък на таблиците (8), съкращения и означения (3 стр.).

Уводът започва с коментар относно някои въпроси, свързани с актуалност и приложимост на многомерните разклоняващи се процеси, като естествено преминава към литературно-исторически обзор, посветен най-вече на модерното развитие на теорията, свързана с Колмогоров и неговата школа в областта на разклоняващите се процеси. Много добро впечатление остава историческата справка, дадена като Приложение А (стр. 154-159), където е представена предкласиката на научните изследвания в тази област, започвайки от Малтус (1798) и завършвайки с Кендал (1949).

Третата част на Увода е посветена на използваните в дисертацията алгоритми, като към това трябва да се добавят съответните: Приложение Б (ЕМ алгоритъм, стр.160-168), Приложение В (Монте Карло алгоритми, стр. 169-173) и Приложение Г (числено намиране на носителя, стр. 174-176), които допринасят много за пълноценното разбиране на получените резултати в дисертацията.

Накрая на Увода е даден списък на публикациите и съответните изнесени доклади.

В Глава 1 (42 стр.) са въведени многотиповите разклоняващи се процеси, които са обект на понатъчно изследване, както и съответните многомерни индивидуални разпределения от тип многомерен степенен ред, които се разглеждат при статистическото оценяване на някои индивидуални характеристики, използвайки елементи на Бейсовски подход.

В §1.1 са добре изложени редица стандартни дефиниции, свързани с многотипов процес на БГУ и най-вече които ще се използват в понататъчните изследвания.

В §1.2 са дефинирани някои многомерни разпределения от тип степенен ред: полиномно, отрицателно полиномно, многомерно Поасоново, многомерно геометрично. Разгледани са някои интересни свойства, а също и частни случаи.

В §1.3 първо е описан общият модел на Бейсовския подход, а в §1.3.1 тази теория се прилага за намиране на апостериорното разпределение на параметрите в случай на мнотипов разклоняващ се процес на БГУ с индивидуално разпределение от тип многомерен степенен ред. Основният резултат тук е представен в Теорема 1.3, която дава спрегнатото апостериорно разпределение, когато разполагаме със съответната структура от наблюдения, достатъчна да се опише многомерната функция на правдоподобие. Някои неща са представени само в двумерния случай, като е посочена съответната аналогия в d -мерен случай. Всъщност в §§1.3.2-1.3.5 се изследват по-подробно двумерните случаи на цитираните по-горе многомерни разпределения от степенен ред, представени в Теореме 1.4 – 1.7.

В Глава 2 (36 стр.) се разглеждат някои въпроси свързани с различни схеми на наблюдение и съответно оценяване, а също и някои робастни процедури.

Анализ на някои съществуващи статистически модели и методи е даден в §2.1. В §2.2, най-общо казано, се изследва въпросът за „възстановяване“ на разпределенията при някои видове частични наблюдения. Тук се предлага параметричен многомерен аналог на някои непараметрични резултати на Гонзалес и съавтори. За различните случаи основните изследвания са представени в Теорема 2.2 – 2.6, като последните две теореми третираат случая на многомерно Поасоново разпределение. Предложен е и един метод за оценяване на потомците от всеки тип в последователни поколения.

В §2.3 и §2.4 се разглеждат някои проблеми, свързани с робастно оценяване. Някои общи резултати в тази област са последователно представени в §2.3, като по-подробно се описва теорията на Вьндев-Нейков относно d -пълнотата и праговата точка на т.н. „претеглена и орязана максимално-правдоподобна оценка от ред k ”, като обикновено се ползва означението $WLTE(k)$. Тази теория се прилага в §2.4 при многотипов разклоняващ се процес с индивидуални разпределения от степенен ред и с наблюдения върху цялото фамилно дърво. Всъщност изложението и основният резултат от Теорема 2.9 са представени в двумерния случай, но нещата могат да се пренесат и в по-висока размерност. Накрая са дадени и някои примери със симулации.

Глава 3 е посветена на задачата за параметрично оценяване, когато могат да се наблюдават само индивидите по всеки тип в последователни поколения. Тъй като тези данни са недостатъчни за максимално-правдоподобно оценяване, то се налага да се използват други методи. В Глава 3 е предложено едно решение на тази задача с помощта на EM алгоритъма. Както вече бе отбелязано, едно добро изложение на теорията на EM оценяване е дадена в Приложение Б. Сега в Глава 3 тази теория успешно се прилага в една статистически сложна задача при многотипови разклоняващи се процеси.

Така, в §3.1 са подробно описани съответните E-стъпка и M-стъпка в случая на разклоняващи се процеси с два типа. По-нататък се работи с конкретни параметрични разпределения от тип степенен ред. В §3.1.1 индивидуалното разпределение е полиномно от ред 3, а в §3.1.2 е отрицателно полиномно от ред 3.

Интересен случай с двумерно поасоново разпределение е разгледан в §3.1.3, където в (3.35) са намерени явно оценките за неизвестните параметри. Тъй като в този случай тези параметри са всъщност

съответните индивидуални математически очаквания, то за тях са в сила добре известните непараметрични оценки на Харис. Намерените оценки в (3.35) са подобни на тези на Харис, но сега в числителя стои математическото очакване. Не е трудно да се види, че в надкритичния случай ЕМ-оценките и тези на Харис са асимптотично еквивалентни. А ако процесът започва с голям брой частици, то това ще бъде вярно във всички случаи.

В §3.2 е описан т.н. ускорен ЕМ алгоритъм предложен от Икеда, който свързва ЕМ-алгоритъма с информационното количество на Фишер. Приведени са петте последователни стъпки на алгоритъма.

В §3.3 са дадени симулационни примери на двата вида ЕМ алгоритъм като се използва триномно индивидуално разпределение. Резултатите са илюстрирани във Фигури 3.2-3.5 и Таблици 3.1-3.2.

В Глава 4 за апостериорните Бейсовски разпределения, намерени в Глава 1, се развиват някои числени оценки, основаващи се на Монте Карло методи и преди всичко на т.н. Gibbs sampler. Тук много добре се използват елементи от общата теория, добре представена в съответните Приложения В и Г.

В §4.1 се разглежда случаят, когато могат да се наблюдават само индивидите по всеки тип в последователни поколения, т.е. с недостатъчна информация за конструиране на функцията на правдоподобие. Показани са стъпките за т.н. семплиране на условните разпределения, които са по-конкретно описани за двутипови разклоняващи се процеси със съответните полиномно разпределение (§4.1.1) и отрицателно полиномно разпределение (§4.1.2).

В §4.2 се изследват въпроси, свързани с оценка на сходимост, като се използва методологията на Gelman-Rubin.

В §4.3 се предлага едно подобрене на ЕМ алгоритъма с използване на Монте Карло методи, съчетани с т.н. importance sampling.

Накрая в §4.4 са представени резултатите от съответните симулации и апроксимации. Резултатите от изследванията са изложени на 17 страници, които обхващат 6 таблици и 17 фигури. Като примери за индивидуално разпределение са разгледани триномно, отрицателно триномно и двумерно поасоново разпределение. Резултатите са интересни и добре илюстрират развитата методология. Всички тези резултати са надлежно коментирани. По-подробно те са представени в приложената дискета, файл Програми_R.

Тук трябва да се отбележи, че реализацията на резултатите се дължи и на добрите информационни умения на дисертанта.

Дисертацията завършва с Глава 5, където са добре описани основните научни, научно-приложни и научно-образователни приноси, които могат да бъдат напълно приети, тъй като съответствуват и на отбелязаното по-горе в съответните глави. Дадено е също кратко, но съдържателно резюме и са посочени някои насоки за бъдещи изследвания.

3. Автореферат, публикации и бележки.

Авторефератът правилно отразява структурата, основните идеи и резултати на дисертацията и отговаря на основните академични изисквания.

Резултатите от дисертацията са публикувани в 4 научни статии, от които 3 в сборници от конференции (2 в чужбина и една у нас), а една работа – в поредицата *Pliska Stud. Math.* Освен това резултатите от дисертационния труд са докладвани общо на 8 конференции, от които 5 са международни (2 в чужбина и 3 в България), а също и на две докторски конференции у нас.

Дисертацията е написана ясно и разбрано, но някои неща биха могли да бъдат представени и по-добре. Например, на стр. 28 от дисертацията и стр. 9 от автореферата векторната дефиниция на процеса не е пълна. На стр. 30 би било по-добре да се въведе общоприетият терминът неразложимост вместо регулярност. Понятието регулярност се използва често в друг смисъл (даже в едномерния случай). Въобще въпросът за разложимост и неразложимост не се дискутира никъде в §1. Всъщност дисертацията изследва неразложими разклоняващи се процеси. На стр. 32 условието 1. (ред 4 отдолу) не следва от това че Пероновият корен е по-голям от 1, а се изисква допълнително.

Освен посочените по-горе забележки, биха могли да се посочат още някои очевидно поправими грешки и недостатъци, които обаче не затрудняват сериозно читателя и не развалят отличното впечатление като цяло от получените резултати и доброто изложение.

4. ЗАКЛЮЧЕНИЕ

Представеният дисертационен труд има всички качества на сериозно научно изследване в областта на Стохастиката. Получените интересни резултати определят ново направление при статистическия анализ на

разклоняващи се стохастични процеси с краен брой типове индивиди и индивидуални разпределения от тип многомерен степенен ред. Една част от приносите имат определено теоретичен характер, свързани с Бейсов анализ и апостериорни разпределения, робастно оценяване, ЕМ оценки и др. От друга страна се предлагат нови методи и алгоритми за приближено пресмятане в една, без съмнение, трудна област, каквата е статистиката на многомерните разклоняващи се процеси. Добро впечатление остава работата със симулации, Монте Карло методи и получаването на крайни резултати. Дисертантката успешно е обладяла и творчески прилага редица сложни съвременни статистически и информационни техники, които я определят като един изграден специалист в областта на изчислителните методи в статистиката (или както се казва накратко – компютърна статистика). Получените резултати са докладвани и публикувани в издания на международни конференции. Всичко това дава пълно основание да заключим, че са налице всички условия от ЗРАСРБ и неговия правилник, както и този на ФМИ - СУ, които се прилагат за получаване на научната степен „Доктор“ по съответната специалност. Ето защо определено предлагам на почитаемото жури да оцени подобаващо високо представения дисертационен труд, а на неговата авторка **Ана Иванова Станева** да бъде присъдена образователната и научна степен „ДОКТОР“ в Област 4. Природни науки, математика и информатика, Професионално направление 4.5. Математика, Научна специалност „Теория на вероятностите и Математическа статистика“ (01-01-13).

28.04.2018

Резензент:

/професор дмн Николай М. Янев/