

## РЕЦЕНЗИЯ

за дисертационния труд на Нина Руменова Даскалова,  
“Алгоритми от тип ЕМ за статистическо оценяване в  
разклоняващи се стохастични процеси”,  
представен за присъждане на образователна и научна степен  
“ДОКТОР” в научното направление 4.5 “Математика”, научна  
специалност 01.01.10 “Теория на вероятностите и математическа  
статистика”.

от проф. дмн Косто Вълов Митов,  
НВУ “Васил Левски”- Факултет Авиационен.

**Основание.** Със заповед на Директора на ИМИ-БАН № 150/15.05.2012г. бях назначен за член на научно жури за защита на дисертационния труд. На първото заседание на научното жури бях избран за рецензент.

**1. Биографична справка.** Нина Даскалова е завършила бакалавърската програма по “Информатика” и магистърска програма по “Приложна статистика” във ФМИ на СУ “Св. Климент Охридски”. Работила е като математик в секция “Вероятности и статистика” на ИМИ-БАН, а от 2008г до сега е асистент в катедра ВОИС на ФМИ. Дисертационният труд разработва като задочен докторант в секция “Вероятности и статистика” на ИМИ-БАН.

**2. Материали по защитата.** Във връзка със защитата са ми предоставени: Дисертация, Автореферат. Списък с публикации по дисертацията, Авторска справка. Копия от публикациите по дисертацията и Автобиография.

Дисертация се състои от Увод, 4 глави, Заключение, Списък на литературата и Приложение и е в обем на 93 страници. Списъкът с литературата съдържа 159 заглавия. Авторефератът е в обем на 25 страници.

**3. Общо за проблемите разглеждани в дисертацията.** Дисертационният труд на Нина Даскалова е в областта на статистиката на разклоняващи се процеси. Статистическото оценяване на параметри на различни класове разклоняващи се процеси се развива интензивно в последните години във връзка с многобройните приложения на тези процеси в биологията, екологията, разпространение на болести, финанси и др. Известни са различни подходи при оценяване на параметрите на един разклоняващ се процес. Най-общо те се разделят на параметрични и непараметрични, като и в двата случая получаването на оценки зависи от наличните данни.

Данните могат да включват:

- цялото дърво от началото на еволюцията до дадения момент на наблюдение;
- само броя на индивидите в отделните поколения;
- броя на индивидите само в някои от поколенията.

В представената дисертация са намерени непараметрични оценки за разпределението на потомството в многотипен процес с дискретно време, в който се предполага че има финални типове частици. При получаването на оценките за пръв път се използва алгоритъм от тип ЕМ (Expectation and Maximization) (Осредняване и Максимизация). Методът ЕМ е обоснован строго от Демпстер, Леард и Рубин през 1977г., но е използван от други статистици и преди това. Методът представлява итеративна процедура за намиране на МПО при предположение, че оценката зависи както от наблюдаваните данни, така и от такива, които не се съдържат в извадката. Процедурата се състои от две стъпки: E(xpectation)-осредняване и M(aximization) - максимизация. При първата стъпка се определя функционал при дадено предполагаемо значение на оценявания параметър (Функционалът е математическо очакване, взето при тази предполагаема стойност на параметъра). На втората стъпка се минимизира функционала, в резултат на което се намира следващата стойност на оценявания параметър и т.н.

Задачата за оценяване на разпределението на потомството на един разклоняващ се процес, когато не е известно цялото еволюционно дърво се оказва подходяща за прилагане на алгоритъм от тип ЕМ.

#### 4. Съдържание на дисертацията.

Уводът на дисертацията е кратък. В него са формулирани целите и задачите, които се решават в дисертацията, представена е структурата на дисертацията и е дадена апробацията на резултатите.

Основната задача, която се решава в дисертацията може да се формулира по следния начин:

*Създаване на алгоритъм от тип ЕМ за оценка на разпределението на потомството на многотипен разклоняващ се процес с финални типове частици по непълни данни.*

Предложени са два подхода за решаване на тази задача:

При първият от тях е доказана еквивалентност между разклоняващия се процес и подходящо конструирана стохастична контекстно-свободна граматика. След това е използван известен алгоритъм от тип ЕМ за оценяване на вероятностното разпределение в тази граматика,

което всъщност се явява и разпределението на потомството в разклоняващия се процес.

Втория се състои в непосредствено създаване на алгоритъм от тип ЕМ за разклоняващ се процес с финални типове частици, като се използват непосредствено неговите свойства.

**Глава 1** на дисертацията има спомагателен характер. Тя е разделена на два параграфа. В първият от тях е направен преглед на голям брой литературни източници, които третираат въпросите за:

- 1) Разклоняващи се процеси с един и няколко типа частици, възможностите за техните приложения и различните подходи за намиране на статистически оценки;
- 2) Стохастични контекстно-свободни граматики;
- 3) Същността на ЕМ метода за намиране на максимално правдоподобни оценки.

Втория параграф съдържа необходимите дефиниции и твърдения за тези три направления, които са необходими по-нататък в изложението на основните резултати от дисертацията.

Главата е кратка и стегната и изпълнява целите, за които е написана. Представеният преглед на литературата, показва че докторантката е запозната както с класическите, така и с най-новите резултати в горните три области.

**Глава 2** се отнася до прилагането на ЕМ алгоритъма, построен за оценяване на разпределението на стохастични контекстно-свободни граматики, за оценка на разпределението в марковски разклоняващ се процес с дискретно време и няколко типа частици.

За целта в параграф 1 е даден ЕМ алгоритъма за оценка на разпределението в дадена стохастична контекстно-свободна граматика като са изведени съответните формули за пресмятане на оценките. Във втория параграф е показано как за даден многотипен разклоняващ се процес с дискретно време, може да се конструира съответна стохастична контекстно-свободна граматика, която да има разпределение, съвпадащо с разпределението на потомството на разклоняващия се процес. Това дава възможност да се използват изведените в параграф 1 формули за оценка на разпределението. Оказва се обаче, че при голям брой типове частици конструираната граматика ще има твърде много символи и изчислителните трудности, които възникват при прилагане на този подход са значителни. Това един от мотивите да се търси непосредствено конструиране на ЕМ алгоритъм за разклоняващия се процес. В послед-

ния параграф предлагания ЕМ алгоритъм е тестван на два примера на разклоняващи се процеси с 3 и 2 типа частици, като във втория случай единият от типовете е терминален.

В Глава 3 е представен ЕМ алгоритъм за оценка на разпределението на потомството в многотипен разклоняващ се процес с терминални типове. Резултатите от тази глава представляват най-съществената част от дисертацията. В параграф 1 е дадена дефиницията на многотипен процес с терминални типове и е изведена формула за МПО оценка на разпределението на потомството при условие, че е известно цялото еволюционно дърво на процеса от началото до дадения момент на наблюдение. Естествено това е най-благоприятната ситуация, но тя рядко се случва в практиката.

Затова във втория параграф се предполага, е известен само корена на дървото  $Z(0)$  и вектора  $Z(t)$ , представящ броя на индивидите в момента на наблюдение. Така скритата информация е всъщност останала част от дървото. Оказва се, че наличието на финални типове частици е съществено за конструиране на алгоритъма, поради това, че наличието на такива типове частици носи информация и за ненаблюдаваната част от фамилното дърво.

При тази постановка в параграф 2 са отделени двете стъпки на ЕМ алгоритъма. Най-напред в Теорема 3.2.1 са дадени формулите за стъпката М (максимизиране). В тези формули участват две математически очаквания, взети при условие, че е налице разпределението намерено в предната Е-стъпка.

Намирането на явни формули за пресмятане на тези математически очаквания, което е Е-стъпката на алгоритъма е направено в параграф 3.2.2, съответно в Теорема 3.2.2, като е използвано съществено наличието на терминални типове частици и разбира се независимостта на размножение на частиците в разклоняващия се процес.

Получените в явен вид рекурентни формули за пресмятане на итерациите на индивидуалните вероятности на разпределението на потомството на процеса дава възможност за непосредствено прилагане на предложения алгоритъм към конкретни реални данни.

Последния параграф 4 съдържа числени примери за прилагане на намерените формули.

Авторът е отбелязал и възможността за намиране на разпределението на потомството и в процеси с непрекъснато време, например на Белман-Харис или Марковски, като се използва вложения (по номерата

на поколенията) процес с дискретно време.

Това е използвано в **Глава 4** която съдържа приложение на конструирания ЕМ алгоритъм към данни от реален експеримент. При този експеримент еволюцията на индивидите (които са клетки) се описва най-добре с процес на Белман-Харис с три типа частици, един от които е терминален. Прилагайки ЕМ алгоритъма са получени оценки за разпределението на потомство. Оказва се, че те съвпадат с оценките получени по метода на максималното правдоподобие при известно цяло еволюционно дърво.

**Приложението** съдържа кода и описанието на процедурите, които са използвани при числените пресмятания в Глави 2 и 3. Те са написани на R.

**5. Забележки.** Според мен не е правилно да се казва многотипен процес на Биенеме - Галтон-Уотсън. По-добре е да се каже многотипен Марковски разклоняващ се процес с дискретно време. Процесът на Биенеме-Галтон-Уотсън е с един тип частици и дискретно време.

В тази точка искам да кажа, че както Дисертацията така и Авторефератът са добре оформени.

Намерил съм две технически грешки:

- на стр. 35, ред -12, би трябвало да е  $p_1 + p_2 + p_3 + p_4 = 1$ .

на стр. 27, ред +2, липсва вертикална черта в условната плътност.

**6. Публикациите** по дисертацията са 3. Всичките са самостоятелни. Две са излезли от печат в Сердика (2010) и в Плиска (2011) и една е приета за печат в Доклади на БАН и ще излезе в том 65, №5 (2012). Те отразяват съдържанието на дисертацията достатъчно пълно, така че то е станало достояние на научната общност.

Сведения за цитиране на работите нямам.

**7. Авторефератът** отразява пълно и точно съдържанието на дисертацията.

Считам, че поставените в дисертацията задачи са решени на много добро ниво. Считам също, че научните, научно-приложни и приложни приноси, заявени в дисертацията и в автореферата са действително такива.

**8. Заключение.** Като имам в предвид изложеното до тук, считам че представения дисертационен труд напълно отговаря на изискванията на ЗРАСРБ и правилника за приложението му, което ми дава право убедено да препоръчам на почитаемото Научно Жури да присъди образователната и научна степен "ДОКТОР" по специалността 01.01.10 "Теория на

вероятностите и математическа статистика” на Нина Руменова Даскалова.

05.08.2012г.  
гр. Плевен

...../  
проф. К. Митов

| Именна карта                       | Адрес на е-поща                                                       |
|------------------------------------|-----------------------------------------------------------------------|
| 1113 София, Акад. Тодор Бонев б.з. | e-mail: dimitov@math.bas.bg<br>GSM:<br>телефон: 979-24-20<br>Държава: |