Софийски университет "Св. Климент Охридски"

Факултет по математика и информатика

Катедра "Вероятности, операционни изследвания и статистика"

Статистически методи за оценяване и анализ на многотипови разклоняващи се процеси

Ана Иванова Станева

Автореферат

за придобиване на образователна и научна степен "Доктор"

Област на висше образование: 4. Природни науки, математика и информатика

Професионално направление: 4.5 Математика

Докторска програма "Теория на вероятностите и математическа статистика"

Научен ръководител

доц. д-р Весела Кирилова Стоименова

Януари 2018

Данни за дисертационния труд:

Обем на дисертацията: 189 стр. Основен текст: 150 стр. Литература: 114 заглавия. Публикации по дисертацията: 4 заглавия. Дисертационният труд е обсъден и препоръчан за започване на процедура по защита на разширено заседание на катедра "Вероятности, операционни изследвания и статистика"при Факултет по математика и информатика - СУ, проведено на 29 януари 2018 г.

Съдържание

ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД

Актуалност на темата

Изучаването на разклоняващите се процеси датира от 19 век. Съвременната теория на разклоняващите се процеси започва от 1947 год. с публикуването на статията на Колмогоров и Дмитриев "Ветвящиеся случайные процессы". Harris описва едно от първите приложения на многотиповите разклоняващи се процеси, което е в генетиката. Днес класическите многотипови разклоняващи се процеси се използват като модел при изследването на биологични системи, при изчисляване на еволюционната динамика в структурата на генотип-фенотипното пространство, в мрежите от генно регулиране, при моделиране на разпространението на вируси, в моделите за прогресиране на туморните клетки, при пропускливост на дървовидните алгоритми, каскадите от космически лъчи. Едно от най-новите приложения на разклоняващите се процеси е в компютърните науки, по-специално при работа с мрежи, в частност сензорни мрежи, както и в облачните технологии, където търсенето на облачни услуги се моделира с разклоняващ се процес.

Цели и задачи на дисертационния труд

Настоящият дисертационен труд има следните цели:

1. Бейсов статистически анализ на многотипови разклоняващи се процеси с индивидуално разпределение от фамилията многомерен степенен ред.

2. Робастно оценяване на двутипови разклоняващи се процеси с индивидуално разпределение от фамилията многомерен степенен ред.

3. Проучване на числени методи, които могат да се приложат към оценяване на многотипови процеси.

   - ЕМ алгоритъм и неговите модификации.
   - Монте Карло методи.

4. Прилагане на числени методи за оценяване на многотипови разклоняващи се процеси с индивидуално разпределение от фамилията многомерен степенен ред.

   - ЕМ алгоритъм;
   - Gibbs семплер;
   - Монте Карло ЕМ.

5. Симулация на многотипови разклоняващи се процеси с полиномно, отрицателно полиномно и Поасоново индивидуално разпределение.

   - да приложи към симулираните процеси *ЕМ алгоритъм*
   - да приложи към симулираните процеси *Gibbs sampler алгоритъм*
   - да приложи към симулираните процеси *Монте Карло ЕМ алгоритъм.*

6. Систематизиране на литературните източници свързани с:

   - *дискретни Многотипови Разклоняващи се Процеси*;
   - *фамилията разпределения Многомерен Степенен Ред*;
   - *статистическо оценяване на многотипови процеси.*

Апробация на резултатите. Публикации

Резултатите в дисертационния труд са докладвани на:

1. 3rd Stochastic Modeling Techniques and Data Analysis International Conference (SMTDA 2014, Lisbon, Portugal);

2. XVI-th International Summer Conference on Probability and Statistics, Seminar on Statistical Data Analysis, Workshop on Branching Processes and Applications, Dedicated to the memory of B. A. Sevastyanov Pomorie (21-28 юни, 2014 г.);

3. Doctoral Conference in Mathematics, Informatics and Education (September 23-25, 2014, Sofia, Bulgaria);

4. V Congress of Mathematicians of Macedonia (September 24-27, 2014, Ochrid, Macedonia);

5. Юбилейната научна конференция по случай 125 години обучение по математика и природни науки в СУ „Св. Климент Охридски", Biomath 2015 (14-19, 06, 2015);

6. Докторантска конференция по математика и информатика (15-18, 10, 2015);

7. Scientific Program XVII-th International Summer Conference on Probability and Statistics, Seminar on Statistical Data Analysis, Workshop on Branching Processes and Applications(25/06-01/07, 2016);

8. Second International Conference "Mathematics Days in Sofia" July 10–14, 2017, Sofia, Bulgaria.

Списък от публикации по темата:

1. Ana Staneva, Vessela Stoimenova, "Statistical estimation in Branching processes with bivariate Poisson offspring distribution Pliska Stud. Mat, том:24, 2015, стр.73-88.

2. Ana Staneva, Vessela Stoimenova, "Modeling and estimating multitype branching processes with negative multinomial offspring distributions Matematiqki Bilten Vol.39 (LXV) No.2 2015 (29–39) , редактори:Aleksa Malceski, издателство:Skopje, Makedonija, 2015, стр.29-39

3. Ana Staneva, "Multitype Branching Processes with Bivariate Multinomial Offspring Distribution – Bayesian Approach Advanced Research in Mathematics and Computer Science; Doctoral Conference in Mathematics, Informatics and Education [MIE 2014] Proceedings, редактори:Peter B. Sloep, Krassen Stefanov , издателство:ResearchGate, 2014, стр.32-45

4. Dimitar Atanasov, Ana Staneva, Vessela Stoimenova, "Robust estimators for the bivariate power series offspring distributions SMTDA2014 Conference Proceedings, 2014, стр.63-75.

Обем и структура на дисертацията

Настоящата дисертация се състои от 4 глави.

В Глава 1 е дадена теоретична справка за многотиповите разклоняващи се процеси и за фамилията разпределения многомерен степенен ред. Приложен е Бейсов подход за двутипови разклоняващи се процеси с индивидуално разпределение многомерен степенен ред, което е принос на дисертанта.

В Глава 2 са разгледани съществуващи резултати от статистическото оценяване на разклоняващи се процеси. Авторът на настоящият труд представя свои резултати, получени при статистическото оценяване на многотипови разклоняващи се процеси с индивидуално разпределение многомерен степенен ред. Специално внимание е отделено на робастните статистически оценки, като е направена теоретична справка, след което е приложено робастно оценяване за двутипови разклоняващи се процеси от разглеждания вид.

Глава 3 е посветена на приложението на EM алгоритъма при статистическото оценяване на многотипови разклоняващи се процеси с индивидуално

разпределение многомерен степенен ред. ЕМ алгоритъмът е приложен за полиномно, Поасоново и отрицателно полиномно разпределение. Направена е симулация на двутипов разклоняващ се процес и е приложен числено ЕМ алгоритъма за двутипов процес с триномиално разпределение. Показани са резултати и графики. Разгледана е и модификация на ЕМ алгоритъма, която ускорява изпълнението му. Този модифициран ЕМ алгоритъм е приложен към пример за двутипов разклоняващ се процес с полиномно индивидуално разпределение.

В Глава 4 се разглежда приложението на Монте Карло алгоритми за статистическо оценяване на многотипови процеси. По-конкретно, в направена симулация на двутипов процес с индивидуално разпределение двумерен степенен ред е приложен Гибс семплер.

Допълнително е разгледан Монте Карло ЕМ алгоритъм за статистическо оценяване на двутипови разклоняващи се процеси с разпределение от фамилията двумерен степенен ред.

В Приложение *А* е дадена историческата справка за възникването и развитието на теорията на разклоняващите се процеси.

В Приложение *Б* е дадена теоретична справка за ЕМ алгоритъма. Специално внимание е отделено на теоремите за сходимост. Разгледан е ЕМ алгоритъма като вариационен метод и като специален случай на проксималните алгоритми.

В Приложение *В* е дадена теоретична справка за Монте Карло алгоритмите и тяхното приложение при статистическото оценяване.

Приложение *Г* разглежда метод за числено определяне на носителя на индивидуално разпределение, базиран на алгоритъм за определяне на модата на такова разпределение. Получени са таблици и графики за носители при различни примери на триномиално, двойно Поасоново и отрицателно триномиално разпределение.

Приложение *Д* включва програмен код на езика R, с помощта на който е направена симулацията и са получени цитираните по-горе резултати.

СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД

Глава 1. Многотипови разклоняващи се процеси с разпределение многомерен степенен ред. Бейсов подход.

Многотиповият разклоняващ се процес е хомогенен във времето векторен процес $\{\mathbf{Z}(0), \mathbf{Z}(1), \ldots, \mathbf{Z}(n), \ldots\}$, където $\mathbf{Z}(n) = \big(Z_1(n), Z_2(n), \ldots, Z_d(n)\big)$, за който е в сила рекурентната зависимост $Z_m(n+1) = \sum\limits_{k=1}^{d} Z_m^k(n)$, където $Z_m(n+1)$ е броят частици от тип m, живеещи в $(n+1)$-во поколение и $Z_m^k(n)$ е броят деца от тип m с родители от тип k, живеещи в $n-$то поколение.

Дефиниция. 1.1 ***Вектор на размера на популацията*** Нека $\{\mathbf{Z}(n),\ n \in \mathbb{N}\}$, е $d-$мерен разклоняващ се процес, с множество от състояния $\mathbb{N}^d$ и начално състояние $\mathbf{Z}(0) = \underbrace{(0, .., 0, 1, 0, .., 0)}_{k-\text{та позиция}} = (\delta_{1,k}, .., \delta_{d,k}) = \boldsymbol{\delta}^k, \delta_{i,j} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}, k = 1..d.$ Векторът $\mathbf{Z}(n) = \big(Z_1(n), Z_2(n), \ldots, Z_d(n)\big)$ е *вектор на размера на популацията*, а $Z_i(n)$ е броят на индивидите от тип i в $n-$то поколение,

Дефиниция. 1.3 ***Индивидуална вероятност*** Нека в $n-$то поколение една частица от тип k да има i_s деца от тип s, $s = 1..d$. Наследниците на частицата се представят с вектора $\boldsymbol{i} = (i_1, i_2, \ldots, i_d)$. Вероятността една частица от тип k да има деца $(i_1, i_2, \ldots, i_d)$ се нарича *индивидуална вероятност* на тази частица и се означава с $p\,^k_{(i_1,i_2,\ldots,i_d)}$.

Нека $\xi_m^k\ (n, l)$ е случайната величина, представяща броят на децата от тип m за $l-$тата поред частица от тип k, живееща в n-то поколение. Тогава, на

тази $l-$та частица от тип k в $n-$то поколение се съпоставя вектора от децата от всеки тип $\vec{\boldsymbol{\xi}}^{k}\ (n,l) = \Big(\xi_1^k\ (n,l), \xi_2^k\ (n,l), \ldots, \xi_d^k\ (n,l)\Big)$.

Следователно, потомството на всички частици от тип k от $n-$то поколение, се изразява от вектора $\vec{\boldsymbol{\xi}}_{k}\ (n) = \big(\sum\limits_{l=1}^{Z_k(n)} \xi_1^k\ (n,l), \sum\limits_{l=1}^{Z_k(n)} \xi_2^k\ (n,l), \ldots, \sum\limits_{l=1}^{Z_k(n)} \xi_d^k\ (n,l)\big)$.

Дефиниция. 1.4 ***Свойство на разклоняване*** (***branching property***)

$$Z_m(n+1) = \sum_{k=1}^{d} \sum_{l=1}^{Z_k(n)} \xi_m^k\ (n,l)$$

Това свойство изразява вектора на размера на популацията на МВР в поколение $(n+1)$ чрез вектора на размера на популацията в n-то поколение:

$$\mathbf{Z(n+1)} = \left(\sum_{k=1}^{d} \sum_{l=1}^{Z_k(n)} \xi_1^k\ (n,l), \sum_{k=1}^{d} \sum_{l=1}^{Z_k(n)} \xi_2^k\ (n,l), \ldots, \sum_{k=1}^{d} \sum_{l=1}^{Z_k(n)} \xi_d^k\ (n,l)\right) = \sum_{k=1}^{d} \sum_{l=1}^{Z_k(n)} \vec{\boldsymbol{\xi}}^{k}$$

Следователно, потомството на многотиповия разклоняващ се процес може да се изрази чрез матрицата

$$\begin{pmatrix} \vec{\boldsymbol{\xi}}^{1}\ (n) \\ \vec{\boldsymbol{\xi}}^{2}\ (n) \\ \vdots \\ \vec{\boldsymbol{\xi}}^{d}\ (n) \end{pmatrix} = \begin{pmatrix} \sum\limits_{l=1}^{Z_1(n)} \xi_1^1\ (n,l) & \sum\limits_{l=1}^{Z_1(n)} \xi_2^1\ (n,l) & \cdots & \sum\limits_{l=1}^{Z_1(n)} \xi_d^1\ (n,l) \\ \sum\limits_{l=1}^{Z_2(n)} \xi_1^2\ (n,l) & \sum\limits_{l=1}^{Z_2(n)} \xi_2^2\ (n,l) & \cdots & \sum\limits_{l=1}^{Z_2(n)} \xi_d^2\ (n,l) \\ \vdots & \vdots & \ddots & \vdots \\ \sum\limits_{l=1}^{Z_d(n)} \xi_1^d\ (n,l) & \sum\limits_{l=1}^{Z_d(n)} \xi_2^d\ (n,l) & \cdots & \sum\limits_{l=1}^{Z_d(n)} \xi_d^d\ (n,l) \end{pmatrix}$$

$k-$тият ред, на която, дава потомството на всички частици от тип k, докато $j-$тия стълб дава броят на децата от тип j.

Основен инструмент при изследване на ВР са пораждащите функции $(p.g.f.)$.

Дефиниция. 1.5 ***Индивидуална пораждаща функция - и.п.ф*** Нека е дадена частица от тип k, с индивидуална вероятност $p^{\,k}_{(i_1,..,i_d)} = P(\xi_1^k = i_1, .., \xi_d^k = i_d)$, където ξ_m^k е случайна величина, даваща броят на наследниците от тип m за тази частица от тип k за едно поколение. За тази частица от тип k се дефинира и.п.ф. $f^{(k)}\ (\mathbf{s}) = \mathbf{E}[\mathbf{s}^{\boldsymbol{\xi_k}}]$ на индивидуалното разпределение за едно поколение:
$f^{(k)}\ (\mathbf{s}) = \sum\limits_{(i_1,i_2,\ldots,i_d)\in N^d} p^k_{(i_1,i_2,\ldots,i_d)}.s_1^{i_1} s_2^{i_2} \ldots s_d^{i_d}$,
където $\mathbf{s} = (s_1, s_2, \ldots, s_d),\ |\mathbf{s}| \le 1, \quad \mathbf{s}^{\boldsymbol{\xi_k}} = s_1^{\xi_1^k} s_2^{\xi_2^k} \ldots s_d^{\xi_d^k}, \quad \boldsymbol{\xi_k} = (\xi_1^k, \xi_2^k, \ldots, \xi_d^k)$.

За потомството, породено от една частица от тип k за n поколения се използва пораждащата функция $\quad F^k(n,\mathbf{s}) = E\big[\mathbf{s}^{\mathbf{Z(n)}} \big|\ \mathbf{Z(0)} = \boldsymbol{\delta^k}\big] =$

$$= \sum_{(i_1,i_2,\ldots,i_d)\in N^d} P\big(\mathbf{Z}(\mathbf{n}) = (i_1,i_2,\ldots,i_d) \mid \mathbf{Z}(\mathbf{0}) = \boldsymbol{\delta^k}\big)\ s_1^{i_1} s_2^{i_2} \ldots s_d^{i_d}.$$

От дефиницията на и.п.ф. се вижда, че $F^k(1,\mathbf{s}) = f^{(k)}(\mathbf{s})$. Образуваме вектора от и.п.ф. с $k-$та компонента - и.п.ф. на частица от тип k за едно поколение

$\mathbf{f}(\mathbf{s}) = \big(\ f^{(1)}(\mathbf{s}), .., f^{(d)}(\mathbf{s})\ \big) = \big(\ F^1(1,\mathbf{s}), .., F^d(1,\mathbf{s})\ \big) = \mathbf{F}(1,\mathbf{s})$.

Твърдение (1.1)**.** $\mathbf{F}^{(k)}(n,\mathbf{s}) = \mathbf{f}_n^{(k)}(\mathbf{s})$,

$\mathbf{f_n}(\mathbf{s}) : \mathbf{f}^n(\mathbf{s}) = \mathbf{f}(\mathbf{f}^{(n-1)}(\mathbf{s})) = \underbrace{\mathbf{f}(\mathbf{f}(\ldots\mathbf{f}(\mathbf{f}(\mathbf{s}))))}_{n-пъти}$

Дефиниция. 1.6 ***Многомерна*** (***векторна***) ***пораждаща функция***

$$\vec{\mathbf{F}}(n;s) = \big(F^1(n;s).F^2(n;s).\ldots.F^d(n;s)\big),\ \text{където}$$

$F^k(n,\mathbf{s}) = E[s_1^{Z_{k1}(n)}, s_2^{Z_{k2}(n)}, \ldots, s_d^{Z_{kd}(n)}]$ и $Z_{kj}(n)$ е броят на частиците от тип j в поколение n, породени от 1 частица от тип k, живееща в нулево поколение.

Следователно, п.ф. е решение на уравнението $\mathbf{F}(n+1;\mathbf{s}) = \mathbf{F}(1,\mathbf{F}(n;\mathbf{s}))$. Нека $m_{ij} = \mathbf{E}\big[\xi_i^j(n;l)\big]$ е очакваният брой деца от тип j на $l-$тата частица от тип i. Ако $\mathbf{M} = ||m_{ij}||$ е матрицата от първите моменти на разпределението на потомството за едно поколение, то $m_{ij} = \frac{\partial}{\partial s_j} f^{(i)}(\mathbf{s})|_{\mathbf{s=1}} = E\big[Z_j(1)|Z_i(0)=1\big]$.

Вторите моменти са

$b_{jk}^i = \frac{\partial^2}{\partial s_j \partial s_k} f^i(\mathbf{s})\ |_{\mathbf{s=1}} = E\big[Z_j(1)\big(Z_k(1) - \delta_{jk}\big)|Z_i(0)=1\big]$, където $i,j,k = \overline{1,d}$

За n поколения, матрицата от първите моменти е $\mathbf{M}(n) := \|m_{ij}(n)\| = \mathbf{M}^n$ където $m_{ij}(n) := E[Z_j(n)|Z_i(0)=1] = \frac{\partial}{\partial s_j} F^i(n,\mathbf{s})|_{\mathbf{s=1}}$.

В този случай вторите моменти са равни на

$b_{jk}^i(n) := \frac{\partial^2}{\partial s_j \partial s_k} F^i(n,\mathbf{s})|_{\mathbf{s=1}} = E[Z_j(n)[Z_k(n) - \delta_{jk}]|Z_i(0)=1],\ \ i,j,k = \overline{1,d}.$

Нека $\mathbf{M} = \|m_{ij}\|,\ \ i,j = \overline{1,d}$ е матрицата на математическото очакване на размера на популацията. Матрицата $\boldsymbol{M}_{d\times d} \geq 0$ се нарича *регулярна*, ако съществува цяло положително число n, за което $\mathbf{M}^n > 0$. Асоциираме матрицата $\mathbf{M} = \|m_{ij}\|$ с насочен граф $G_{\boldsymbol{M}}$, който има d възела и $m_{ij} > 0$, когато има ребро от връх i до връх j. Матрицата $\mathbf{M} = \|m_{ij}\|$ е регулярна $\Longleftrightarrow$ когато нейният граф G_M е силно свързан, т.е. за всяка двойка различни върхове i и j съществува път от i до j и от j до i, независимо с каква дължина $n < \infty$.

Дефиниция. 1.8 Многотиповият разклоняващ се процес $\mathbf{Z}(\mathbf{n})$ се нарича *положително регулярен*, ако матрицата $\mathbf{M} = \|m_{ij}\|$ е положителна и регулярна.

Следователно, МВР е регулярен, ако за всяко $k = \overline{1,d}$, частица от тип k ще породи частици от всеки друг тип, независимо след колко поколения.

Класификацията на МВР се прави въз основа на Теоремата на Перон-Фробениус. Ако $\alpha_1, \ldots, \alpha_n$ са собствените стойности на матрицата $\boldsymbol{M}$, то $r(\boldsymbol{M}) = \max\limits_{i=1\ldots n} |\alpha_i|$ се нарича *спектрален радиус на матрицата* $\boldsymbol{M}$, а най-голямата по модул реална собствена стойност на матрицата $\boldsymbol{M} = ||m_{ij}(n)||$ се нарича *Перонов корен* и той се означава с ρ.

Дефиниция. 1.10 *Израждане* (*extinction*) МВР се нарича *изроден*, ако $\exists$ естествено число N, такова, че за всяко $n \geq N$ е изпълнено $\mathbf{Z}(n) = (0, 0, \ldots, 0)$.

Дефиниция. 1.12 Нека P^k е вероятността за израждане, ако процесът стартира с една частица от тип k. Векторът $\mathbf{P} = (P^1, P^2, \ldots, P^d) = \lim\limits_{n\to+\infty} \mathbf{f}_n(\mathbf{0})$ се нарича *вероятност за израждане на процеса* и се получава като решение на векторното уравнение $\mathbf{P} = \lim\limits_{n\to+\infty} \mathbf{f}_{n+1}(\mathbf{0}) = \mathbf{f}(\lim\limits_{n\to+\infty} \mathbf{f}_n(\mathbf{0})) = \mathbf{f}(\mathbf{P})$.

Дефиниция. 1.13 Многотиповите разклоняващи се процеси се наричат:

- надкритични (supercritical) процеси, ако $\rho > 1$;
- критични (critical) процеси, ако $\rho = 1$;
- докритичени (subcritical) процеси, ако $\rho < 1$.

Бейсов подход за многотипови разклоняващи се процеси

При *Бейсовия подход* за анализиране на данни се определя както моделът за наблюдение на данните $\boldsymbol{y} = \{y_1 \ldots y_n\}$, така и векторът на неизвестните параметри $\boldsymbol{\theta}$, зададен с разпределение $\pi(\boldsymbol{\theta})$, наречено априорно разпределение. Моделът се задава във формата на условно вероятностно разпределение $f(\boldsymbol{y} \mid \boldsymbol{\theta})$, а изводите относно параметъра $\boldsymbol{\theta}$ се базират на неговото апостериорно разпределение. *Апостериорната плътност* се изчислява чрез априорното разпределение, функцията на правдоподобие и правилото на Бейс:

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{\theta}, \boldsymbol{y})}{p(\boldsymbol{y})} = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{u})\pi(\boldsymbol{u})d\boldsymbol{u}}. \quad \text{или} \quad p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto f(\boldsymbol{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

Априорното разпределение може да се разглежда като представяне на текущото състояние от знания или текущото състояние на несигурност при моделиране на параметрите, преди да се наблюдават данните.

Избираме априорното разпределение да е от спрегната (conjugate prior) фамилия с фамилията на даденото разпределение $f(\boldsymbol{y} \mid \boldsymbol{\theta})$, което води до това *апостериорното разпределение да бъде от фамилията на априорното разпределение*. Бейсовият подход предоставя повече информация при вземане на решения. Използването на априорна информация допринася за увеличаване на точността и за редуциране на размера на извадката.

Разглеждаме $d-$типов разклоняващ се процес. Нека разпределението на потомството му е многомерен степенен ред. Тогава:

$$L(\tilde{\mathcal{J}}_n \mid \boldsymbol{\theta}) = \prod_{k=1}^{d} \prod_{\vec{i} \in S_k} \left(\frac{a_k(i_1, i_2, \ldots, i_d)\ \theta_{1k}^{i_1}\ \theta_{2k}^{i_2} \ldots \theta_{dk}^{i_d}}{A_k(\theta_{1k}, \theta_{2k}, \ldots, \theta_{dk})} \right)^{Z_k(n,\vec{i})}, \quad \theta_{jk} \in \mathcal{R}^+$$

$$A_k(\theta_{1k}, \theta_{2k}, .., \theta_{dk}) = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{\infty} \ldots \sum_{i_d=0}^{\infty} a_k(i_1, i_2, \ldots, i_d)\ \theta_{1k}^{i_1}\ \theta_{2k}^{i_2} \ldots \theta_{dk}^{i_d}.$$

Фамилията разпределения многомерен степенен ред е от класа на експоненциалната фамилия, защото $f(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{a(\boldsymbol{x})}{A(\boldsymbol{\theta})} \prod_{j=1}^{n} \theta_j^{x_j} = \big(A(\boldsymbol{\theta})\big)^{-1} a(\mathbf{x})\ exp\big(\sum_{j=1}^{n} x_j \log \theta_j\big)$. Използваме следната теорема:

Теорема. ***1.2*** *Функцията на правдоподобие за експоненциалната фамилия разпределения има спрегнато априорно разпределение (*conjugate prior*), ако априорното разпределение е от вида* $p(\boldsymbol{\theta}) \propto C(\boldsymbol{\theta})^a exp\big(\sum_{j=1}^{n} \phi_j(\boldsymbol{\theta}) b_j\big)$

В следващите теореми сме намерили апостериорните разпределения за фамилията многомерен степенен ред.

Теорема. ***1.3*** *Нека априорното разпределение на параметрите на индивидуалното разпределение за d-типов разклоняващ се процес с индивидуално разпределение многомерен степенен ред е*

$$\pi(\boldsymbol{\theta}) = \frac{\prod_{k=1}^{d} \theta_{1k}^{\alpha_{1k}}\ {\theta_{2k}}^{\alpha_{2k}}\ {\theta_{3k}}^{\alpha_{3k}}, \ldots, {\theta_{dk}}^{\alpha_{dk}}}{C(\boldsymbol{\alpha}, \boldsymbol{\beta}) \prod_{k=1}^{d} A_k(\theta_{1k}, \theta_{2k}, \ldots, \theta_{dk})^{\beta_k}}.$$

Тогава при дадена извадка в n-то поколение

$$\widetilde{\mathcal{J}}_n = \Big\{ Z_k(n, \boldsymbol{i}) = z_k(n, \boldsymbol{i}),\ k = 1, 2, \ldots, d \Big\},\ \textit{където}\ \boldsymbol{i} = (i_1, i_2, \ldots, i_d) \in \mathcal{S}_k$$

а $Z_k(n,\boldsymbol{i})$ е броят частици от тип k, които в $n-$то поколение имат деца $\boldsymbol{i}$, спрегнатото апостериорно разпределение е

$$f(\boldsymbol{\theta}|\widetilde{\mathcal{J}}_n) \propto \prod_{k=1}^{d} \frac{\theta_{1k}^{\sum_{n=0}^{N-1} Z_1^k(n+1)+\alpha_{1k}} . \theta_{2k}^{\sum_{n=0}^{N-1} Z_2^k(n+1)+\alpha_{2k}} \ldots \theta_{dk}^{\sum_{n=0}^{N-1} Z_d^k(n+1)}}{A_k(\theta_{1k},\theta_{2k},\ldots,\theta_{dk})^{\sum_{n=0}^{N-1} Z_k(n)}}$$

където $Z_j^k(n+1)$ е броят частици от тип j, чийто родител от $n-$то поколение е от тип k, а $Z_k(n)$ е броят частици от тип k в $n-$то поколение.

Аналогични теореми са доказани за всяко от разглежданите конкретни разпределения от фамилията многомерен степенен ред.

Теорема. ***1.4*** *Апостериорното разпределение на параметрите на индивидуалното разпределение за d-типов разклоняващ се процес с полиномно разпределение на потомството, при дадена извадка*

$$\widetilde{\mathcal{J}_{\mathcal{N}}} = \Big\{\{Z_k(n,\vec{\boldsymbol{i}}) = z_k(n,\vec{\boldsymbol{i}})\},\ \vec{\boldsymbol{i}} = (i_1,i_2,\ldots,i_d) \in \mathcal{S}_k, k = 1,2,..,d,\ n = \overline{0,N}\Big\}.$$

и при спрегнато априорно разпределение на Дирихле - $Dirichlet(\alpha_{1k},\alpha_{2k},\ldots,\alpha_{dk},\alpha_{(d+1)k})$, е произведение от разпределения на Дирихле и е от вида

$$\prod_{k=1}^{d} Dirichlet\left(\sum_{n=1}^{N} Z_1^k(n+1)+\alpha_{1k},..\sum_{n=1}^{N} Z_d^k(n+1)+\alpha_{dk},\sum_{n=0}^{N-1}(M_k Z_k(n) - Z_k(n+1))+\alpha_{(d+1)k}\right),$$

където $Z_j^k(n+1)$ е брой частици от тип j с родител в $n-$то поколение от тип k, а $Z_k(n)$ е брой частици от тип k в $n-$то поколение.

Теорема. ***1.5*** *Апостериорното разпределение на параметрите на индивидуалното разпределение за d-типов разклоняващ се процес с отрицателно полиномно разпределение на потомството, при дадена извадка*

$$\widetilde{\mathcal{J}_{\mathcal{N}}} = \Big\{\{Z_k(n,\vec{\boldsymbol{i}}) = z_k(n,\vec{\boldsymbol{i}})\},\ \vec{\boldsymbol{i}} = (i_1,i_2,\ldots,i_d) \in \mathcal{S}_k, k = 1,2,..,d,\ n = \overline{0,N}\Big\}.$$

и при спрегнато априорно разпределение - обратното Дирихле разпределение $InvertedDirichlet(\alpha_{1k},\alpha_{2k},\ldots,\alpha_{dk})$, е произведение от обратни Дирихле разпределения и е от вида

$$\prod_{k=1}^{d} InvertedDirichlet\Big(\sum_{n=1}^{N} Z_1^k(n)+\alpha_{1k},\ldots,\ \sum_{n=1}^{N} Z_2^k(n)+\alpha_{(d-1)k},\ M\sum_{n=0}^{N-1} Z_k(n)+\alpha_{dk}\Big)$$

където $Z_j^k(n+1)$ е брой частици от тип j в $(n+1)-$во поколение,с родител

в $n-$то поколение е от тип k, а $Z_k(n)$ е брой частици от тип k в $n-$то поколение.

Теорема. *1.6* *Апостериорното разпределение на параметрите на индивидуалното разпределение за d-типов разклоняващ се процес с многомерно логаритмично разпределение на потомството, при дадена извадка*

$$\widetilde{\mathcal{J}_{\mathcal{N}}} = \Big\{\{Z_k(n,\vec{\boldsymbol{i}}) = z_k(n,\vec{\boldsymbol{i}})\},\ \vec{\boldsymbol{i}} = (i_1, i_2, \ldots, i_d) \in \mathcal{S}_k, k = 1,2,..,d,\ n = \overline{0,N}\Big\}.$$

и при спрегнато априорно рапзределение $\pi(\theta_{1k}, \theta_{2k}, \ldots, \theta_{dk}) \propto \dfrac{\prod\limits_{j=1}^{d} \theta_{jk}^{\alpha_j}}{\big(-\log(1-\sum\limits_{j=1}^{d}\theta_{jk})\big)^{\beta}}$

е разпределения от вида $\prod\limits_{k=1}^{d} \dfrac{\prod\limits_{j=1}^{d} \theta_{jk}^{\sum\limits_{n=1}^{N-1} Z_j^k(n+1)+\alpha_j}}{\Big(-\log(1-\sum\limits_{j=1}^{d}\theta_{jk})\Big)^{\sum\limits_{n=1}^{N-1} Z_k(n)+\beta}}$, *където*

$Z_j^k(n+1)$ е брой частици от тип j в $(n+1)-$во поколение, чиито родител в $n-$топоколение е от тип k, а $Z_k(n)$ е брой частици от тип k в $n-$то поколение.

Теорема. *1.7* *Апостериорното разпределение на параметрите на индивидуалното разпределение за двутипов разклоняващ се процес с двумерно Поасоново разпределение на потомството, при дадена извадка*

$$\widetilde{\mathcal{J}_{\mathcal{N}}} = \Big\{\{Z_k(n,\vec{\boldsymbol{i}}) = z_k(n,\vec{\boldsymbol{i}})\},\ \vec{\boldsymbol{i}} = (i_1, i_2, \ldots, i_d) \in \mathcal{S}_k, k = 1,2,..,d,\ n = \overline{0,N}\Big\}.$$

и при спрегнато априорно разпределение $\prod\limits_{k=1}^{2}\prod\limits_{j=1}^{2} Gamma(\alpha_{jk}, \beta_{jk})$, *е произведение от Гама разпределения от вида*

$$Gamma(\sum_{n=0}^{N-1} Z_j^k(n+1) + \alpha_{jk}, \sum_{n=0}^{N-1} Z_k(n) + \beta_{jk}), \quad j,k = 1,2.$$

където $Z_j^k(n+1)$ е брой частици от тип j в $(n+1)-$во поколение, с родител в $n-$то поколение е от тип k, а $Z_k(n)$ е брой частици от тип k в $n-$то поколение.

Глава 2. Статистическо оценяване на МВР

Наблюдават се първите N поколения от фамилното дърво на многотипов разклоняващ се процес. Разглеждат се следните извадъчни схеми:

1. $\mathcal{J}_N = \{Z(0), \ldots, Z(N)\}$ размерът на популацията в първите N поколения
2. $\widetilde{\mathcal{J}}_N = \Big\{ Z_k\big(n, (i_1, i_2..i_d)\ \big) = z_k\big(n,\ (i_1, i_2..i_d)\big),\ n = \overline{0..N},\ (i_1, i_2..i_d) \in S_k \Big\}$ броят на индивидите от всеки тип k, в дадено поколение n, които имат точно $(i_1, i_2..i_d) \in \mathcal{S}_k$ деца и $\mathcal{S}_k$ е носителят за частица от тип k.
3. $\widetilde{\widetilde{\mathcal{J}}}_N = \Big\{ \xi_{is}^k(n) :\ s = 1, 2, .., Z_i(n);\ i, k = \overline{1, d},\ \ n = \overline{0..N} \Big\}$ цялото фамилно дърво, $\xi_{is}^k(n)$ е броят деца от тип k в $n-$то поколение с родител $s-$тата частица от тип i.

Наблюдаването на цялото фамилно дърво е най-удобно от статистическа гледна точка, но събирането на информация за поведението на всяка отделна частица представлява проблем. Затова често се налага "да се възстанови" цялото фамилно дърво на базата на първите две извадъчни схеми.

Едно решение на този проблем, в *непараметричния случай*, дават Gonzalez, Martin, Martinez и Mota, които предлагат да се възстанови разпределението на броя на частиците с определен брой наследници до N-тото поколение на фамилното дърво, при условие, че са наблюдавани единствено големините на поколенията, като произведение от условните разпределения на броя на частиците с определен брой наследници във всяко едно от поколенията $n < N$, при условие, че са наблюдавани само големините на текущото поколение n и на следващото $(n+1)$-во поколение.

Теорема. 2.1 *Нека* $\boldsymbol{p} = (p_i,\ i = \overline{1, d})$ *е векторът от индивидуалните вероятности на* $d-$*типов МВР, където* $p_i = (p_{i\boldsymbol{k}}\ :\ \boldsymbol{k} \in S_i\ )$ *и* $\boldsymbol{k}$ *е векторът от потомството на частица от тип* i*, чийто носител е* S_i*.*

Тогава $$P\big(\widetilde{\mathcal{J}}_N \mid \mathcal{J}_N, \boldsymbol{p}\big) = \prod_{n=0}^{N-1} P\Big(\widetilde{\mathcal{J}}_n \mid \boldsymbol{Z}(n) = \boldsymbol{z}(n), \boldsymbol{Z}(n+1) = \boldsymbol{z}(n+1), \boldsymbol{p} \Big),$$

където $n = \overline{1, (N-1)}$ *са първите* N *поколения от* $d-$*типовия МВР.*

В настоящата дисертация се разглеждат многотипови разклоняващи се процеси с фиксирано индивидуално разпределение, принадлежащо на класа

от разпределения многомерен степенен ред. Самото разпределение е определено с точност до неизвестен параметричен вектор $\boldsymbol{\theta}$.

В така разгледаният модел на процеса, в настоящата работа е направен *параметричен Бейсов анализ.*

Поради тясната връзка между стойностите на параметрите и индивидулните вероятности, прилагаме горната теорема:

Следствие. ***2.1*** *За многотипов разклоняващ се процес при използване на извадъчна схема $\mathcal{J}_N$, когато не са известни стойностите на сл. в. от множеството $\widetilde{\mathcal{J}}_N$, при $n = \overline{1,(N-1)}$, за условното разпределение за фамилното дърво е в сила формулата*

$$P\big(\widetilde{\mathcal{J}}_N \mid \mathcal{J}_N, \boldsymbol{\theta}\big) = \prod_{n=0}^{N-1} P\Big(\widetilde{\mathcal{J}}_n \mid \boldsymbol{Z}(n) = \boldsymbol{z}(n), \boldsymbol{Z}(n+1) = \boldsymbol{z}(n+1), \boldsymbol{\theta}\Big), \qquad \textit{където}$$

$$\widetilde{\mathcal{J}}_n = \big\{\{Z_k(n,\vec{\boldsymbol{i}}) = z_k(n,\vec{\boldsymbol{i}})\}, \vec{\boldsymbol{i}} = (i_1, i_2, \ldots, i_d) \in \mathcal{S}_k, k = \overline{1,d}\big\}.$$

Така можем да съсредоточим изследванията си за едно поколение.

За да се възстановят първите N поколения на цялото фамилно дърво на многотипов разклоняващ се процес, т.е. извадъчна схема $\widetilde{\widetilde{\mathcal{J}}}_N$, е достатъчно до се знае разпределението на броя частици с определен брой наследници, до N-тото поколение на фамилното дърво или извадъчна схема $\widetilde{\mathcal{J}}_N$.

Теорема. ***2.2*** *За двутипов разклоняващ се процес с индивидуално разпределение двумерен степенен ред, при наблюдаване на цялото фамилно дърво, за функцията на правдоподобие са в сила твърденията:*

- $L(\widetilde{\widetilde{\mathcal{J}}}_N \mid \boldsymbol{\theta}) = L(\widetilde{\mathcal{J}}_N \mid \boldsymbol{\theta}), \quad$ *където* $\quad \boldsymbol{\theta} = (\theta_{11}, \theta_{21}, \theta_{12}, \theta_{22})$
- $L(\widetilde{\widetilde{\mathcal{J}}}_n \mid \boldsymbol{\theta}) = L(\widetilde{\widetilde{\mathcal{J}}}_n \mid \theta_{11},\ \theta_{21})\ .\ L(\widetilde{\widetilde{\mathcal{J}}}_n \mid \theta_{12},\ \theta_{22})$

В случай, че се наблюдава единствено размера на популацията в първите N поколения (извадъчна схема $\mathcal{J}_N$), изчисляването на функцията на правдоподобие е невъзможно, поради липса на наблюдения. В такъв случай се използва условното разпределение на случайните величини $Z_k(n,(i,j)),\ \ k = 1,..,d,\ \ (i,j) \in S_k$, формиращи извадката $\widetilde{\mathcal{J}}_n$ за всяко поколение $n = 1,..,N$ от фамилното дърво.

За това условно разпределение, участващо като множител в Следствие 2.1, *доказваме следната основна теорема* [4], публикувана от Atanasov, Staneva, Stoimenova в статията "Robust estimators for the bivariate power series offspring distributions" през 2014 г., чийто аналог за контролируеми разклоняващи се процеси е разгледан от Gonzalez.

Теорема. *2.4* *(Основна теорема) За многотипови разклоняващи се процеси с разпределение от тип многомерен степенен ред* $p^k_{\vec{\boldsymbol{i}}} = \frac{a(x_1,\ldots,x_d)}{A(\theta_1,\ldots,\theta_d)} \prod_{l=1}^{d} {\theta_l}^{x_l}$ *и краен носител* $\vec{\boldsymbol{i}} = (i_1,..,i_d) \in \mathcal{S}_k,\ |\mathcal{S}_k| < \infty,\ k = \overline{1,d},$ *е изпълнено*

$$P\big(\widetilde{\mathcal{J}}_n \mid \boldsymbol{Z}(n) = \boldsymbol{z}(n), \boldsymbol{Z}(n+1) = \boldsymbol{z}(n+1), \boldsymbol{\theta}\big) =$$

$$\frac{1}{P\big(\boldsymbol{Z}(n+1) = \boldsymbol{z}(n+1) \mid Z(n) = z(n), \boldsymbol{\theta}\big)} \prod_{k=1}^{2} \frac{z_k(n)!}{\prod\limits_{\vec{\boldsymbol{i}}} z_k(n,\vec{\boldsymbol{i}})!} \prod_{\vec{\boldsymbol{i}}=(i_1,..,i_d)} (p^k_{\vec{\boldsymbol{i}}})^{z_k(n,\vec{\boldsymbol{i}})}$$

Като пример на *степенен ред с безкраен носител е разгледано многомерното Поасоново разпределение*, за което са доказани следните теореми (Staneva[81]).

Теорема. *2.5* *За многотипов разклоняващ се процес с многомерно Поасоново индивидуално разпределение е в сила*

$$P\left(\mathbf{Z}(n+1) = \mathbf{z}^*(n+1) | \mathbf{Z}(n) = \mathbf{z}^*(n)\right) = \prod_{k=1}^{d} \frac{e^{-\left(\sum\limits_{m=1}^{d} z^*_m(n)\ \theta_{mk}\right)} \left(\sum\limits_{m=1}^{d} z^*_m(n)\ \theta_{mk}\right)^{z^*(n+1)}}{z^*(n+1)\ !}$$

Следващата теорема е доказана в двумерния случай, но може да се обобщи.

Теорема. *2.6* *За двутипов разклоняващ се процес с двумерно Поасоново индивидуално разпределение е в сила*

$$P\big(\widetilde{\mathcal{J}}_n \mid \mathcal{J}_n, \boldsymbol{\theta}\big) = \prod_{k=1}^{2} \prod_{j=1}^{2} \frac{\left(\sum\limits_{s=1}^{z^*_j(n)} \xi^k_{js}(n)\right)!}{\prod\limits_{s=1}^{z^*_j(n)} \xi^k_{js}(n)!} \cdot \left(\frac{1}{z^*_j(n)}\right)^{\sum\limits_{s=1}^{z^*_j(n)} \xi^k_{js}(n)} \times$$

$$\prod_{k=1}^{2} \frac{z^*_k(n+1)!}{z^1_k(n)!\ \ z^2_k(n)!} \left(\frac{z^*_1(n)\theta_{k1}}{z^*_1(n)\theta_{k1} + z^*_2(n)\theta_{k2}}\right)^{z^1_k(n)} \left(\frac{z^*_1(n)\theta_{k1}}{z^*_1(n)\theta_{k1} + z^*_2(n)\theta_{k2}}\right)^{z^2_k(n)}.$$

$$\textit{където } z^t_k(n) = \sum_{s=1}^{z^*_t(n)} \xi^k_{ts}(n), \quad Z_k(n+1) = \sum_{s=1}^{Z_1(n)} \xi^k_{1s}(n) + \sum_{s=1}^{Z_2(n)} \xi^k_{2s}(n)$$

Следствие. ***2.2*** *При двумерно Поасоново индивидуално разпределение, условното разпределение* $\left(\widetilde{\mathcal{J}}_n \mid \mathcal{J}_n, \boldsymbol{\theta}\right)$ *може да се представи като произведение от полиномни разпределения*[1]

Метод за симулиране на броя на децата от всеки тип във всяко поколение

1) Генерираме в $(n+1)$-во поколение броя на частиците от първи тип $\sum\limits_{s=1}^{Z_1(n)} \xi_{1s}^1(n)$ и $\sum\limits_{s=1}^{Z_2(n)} \xi_{2s}^1(n)$, потомци съответно на частиците от първи и втори тип от n-то поколение. Използваме полиномно разпределение с брой опита $Z_1(n+1) = \sum\limits_{s=1}^{Z_1(n)} \xi_{1s}^1(n) + \sum\limits_{s=1}^{Z_2(n)} \xi_{2s}^1(n)$ и вероятности за събитията $(\frac{Z_1(n)\theta_{11}}{Z_1(n)\theta_{11}+Z_2(n)\theta_{21}}, \frac{Z_2(n)\theta_{21}}{Z_1(n)\theta_{11}+Z_2(n)\theta_{21}})$.

2) Генерираме в $(n+1)$-во поколение броя на частиците от първи тип $\sum\limits_{s=1}^{Z_1(n)} \xi_{1s}^2(n)$ и $\sum\limits_{s=1}^{Z_2(n)} \xi_{2s}^2(n)$, потомци съответно на частиците от първи и втори тип от n-то поколение. Използваме полиномно разпределение с брой опита $Z_2(n+1) = \sum\limits_{s=1}^{Z_1(n)} \xi_{1s}^2(n) + \sum\limits_{s=1}^{Z_2(n)} \xi_{2s}^2(n)$ и вероятности за събитията $(\frac{Z_1(n)\theta_{12}}{Z_1(n)\theta_{12}+Z_2(n)\theta_{22}}, \frac{Z_2(n)\theta_{22}}{Z_1(n)\theta_{12}+Z_2(n)\theta_{22}})$.

3) Използвайки генерирания брой $\sum\limits_{s=1}^{Z_1(n)} \xi_{1s}^1(n)$, генерираме броя наследници $\xi_{1s}^1(n)$, където $s = 1, .., Z_1(n)$, за всяка една от общо $Z_1(n)$ броя частици от първи тип, чието потомство е от тип 1, използвайки полиномно разпределение със $\sum\limits_{s=1}^{Z_1(n)} \xi_{1s}^1(n)$ опита и $Z_1(n)$ събития с равни вероятности, равни на $\frac{1}{Z_1(n)}$.

4) Използвайки генерирания брой $\sum\limits_{s=1}^{Z_2(n)} \xi_{2s}^1(n)$, генерираме броя наследници $\xi_{2s}^1(n)$, където $s = 1, .., Z_2(n)$, за всяка от $Z_2(n)$ броя частици от тип 2, чието потомство е от тип 1, използвайки полиномно разпределение със $\sum\limits_{s=1}^{Z_2(n)} \xi_{2s}^1(n)$ опита и $Z_2(n)$ събития с равни вероятности, равни на $\frac{1}{Z_2(n)}$.

5) Използвайки генерирания брой $\sum\limits_{s=1}^{Z_1(n)} \xi_{1s}^2(n)$, генерираме броя наследници $\xi_{1s}^2(n)$, където $s = 1, .., Z_1(n)$, за всяка от $Z_1(n)$ броя частици от тип 1, чието потомство е от тип 2, използвайки полиномно разпределение със $\sum\limits_{s=1}^{Z_1(n)} \xi_{1s}^2(n)$ опита и $Z_1(n)$ събития с равни вероятности, равни на $\frac{1}{Z_1(n)}$.

---

[1] За по разбираем вид формулата може да се запише така: $\left(\widetilde{\mathcal{J}}_n \mid \mathcal{J}_n, \theta\right) \sim$

$$\prod_{k=1}^{2} Multinomial\Big(\sum_{s=1}^{Z_1(n)} X_{1s}^k,\ \Big\{\frac{1}{Z_1(n)}\Big\}\Big).Multinomial\Big(\sum_{s=1}^{Z_2(n)} X_{2s}^k,\ \Big\{\frac{1}{Z_2(n)}\Big\}\Big) \prod_{k=1}^{2} Multinomial\Big(Z_k(n+1), \Big\{\frac{Z_1(n)\theta_{k1}}{Z_1(n)\theta_{k1}+Z_2(n)\theta_{k2}}, \frac{Z_2(n)\theta_{k2}}{Z_1(n)\theta_{k1}+Z_2(n)\theta_{k2}}\Big\}\Big)$$

6) Използвайки генерирания брой $\sum_{s=1}^{Z_2(n)} \xi_{2s}^2(n)$, генерираме броя наследници $\xi_{2s}^2(n)$, където $s = 1, 2, \ldots, Z_2(n)$, за всяка от $Z_2(n)$ броя частици от тип 2, чието потомство е от тип 2, използвайки полиномно разпределение със $\sum_{s=1}^{Z_2(n)} \xi_{2s}^2(n)$ опита и $Z_2(n)$ събития с равни вероятности, равни на $\frac{1}{Z_2(n)}$.

Робастно оценяване

Дефиниция. 2.3 Нека $\boldsymbol{X} = (x_1, \ldots, x_n)$ е крайна извадка с обем n.

$$BP(T) = \frac{1}{n} \max\{m : \sup_{\boldsymbol{X}'} \|T(\boldsymbol{X}') - T(\boldsymbol{X})\| < \infty\} \quad \text{наричаме}$$

прагова точка на статистиката T, където X' е коя да е извадка, получена от X чрез заместване на кои да е m от стойностите на X с произволни стойности.

Робастно разширение на MLE, което притежава висока прагова точка, е *орязаното и претеглено правдоподобие* (WTL), въведено от Vandev и Neykov(1993). Във функцията на правдоподобие, от извадката с размер n, те включват само k на брой наблюдения с най-голяма вероятностна плътност. Останалите $n - k$ наблюдения се разглеждат като аутлаери.

Дефиниция. 2.4 Числото $k \in ([\frac{n}{2}, n]$ се нарича *фактор на орязване.*

Факторът на орязване k определя броя на наблюденията, участващи в оценката и нивото на устойчивост на тази оценка. Малката стойност на k се свързва с висока прагова точка, но тогава участват малък брой наблюдения и функцията на правдоподобие има по-плоска графика.

През 1992г. Vandev въвежда понятието $d-$пълнота, а през 1993г. Vandev и Neykov изучават връзката между прагова точка при крайна извадка и $d-$пълнотата.

Дефиниция. 2.5 (Vandev) *Крайното множество F от n функции се нарича $d-$пълно, ако за всяко подмножество на F, с кардиналност d, супремумът на това подмножество е субкомпактна функция.*

Дефиниция. 2.6 (Vandev, Vandev & Neykov) Реалнозначната функция $g(\theta)$, дефинирана в топологичното пространство Θ, се нарича субкомпактна, ако нейните Лебегови множества $L_g(C) = \{\theta : g(\theta) \leq C\}$ са компактни за произволна константа C.

Vandev(1993) доказва, че ако едно множество е $d-$пълно, то е и $d+1-$пълно.

В дисертацията е доказано следното твърдение:

Твърдение. 2.1 *Всяка непрекъсната функция $g(x)$, дефинирана в компактно множество D е субкомпактна.*

През 1996г. Neykov доказва, че сума на константа и субкомпактна функция е субкомпактна функция, както и че произведение на положителна константа и субкомпактна функция е субкомпактна функция.

В дисертацията са използвани и следните две свойства:

Твърдение. 2.4 *(Neykov) Ако функцията $f(x)$ е субкомпактна и $f(x) \leq g(x)$ за всяко x от дефиниционната област на непрекъснатата функция $g(x)$, то $g(x)$ също е субкомпактна функция.*

Теорема. 2.7 *(Atanasov&Neykov) Реалнозначната непрекъсната функция $g(\theta)$, дефинирана в отвореното множество $\Theta \subseteq R^n$ е субкомпактна $\iff$ за всяка редица от точки $\{\theta_i\} \in \Theta$, сходяща към точка от границата $\theta_0 \in \partial\Theta$ е изпълнено $g(\theta_i) \xrightarrow{i\to\infty} \infty$ при $\theta_i \xrightarrow{i\to\infty} \theta_0$.*

Дефиниция. 2.7 (WLTE) Нека $x_1, x_2, \ldots, x_n$ са n на брой независими и еднакво разпределени наблюдения с вероятностна плътност $\varphi(x, \boldsymbol{\theta})$, където $\boldsymbol{\theta}$ е неизвестен параметричен вектор. Претеглена и орязана максимално правдоподобна оценка от ред k $-WLTE(k)$ за параметричния вектор $\boldsymbol{\theta}$ се нарича

$$WLTE(k) = arg \min_{\boldsymbol{\theta}\in\Theta^p} \sum_{1}^{k} w_i f_{\nu(i)}(\boldsymbol{\theta}), \tag{1}$$

където $f_{\nu(1)} \leq f_{\nu(2)}, \leq \ldots, \leq f_{\nu(n)}$ са подредени стойности на $f_i = -\log\varphi(x_i, \boldsymbol{\theta})$ в $\boldsymbol{\theta}$, а $\nu(1) \leq \nu(2), \ldots, \nu(n)$ е съответната пермутация на индексите, които могат да зависят от $\boldsymbol{\theta}$. Теглата $w_i \geq 0,\ i = 1, .., k$ са такива, че съществува индекс $k = \max\{i\ :\ w_i > 0\}$.

Намирането на праговата точка на оценката WLTE(k), изисква да се намери индекса на пълнота d на множеството $F = \{-\log f(x_i, \boldsymbol{\theta})\}, \quad i = \overline{1, k}$.

Въвеждаме означението $R(k) = \big\{\boldsymbol{\theta} : \boldsymbol{\theta} = arg\min\big(\sum_{i=1}^{k} w_i . f_{\nu(i)}(\boldsymbol{\theta})\big)\big\}$.

В дисертацията използваме следващата основна теорема:

Теорема. 2.8 *(Vandev & Neykov) Праговата точка на $WLTE(k)$, върху извадката $(x_1, x_2, \ldots, x_n)$, е не по малка от $\frac{n-k}{n}$, ако множеството функции $F = \left\{ f_i(\boldsymbol{\theta}) = -\log \varphi(x_i, \boldsymbol{\theta}),\ i = \overline{1,n},\ \boldsymbol{\theta} \in \Theta \right\}$ е d- пълно, за $n \geq 3d$ и $\frac{n+d}{2} \leq k \leq n-d$. Тук $f_{\nu(1)} \leq \cdots \leq f_{\nu(n)}$ са подредените стойности на f, а $\nu(1) \leq \cdots \leq \nu(n)$ е пермутацията на индексите, които зависят от $\boldsymbol{\theta}$.*

За следващия резултат наблюдаваме цялото фамилно дърво на двумерен разклоняващ се процес с индивидуално разпределение двумерен степенен ред и въвеждаме означенията:

- $\mathcal{S} = (\mathcal{S}^1, \mathcal{S}^2)$ е множеството от възможните стойности на сл. вектор (X, Y);
- $\mathcal{E}$ е областта на сходимост на степенния ред $A(\theta_{1k}, \theta_{2k})$, а $\partial\mathcal{E}$ е нейната граница;
- $MaxVX \in \mathcal{S}^1$ и $MaxVY \in \mathcal{S}^2$ са максималната стойност съответно на случайната променлива X и случайната променлива Y;
- $MinVX \in \mathcal{S}^1$ и $MinxVY \in \mathcal{S}^2$ са минималната стойност съответно на случайната променлива X и случайната променлива Y;
- Ако $(MinVX, MinVY) \in \mathcal{S}$, тогава $N_{MinVX,\, MinVY}$ е наблюдавания брой вектори в извадката с минимални стойности на координатите;
- Ако $MaxVX < \infty$, $MaxVY < \infty$ и $(MaxVX, MaxVY) \in \mathcal{S}$, тогава означаваме с $N_{MaxVX,\, MaxVY}$ наблюдавания брой вектори в извадката с максимална стойности на координатите;

В следващата Теорема, прилагаме Теорема 2.8 и намираме долна граница на праговата точка на оценката на параметъра на индивидуалното двумерно разпределение от тип двумерен степенен ред, в зависимост от стойността на въведените величини. Теорема е публикувана от Atanasov, Staneva, Stoimenova.

Теорема. 2.9 *Разглеждаме извадка от n независими двумерни наблюдения с разпределение двумерен степенен ред.*

$$\{(x_1, y_i), .., (x_n, y_n)\}, \quad \{(x_k, y_k)\} \sim \frac{{\theta_1}^{x_k}\, {\theta_2}^{y_k}}{A(\theta_1, \theta_2)}$$

За праговата точка на оценката WLTE(k) са в сила твърденията:

1) *Нека* $|\mathcal{S}| = \infty, \quad \theta_1 > 0, \ \theta_2 > 0 \quad$ *и* $\quad (\theta_1, \theta_2) \in \mathcal{E},$

$$\textit{като} \lim_{\theta_1, \theta_2 \to \partial\mathcal{E}} \frac{A(\theta_1, \theta_2)}{\theta_1^{\,i}\ \theta_2^{\,j}} = \infty, \quad \forall (i,j) \in \ \mathcal{S} \ \setminus \ (MinVX, MinVY).$$

Тогава $\exists$ *оценката* $WLTE(k)$ *и праговата точка не е по-малка от* $\dfrac{[n-k]}{n}$,

$$\textit{ако} \qquad n \geq 3\ (\gamma + \ 1\ ) \qquad \textit{и} \qquad \frac{\big[n+\gamma+1\big]}{2} \leq k \leq n-\gamma-1,$$

където $\gamma = N_{MinVX,\, MinVY}$

2) *Нека* $|\mathcal{S}| < \infty, \quad \theta_1 \in (0, \infty) \quad$ *и* $\quad \theta_2 \in (0, \infty)$

Тогава $\exists$ *оценката* $WLTE(k)$ *и праговата точка не е по-малка от* $\dfrac{[n-k]}{n}$,

$$\textit{ако} \qquad n \geq 3\ \big(\nu + 1\big) \qquad \textit{и} \qquad \frac{1}{2}\big(n+\nu+1\big) \leq k \leq n-\nu-1,$$

където $\nu = \max(N_{MinVX,MinVY}, N_{MaxVX,MaxVY})$.

Ако $(MaxVX, MaxVY) \notin \mathcal{S}$, *тогава е в сила (1)*

3) *Нека* $(\theta_1, \theta_2) \in \Phi \quad$ *и* $\quad \Phi$ *е компактно подмножество на* $\mathcal{E}$.

Тогава $\exists$ *оценката* $WLTE(k)$ *и праговата точка не е по-малка от* $\dfrac{[n-k]}{n}$,

$$\textit{ако} \qquad n \geq 3 \qquad \textit{и} \qquad \big[\frac{n+1}{2}\big] \leq k \leq n-1$$

Глава 3. EM алгоритъм за статистическа оценка на MBP

С помощта на EM алгоритъма е получена MLE за параметрите на индивидуалното разпределение на двутипов разклоняващ се процес, при *непълни наблюдавани данни*. Генерира се редица от приближения $\big\{\boldsymbol{\theta}^{(t)}\big\}_{t=1}^{\infty}$, при случайно начално $\boldsymbol{\theta}^{(0)}$. Нека $\boldsymbol{\theta}^{(t-1)}$ е дадено.

EM алгоритъмът има две стъпки E-стъпка и M-стъпка:

E step

Търси се условното очакване на скритите данни. Ако $\ell\big(\boldsymbol{\theta} \mid \widetilde{\mathcal{J}}_N, \mathcal{J}_N\big)$ е функцията на лог-правдоподобие при наблюдаване на данните от цялото дърво $\widetilde{\mathcal{J}}_N$ и $\mathcal{J}_N$, то ненаблюдаваните данни във функцията на правдоподобие се заместват

с условното очакване.

$$Q\big(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t-1)}\big) = E_{\widetilde{\mathcal{J}}_N|\mathcal{J}_N}\big[\ell\big(\boldsymbol{\theta} \mid \widetilde{\mathcal{J}}_N, \mathcal{J}_N\big)\big].$$

M-step

Новото приближение $\boldsymbol{\theta}^{(t)}$ се получава от $\boldsymbol{\theta}^{(t)} = arg\max_\theta Q\big(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t-1)}\big)$.

За да се изрази условното разпределение за едно поколение, се прилага следствието от Теорема 2.1. Прилага се и основната Теорема 2.4 и се получава условното разпределение за n-то поколение:

$$P(\widetilde{\mathcal{J}}_n|\mathcal{J}_n, \boldsymbol{\theta}) = \frac{\prod_{k=1}^{2} Mn\big(z_k(n), \{p^k_{(i_k,j_k)}\}_{(i_k,j_k)\in\mathcal{S}_k}\big)}{\sum_{s=1}^{r} \alpha_s(n).\beta_s(n)}, \text{ където } z_1(n) \text{ и } z_2(n)$$

са наблюдаваните данни и $Mn(x, prob)$ е полиномна вероятност.

За МВР с индивидуално разпределение MPSOD, Е-стъпката е:

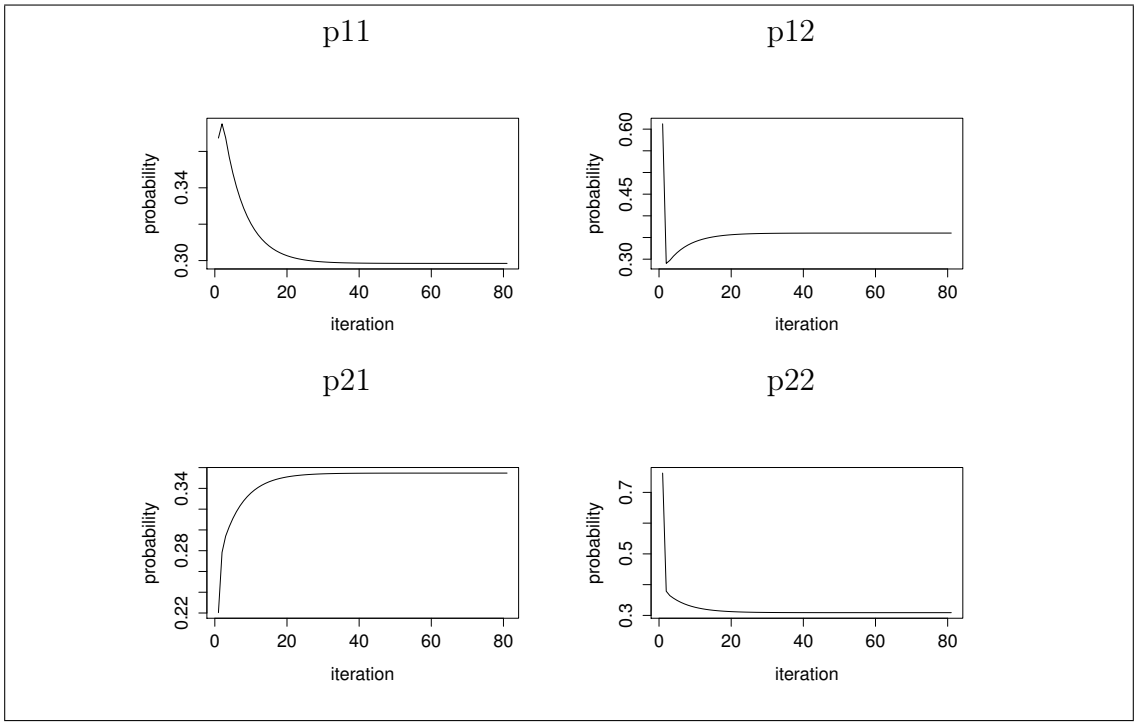
$$E_{\widetilde{\mathcal{J}}_N|\mathcal{J}_N}\big[\ell(\boldsymbol{\theta} \mid \widetilde{\mathcal{J}}_N)\big] = \sum_{k=1}^{2}\sum_{n=1}^{N-1}\sum_{(i_k,j_k)\in\mathcal{S}_k} E_{\widetilde{\mathcal{J}}_N|\mathcal{J}_N}\big[Z_k\big(n,(i_k,j_k)\big)\big] log\ p^k_{(i,j)}$$

Така за двутипов разклоняващ се процес с триномиално индивидуално разпределение, на М-стъпката получаваме приближенията:

$$\widehat{\theta_{ks}} = \frac{\sum_{n=0}^{N-1} E\big[Z^s_k(n+1)\big]}{M_s \sum_{n=0}^{N-1} Z_s(n) - \sum_{n=0}^{N-1} E\big[Z^s_1(n+1)\big] - \sum_{n=0}^{N-1} E\big[Z^s_2(n+1)\big]}, \ s,k = 1,2.$$

При отрицателно полиномно разпределение тези приближения са:

$$\widehat{\theta_{sk}} = \frac{\sum_{n=0}^{N-1} E\left[Z^k_s(n+1)\right]}{M\sum_{n=0}^{N-1} Z_k(n) + \sum_{n=0}^{N-1} E\left[Z^k_1(n+1)\right] + \sum_{n=0}^{N-1} E\left[Z^k_2(n+1)\right]} \quad s,k = 1,2. \quad (2)$$



Фигура 0.1: Класически EM алгоритъм, 80 итерации за 10 поколения от двутипов разклоняващ процес с полиномното разпределение.

При двумерно Поасоново разпределение за параметрите получаваме:

$$\widehat{\theta}_{sk} = \frac{\sum\limits_{n=0}^{N-1} E\Big[Z_s^k(n+1)\Big]}{\sum\limits_{n=0}^{N-1} Z_k(n)}, \quad s=1,2; \;\; k=1,2.$$

През 1999 г. Shiro Ikeda прави връзка между EM алгоритъма и *Fisher scoring* алгоритъма и използва EM алгоритъма рекурсивно.

Итерационната стъпка при EM алгоритъма той я представя във вида: $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}_t + {I_X}^{-1}(\boldsymbol{\theta}^{(t)})\, \partial\ell(x;\boldsymbol{\theta})$, където $\ell(x;\boldsymbol{\theta}) = \log L(x;\boldsymbol{\theta})\,, \qquad \partial = \big(\frac{\partial}{\partial\theta^1}, \frac{\partial}{\partial\theta^2}, \ldots, \frac{\partial}{\partial\theta^d}\big)^T,$ и $I_X(\boldsymbol{\theta}^{(t)})$ е информационната матрица на Фишер за $L(x;\boldsymbol{\theta})$:

$$I_X(\boldsymbol{\theta}^{(t)}) = \big(E_{p(x;\boldsymbol{\theta})}\big[\partial_i\ell(x;\boldsymbol{\theta})\partial_j\ell(x;\boldsymbol{\theta})\big]\big)_{i,j} = \big(-E_{p(x;\boldsymbol{\theta})}\big[\partial_i\partial_j\ell(x;\boldsymbol{\theta})\big]\big)_{i,j}.$$

Ikeda разширява информационната матрица на Фишер
${I_Y}^{-1} = \big(I + \sum\limits_{i=1}^{\infty}\big({I_X}^{-1}I_{Z|Y}\big)^i\big){I_X}^{-1}(\boldsymbol{\theta})$ и доказва следната теорема:

Теорема. *3.1* *(Ikeda) Нека* $\widehat{\boldsymbol{\theta}}^{(t+1)}$ *е оценката на параметъра* $\boldsymbol{\theta}$*, получена чрез*

прилагане на една EM стъпка от EM-алгоритъма при начална стойност $\boldsymbol{\theta}^{(t)}$ *и целево разпределение* $p(y;\boldsymbol{\theta})$. *Тогава* $\widehat{\boldsymbol{\theta}}^{(t+1)} \simeq \boldsymbol{\theta}^{(t)} + {I_X}^{-1} I_Y {I_X}^{-1} \partial\ell(x;\boldsymbol{\theta}^{(t)})$

Следствие от това е, че зависимостта $\boldsymbol{\theta}^* = 2\ \boldsymbol{\theta}^{(t+1)} - \widehat{\boldsymbol{\theta}}^{(t+1)}$ дава приближение от по-висок ред на Fisher scoring алгоритъма.

Ускореният EM алгоритъм, на Ikeda се състои от следните стъпки:

[Step 0:] Задаваме начална стойност: $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$, при $t = 0$.

[Step 1:] С начална стойност $\boldsymbol{\theta}^{(t)}$ и използвайки данните, изпълняваме една EM стъпка: $\boldsymbol{\theta}^{*(t+1)} = \operatorname{argmax}\left\{E_{p(z|y;\boldsymbol{\theta}^{(t)})}\left[\ell(y,z;\boldsymbol{\theta}_t)\right]\right\}$

[Step2:] С получената стойност на параметъра $\boldsymbol{\theta}^{*(t+1)}$ генерираме нови данни $\mathbf{y_1} \sim p(y;\boldsymbol{\theta}^{*(t+1)})$ от известното ни разпределение

[Step 3:] С начална стойност $\boldsymbol{\theta}^{(t)}$ и с генерираните в предната стъпка данни $\mathbf{y_1}$, изпълняваме още една EM стъпка: $\widehat{\boldsymbol{\theta}}^{(t+1)} = \operatorname{argmax}\left\{E_{p(z|y_1;\boldsymbol{\theta}^{(t)})}\left[\ell(y_1,z;\boldsymbol{\theta}^{(t)})\right]\right\}$

[Step 4:] $\boldsymbol{\theta}^{(t+1)} = 2\widehat{\boldsymbol{\theta}}^{(t+1)} - \boldsymbol{\theta}^{*(t+1)}$

[Step 5:] $t = t + 1$ и се връщаме на стъпка *Step 1.*

Глава 4. Monte Carlo алгоритми за оценка на многотипови разклоняващи се процеси

В дисертацията прилагаме Бейсов анализ за двутипови разклоняващи се процеси с индивидуално разпределение двумерен степенен ред и използваме *Gibbs sampler* за да апроксимираме апостериорното разпределение на незивестните за модела параметри. В симулиран двутипов разклоняващ се процес, приемаме, че се наблюдават само $\mathcal{J}_n = \{\boldsymbol{Z}(0),\ldots,\boldsymbol{Z}(n)\}$. Сканирането на Gibbs семплера се състои от следните стъпки:

[Стъпка 1.] Семплиране на ненаблюдаемите данни $\widetilde{\mathcal{J}}_n = \{\boldsymbol{x},\boldsymbol{y}\}$

$$\boldsymbol{x}^{(k)} \sim \frac{z_1(n)!}{\prod_{(i,j)\in S_1} z_1(n,(i,j))!} \prod_{(i,j)\in S_1} (p^1_{(i,j)})^{z_1(n,(i,j))} = Mn(z_1(n),\{p^1_{(i,j)}\})$$ [2]

$$\boldsymbol{y}^{(k)} \sim \frac{z_2(n)!}{\prod_{(i,j)\in S_2} z_2(n,(i,j))!} \prod_{(i,j)\in S_2} (p^2_{(i,j)})^{z_2(n,(i,j))} = Mn(z_2(n),\{p^2_{(i,j)}\}),$$

[2] Mn е съкращение на полиномно разпределение

като $\boldsymbol{x} = \{Z_1(n,(i,j)) : (i,j) \in S_1\}, \boldsymbol{y} = \{Z_2(n,(i,j)) : (i,j) \in S_2\}$

[Стъпка 2] Генериране на ново приближение за параметъра

$$\boldsymbol{\theta}^{(t+1)} \sim f(\boldsymbol{\theta}^{(t)} \mid \widetilde{\mathcal{J}}) \propto \pi(\boldsymbol{\theta}^{(t)}).L(\widetilde{\mathcal{J}} \mid \boldsymbol{\theta}^{(t)}), \quad \boldsymbol{\theta} = (\theta_{11}, \theta_{21}, \theta_{12}, \theta_{22})$$

- $f(\boldsymbol{\theta} \mid \widetilde{\mathcal{J}})$ е апостериорното разпределение; - $\pi(\boldsymbol{\theta})$ е приорното му разпределение, - $L(\widetilde{\mathcal{J}} \mid \boldsymbol{\theta})$ е функцията на правдоподобие.

Изборът на началните стойности за $\boldsymbol{\theta}$ не влияе на оценката.

Gibbs sampler прилагаме за:

-полиномното разпределение, където при $s = 1, 2$

$$\boldsymbol{p}_s^{(l)} \sim Dirichlet\big(\sum_{n=1}^{N} Z_1^s(n) + \alpha_{1s}, \sum_{n=1}^{N} Z_2^s(n) + \alpha_{2s}, \sum_{n=0}^{N-1} \big(MZ_s(n) - Z_s(n+1)\big) + \alpha_{3s}\big)$$

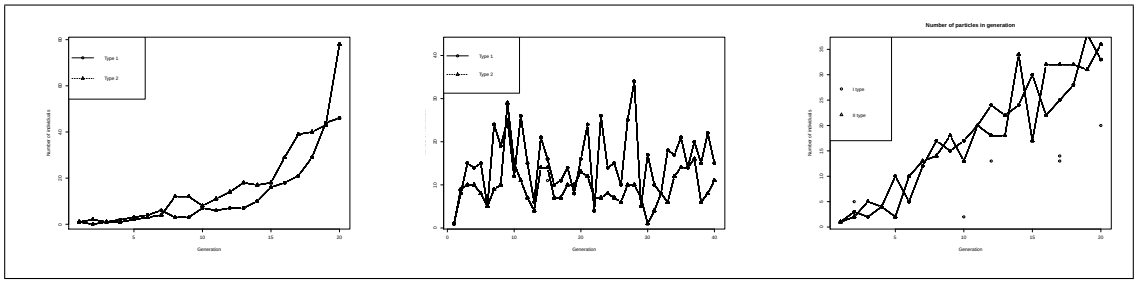
-отрицателното полиномно разпределение, където при $k = 1, 2$

$$\boldsymbol{p}_k^{(l)} \sim InvertedDirichlet\left(\sum_{n=1}^{N} Z_1^k(n) + \alpha_k, \sum_{n=1}^{N} Z_2^k(n) + \beta_k, \sum_{n=0}^{N-1} MZ_k(n) + \gamma_k\right)$$

-двумерното Поасоново разпределение, където при $k, s = 1, 2$

$$\theta_{ks}^{(l)} \sim Gamma\,\big(\sum_{n=0}^{N-1} Z_k^s(n+1) + \alpha_{ks}, \sum_{n=0}^{N-1} Z_s(n) + \beta_{ks}\big).$$

Сходимостта на алгоритъма е оценена, като са използвани *диагностиките за сходимост* на Gelman&Rubin, с които се оценява сходимостта на MCMC алгоритмите. Методът е приложен в симулирани процеси с полиномно, отрицателно полиномно и Поасоново разпределение. На фиг.(0.2) е показана популацията на двутипов процес съответно с полиномно, с отрицателно полиномно и с двумерно Поасоново индивидуално разпределение. На фигура (0.3) са показани графиките от диагностиката за сходимост на Gelman & Rubin, отнасящи се за двутипови процеси със същите индивидуални разпределения - полиномно и отрицателно полиномно и Поасоново, които са получени при използване на няколко Марковски вериги. Тук използваме описаният в приложението числен метод за определяне на носителя на отрицателното полиномно разпределение, което както е известно има безкраен носител.



Фигура 0.2: Размерът на популациите на двутипов разклоняващ процес с полиномно индивидуално разпределение, с отрицателно полиномно индивидуално и с Поасоново индивидуално разпределение

В настоящата дисертация е разгледан и Монте Карло ЕМ алгоритъма. При прилагането на Е-стъпката в общия случай е много трудно да се изчисли условното очакване, защото се получават твърде големи суми и много обемни изчисления, изискващи много процесорно време. Това налага използването на Монте Карло алгоритмите. При прилагането на MCMC методите се използват фиксирани стойности на параметрите, за да се семплира от условното разпределение на ненаблюдаваните данни относно наблюдаваните данни, защото не може да се получат извадки директно от разпределението, което трябва да се оцени.

Основните стъпки на Монте Карло ЕМ алгоритъма са:

[Step 0:] Инициализиране на параметрите $\theta_{11}^{(0)},\ \theta_{21}^{(0)},\ \theta_{12}^{(0)},\ \theta_{22}^{(0)}$;

[Step 1:] Изчисляване на индивидуалните вероятности за двата типа частици $p^1_{(i_1,j_1)}$ и $p^2_{(i_2,j_2)}$, където $(i_k, j_k) \in S_k$. За да се получат ненаблюдаваните данни $z_k(n,(i,j))$, се семплира от условното разпределение $p\big(\widetilde{\mathcal{J}}_n|\mathcal{J}_n,\boldsymbol{\theta}\big)\ \{z_k(n,(i,j))\} \sim p\big(\widetilde{\mathcal{J}}_n|\mathcal{J}_n,\boldsymbol{\theta}\big)$, като се използва удобен MCMC алгоритъм, например Gibbs Sampler.

[Step 2 - E стъпка:] Монте Карло изчисления. На тази стъпка се апроксимира условното очакване

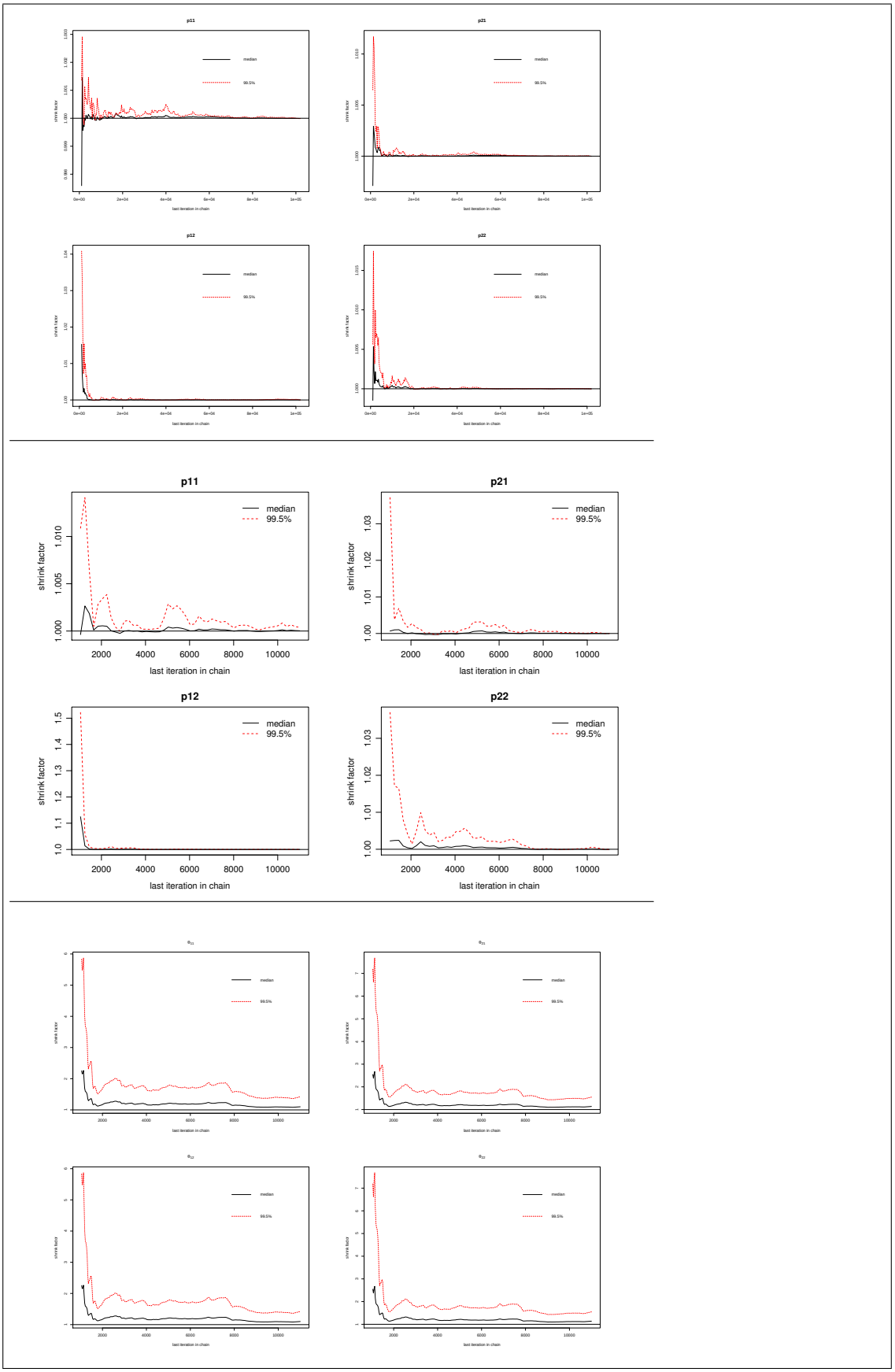
$$E_{\widetilde{\mathcal{J}_\mathcal{N}}|J}\Big[\ell(\boldsymbol{\theta}\mid \widetilde{\mathcal{J}_\mathcal{N}},\mathcal{J}_\mathcal{N},)\Big] = \sum_{y\in\widetilde{\mathcal{J}_\mathcal{N}}} \ell(\boldsymbol{\theta}|y,\mathcal{J}_\mathcal{N}).p\big(\widetilde{\mathcal{J}_\mathcal{N}}|\mathcal{J}_\mathcal{N},\boldsymbol{\theta}\big)$$

с функцията

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(0)}) = \frac{1}{m}\sum_{j=1}^{m} log\ p(y^j|\mathcal{J}_\mathcal{N},\boldsymbol{\theta})$$

където $\{y^j\}_{j=1}^m \sim p\big(\widetilde{\mathcal{J}_\mathcal{N}}|\mathcal{J}_\mathcal{N},\boldsymbol{\theta}\big)$ и m е броят на извадките

[Step 3 - M стъпка:] Максимизиране $\boldsymbol{\theta}^{(i)} = arg\max_{\boldsymbol{\theta}} \widehat{Q}(\boldsymbol{\theta})$



Фигура 0.3: Диагностиката на Gelman-Rubin за двумерното Поасоново разпределение и за отрицателното полиномно индивидуално разпределение на двутипов ВР

ОСНОВНИ ПРИНОСИ В ДИСЕРТАЦИЯТА

В настоящата дисертация могат да се отбележат следните **приноси:**

1. *Бейсово оценяване на многотипови разклоняващи се процеси.*
   Направен е Бейсов параметричен анализ на неизвестните параметри за многотипови процеси с разпределение на потомството многомерен степенен ред. Получените *оценки и изчисления са приложени при методите Монте Карло* и по-конкретно Gibbs sampler и Монте Карло ЕМ.

2. *Робастното оценяване на двутипови разклоняващи се процеси. Доказана е Теорема* за праговата точка на претеглената и орязана правдоподобна оценка - WLTE(k), отнасяща се за двутипови процеси с разпределение двумерен степенен ред. Полученият резултат е публикуван.

3. *Приложен е ЕМ алгоритъм за двутипови разклоняващи се процеси* в случаите на полиномно, отрицателно полиномно и двумерно Поасоново индивидуално разпределение.

4. *Направени са изследвания за численото намиране на носителя.* Описан е итеративен алгоритъм, който използвайки модата на разпределенията, изчислява носителя на полиномното, отрицателното полиномно и двумерното Поасоново разпределение. Разгледани са примери.

5. *Разгледана е ускорена модификация на ЕМ алгоритъма*, наречена *ускорен ЕМ алгоритъм.* Алгоритъмът е приложен към двутипови процеси с полиномно индивидуално разпределение. Резултатите от изпълнението му са сравнени с изпълнението на класическия ЕМ алгоритъм.

6. *Направена е симулация на двутипови разклоняващи се процеси.* Към симулирани процеси с полиномно индивидуално разпределение са приложени алгоритмите ЕМ и Ускорен ЕМ. За симулирани двутипови разклоняващи се процеси с полиномно, отрицателно полиномно и Поасоново разпределение е приложен Gibbs sampler. Изследвана и сходимостта, като е използвана диагностика за сходимост на МСМС алгоритми.

7. *Дисертацията има научно-образователен принос.* Използван е голям обем литература, която е представена систематизирано. На едно място е описана теорията за статистическо оценяване на разклоняващи се процеси, приложена е тази теория и е използвана в симулации чрез примери.

8. *Създадено е улеснение за бъдещи изследвания*, като е показано, че към многотиповите разклоняващи се процеси могат да се приложат и по-сложни методи.

ОБЩИ ИЗВОДИ И ПРЕДЛОЖЕНИЯ

Настоящата дисертация показва използването на компютърни методи при статистическа оценка на многотиповите разклоняващи се процеси. Едно от предизвикателствата при анализирането на многотиповите разклоняващи се процеси е големият обем данни, които трябва да се обработят, както и получаването на твърде големи числови стойности. Освен това, статистическото оценяване на многотиповите разклоняващи се процеси чрез ЕМ алгоритъм е съпроводено с много обемни изчисления, изискващи много процесорно време. Бейсовият подход, приложен в настоящата работа, позволява използването на методите Монте Карло, в които за да се направи статистическа оценка на неизвестните параметри, се търси апостериорното им разпределение. Използването на МСМС алгоритмите води до редуциране на сложността на изчисленията.

Предмет на бъдещи изследвания могат да бъдат:

- робастно оценяване и намиране на праговата точка на друг клас многотипови разклоняващи се процеси;
- прилагане на други МСМС алгоритми, ЕМ модификации и проксимални алгоритми за статистическа оценка на МВР.

Благодарности

Благодаря на научния си ръководител доц. д-р Весела Стоименова за подкрепата и насоките, за това, че ме въведе в научната област и по този начин ми предостави възможността да работя върху тези проблеми. Благодаря на проф. д.м.н. д-р Николай Янев за вниманието и съветите. Участието ми в научни конференции бе подпомогнато от Европейския фонд за развитие на човешките ресурси[3]. Благодаря на катедра ВОИС към ФМИ на СУ, затова че ми предостави тази възможност. Благодаря на всички, които ме подкрепяха в това начинание.

Декларация за оригиналност

Аз декларирам, че получените резултати са оригинални.

Резултатите, които са получени, описани и/или публикувани от други учени, са надлежно и подробно цитирани в библиографията.

Настоящата дисертация не е прилагана за придобиване на научна степен в друго висше училище, университет или научен институт.

[3]European Social Fund through the Human Resource Development Operational Programme BG051PO001-3.3.06-0052 (2012/2014)

Библиография

[1] E. Altman and D. Fiems. *Branching Processes, the Max-Plus Algebra and Network Calculus.* In 19th International Conference, ASMTA, Grenoble, France, 2012.

[2] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. *An introduction to MCMC for machine learning.* Mach. Learn., 50(1-2):5–43, 2003.

[3] D. Atanasov. *About the concept of weights of wlte(k) estimators.* Pliska Studia Mathematica Bulgarica, Vol. 14, (2003), pages 5-13, 2003.

[4] D. Atanasov, A. Staneva and V. Stoimenova. *Robust estimators for the bivariate power series offspring distributions.* 3rdSMTDA Conference Proceedings, 11-14 June 2014, Lisbon Portugal, 2014.

[5] D. Atanasov, V. Stoimenova and N. Yanev. *Estimators in branching processes with immigration.* Pliska Studia Mathematica Bulgarica, Vol. 14 (2007), pages 19-40, 2007.

[6] D. Atanasov, V. Stoimenova and N. Yanev. *Offspring mean estimators in branching processes with immigration.* Pliska Studia Mathematica Bulgarica, Vol. 19 (2009), pages 69-82, 2009.

[7] D.V. Atanasov and N.M. Neykov. *On the Finite Sample Breakdown Point of the WLTE (k) and d-fullness of a Set of Continuous Functions.* Proceedings of the VI International Conference" Computer Data Analysis And Modeling", Minsk, Belarus, 2001.

[8] K.B. Athreya and P.E. Ney. *Branching processes.* Springer-Verlag Berlin, New York, 287 p, 1972.

[9] L.E. Baum, T. Petrie, G. Soules and N. Weiss. *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains.* Ann. Math. Statist., vol. 41, p.164–171, 1970.

[10] E.M.L. Beale and R.J.A. Little. *Missing values in multivariate analysis.* J. R. Stat. Soc., Ser. B, vol. 37, p.129–145, 1975.

[11] José M. Bernardo and Adrian F.M. Smith. *Bayesian theory.* John Wiley & Sons, Inc., 2008

[12] D.M. Blei, A. Kucukelbir and J.D. McAuliffe. *Variational Inference: A Review for Statisticians.* Journal of the American Statistical Association, vol. 112, no 518, p. 859–877, 2017

[13] S. P. Brooks and A. Gelman. *General methods for monitoring convergence of iterative simulations.* Journal of Computational and Graphical Statistics, vol. 7, no 4, pp.434–455, 1998.

[14] S. P. Brooks and G. O. Roberts. *Assessing convergence of Markov Chain Monte Carlo algorithms.* Statistics and Computing, vol. 8, pp. 319–335, 1997.

[15] B. P. Carlin and T. A. Louis. *Bayes and empirical Bayes methods for data analysis.* Chapman & Hall, no. XVI, pp.399, London 1996

[16] G. Casella and E. I. George. *Explaining the Gibbs Sampler.* The American Statistician, vol. 46, no. 3, pp. 167–174, 1992.

[17] C. Chatfield. *On estimating the parameters of the logarithmic series and negative binomial distributions.* Biometrika, 56(2):411–414, 1969.

[18] C. Chatfield, A. S. C. Ehrenberg and G. J. Goodhardt. *Progress on a simplified model of stationary purchasing behaviour.* Journal of the Royal Statistical Society. Series A (General), vol. 129, no 3, pp 317–367, 1966.

[19] S. Chrétien and A.O. Hero. *On EM algorithms and their proximal generalizations.* ESAIM, Probab. Stat., vol 12, pp.308–326, 2008.

[20] S. Chrétien and A. O. Hero. *Kullback proximal algorithms for maximum likelihood estimation.* https://hal.inria.fr/inria-00072906, Research Report, no RR-3756, INRIA, 1999

[21] D.R. Cox and D.V. Hinkley *Theoretical Statistics.* newblock Chapman and Hall London , isbn 0412124203 , pp xii, 511 p., 1974

[22] M. K. Cowles and B. P. Carlin. *Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review.* Journal of the American Statistical Association, 91:883–904, 1996.

[23] K. S. Crump and C.J. Mode. *A general age-dependent branching process. I.II.* J. Math. Anal. Appl., 24:494–508, 1968.

[24] A. Davis, R Gao and N. Navin. *Tumor evolution: Linear, branching, neutral or punctuated?* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, vol.1867, (2), (151 - 161), 2017.

[25] A. P. Dempster, N.M. Laird and D.B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm. Discussion.* J. R. Stat. Soc., Ser. B, 39:1–38, 1977.

[26] A. P. Dempster, N.M. Laird and D.B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm. Discussion.* J. R. Stat. Soc., Ser. B, 39:1–38, 1977.

[27] J. M. Drake, R. B. Kaul, L.W. Alexander, S.M. O'Regan, A.M. Kramer, J. T. Pulliam, MJ Ferrari and AW Park. *Ebola Cases and Health System Demand in Liberia*, Riley S, ed. PLoS Biology. 13(1), 2015.

[28] A. S. C. Ehrenberg. *The practical meaning and usefulness of the nbd/lsd theory of repeat-buying.* Applied Statistics, 17:7–10, 1968.

[29] D. Fiems and E. Altman. *Applying Branching Processes to Delay-Tolerant Networks. Bioinspired Models of Network, Information, and Computing Systems.* Springer Berlin Heidelberg , pp 117–125, 2010.

[30] R. A. Fisher, A. S. Corbet and C. B. Williams. *The relation between the number of species and the number of individuals in a random sample of an animal population.* Journal of Animal Ecology, 12(1):42–58, 1943.

[31] F.Le Gall. *The modes of a negative multinomial distribution.* Stat. Probab. Lett., 76(6):619–624, 2006.

[32] F. Le Gall. *Determination of the modes of a multinomial distribution.* Stat. Probab. Lett., 62(4):325–333, 2003.

[33] A. E. Gelfand and A. F.M. Smith. *Sampling-based approaches to calculating marginal densities.* J. Am. Stat. Assoc., 85(410):398–409, 1990.

[34] S. Geman and D. Geman. *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images.* IEEE Trans. Pattern Anal. Mach. Intell., 6:721–741, 1984.

[35] A. Gelman and D. B. Rubin. *Inference from Iterative Simulation Using Multiple Sequences.* Statistical Science, 7(4):457-472, 1992.

[36] T. Gerstenkorn. *On multivariate power series distributions.* Rev. Roum. Math. Pures Appl., 26:247–266, 1981.

[37] C. J. Geyer. *Introduction to Markov chain Monte Carlo.* In Handbook of Markov chain Monte Carlo, pp 3–48. Boca Raton, FL: CRC Press , 2011.

[38] M. González, C. Gutiérrez, R. Martínez and I. M. del Puerto. *Bayesian inference for controlled branching processes through MCMC and ABC methodologies.* Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas, vol. 107, 2013.

[39] M. González, J. Martín, R. Martínez and M. Mota. *Non-parametric Bayesian estimation for multitype branching processes through simulation-based methods.* Comput. Stat. Data Anal., 52(3):1281–1291, 2008.

[40] M. Gonzalez, C. Minuesa and I. del Puerto. *Maximum likelihood estimation and Expectation-Maximization algorithm for controlled branching processes.* Comput. Stat. Data Anal., vol. 93, no C, pp 209–227, January 2016.

[41] M. González and I. M. del Puerto. *Statistical inference for controlled multitype branching processes.* University of Extremadura, 06006-Badajoz, Spain, 2009

[42] P. J. Green. *Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.* Biometrika, 82(4):711–732, 1995.

[43] M. R. Gupta and Y. Chen. *Theory and use of the EM algorithm.* Found. Trends Signal Process., 4(3):223–296, 2010.

[44] P. Guttorp. *Statistical inference for branching processes.* New York etc.: John Wiley & Sons, 1991.

[45] P. Haccou, P. Jagers and V. A. Vatutin. *Branching processes. Variation, growth, and extinction of populations.* Cambridge: Cambridge University Press , 2005.

[46] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel. *Robust statistics. The approach based on influence functions.* Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley & Sons. XXI, 502 p., 1986.

[47] T.E. Harris. *The theory of branching processes.* Die Grundlehren der mathematischen Wissenschaften. 119. Berlin- Göttingen-Heidelberg: Springer-Verlag. XIV, 230 p, 1963.

[48] H. O. Hartley. *Maximum likelihood estimation from incomplete data.* Biometrics, 14:174–194, 1958.

[49] M. Healy and M. Westmacott. *Missing values in experiments analysed on automatic computers.* j-APPL-STAT, 5(3):203–206 , November 1956.

[50] P. J. Huber and E.. M. Ronchetti. *Robust statistics.* Hoboken, NJ: John Wiley & Sons, 2nd revised ed. edition, 2009.

[51] M. Hubert and P. J. Rousseeuw and V. A. Stefan. *High-Breakdown Robust Multivariate Methods.* Statist. Sci., 23(1):92–119, 02 2008.

[52] S. Ikeda. *Acceleration of the EM algorithms.* Systems and Computers in Japan, vol. 31(2), pp 10–18, 2000.

[53] C. Jacob, *Branching processes: their role in epidemiology.* Int. J. Environ. Res. Public Health 7, 1186-1204, 2010.

[54] N. L. Johnson, S. Kotz and N. Balakrishnan. *Discrete multivariate distributions.* New York, NY: Wiley, 1997.

[55] C.G. Khatri. *On certain properties of power-series distributions.* Biometrika, 46:486–490, 1959.

[56] C.G. Khatri. *On certain problems in Multivariate analysis.* Phd. thesis, 1960.

[57] C.G. Khatri. *On multivariate contagious distributions.* Sankhyia, Series B, 33, 197-216, 1971.

[58] S. Kocherlakota and K. Kocherlakota. *Bivariate Discrete Distributions.* John Wiley & Sons, Inc., 2004.

[59] R. A. Levine and G. Casella. *Implementations of the Monte Carlo EM Algorithm.* Journal of Computational and Graphical Statistics, 10(3):422–439, 2001.

[60] K. Li and Z. Geng. *Convergence rate of Gibbs sampler and its application.* Sci. China, Ser. A, 48(10):1430–1439, 2005.

[61] F. Liang, C. Liu and R. J. Carroll. *Advanced Markov chain Monte Carlo methods. Learning from past samples.* Chichester: John Wiley & Sons, 2010.

[62] C. A. Macken and A. S. Perelson. *Stem cell proliferation and differentiation: a multitype branching process model / Catherine A. Macken, Alan S. Perelson.* Springer-Verlag Berlin; New York, 1988.

[63] E. I. Marcelo. *Evolutionary dynamics of populations with genotype-phenotype map.* phd - thesis, Universitat Polite'cnica de Catalunya - Barcelona Tech, 2014.

[64] G. J. McLachlan and T. Krishnan. *The EM algorithm and extensions.* Wiley series in probability and statistics. John Wiley & Sons, New York, 1997.

[65] X. L. Meng and D. van Dyk. *The EM algorithm - an old folk-song sung to a fast new tune.* J. R. Stat. Soc., Ser. B, 59(3):511–567, 1997.

[66] M. Molina, M. González and M. Mota. *Bayesian inference for bisexual galton-watson processes.* Communications in Statistics - Theory and Methods, 27(5):1055–1070, 1998.

[67] T. Man-Lai, Ng. K. Wang and T. Guo-Liang. *Inverted Dirichlet Distribution.* John Wiley & Sons, Ltd, 2011.

[68] T. Nguyen, A.K. Gupta, D.M. Nguyen and Y. Wang. *Characterizations of negative multinomial distributions based on conditional distributions.* Metrika, 66(3):315–322, 2007.

[69] J. Neyman and G. Bates. *Contributions to the theory of accident proneness.* Cambridge University Press London, England, vol 1, 1952.

[70] T. Orchard and M. A. Woodbury. *A missing information principle: Theory and applications.* Proceedings of the 6 $^{\text{th}}$ Berkeley Symposium on Mathematical Statistics and Probability, vol 1, pages 697–715. University of California Press, 1972.

[71] G. P. Patil and S. Bildikar. *Multivariate logarithmic series distribution as a probability model in population and community ecology and some of its statistical properties.* Journal of the American Statistical Association, 62:655–674, 1967.

[72] G.P. Patil. *On multivariate generalized power series distribution and its application to the multinomial and negative multinomial.* Sankhyā, Ser. A, 28:225–238, 1966.

[73] C. Robert and G. Casella. *A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data.* Stat. Sci., 26(1):102–115, 2011.

[74] J.A. Roderick and D. B. Rubin. *Statistical analysis with missing data.* Chichester: Wiley, 2nd ed., 2002.

[75] D. B. Rubin. *Characterizing the estimation of parameters in incomplete-data problems.* Journal of the American Statistical Association, 69(346):467–474, 1974.

[76] M. J. Rufo, C. J. Pérez and J. Martín. *A bayesian approach to aggregate experts' initial information.* Electron. J. Statist., 6:2362–2382, 2012.

[77] B. A. Sevastyanov. *The theory of branching random processes.* Usp. Mat. Nauk, 6(6(46)):47–99, 1951.

[78] B. A. Sevastyanov and A.M. Zubkov. *Controlled branching processes.* Theory Probab. Appl., 19:15–24, 1974.

[79] J. Shao. *Mathematical statistics.* New York, NY: Springer, 2nd ed., 2003.

[80] S. Singh, D. J. Schneider and C. R. Myers. *Using multitype branching processes to quantify statistics of disease outbreaks in zoonotic epidemics.* Phys. Rev. E, 89:032702, Mar 2014.

[81] A. Staneva and V. Stoimenova. *Statistical estimation in branching processes with bivariate poisson offspring distribution. Pliska Studia Mathematica*, 24:73–88, 2015.

[82] V. Stoimenova. *Robust parametric estimation of branching processes with a random number of ancestors.* Serdica Mathematical Journal, 31(3):243–262, 2005.

[83] V. Stoimenova, D. Atanasov and N. Yanev. *Robust estimation and simulation of branching processes.* C. R. Acad. Bulg. Sci., 57(5):19–22, 2004.

[84] V. Stoimenova and N. Yanev *Parametric Estimation in Branching Processes with an Increasing Random Number of Ancestors.* Pliska Stud. Math. Bulgar. 17, 295-312, 2005.

[85] R. Sundberg. *Maximum likelihood theory for incomplete data from an exponential family.* Scandinavian Journal of Statistics, 1(2):49–58, 1974.

[86] Y. J. Sung and C. J. Geyer. *Monte Carlo likelihood inference for missing data models.* Ann. Stat., 35(3):990–1011, 2007.

[87] L.E. Sánchez, D.K. Nagar and A.K. Gupta. *Properties of noncentral dirichlet distributions.* Computers & Mathematics with Applications, 52(12):1671 – 1682, 2006.

[88] M. A. Tanner. *Tools for statistical inference. Observed data and data augmentation methods.* New York etc.: Springer-Verlag, 1991.

[89] M. A. Tanner and W. H. Wong. *The calculation of posterior distributions by data augmentation.* J. Am. Stat. Assoc., 82:528–541, 1987.

[90] L. Tierney. *Markov chains for exploring posterior distributions (With discussion).* Ann. Stat., 22(4):1701–1762, 1994.

[91] P. Tseng. *An analysis of the EM algorithm and entropy-like proximal point methods.* Math. Oper. Res., 29(1):27–44, 2004.

[92] D. L. Vandev and N. M. Neykov. *About regression estimators with high breakdown point.* Statistics, 32(2):111–129, 1998.

[93] D. Vandev and N. Neykov. *Robustified Maximum Likelihood.*, $https : /store.fmi.uni - sofia.bg/fmi/statist/personal/vandev/papers/talk_c.pdf$, 2001.

[94] D. L. Vandev and N.M. Neykov. *Robust maximum likelihood in the Gaussian case.* Proceedings of a workshop, held at the Centro Stefano Franscini in Ascona, Switzerland, June 28-July 4, 1992, pp 259–264. Basel: Birkhäuser, 1993.

[95] D. L. Vandev. *A Note on Breakdown Point of the Least Median of Squares and Least Trimmed Estimators.* Statistics & Probability Letters, vol 16, no 2, pp 117 - 119, 1993.

[96] D. L. Vandev and N. M.Neykov. *About regression estimators with high breakdown point.* Statistics, 32(2):111–129, 1998.

[97] V. A. Vatutin. *Vetvyashchiesya protsessy Bellmana-Kharrisa.* Matematicheskiĭ Institut im. V. A. Steklova, RAN, 2009.

[98] M. Watanabe and K. Yamaguchi, editors. *The EM algorithm and related statistical models.* New York, NY: Marcel Dekker, 2004.

[99] G C G Wei and M A Tanner. *A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms.* Journal of the American Statistical Association, 85:699–704, 1990.

[100] Wilson J.R. *Logarithmic series distribution and its use in analyzing discrete data.* Arizona State University, pp 275–279 , 1988

[101] C.J.J. Wu. *On the convergence properties of the EM algorithm. Ann. Stat.*, 11:95–103, 1983.

[102] A. M. Yaglom. *Certain limit theorems of the theory of branching processes.* УМН, 5:5–41, 1938.

[103] A.Y. Yakovlev, V.K. Stoimenova and N.M. Yanev, *Branching Processes as Models of Progenitor Cell Populations and Estimation of the Offspring Distributions.* J. American Statistical Assoc, 103, 1357-1366, 1998.

[104] A. Y. Yakovlev and N. M. Yanev, *Transient Processes in Cell Proliferation Kinetics.* Springer Verlag, Berlin, 1989.

[105] A. Y. Yakovlev and N. M. Yanev. *Relative frequencies in multitype branching processes.* Ann. Appl. Probab., 19(1):1–14, 02, 2009.

[106] N. M. Yanev. *Statistical inference for branching processes.* Records and Branching Processes, NOVA, Science Publishers Inc., New York, Ch.7:143–168, 2008.

[107] М. С. Божкова и Н. Янев. *Разклоняващи се стохастични процеси.* УИ "Св. Климент Охридски 2008.

[108] Н. Р. Даскалова. *Алгоритми от тип ЕМ за статистическо оценяване в разклоняващи се стохастични процеси.* PhD thesis, БАН, 2012.

[109] А. Н. Колмогоров. *К решению одной биологической задачи. Изв. НИИ математики и механики, Том. ун-та*, Т. 2, вып. 1.:7, 1938.

[110] А. Н. Колмогоров. *Об аналитических методах в теории вероятностей. Dokl. Acad. Nauk.SSSR*, Т.56.:795–798, 1947.

[111] А. Н. Колмогоров и Н.А. Дмитриев. *Ветвящиеся случайные процессы. ДАН СССР.*, Т.56. - Вып.1.:17–32, 1947.

[112] А. Н. Колмогоров и Б. А. Севастьянов. *Вычисление финальных вероятностей для ветвящихся случайных процессов. Докл. АН СССР*, 56,8.:783–786, 1947.

[113] В. К. Стоименова. *Статистическо оценяване на разклоняващи се стохастични процеси.* PhD thesis, БАН, 2005.

[114] А. Н. Ширяев. *Колмогоров. Юбилейное издание в 3-х кн.*, volume 22. ФИЗМАТЛИТ, 2003.