УНИВЕРЗИТЕТ У БЕОГРАДУ

МАТЕМАТИЧКИ ФАКУЛТЕТ

Марко Х. Обрадовић

КАРАКТЕРИЗАЦИЈЕ НЕКИХ РАСПОДЕЛА И БАХАДУРОВА АСИМПТОТСКА ЕФИКАСНОСТ ТЕСТОВА САГЛАСНОСТИ

— докторска дисертација —

Београд, 2015.

UNIVERSITY OF BELGRADE
FACULTY OF MATHEMATICS

Marko H. Obradović

CHARACTERIZATIONS OF SOME DISTRIBUTIONS AND BAHADUR ASYMPTOTIC EFFICIENCY OF GOODNESS-OF-FIT TESTS

— Doctoral Dissertation —

Belgrade, 2015.

Ментор:

др Слободанка Јанковић, редовни професор
Математичког факултета Универзитета у Београду

Чланови комисије:

др Весна Јевремовић, ванредни професор
Математичког факултета Универзитета у Београду

др Јаков Јурјевич Никитин, редовни професор
Факултета за математику и механику
Државног универзитета у Санкт Петербургу

Датум одбране:

Рад на овој дисертацији као и ономе што јој је претходило било ми је драгоцено и пријатно искуство. Многи су ми пружали, како помоћ у раду, тако и моралну подршку, па користим ову прилику да се захвалим.

Захваљујем се првенствено професору Јакову Никитину, из чијег сам најпре научног рада, а затим и из преписке с њим научио много о овој области из које је као финални производ настала ова дисертација. Поред стручне помоћи, захваљујем му се и на љубазности и предусретљивости.

Захваљујем се такође и свом ментору, професорки Слободанки Јанковић, на несебичној подршци и значајном доприносу реализацији ове дисертације.

Велико хвала и професорки Весни Јевремовић на свој стручној и пријатељској помоћи свих ових година.

Нарочиту захвалност дугујем својим сарадницима, Милану и Бојани, резултати ове дисертације су у великој мери плод сарадње с њима. Посебно Бојани хвала на свему са жељом да настави истим путем.

Хвала и колегама с факултета на подршци и саветима, као и мојим студентима, бившим и садашњим, који су свој оптимизам преносили на мене.

На крају велику захвалност дугујем својој породици, Миши, тати и нарочито мами, којима посвећујем ову дисертацију.

Београд, фебруар 2015.

Марко Обрадовић

Наслов докторске дисертације: *Карактеризације неких расподела и Бахадурова асимптотска ефикасност тестова сагласности*

Резиме: Прве карактеризације расподела датирају од тридесетих година прошлог века. Ова област, која је на граници теорије вероватноће и математичке статистике, привлачи интересовање великог броја истраживача, а у последње време све је већи број радова на ову тему.

Тестови сагласности с расподелом су међу најважнијим непараметарским тестовима. Велики број њих заснован је на емпиријској функцији расподеле. Примена карактеризационих теорема за конструкцију тестова сагласности појавила се средином двадесетог века а последњих година постала је један од најпопуларнијих праваца у тој области. Предност оваквих тестова је што су често слободни од параметара расподеле па омогућавају тестирање сложених нултих хипотеза.

Циљеви ове дисертације су формулација нових карактеризација експоненцијалне и Паретове расподеле, као и примена теорије U-статистика, великих одступања и Бахадурове ефикасности за формирање и проучавање асимптотике тестова сагласности с наведеним расподелама. Дисертација се састоји из шест поглавља.

У првом поглављу дат је приказ различитих типова карактеризација с наглашавањем њихове многобројности и разноликости. Посебно су истакнуте карактеризације помоћу једнако расподељених функција узорка. У њему су такође приказане две нове карактеризације Паретове расподеле.

Друго поглавље посвећено је новим карактеризацијама ескпоненцијалне расподеле изложених у радовима [65] и [53]. Дато је шест карактеризација које су засноване на статистикама поретка. Специјалан случај једне од њих (теорема 2.4.3) представља решење отвореног проблема који су поставили Арнолд и Виљасењор [9].

У трећем поглављу дати су основни појмови о U-статистикама, класи статистика која је значајна у теорији непристрасних оцена параметара. Изложена су нека њихова асимптотска својства. Поред тога, дефинисане су и U-емпиријске функције расподеле, уопштење стандардних емпириј-

ских функција расподеле. Четврто поглавље посвећено је асимптотској ефикасности тестова, у првом реду Бахадуровој асимптотској ефикасност, тј. асимптотској ефикасности теста када праг значајности тежи нули. Приказани су неки теоријски резултати објављени у монографији Никитина [57], те радовима [61], [59], итд.

У петом поглављу приказани су нови резултати на пољу тестова сагласности с Паретовом расподелом. На основу три карактеризационе теореме наведене у одељку 1.1.2 направљено је шест тестова сагласности, три интегралног типа и три Колмогоровљевог типа. У свим случајевима тестира се сложена нулта хипотеза јер је расподела тест статистике слободна од параметра Паретове расподеле. За сваки од тестова испитана је асимптотска расподела под нултом хипотезом, као и асимптотско понашање репа расподеле (велика одступања) под блиским алтернативама. За неке стандардне блиске алтернативе одређена је локална Бахадурова асимптотска ефикасност и одређени су домени локалне асимптотске оптималности у свим случајевима. Резултати из овог поглавља објављени су у [66] и [64].

Шесто поглавље посвећено је новим тестовима сагласности с експоненцијалном расподелом. На основу решене хипотезе Арнолда и Виљасењора направљене су две класе тестова, интегралних и Колмогоровљевих, у зависности од броја сабирака у карактеризацији. Проучавање асимптотских својстава аналогних оним у петом поглављу урађено је у случајевима два и три сабирка, за које тестови имају практични значај. Резултати овог поглавља изложени су у [39].

Кључне речи: карактеризације расподела, експоненцијална расподела, Паретова расподела, тестови сагласности, Бахадурова ефикасност, теорија великих одступања, U-статистике

Научна област: Математика

Ужа научна област: Вероватноћа и статистика

УДК: 519.224+519.234.3(0433)

AMS Classification: 60F10, 62E10, 62G10, 62G20, 62G30, 62H05

Doctoral Dissertation Title: *Characterizations of some Distributions and Bahadur Asymptotic Efficiency of Goodness-of-Fit Tests*

Abstract: First characterizations of probability distributions date to the thirties of last century. This area, which lies on the borderline of probability theory and mathematical statistics, attracts large number of researchers, and in recent times the number of papers on the subject is increasing.

Goodness-of-fit tests are among the most important nonparametric tests. Many of them are based on empirical distribution function. The application of characterization theorems for construction of goodness-of-fit tests dates to the middle of last century, and recently has become one of the main directions in this field. The advantage of such tests is that they are often free of distribution parametres and hence enable testing of composite hypotheses.

The goals of this dissertation are the formulation of new characterizations of exponential and Pareto distribution, as well as the application of the theory of U-statistics, large deviations and Bahadur efficiency to construction and examination of asymptotics of goodness-of-fit tests for aforementioned distributions. The dissertation consists of six chapters.

In the first chapter a review of different types of characterizations is presented, pointing out their abundance and variety. The special emphasis is given to the characterizations based on equidistribution of functions of the sample. Besides, two new characterizations of Pareto distribution are presented.

The second chapter is devoted to some new characterizations of the exponential distributions presented in papers [65] and [53]. Six characterizations based on order statistics are presented. A special case of one of them (theorem 2.4.3) represents the solution of open problem stated by Arnold and Villasenor [9].

In the third chapter there are basic concepts on U-statistics, the class of statistics important in the theory of unbiased estimation. Some of their asymptotic properties are given. U-empirical distribution functions, a generalization of standard empirical distribution functions, are also defined. The fourth chapter is dedicated to the asymptotic efficiency of statistical tests,

primarily to Bahadur asymptotic efficiency, i.e. asymptotic efficiency of the test when the level of significance approaches zero. Some theoretical results from the monograph by Nikitin [57], and papers [61], [59], etc. are shown.

In the fifth chapter new results in the field of goodness-of-fit tests for Pareto distribution are presented. Based on three characterizations of Pareto distribution given in section 1.1.2. six goodness-of-fit tests, three of integral, and three of Kolmogorov type, are proposed. In each case the composite null hypothesis is tested since the test statistics are free of the parameter of Pareto distribution. For each test the asymptotic distribution under null hypothesis, as well as asymptotic behaviour of the tail (large deviations) under close alternatives is derived. For some standard alternatives, the local Bahadur asymptotic efficiency is calculated and the domains of local asymptotic optimality are obtained. The results from this chapter are published in [66] and [64].

The sixth chapter brings new goodness-of-fit tests for exponential distribution. Based on the solved hypothesis of Arnold and Villasenor two classes of tests, integral and Kolmogorov type, are proposed, depending on the number of summands in the characterization. The study of asymptotic properties, analogous to the ones in the fifth chapter is done in case of two and three summands, for which the tests have practical importance. The results of this chapter are presented in [39].

Keywords: characterizations of probability distributions, exponential distribution, Pareto distribution, goodness-of-fit tests, Bahadur efficiency, large deviation theory, U-statistics

Scietific Area: Mathematics

Scientific Sub-area: Probability and Statistics

UDC: 519.224+519.234.3(0433)

AMS Classification: 60F10, 62E10, 62G10, 62G20, 62G30, 62H05

Садржај

Поглавље 1

Карактеризације расподела

Проблем карактеризације расподела припада области математике која је смештена на граници теорије вероватноће и математичке статистике. Карактеризациони проблем може се дефинисати на следећи начин [40]: нека за случајни вектор X и функцију $\mathcal{L}$ постоји фамилија расподела $\mathcal{F}$ таква да $\mathcal{L}(X) \in \mathcal{F}$ повлачи да X има неку особину $\mathcal{P}$. Карактеризациони проблем је тада "други смер" овог тврђења, наиме, да ако X има особину $\mathcal{P}$, тада мора да $\mathcal{L}(X) \in \mathcal{F}$.

На пример, за случајни вектор (X_1, X_2) чије су компоненте независне једнако расподељене нормалне случајне величине с нултом средњом вредношћу, важи да $\mathcal{L}(X_1, X_2) = \frac{1}{\sqrt{2}}(X_1 + X_2)$ има идентичну расподелу као и компоненте случајног вектора. Важи и обрнуто, тј. да уколико имамо вектор (X_1, X_2) чије су компоненте независне једнако расподељене случајне величине с нултим математичким очекивањем тако да важи

$$\mathcal{L}(X_1, X_2) \stackrel{d}{=} X_1,$$

тада X_1 има нормалну расподелу. Знак $\stackrel{d}{=}$ у горњој формули а и у даљем тексту означава "једнакост у расподели". Наведено тврђење познато је као Пољина карактеризација нормалне расподеле (в. [68]). Ова карактеризациона теорема од велике је историјске важности и сматра се пионирским радом у овој области.

Карактеризацијама расподела посвећене су неке монографије (нпр. [40], [22]) и поглавља у монографијама (нпр. [8], [4], [14]).

Како се теореме карактеризације с пробабилистичког лако преводе на језик функционалне анализе, то су за њихово извођење често потребни алати математичке анализе, комплексних функција, различитих типова диференцијалних и интегралних једначина, редова и теорије функционалних једначина.

Својство $\mathcal{P}$ које карактерише расподелу може бити различите врсте и према тим врстама можемо направити поделу карактеризационих теорема. У наредним одељцима даћемо примере за неке врсте карактеризација. Наравно, изузетно је тешко направити темељан преглед резултата у овој преширокој области, па је избор направљен тако да се обухвати што више различитих врста карактеризација. Како је главни акценат ове дисертације на карактеризацијама на основу једнако расподељених статистика, њима ће бити посвећено највише пажње. Наравно, ова подела није једина могућа, а постоје и карактеризације које припадају у више група.

У тексту су наведени само докази оних теорема које је аутор формулисао. Под термином узорак подразумева се прост случајан узорак. У исказима неких теорема говори се о узорку, а у другим о независним и једнако расподељеним случајним величинама. Ово су, наравно, еквивалентни термини, а у конкретним случајевима биран је један или други због погодности или у складу с извором из ког су преузети.

1.1 Карактеризације засноване на једнако расподељеним функцијама

1.1.1 Карактеризације експоненцијалне расподеле

Не само из разлога што је експоненцијалној расподели посвећен вероватно највећи број карактеризација, већ и због тога што се и овај рад бави том проблематиком, највећи простор дајемо управо њој.

Под експоненцијалном расподелом подразумевамо функцију расподеле

$$F(x) = 1 - e^{-\lambda x}, x \geq 0, \ \ \lambda > 0. \tag{1.1}$$

Значајан део карактеризација заснива се на неким односима међу статистикама поретка. Ово проистиче из једноставног разлога што размаци међу суседним статистикама поретка узорка из експоненцијалне расподеле имају такође неку експоненцијалну расподелу. То својство је својеврсни рудник идеја за нове карактеризације.

Користићемо ознаку $X_{(k;n)}$ за статистику поретка реда k у узорку обима n. Опште је познато својство експоненцијалне расподеле да $nX_{(1;n)}$ има експоненцијалну расподелу с истим параметром као и X. Десу [21] је показао да ово својство, уколико је испуњено за свако n, карактерише експоненцијалну расподелу.

Пури и Рубин су у раду [69] доказали следећу теорему.

Теорема 1.1.1 (Пури и Рубин, 1970)**.** *Нека су X и Y две независне и једнако расподељене ненегативне апсолутно непрекидне случајне величине. Ако X и $|X - Y|$ имају исту расподелу, тада X има експоненцијалну расподелу.*

Занимљиво је да "слично" тврђење овом, у коме се тврди да статистика $|X - Y|$ уместо исте расподеле као X, има експоненцијалну расподелу, није карактеризација. Розберг [73] је показао постојање других расподела које задовољавају ово тврђење.

Следеће две теореме су карактеризације засноване на размацима између суседних статистика поретка. Докази се могу наћи у [71] и [2].

Теорема 1.1.2 (Розберг, 1972)**.** *Нека је $X_1, \ldots, X_n$ узорак из ненегативне расподеле F. Ако, за неко j, статистике $X_{(j+s;n)} - X_{(j;n)}$ и $X_{s;n-j}$ имају исту расподелу, тада је F експоненцијална расподела.*

Теорема 1.1.3 (Ахсанулах, 1978)**.** *Нека је $X_1, \ldots, X_n$ узорак из ненегативне апсолутно непрекидне расподеле чија је функција расподеле F строго монотона и хазардна функција $f(t)/(1 - F(t))$ је монотоно* *растућа или опадајућа за свако $t \geq 0$. Ако, за неко j и k, $1 \leq j \leq k < n$ важи*

$$(n-j)(X_{(j+1;n)} - X_{(j;n)}) \stackrel{d}{=} (n-k)(X_{(k+1;n)} - X_{(k;n)}),$$

тада је расподела F експоненцијална.

Својство експоненцијалности размака може се искористити и на други начин. Познато је представљање статистике поретка из експоненцијалне расподеле преко збира елемената узорка уз одговарајуће тежинске коефицијенте на следећи начин (в. нпр. [8], стр. 73)

$$X_{(k;n)} \stackrel{d}{=} \sum_{j=n-k+1}^{n} \frac{X_j}{j},\ 1 \leq k \leq n. \tag{1.2}$$

Ахсанулах и Рахман [5] доказали су да ако својство (1.2) важи за свако n онда оно карактерише експоненцијалну расподелу. Касније је Хуанг [33] показао да се услов може ослабити тако да важи за две узастопне вредности k, али и да се не може ослабити тако да важи само за једну вредност k.

Поред наведених теорема у којима учествују само функције узорка расподеле која се карактерише, постоје и нешто другачија тврђења у којима поред њих налазимо и друге случајне величине с познатом расподелом. У свом раду [44] Коц и Стутел дали су следећу карактеризацију.

Теорема 1.1.4. *Нека су X_1 и X_2 две независне и једнако расподељене ненегативне случајне величине и нека је U случајна величина независна од њих која има униформну $\mathcal{U}[0,1]$ расподелу. Ако важи*

$$X_1 \stackrel{d}{=} U(X_1 + X_2), \tag{1.3}$$

тада X_1 има експоненцијалну расподелу.

У истом раду дато је и уопштење у коме уместо две имамо произвољан број n случајних величина. Даља уопштења и модификације ове теореме дата су у радовима [89] и (за гама расподелу) [34].

Карактеризације чија је идеја добијена из својства (1.2) такође су многобројне и у овој другој варијанти. Наводимо теорему из рада [17].

Теорема 1.1.5. *Нека је $X_1, \ldots, X_n$ узорак из ненегативне расподеле F.*

а) Ако за природне бројеве $1 \leq i < j \leq n$ важи

$$X_{(j;n)} \stackrel{d}{=} X_{(i;n)} + T,$$

где је $T \stackrel{d}{=} -\ln W$, а W има бета $B(n-j+1, j-i)$ независна од $X_{(i;n)}$, тада је F експонецијална $\mathcal{E}(1)$ расподела.

б) Ако за природне бројеве $2 \leq r+1 \leq j \leq n$ важи

$$X_{(j;n)} \stackrel{d}{=} X_{(j-r;n-r)} + T,$$

где је $T \stackrel{d}{=} -\ln W$, а W има бета $B(n-r+1, r)$ независна од $X_{(j-r;n-r)}$, тада је F експонецијална $\mathcal{E}(1)$ расподела.

1.1.2 Карактеризације Паретове расподеле

Трансформацијом $Y = e^X$ случајне величине X која има експоненцијалну расподелу добија се Паретова расподела. Стога се велики број карактеризација експоненцијалне расподеле може модификовати тако да буду карактеризације Паретове расподеле. Из овог разлога у литератури ретко наилазимо на "оригиналне" карактеризације Паретове расподеле, већ се оне углавном наводе као последице постојећих теорема.

Овде наводимо три карактеризације Паретове расподеле које су последице теорема Пури-Рубина, Розберга и Ахсанулаха наведених у претходном одељку. Теорема Розберга наведена је у раду [73], док су преостале две теореме први пут наведене у радовима аутора ([66] и [64]) па их овде наводимо с доказима.

Фамилија Паретових расподела с параметром облика α дата је функцијом расподеле

$$F(x) = 1 - x^{-\alpha}, x \geq 1, \quad \alpha > 0. \tag{1.4}$$

Теорема 1.1.6. *Нека су X и Y једнако расподељене ненегативне апсолутно непрекидне случајне величине. Тада, X и $\max\{\frac{X}{Y}, \frac{Y}{X}\}$ имају исту расподелу ако и само ако случајна величина X има Паретову расподелу.*

Доказ. Нека је $\widetilde{X} = \ln X$ и $\widetilde{Y} = \ln Y$. Тада је

$$\max\left\{\frac{X}{Y}, \frac{Y}{X}\right\} = \max\left\{e^{\widetilde{X}-\widetilde{Y}}, e^{\widetilde{Y}-\widetilde{X}}\right\} = e^{|\widetilde{X}-\widetilde{Y}|}$$

и

$$\ln\left(\max\left\{\frac{X}{Y}, \frac{Y}{X}\right\}\right) = |\widetilde{X} - \widetilde{Y}|.$$

Како је логаритмовање монотона трансформација, тврђење да X и $\max\{\frac{X}{Y}, \frac{Y}{X}\}$ имају исту расподелу еквивалентно је тврђењу да су $\widetilde{X}$ и $|\widetilde{X} - \widetilde{Y}|$ једнако расподељене. На основу теореме 1.1.1 једине непрекидне расподеле које задовољавају овај услов су експоненцијалне расподеле, па $\widetilde{X}$ мора имати експоненцијалну расподелу с неким параметром α. Пошто је $X = e^{\widetilde{X}}$, следи да X има Паретову расподелу с истим параметром α.

$\square$

Теорема 1.1.7 (Розберг, 1972). *Нека је $X_1, \ldots, X_n$ узорак из ненегативне расподеле. Ако за неко j статистике $X_{(j+1;n)}/X_{(j;n)}$ и $X_{(1;n-j)}$ имају исту расподелу, тада X_1 има Паретову расподелу.*

Доказ се може наћи у [73].

Теорема 1.1.8. *Нека је $X_1, \ldots, X_n$ узорак из ненегативне апсолутно непрекидне расподеле са строго монотоном функцијом расподеле и монотоно растућом или опадајућом хазардном функцијом. Тада, $(X_{(j+1;n)}/X_{(j;n)})^{n-j}$ и $(X_{(k+1;n)}/X_{(k;n)})^{n-k}$, $1 \leq j \leq k < n$, имају исту расподелу ако и само ако X_1 има Паретову расподелу.*

Доказ. Нека је $Y_l = \ln X_l$, $l = 1, \ldots, n$. С обзиром да је логаритам монотона трансформација, тврђење теореме може се преформулисати да $(n-j)(Y_{(j+1;n)} - Y_{(j;n)})$ и $(n-k)(Y_{(k+1;n)} - Y_{(k;n)})$ имају исту расподелу. Ово је тврђење теореме 1.1.3 која карактерише експоненцијалну расподелу с неким параметром α. Дакле, Y_1 има експоненцијалну, па X_1 има Паретову расподелу с истим параметром α. $\square$

У раду [17] можемо наћи и једну карактеризациону теорему "друге врсте" сличну оној у теореми 1.1.5.

1.1.3 Карактеризације неких других расподела

Поред већ поменуте Пољине карактеризације у којој $\frac{1}{\sqrt{2}}(X+Y)$ и X имају исту расподелу, навешћемо и следећу каркатеризацију нормалне расподеле на основу такозваног Шеповог својства коју су доказали Галамбош и Симонели 2003. године у [23].

Теорема 1.1.9. *Нека су X и Y независне случајне величине с истом расподелом F таквом да је $0 < F(x) < 1$ за свако x и да је функција $F(x) - F(-x)$ у нули правилно промењлива с индексом 1. Тада ако важи*

$$\frac{2XY}{\sqrt{X^2+Y^2}} \stackrel{d}{=} X,$$

X има нормалну $\mathcal{N}(0, \sigma^2)$ расподелу за неко $\sigma^2 > 0$.

Занимљива је и следећа карактеризација Клебанова [41].

Теорема 1.1.10. *Нека је $\mathbf{X} = (X_1, \dots, X_n)$ узорак обима $n \geq 4$ и нека је $\mathbf{Y} = \Gamma\mathbf{X}$, где је Γ ортогонална матрица реда $n \times n$. Тада X_1 има нормалну расподелу ако и само ако су статистике $X_{(n;n)} - X_{(1;n)}$ и $Y_{(n;n)} - Y_{(1;n)}$ једнако расподељене.*

Касније је Су [87] показао уз неке додатне услове за функцију расподеле узорка да карактеризација важи и када је $n = 2$ или $n = 3$.

Као пример карактеризација Кошијеве расподеле наводимо неколико карактеризација из Арнолдовог рада [7].

Теорема 1.1.11. *Нека је X апсолутно непрекидна случајна величина таква да $2X/(1 - X^2)$ и X имају исту расподелу. Тада X има Кошијеву $\mathcal{C}(0, 1)$ расподелу.*

Теорема 1.1.12. *Нека је X апсолутно непрекидна случајна величина таква да $\frac{1}{2}(X + X^{-1})$ и X имају исту расподелу. Тада X има Кошијеву $\mathcal{C}(0, 1)$ расподелу.*

Степена расподела, слично као и Паретова, може се изразити преко експоненцијалне расподеле, трансформацијом $Y = e^{-X}$. Стога се све карактеризације експоненцијалне и Паретове расподеле могу превести на њу. Уопште у научној литератури карактеризације степене расподеле наводе се у облику последица, као и у случају Паретове расподеле. Као пример можемо навести да је у раду Пурија и Рубина [69] наведена карактеризација степене расподеле у којој X и $\min\{X/Y, Y/X\}$ имају исту расподелу.

Међутим, постоје и радови у којима је карактеризација изведена баш за степену расподелу. Као пример навешћемо теорему коју је 2014. навео Арслан [10]. Она је применљива у такозваном узорковању помоћу рангираних скупова а доказ јој се заснива на примени Шоке-Денијеве теореме из функционалне анализе.

Теорема 1.1.13. *Нека је $X_1, \dots, X_n$ узорак из непрекидне расподеле F, и нека су $X_1^i, \dots, X_i^i$, $k \leq i \leq n$, где је $1 \leq k \leq n-1$, независни скупови узорака величине i. Тада је F степена расподела, тј.*

$$F(x) = x^{\alpha}, \;\; x \in [0,1],$$

с неким параметром $\alpha > 0$ ако и само ако за неко k, $1 \leq k \leq n-1$, важи

$$X_{(k;n)} \stackrel{d}{=} X_{(k;k)} X_{(k+1;k+1)} \cdots X_{(n;n)}.$$

Карактеризације степене расподеле "друге врсте" с идејом сличном као у теореми 1.1.5 можемо наћи у [17] и [85].

1.2 Други типови карактеризација

Вероватно најпознатија и најчешће коришћена карактеризација експоненцијалне расподеле, особином одсуства меморије, дата је следећом теоремом.

Теорема 1.2.1. *Нека је $F(x)$ недегенерисана функција расподеле. Претпоставимо да $F(x)$ задовољава следећу функционалну једначину*

$$1 - F(x+y) = (1 - F(x))(1 - F(y)),$$

за свако $x \geq 0$ и $y \geq 0$ скоро сигурно. Тада је $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$, где је λ позитивна константа.

Наведена карактеризација спада у такозвану групу карактеризација преко засечених расподела. Поред ње навешћемо и следећу карактеризацију логистичке расподеле која се може наћи у [22].

Теорема 1.2.2. *Нека случајна величина X има апсолутно непрекидну расподелу $F(x)$ која је симетрична око нуле. Тада X има логистичку $\mathcal{L}og(\lambda)$ расподелу ако и само ако за свако $x, y > 0$ важи*

$$\frac{1 - F(x+y)}{(1 - F(x))(1 - F(y))} = \frac{F(x+y)}{F(x)F(y)}.$$

Особеност ове каракатеризације је та што је њен најпогоднији облик управо за логистичку расподелу. Може се одговарајућом трансформацијом модификовати да буде карактеризација експоненцијалне расподеле, али би у том случају била прилично компликована.

Следећи тип карактеризација које наводимо су карактеризације на основу расподеле неке статистике. Прва теорема овог типа коју ћемо навести изведена је из познате особине, која се у неким изворима наводи и као дефиниција χ^2_n расподеле, да је она збир n независних нормалних $\mathcal{N}(0,1)$ расподела. Теорема коју је навео Круглов [47] тврди да, ако збир квадрата бесконачно дељивих једнако расподељених случајних величина има χ^2_n расподелу, тада оне морају имати стандардну нормалну расподелу.

Следећа теорема ове врсте коју наводимо дата је у раду [3]

Теорема 1.2.3. *Нека је X позитивна случајна величина с непрекидном функцијом расподеле F. Ако статистика $\sum_{i=1}^{n} \frac{X_i}{X_{n;n}} - 1$ има Ирвин-Холову расподелу са параметром $n-1$ за неке две узастопне вредности*

n, тада X има униформну $\mathcal{U}[0, a]$ расподелу за неко $a > 0$.

Познато је својство да количник две независне гама расподеле с истим параметром скалирања има бета расподелу друге врсте, али то тврђење није карактеризација гама расподеле (в. [43]). Међутим, у истом раду наведена је следећа теорема у којој расподела аналогне дводимензионе статистике карактерише гама расподелу.

Теорема 1.2.4. *Нека су X_1, X_2 и X_3 независне позитивне случајне величине. Уколико дводимензиона статистика $(X_1/X_2, X_2/X_3)$ има дводимензиону бета расподелу друге врсте с густином*

$$f(y_1, y_2) = \frac{\Gamma(p_1 + p_2 + p_3)}{\Gamma(p_1)\Gamma(p_2)\Gamma(p_3)} \frac{y_1^{p_1-1} y_2^{p_2-1}}{(1 + y_1 + y_2)^{p_1+p_2+p_3}}, \quad y_1, y_2 > 0,$$

тада X_1, X_2 и X_3 имају гама расподеле с параметрима облика p_1, p_2 и p_3 и истим параметром скалирања.

Велики број карактеризација заснован је на независности две статистике. Најпознатија од њих је свакако да независност узорачке средине од узорачке дисперзије карактерише нормалну расподелу. Својство је наравно опште познато и налази се у свим уџбеницима статистике, а карактеризациону теорему објавио је Лукач [50].

Такође за нормалну расподелу имамо у чувену Бернштајнову теорему [16].

Теорема 1.2.5 (Бернштајн, 1941)**.** *Нека су X и Y независне и једнако расподељене случајне величине и нека су $X + Y$ и $X - Y$ независне случајне величине. Тада X и Y имају нормалну расподелу.*

Уопштење ове теореме за произвољне линеарне комбинације, теорема Скитович-Дармуа [77], један је од најзначајнијих и изненађујућих резултата на пољу карактеризација.

Следећа карактеризација гама расподеле [51] занимљива је и по томе што се за X и Y не поставља услов да су једнако расподељене.

Теорема 1.2.6. *Нека су X и Y две независне недегенерисане и позитивне случајне величине. Случајне величине $U = X + Y$ и $V = X/Y$*

независне су ако и само ако X и Y имају гама расподеле с истим параметром скалирања.

У раду [7] Арнолд је дао наредну карактеризацију Кошијеве расподеле.

Теорема 1.2.7. *Нека су X и Y независне и једнако расподељене апсолутно непрекидне случајне величине и нека су X и $(X+Y)/(1-XY)$ такође независне случајне величине. Тада X има Кошијеву $\mathcal{C}(0,1)$ расподелу.*

Три карактеризације унинформне расподеле у којима се комбинује расподела статистике и независност дате су у раду Денга [20].

Теорема 1.2.8. *Нека су U и V независне апсолутно непрекидне случајне величине с носачем $(0,1)$. Тада су следећа тврђења еквивалентна:*

(i) *U има униформну $\mathcal{U}(0,1)$ расподелу.*

(ii) *$W_1 = \min\{\frac{U}{V}, \frac{1-U}{1-V}\}$ има униформну $\mathcal{U}(0,1)$ расподелу и независна је од V.*

(iii) *$W_2 = |U-V|(\frac{\Delta}{V}+\frac{1-\Delta}{1-V})$ има униформну $\mathcal{U}(0,1)$ расподелу и независна је од V, где је $\Delta = \mathrm{I}(V>U)$.*

(iv) *$W_3 = U+V \mod 1$ има униформну $\mathcal{U}(0,1)$ расподелу и независна је од V.*

Наредна врста карактеризационих теорема које ћемо навести су карактеризације засноване на моментима. Поменућемо најпре резултат Хефдинга [32] у коме низ момената статистике поретка фиксираног реда r за свако $n \geq r$ једнозначно одређује расподелу. Специјално, добијамо да низ $EX_{(1;n)} = \frac{1}{n}$, $n \geq 1$ карактерише експоненцијалну $\mathcal{E}(1)$, док $EX_{(1;n)} = \frac{1}{n+1}$ карактерише униформну $\mathcal{U}[0,1]$ расподелу.

Карактеризација степене расподеле преко рекурентне везе момената дата је у раду Туа и Лина [78] следећом теоремом.

Теорема 1.2.9 (Ту и Лин, 1989)**.** *Нека су $X_1, \dots X_n$ независне једнако расподељене случајне величине чији моменти одговарајућих редова постоје. Тада важи*

$$\left(r\binom{n}{r}\right)^{-1} EX^2_{(r;n)} \quad - \quad 2\left((r+p)\binom{n+p}{r+p}\right)^{-1} EX_{(r+p;n+p)} \\ + \quad \left((r+2p)\binom{n+2p}{r+2p}\right)^{-1} = 0$$

ако и само ако X_1 има степену расподелу с параметром $1/p$.

Специјалан случај ове теореме за $n = 1$, $r = 1$ и $p = 1$ своди се на карактеризацију униформне $\mathcal{U}[0,1]$ расподеле својством $2EX^2 = EX_{(2;2)} = \frac{2}{3}$. Ова карактеризација (али за произвољну униформну расподелу) наведена је и у [67] следећом теоремом

Теорема 1.2.10 (Папатанасију, 1990)**.** *Нека су X_1 и X_2 независне и једнако расподељене случајне величине. Важи неједнакост*

$$\mathrm{Cov}(X_{1;2}, X_{2;2}) \leq \frac{1}{3} DX,$$

где се једнакост достиже ако и само ако X_1 има униформну $\mathcal{U}[a,b]$ расподелу.

Поред наведених, многобројне су и карактеризације преко условних момената које се налазе у пресеку карактеризација на основу момената и засечених расподела. Занимљив пример ове врсте је карактеризација генералисане Паретове расподеле из рада [74] која представља уопштење особине одсуства меморије.

Теорема 1.2.11. *Нека је X ненегативна случајна величина с функцијом расподеле F и коначним математичким очекињем μ. Нека је $\mu(x) = E(X - x|X > x)$. Тада ако важи*

$$1 - F\left(x + \frac{y\mu(x)}{\mu}\right) = (1 - F(x))(1 - F(y)), \quad \text{за свако } x, y \geq 0,$$

онда X има генералисану Паретову расподелу, тј. њена функција расподеле је једна од следеће три:

$$F_1(x) = 1 - e^{-\lambda x}, \ \ x \geq 0, \ \textit{за неко}\ \lambda > 0;$$
$$F_2(x) = 1 - \left(1 + \frac{x}{\lambda}\right)^{-\alpha}, \ \ x \geq 0, \ \textit{за неко}\ \lambda > 0 \ \textit{и}\ \alpha > 1;$$
$$F_3(x) = 1 - \left(1 - \frac{x}{\lambda}\right)^{\alpha}, \ \ x \in [0, \lambda], \ \textit{за неко}\ \lambda > 0 \ \textit{и}\ \alpha > 0.$$

Низ карактеризација различитих расподела на основу максималне ентропије дат је у (в. [48]). Међу расподелама чији је носач коначни интервал максимална ентропија карактерише униформну, а уз додатни услов да је математичко очекивање логаритма фиксирано, бета расподелу. Код расподела дефинисаних на позитивној полуоси ово својство уз фиксирано математичко очекивање карактерише експоненцијалну, док код расподела чији је носач цео скуп реалних бројева, уз фиксирано математичко очекивање и дисперзију, својство максималне ентропије карактерише нормалну расподелу.

Навешћемо још да постоје и карактеризације на основу различитих статистичких својстава (максимална веродостојност, регресија), а у новије време у литератури веома су заступљене карактеризације на основу рекорда.

Поглавље 2

Нове карактеризације експонецијалне расподеле

2.1 Увод

У овом поглављу биће приказане нове карактеризације експоненцијалне расподеле које већином представљају резултате аутора из ове области Све теореме базирају се на статистикама поретка, односно на разлагању преко израза (1.2).

Године 2013. Арнолд и Виљасењор у свом раду [9] навели су неколико карактеризационих теорема експоненцијалне расподеле који се заснивају на узорцима обима два. Већина њих захтева да густина расподеле која се карактерише буде аналитичка. Стога ћемо пре него што кренемо у даље разматрање најпре дефинисати следећу класу.

Дефиниција 2.1.1. *Нека је $\mathcal{F}$ класа апсолутно непрекидних функција расподеле F таквих да важи $F(0) = 0$ и чија се густина расподеле f може развити у Маклоренов ред за свако $x > 0$.*

Прва међу наведеним теоремама је специјални случај карактеризације преко својства (1.2) за $k = n = 2$.

Теорема 2.1.1 (Арнолд и Виљасењор, 2013)**.** *Ако су X_1 и X_2 независне* *и једнако расподељене случајне величине с расподелом $F \in \mathcal{F}$, и ако важи*

$$X_1 + \frac{1}{2}X_2 \stackrel{d}{=} \max\{X_1, X_2\},$$

тада X_1 има експоненцијалну расподелу с неким $\lambda > 0$.

Ова карактеризација захтева аналитичност функције густине f, што се огледа у доказу теореме који користи следећу лему

Лема 2.1.1. *Нека је F функција расподеле из класе $\mathcal{F}$. Ако за свако природно k важи:*

$$f^{(k)}(0) = \left[\frac{f'(0)}{f(0)}\right]^{k-1} f'(0), \tag{2.1}$$

тада је $f(x) = \lambda e^{-\lambda x}$ за неко $\lambda > 0$.

Због доказа преко развоја функције у ред лему је могуће применити само када је функција густине аналитичка. На крају рада наведене су хипотезе о уопштењу ове теореме.

Прва, заснована на репрезентацији максимума у узорку обима три, доказана је у раду Јанева и Чакрабортија [88], а затим и карактеризација преко везе узастопних максимума, у раду истих аутора [18].

Ова идеја може се проширити на произвољне статистике поретка. Следе резултати аутора на ову тему који се могу наћи у [65] и [53].

2.2 Карактеризације на узорцима малог обима

Приказаћемо најпре три карактеризације експоненцијалне расподеле за које је заједничко то што се у њима јавља медијана узорка обима три.

Следећа лема погоднија је од леме 2.1.1 јер се и први извод такође изражава преко $f(0)$.

Лема 2.2.1. *Нека је F функција расподеле из класе $\mathcal{F}$. Ако за свако природно q важи:*

$$f^{(q)}(0) = (-1)^q f^{q+1}(0), \tag{2.2}$$

тада је $f(x) = \lambda e^{-\lambda x}$ *за неко* $\lambda > 0$.

Доказ. Развијајући функцију f у Маклоренов ред за позитивне вредности x имамо

$$f(x) = \sum_{q=0}^{\infty} f^{(q)}(0)\frac{x^q}{q!} = \sum_{q=0}^{\infty}(-1)^q f^{q+1}(0)\frac{x^q}{q!} = f(0)e^{-f(0)x}.$$

За $f(0) > 0$ ово је густина експоненцијалне расподеле за $\lambda = f(0)$. □

С обзиром да ће нам у доказима наредних теорема бити потребни изводи функција $F(x)f(x)$ и $F^2(x)f(x)$, њихово представљање преко $f(0)$ дајемо следећом лемом.

Лема 2.2.2. *Нека је* F *расподела из* $\mathcal{F}$. *Нека је* $G(x) = F(x)f(x)$ *и* $H(x) = F^2(x)f(x)$. *Нека је услов* (2.2) *задовољен за* $q \leq r-2$. *Тада важе следеће једнакости*

$$\begin{aligned} G^{(q)}(0) &= (-1)^{q-1}f^{q+1}(0)(2^q - 1), \;\; 1 \leq q \leq r-1. \\ H^{(q)}(0) &= (-1)^{q-2}f^{q+1}(0)(3^q - 2^{q+1} + 1), \;\; 1 \leq q \leq r. \end{aligned} \tag{2.3}$$

Доказ. Применом Лајбницове формуле за изводе производа на $G(x)$ добијамо

$$G^{(q)}(x) = \sum_{j=0}^{q} \binom{q}{j} F^{(j)}(x) f^{(q-j)}(x).$$

Када заменимо $x = 0$, анулирају се сви сабирци који садрже $F^{(0)}(x)$, па имамо

$$\begin{aligned} G^{(q)}(0) &= \sum_{j=1}^{q} \binom{q}{j} f^{(j-1)}(0) f^{(q-j)}(0) \\ &= \sum_{j=1}^{q} \binom{q}{j} (-1)^{j-1} f^{j}(0) (-1)^{q-j} f^{q-j+1}(0) \\ &= (-1)^{q-1} f^{q+1}(0) \sum_{j=1}^{q} \binom{q}{j} \\ &= (-1)^{q-1} f^{q+1}(0)(2^q - 1). \end{aligned}$$

Приметимо да је израз на десној страни (2.3) једнак нули за $q = 1$. Како је $H(x) = F^2(x)f(x) = F(x)F(x)f(x)$, применом Лајбницове формуле за изводе производа три функције добијамо

$$H^{(q)}(x) = \sum_{s=0}^{q} \sum_{j=0}^{s} \binom{q}{j, s-j, q-s} F^{(j)}(x) F^{(s-j)}(x) f^{(q-s)}(x).$$

Замењујући $x = 0$ на исти начин као у претходном случају добијамо да је $H'(0) = 0$, а за $q \geq 2$

$$\begin{aligned} H^{(q)}(0) &= \sum_{s=2}^{q} \sum_{j=1}^{s} \binom{q}{j, s-j, q-s} f^{(j-1)}(0) f^{(s-j-1)}(0) f^{(q-s)}(0) \\ &= \sum_{s=2}^{q} \sum_{j=1}^{s-1} \binom{q}{j, s-j, q-s} (-1)^{j-1} f^{j}(0) (-1)^{s-j-1} f^{s-j}(0) (-1)^{q-s} f^{q-s+1}(0) \\ &= (-1)^{q-2} f^{q+1}(0) \sum_{s=2}^{q} \sum_{j=1}^{s-1} \binom{q}{j, s-j, q-s} \\ &= (-1)^{q-2} f^{q+1}(0) \left(3^q - \sum_{s=0}^{1} \sum_{j=0}^{s} \binom{q}{j, s-j, q-s} - 2 \sum_{s=2}^{q} \binom{q}{s} \right) \\ &= (-1)^{q-2} f^{q+1}(0)(3^q - 2^{q+1} + 1). \end{aligned}$$

$\square$

Напомена 2.2.1. *Приметимо да су изрази* $G'(0) = f^2(0)$ *и* $H''(0) = 2f^3(0)$ *увек тачни без обзира на услов теореме. Поред тога важи* $G(0) = H(0) = H'(0) = 0$.

Сада ћемо навести и доказати карактеризационе теореме.

Теорема 2.2.1. *Нека су* X_1, X_2, X_3 *независне једнако расподељене случајне величине с расподелом* $F \in \mathcal{F}$. *Ако важи*

$$\frac{1}{3}X_1 + \frac{1}{2}X_2 \stackrel{d}{=} X_{(2;3)} \tag{2.4}$$

тада X_1 *има експоненцијалну расподелу с неким* $\lambda > 0$.

Доказ. Изједначавањем густина с леве и десне стране (2.4) добијамо

$$\int_0^x 3f(3y)2f(2(x-y))dy = 6F(x)(1-F(x))f(x)$$

$$\int_0^x f(3y)f(2(x-y))dy = f(x)\int_0^x f(y)(1-2F(y))dy$$

$$\int_0^x f(3y)f(2(x-y))dy = f(x)\int_0^x (f(y)-2G(y))dy. \tag{2.5}$$

Доказаћемо индукцијом да из (2.5) следи (2.2). Диференцирањем по x два пута добијамо

$$\begin{aligned} f'(3x)3f(0) + 2f(3x)f'(0) + 4\int_0^x f(3y)f''(2(x-y))dy \\ = 2f(x)f'(x) - 2G'(x)f(x) - 2G(x)f'(x) + (f(x)-2G(x))f'(x) \\ + f''(x)\int_0^x (f(y)-2G(y))dy. \end{aligned}$$

Замењујући $x = 0$ имамо

$$5f'(0)f(0) = 3f(0)f'(0) - 2f(0)G'(0).$$

Применом леме 2.2.2 добијамо

$$f'(0) = -f^2(0),$$

па (2.2) важи за $q = 1$. Претпоставимо сада да је (2.2) задовољено за $1 \leq q \leq r-2$. Диференцирањем обе стране (2.5) r пута добијамо

$$\begin{aligned}&\sum_{j=0}^{r-1} 3^{r-1-j} f^{(r-1-j)}(3x) 2^j f^{(j)}(0) + \int\limits_0^x f(3y) 2^r f^{(r)}(2(x-y))dy \\ &\quad = \sum_{j=1}^{r} \binom{r}{j} f^{(r-j)}(x)(f^{(j-1)}(x) - 2G^{(j-1)}(x)) + f^{(r)}(x) \int\limits_0^x (f(y) - 2G(y))dy.\end{aligned}$$

Заменом $x = 0$ и елиминацијом чланова једнаких нули добијамо

$$\sum_{j=1}^{r} 3^{r-j} f^{(r-j)}(0) 2^{j-1} f^{(j-1)}(0) = \sum_{j=1}^{r} \binom{r}{j} f^{(r-j)}(0)(f^{(j-1)}(0) - 2G^{(j-1)}(0)).$$

Извлачећи из суме чланове који садрже $f^{(r-1)}(0)$ и групишући их с једне стране добијамо

$$\begin{aligned}&f^{(r-1)}(0) f(0)(r + 1 - 3^{r-1} - 2^{r-1}) \\ &\quad = \sum_{j=2}^{r-1} f^{(r-j)}(0) 3^{r-j} 2^{j-1} f^{(j-1)}(0) - \sum_{j=2}^{r-1} \binom{r}{j} f^{(r-j)}(0)(f^{(j-1)}(0) - 2G^{(j-1)}(0)).\end{aligned}$$

Применом индуктивне хипотезе и лема 2.2.1 и 2.2.2 добијамо

$$\begin{aligned}&f^{(r-1)}(0)(r + 1 - 3^{r-1} - 2^{r-1}) \\ &\qquad\qquad = (-1)^{r-1} f^r(0) \Bigg(\sum_{j=2}^{r-1} \Big(3^{r-j} 2^{j-1} - \binom{r}{j} 2^{j-1} \Big) - 2(2^{r-1} - 1) \Bigg).\end{aligned}$$

Да бисмо доказали да је (2.2) задовољено за $q = r-1$ треба још показати да је

$$r + 1 - 3^{r-1} - 2^{r-1} = \sum_{j=2}^{r-1} \left(3^{r-j}2^{j-1} - \binom{r}{j}2^{j-1}\right) - 2(2^{r-1} - 1),$$

односно

$$\sum_{j=1}^{r} 3^{r-j}2^{j-1} = \sum_{j=1}^{r} \binom{r}{j}(2^j - 1).$$

Једноставним рачуном добија се да су обе суме једнаке $3^r - 2^r$ чиме је доказ завршен. □

Теорема 2.2.2. *Нека су X_0, X_1, X_2, X_3 независне једнако расподељене случајне величине с расподелом $F \in \mathcal{F}$. Ако је*

$$X_0 + X_{(2;3)} \stackrel{d}{=} X_{(3;3)}$$

тада X_1 има експоненцијалну расподелу с неким $\lambda > 0$.

Доказ. Изједначавајући одговарајуће густине имамо

$$\begin{aligned}
\int_0^x f(y)6F(x-y)(1-F(x-y))f(x-y)dy &= 3F^2(x)f(x) \\
6\int_0^x f(y)F(x-y)(1-F(x-y))f(x-y)dy &= 6f(x)\int_0^x F(y)f(y)dy \\
\int_0^x f(y)G(x-y)dy - \int_0^x f(y)H(x-y)dy &= f(x)\int_0^x G(y)dy. \qquad (2.6)
\end{aligned}$$

Као и у претходном доказу показаћемо индукцијом да из (2.6) следи (2.2).

Диференцирајући обе стране три пута добијамо

$$f''(x)G(0) + f'(x)G'(0) + f(x)G''(0) + \int\limits_0^x f(x)G^{(3)}(x-y)dy$$
$$- \Big(f''(x)H(0) + f'(x)H'(0) + f(x)H''(0) + \int\limits_0^x f(x)H^{(3)}(x-y)dy\Big)$$
$$= 3f''(x)G(0) + 3f'(x)G'(0) + f(x)G''(0) + f^{(3)} \int\limits_0^x G(y)dy.$$

Заменом $x = 0$ добијамо

$$f'(0)G'(0) - f(0)H''(0) = 3f'(0)G'(0),$$

односно

$$f'(0) = -f^2(0).$$

Значи да (2.2) важи за $q = 1$. Претпоставимо сада да (2.2) важи за $1 \leq q \leq r-3$. Диференцирањем обе стране израза (2.6) r пута добијамо

$$\sum_{j=0}^{r-1} f^{(r-1-j)}(x)G^{(j)}(0) + \int\limits_0^x f(y)G^{(r)}(x-y)dy$$
$$- \sum_{j=0}^{r-1} f^{(r-1-j)}(x)H^{(j)}(0) - \int\limits_0^x f(y)H^{(r)}(x-y)dy$$
$$= \sum_{j=1}^{r} \binom{r}{j} f^{(r-j)}(x)G^{(j-1)}(x) + f^{(r)}(x) \int\limits_0^x G(y)dy.$$

Стављајући $x = 0$ и уклањајући чланове једнаке нули добијамо

$$\sum_{j=2}^{r} f^{(r-j)}(0)G^{(j-1)}(0) - \sum_{j=2}^{r-1} f^{(r-1-j)}(0)H^{(j)}(0) = \sum_{j=2}^{r} \binom{r}{j} f^{(r-j)}(0)G^{(j-1)}(0)$$

Чланови за $j = r$ су једнаки па се могу скратити. Премештањем чланова који садрже $f^{(r-2)}(0)$ на једну страну добијамо

$$
\begin{aligned}
& f^{(r-2)}(0)G'(0)\left(1-\binom{r}{2}\right) \\
= \; & \sum_{j=3}^{r-1}\left(\binom{r}{j}-1\right)f^{(r-j)}(0)G^{(j-1)}(0)+\sum_{j=2}^{r-1}f^{(r-1-j)}(0)H^{(j)}(0)
\end{aligned}
$$

Применом индуктивне хипотезе имамо

$$
\begin{aligned}
& f^{(r-2)}(0)\left(1-\binom{r}{2}\right) \\
& = (-1)^{r-2}f^{r-1}(0)\left(\sum_{j=3}^{r-1}(2^{j-1}-1)\left(\binom{r}{j}-1\right)-\sum_{j=2}^{r-1}(3^j-2^{j+1}+1)\right),
\end{aligned}
$$

Да бисмо показали индуктивни корак остаје још да се докаже

$$
1-\binom{r}{2}=\left(-2+\sum_{j=3}^{r-1}(2^{j-1}-1)\left(\binom{r}{j}-1\right)-(3^j-2^{j+1}+1)\right),
$$

односно

$$
\sum_{j=2}^{r-1}(2^{j-1}-1)\left(\binom{r}{j}-1\right)=\sum_{j=2}^{r-1}(3^j-2^{j+1}+1). \tag{2.7}
$$

Лако се може израчунати да су обе стране једнакости $\frac{3^r}{2}-2^{r+1}+r+\frac{3}{2}$, чиме је доказ завршен. $\square$

Теорема 2.2.3. *Нека су X_1, X_2, X_3, X_4 независне једнако расподељене случајне величине с расподелом $F \in \mathcal{F}$. Ако је*

$$
X_{(2;3)}+\frac{1}{4}X_4 \overset{d}{=} X_{(3;4)}
$$

онда X_1 има експоненцијалну расподелу с неким $\lambda > 0$.

Доказ. Изједначавајући одговарајуће густине имамо

$$\int_0^x 6F(x-y)(1-F(x-y))f(x-y)4f(4y)dy = 12F^2(x)(1-F(x))f(x)$$

$$2\int_0^x f(4y)F(x-y)(1-F(x-y))f(x-y)dy = f(x)\int_0^x (F^2(y)-F^3(y))f(y)dy$$

$$2\int_0^x f(4y)(G(x-y)-H(x-y))dy = f(x)\int_0^x (2G(y)-3H(y))dy. \tag{2.8}$$

Као и у претходним случајевим, доказаћемо да из (2.8) следи (2.2).

После израчунавања трећег извода и елиминисања чланова једнаких нули добијамо

$$8f'(0)G'(0) - 2f(0)H''(0) = 6f'(0)G'(0) - 3f(x)H''(0)$$

односно

$$f'(0) = -f^2(0).$$

Услов (2.2) дакле важи за $q = 1$. Претпоставимо да важи за $1 \leq q \leq r-3$. Диференцирањем израза (2.8) r пута добијамо

$$\begin{aligned} & 2\sum_{j=0}^{r-1} 4^{r-1-j} f^{(r-1-j)}(4x)(G^{(j)}(0) - H^{(j)}(0)) \\ & + 2\int_0^x 4^r f^{(r)}(4y)(G(x-y) - H(x-y))dy \\ & = \sum_{j=1}^{r} \binom{r}{j} f^{(r-j)}(x)(2G^{(j-1)}(x) - 3H^{(j-1)}(x)) \\ & + f^r(x)\int_0^x (2G(y) - 3H(y))dy. \end{aligned}$$

Стављајући $x = 0$ и елиминишући чланове једнаке нули имамо

$$
\begin{aligned}
&2\sum_{j=2}^{r-1} 4^{r-j} f^{(r-j)}(0)G^{(j-1)}(0) - 2\sum_{j=2}^{r-1} 4^{r-1-j} f^{(r-1-j)}(0)H^{(j)}(0)\\
&\qquad = 2\sum_{j=2}^{r} \binom{r}{j} f^{(r-j)}(0)G^{(j-1)}(0) - 3\sum_{j=2}^{r-1} \binom{r}{j+1} f^{(r-1-j)}(0)H^{(j)}(0).
\end{aligned}
$$

Чланови за $j = r$ у првој и трећој суми поклапају се па се могу скратити. Извлачећи на једну страну чланове који садрже $f^{(r-2)}(0)$ добијамо

$$
\begin{aligned}
2f^{(r-2)}(0)f^2(0)\Big(4^{r-2} - \binom{r}{2}\Big) &= 2\sum_{j=3}^{r-1} f^{(r-j)}(0)G^{(j-1)}(0)\Big(\binom{r}{j} - 4^{r-j}\Big)\\
&\quad - \sum_{j=2}^{r-1} f^{(r-1-j)}(0)H^{(j)}(0)\Big(3\binom{r}{j+1} - 2\cdot 4^{r-1-j}\Big).
\end{aligned}
$$

Применом индуктивне хипотезе на десну страну претходног израза добијамо

$$
\begin{aligned}
2f^{(r-2)}(0)\Big(4^{r-2} - \binom{r}{2}\Big) &= (-1)^{r-2} f^{r-1}(0)\Big(2\sum_{j=3}^{r-1}(2^{j-1}-1)\Big(\binom{r}{j} - 4^{r-j}\Big)\\
&\quad + \sum_{j=2}^{r-1}(3^j - 2^{j+1} + 1)\Big(3\binom{r}{j+1} - 2\cdot 4^{r-1-j}\Big)\Big).
\end{aligned}
$$

Остаје још да се покаже

$$
\sum_{j=2}^{r-1}(2^j - 2)\Big(4^{r-j} - \binom{r}{j}\Big) = \sum_{j=2}^{r-1}(3^j - 2^{j+1} + 1)\Big(3\binom{r}{j+1} - 2\cdot 4^{r-j-1}\Big). \tag{2.9}
$$

Лако се може израчунати да су обе суме једнаке $\frac{4^r}{3} - 3^r + 2^r - \frac{1}{3}$, чиме је доказ завршен. $\square$

Све три наведене теореме имају своја природна уопштења која ћемо у даљем излагању представити.

2.3 Неки идентитети са Стирлинговим бројевима друге врсте

У овом одељку представићемо четири комбинаторна идентитета која ће бити коришћена у доказима наредних карактеризационих теорема. У свим тим идентитетима појављују се Стирлингови бројеви друге врсте.

Стирлинови бројеви друге врсте, које обележавамо са $\left\{{a \atop b}\right\}$, представљају број начина да се скуп од a елемената подели на b непразних подскупова. Више о Стирлинговим бројевима може се наћи нпр. у [25]. Овде наводимо четири елементарна идентитета за Стирлингове бројеве друге врсте:

$$\left\{{a \atop b}\right\} = \left\{{a-1 \atop b-1}\right\} + b\left\{{a-1 \atop b}\right\}, \tag{2.10}$$

$$\left\{{a+1 \atop b+1}\right\} = \sum_{l=0}^{a} \binom{a}{l} \left\{{l \atop b}\right\}, \tag{2.11}$$

$$\left\{{a+b+1 \atop b}\right\} = \sum_{l=0}^{b} l\left\{{a+l \atop l}\right\}, \tag{2.12}$$

$$a^b = \sum_{l=0}^{b} \left\{{b \atop l}\right\} a(a-1)\cdots(a-l+1). \tag{2.13}$$

Прелазимо сада на идентитете потребне за доказивање карактеризационих теорема.

Лема 2.3.1. *За целе бројеве k, n, r такве да је $1 < k \leq n$ и $r \geq 0$ важи*

$$\sum_{j=k-2}^{k+r-1} \sum_{i=0}^{j-k+2} \binom{n-k}{i} (i+k-2)! \left\{{j+1 \atop i+k-1}\right\} (k-1) n^{k+r-1-j}$$
$$= \sum_{i=0}^{r+1} \binom{n-k}{i} (i+k-1)! \left\{{k+r+1 \atop i+k}\right\}. \tag{2.14}$$

Доказ. Доказаћемо тврђење индукцијом по r. За $r = 0$ једнакост (2.14) своди се на

$$(k-1)!\left(n+\left\{{k \atop k-1}\right\}+(k-1)(n-k)\right)=(k-1)!\left(\left\{{k+1 \atop k}\right\}+k(n-k)\right),$$

што важи на основу идентитета (2.10). Стога тврђење леме важи за $r=0$ и за свако $1<k\leq n$.

Претпоставимо сада да (2.14) важи за $r-1$ за свако $1<k\leq n$. Лева страна једнакости (2.14) може се раставити као

$$\sum_{j=k-2}^{k+r-2}\sum_{i=0}^{j-k+2}\binom{n-k}{i}(i+k-2)!\left\{{j+1 \atop i+k-1}\right\}(k-1)n^{k+r-1-j} \\ +\sum_{i=0}^{r+1}\binom{n-k}{i}(i+k-2)!\left\{{k+r \atop i+k-1}\right\}(k-1).$$

Применом индуктивне хипотезе на први сабирак добијамо да је горњи израз једнак

$$\sum_{i=0}^{r}\binom{n-k}{i}(i+k-1)!\left\{{k+r \atop i+k}\right\}n+\sum_{i=0}^{r+1}\binom{n-k}{i}(i+k-2)!\left\{{k+r \atop i+k-1}\right\}(k-1). \tag{2.15}$$

Остаје још да се докаже да је израз (2.15) једнак

$$\sum_{i=0}^{r+1}\binom{n-k}{i}(i+k-1)!\left\{{k+r+1 \atop i+k}\right\},$$

што се може записати као

$$\sum_{i=0}^{r+1}\binom{n-k}{i}(i+k-1)!\left\{{k+r \atop i+k-1}\right\}+\sum_{i=0}^{r}\binom{n-k}{i}(i+k)!\left\{{k+r \atop i+k}\right\}. \tag{2.16}$$

Груписањем одговарајућих сабирака из израза (2.15) и (2.16) добијамо

$$\sum_{i=0}^{r}\binom{n-k}{i}(i+k-1)!\left\{{k+r \atop i+k}\right\}(n-k-i)$$
$$=\sum_{i=0}^{r+1}\binom{n-k}{i}(i+k-2)!\left\{{k+r \atop i+k-1}\right\}i.$$

Последња једнакост лако се доказује заменом индекса $j=i+1$ у првој суми. $\square$

Лема 2.3.2. *За целе бројеве k,n,r такве да је $1<k\leq n$ и $r\geq 0$ важи*

$$\sum_{j=k-2}^{k+r-1}\sum_{i=0}^{j-k+2}\binom{n-k+1}{i}(i+k-2)!\left\{{j+1 \atop i+k-1}\right\}(k-1)(n-k+1)^{k+r-1-j}$$
$$=\sum_{i=0}^{r+1}\binom{n-k}{i}(i+k-1)!\left\{{k+r+1 \atop i+k}\right\}. \tag{2.17}$$

Доказ ове леме је потпуно аналоган доказу претходне леме 2.3.1 па га изостављамо.

Лема 2.3.3. *За целе бројеве k,n,r такве да је $1<k\leq n$ и $r\geq 0$ важи*

$$\sum_{i=0}^{r+1}\binom{n-k}{i}\sum_{s=1}^{k}(n-k+s)\frac{(i+s-1)!}{(s-1)!}\left\{{s+r \atop i+s}\right\}$$
$$=\sum_{i=0}^{r+1}\binom{n-k}{i}\frac{(i+k-1)!}{(k-1)!}\left\{{k+r+1 \atop i+k}\right\}. \tag{2.18}$$

Доказ. Доказаћемо тврђење леме индукцијом по n. За свако r и k и $n=k$ израз (2.18) своди се на идентитет (2.12).

Претпоставимо сада да је једнакост (2.18) испуњена за свако k, свако r и $n-1$. Доказаћемо да је такође испуњена за n. Трансформацијом леве стране једнакости (2.18) добијамо

$$\sum_{i=0}^{r+1}\binom{n-k}{i}\left((n-k-i)\sum_{s=1}^{k}\frac{(i+s-1)!}{(s-1)!}\left\{ {s+r \atop i+s} \right\}+\sum_{s=1}^{k}\frac{(i+s)!}{(s-1)!}\left\{ {s+r \atop i+s} \right\}\right)$$
$$=\sum_{i=0}^{r+1}(n-k)\binom{n-k-1}{i}\sum_{s=1}^{k}\frac{(i+s-1)!}{(s-1)!}\left\{ {s+r \atop i+s} \right\}$$
$$+\sum_{i=0}^{r+1}\binom{n-k-1}{i}\sum_{s=1}^{k}\frac{(i+s)!}{(s-1)!}\left\{ {s+r \atop i+s} \right\}+\sum_{i=0}^{r+1}\binom{n-k-1}{i-1}\sum_{s=1}^{k}\frac{(i+s)!}{(s-1)!}\left\{ {s+r \atop i+s} \right\}$$
$$=\sum_{i=0}^{r+1}(n-k)\binom{n-k-1}{i}\sum_{s=1}^{k}\frac{(i+s-1)!}{(s-1)!}\left\{ {s+r \atop i+s} \right\}$$
$$+\sum_{i=0}^{r+1}\binom{n-k-1}{i}\sum_{s=1}^{k}\frac{(i+s)!}{(s-1)!}\left(\left\{ {s+r \atop i+s} \right\}+(i+s+1)\left\{ {s+r \atop i+1+s} \right\}\right).$$

Применом идентитета (2.10), померањем индекса s у последњој унутрашњој суми, и издвајањем члана за $s=k+1$, горњи израз постаје

$$\begin{aligned}\sum_{i=0}^{r+1}(n-k)\binom{n-k-1}{i}\sum_{s=1}^{k}\frac{(i+s-1)!}{(s-1)!}\left\{ {s+r \atop i+s} \right\}&\\ +\sum_{i=0}^{r+1}\binom{n-k-1}{i}\sum_{s=2}^{k}\frac{(i+s-1)!}{(s-2)!}\left\{ {s+r \atop i+s} \right\}&\\ +\sum_{i=0}^{r+1}\binom{n-k-1}{i}\frac{(i+k)!}{(k-1)!}\left\{ {k+1+r \atop i+k+1} \right\}&.\end{aligned}$$

Груписањем прва два сабирка и применом индуктивне хипотезе добијамо десну страну израза (2.18), чиме је доказ завршен. □

Лема 2.3.4. *За целе бројеве k,n,r такве да је $1<k\le n$ и $r\ge 0$ важи*

$$\sum_{\substack{j_1,\ldots,j_k\ge 0\\ j_1+\cdots+j_k=r+1}} n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k}=\sum_{i=0}^{r+1}\binom{n-k}{i}\frac{(i+k-1)!}{(k-1)!}\left\{ {i+r+1 \atop i+k} \right\}. \tag{2.19}$$

Доказ. Тврђење доказујемо јаком индукцијом по r. За свако k и n и

$r = 0$ имамо

$$n + (n-1) + \cdots + (n-k+1) = \left\{ {k+1 \atop k} \right\} + (n-k)k,$$

што очигледно важи јер је $\left\{ {k+1 \atop k} \right\} = \frac{k(k+1)}{2}$. Претпоставимо сада да једнакост (2.19) важи за све вредности до $r-1$. Остаје још доказати да важи и за r.

Раздвајањем суме с леве стране једнакости (2.19) на два дела: за $j_1 = 0$ и $j_1 \geq 0$, добијамо

$$\begin{aligned} \sum_{\substack{j_1,\ldots,j_k \geq 0 \\ j_1+\cdots+j_k=r+1}} n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k} &= \sum_{\substack{j_1,\ldots,j_k \geq 0 \\ j_1+\cdots+j_k=r+1}} (n-1)^{j_2}\cdots(n-k+1)^{j_k} \\ &+ \sum_{j_1=1}^{r+1} n^{j_1} \sum_{\substack{j_2,\ldots,j_k \geq 0 \\ j_2+\cdots+j_k=r+1-j_1}} (n-1)^{j_2}\cdots(n-k+1)^{j_k}. \end{aligned} \tag{2.20}$$

Збир индексā унутрашње суме другог сабирка израза (2.20) мањи је од $r+1$ па се индуктивна хипотеза може применити. (у овом случају за $n-1$, $k-1$ и $r+1-j_1$). Први сабирак може се рекурзивно даље разбијати на исти начин док сви индекси изузев последњег не буду једнаки нули. Након овог процеса, укључујући и примену индуктивне хипотезе, добијамо

$$\begin{aligned} \sum_{\substack{j_1,\ldots,j_k \geq 0 \\ j_1+\cdots+j_k=r+1}} n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k} = \ & (n-k+1)^{r+1} + \\ + \sum_{l=1}^{k-1} \sum_{j=1}^{r+1} (n-l+1)^j \sum_{i=0}^{r+1-j} \binom{n-k}{i} & \frac{(i+k-l-1)!}{(k-l-1)!} \left\{ {k-l+r+1-j \atop i+k-l} \right\}. \end{aligned} \tag{2.21}$$

Заменом индекса j индексом $m = k+r-1-l-j$ и, након тога, индекса l индексом $s = k-l+1$, а затим применом идентитета (2.13) на $(n-k+1)^{r+1}$, израз (2.21) постаје

$$\sum_{s=2}^{k}\sum_{i=0}^{r}\sum_{m=i+s-2}^{r+s-2}\binom{n-k}{i}(n-k+s)^{r+s-1-m}\frac{(i+s-2)!}{(s-2)!}\left\{{m+1 \atop i+s-1}\right\}$$
$$+\sum_{i=0}^{r+1}\frac{(n-k+1)!}{(n-k+1-i)!}\left\{{r+1 \atop i}\right\}$$
$$=\sum_{s=2}^{k}\frac{(n-k+s)}{(s-1)!}\sum_{i=0}^{r}\binom{n-k}{i}\sum_{m=i+s-2}^{r+s-2}\Bigg((n-k+s)^{r+s-2-m}(s-1)$$
$$(i+s-2)!\left\{{m+1 \atop i+s-1}\right\}\Bigg)+(n-k+1)\sum_{i=1}^{r+1}\binom{n-k}{i-1}(i-1)!\left\{{r+1 \atop i}\right\}.$$

Применом леме 2.3.1 на две унутрашње суме и груписањем сабирака добијамо

$$\sum_{s=1}^{k}\sum_{i=0}^{r}(n-k+s)\binom{n-k}{i}\frac{(i+s-1)!}{(s-1)!}\left\{{r+s \atop i+s}\right\}.$$

Примењујући сада лему 2.3.3 добијамо десну страну једнакости (2.19), чиме је доказ завршен. $\square$

Напомена 2.3.1. *Тврђења лема 2.3.3 и 2.3.4 такође важе за* $k = 1$. *Докази су аналогни, али доста једноставнији, па их нећемо наводити.*

2.4 Карактеризације у којима је узорак произвољног обима

Као и у случају карактеризација на узорцима малог обима, и овде је потребно изразити произвољне изводе помоћне фунцкије преко $f(0)$. Означимо $A_m(x) = F^m(x)f(x)$. Њено изражавање преко $f(0)$ дајемо следећом лемом.

Лема 2.4.1. *Нека је* F *функција расподеле која припада класи* $\mathcal{F}$. *Ако је услов* (2.2) *испуњен за свако* $0 \le q \le r-m$, $r > m$, *онда је*

$$A_m^{(r)}(0) = (-1)^{r-m} f^{r+1}(0)\left\{{r+1 \atop m+1}\right\} m!. \tag{2.22}$$

Напомена 2.4.1. *За $r \leq m$ тврђење је тачно без икаквих претпоставки о изводима функције f.*

Доказ. Извод реда r функције $A_m(x)$ преко Лајбницове формуле за извод производа може се записати као

$$A_m^{(r)}(x) = \sum_{\substack{j_1,\dots,j_{m+1}\geq 0 \\ j_1+\dots+j_{m+1}=r}} \binom{r}{j_1,\dots,j_{m+1}} F^{(j_1)}(x)\cdots F^{(j_m)}(x) f^{(j_{m+1})}(x).$$

Користећи чињеницу да је $F(0) = 0$ добијамо

$$A_m^{(r)}(0) = \sum_{\substack{j_1,\dots,j_m\geq 1,\ j_{m+1}\geq 0 \\ j_1+\dots+j_{m+1}=r}} \binom{r}{j_1,\dots,j_{m+1}} f^{(j_1-1)}(0)\cdots f^{(j_m-1)}(0) f^{(j_{m+1})}(0). \tag{2.23}$$

С обзиром да су сви изводи који се појављују у изразу (2.23) реда мањег или једнаког $r - m$, користећи (2.2) добијамо

$$\begin{aligned}
A_m^{(r)}(0) &= \sum_{\substack{j_1,\dots,j_m\geq 1,\ j_{m+1}\geq 0 \\ j_1+\dots+j_{m+1}=r}} \binom{r}{j_1,\dots,j_{m+1}} (-1)^{r-m} f^{r+1}(0) \\
&= (-1)^{r-m} f^{r+1}(0) \sum_{\substack{j_1,\dots,j_m,j_{m+1}\geq 1 \\ j_1+\dots+j_{m+1}=r}} \binom{r}{j_1,\dots,j_{m+1}} \\
&+ (-1)^{r-m} f^{r+1}(0) \sum_{\substack{j_1,\dots,j_m\geq 1,\ j_{m+1}=0 \\ j_1+\dots+j_{m+1}=r}} \binom{r}{j_1,\dots,j_m} \\
&= (-1)^{r-m} f^{r+1}(0) \left(\left\{ {r \atop m+1} \right\} (m+1)! + \left\{ {r \atop m} \right\} m! \right) \\
&= (-1)^{r-m} f^{r+1}(0) m! \left\{ {r+1 \atop m+1} \right\}.
\end{aligned}$$

У последњем кораку искоришћен је идентитет (2.10). $\square$

Прелазимо сада на карактеризационе теореме. Најпре наводимо уопштење теореме 2.2.3 за произвољну статистику поретка у узорку произвољног обима.

Теорема 2.4.1. *Нека је* $X_1, \dots, X_n$ *узорак из расподеле* F *која припада* $\mathcal{F}$. *Нека је* k *фиксирани број такав да је* $1 < k \leq n$. *Ако је*

$$X_{(k-1;n-1)} + \frac{1}{n} X_n \stackrel{d}{=} X_{(k;n)}, \tag{2.24}$$

онда је узорак из експоненцијалне расподеле с неким параметром $\lambda > 0$.

Доказ. Изједначавајући одговарајуће густине из (2.24) добијамо

$$\begin{aligned}
&\int_0^x \frac{(n-1)!}{(k-2)!(n-k)!} F^{k-2}(x-y)(1-F(x-y))^{n-k} f(x-y) n f(ny) dy \\
&= \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x)(1-F(x))^{n-k} f(x),
\end{aligned}$$

односно

$$\begin{aligned}
&(k-1)\sum_{i=0}^{n-k}(-1)^i \binom{n-k}{i} \int_0^x A_{i+k-2}(x-y) f(ny) dy \\
= \quad & f(x) \sum_{i=0}^{n-k}(-1)^i \binom{n-k}{i} \int_0^x A_{i+k-2}(y) dy. \qquad (2.25)
\end{aligned}$$

Доказаћемо индукцијом да једнакост (2.2) важи за свако природно q што преко леме 2.2.1 повлачи да је $f(x)$ густина експоненцијалне расподеле.

Диференцирањем интегралне једначине (2.25) k пута добијамо

$$
\begin{aligned}
&(k-1)\sum_{i=0}^{n-k}(-1)^i\binom{n-k}{i}\Bigg(\sum_{j=0}^{k-1} n^{k-1-j} f^{(k-1-j)}(nx)A_{i+k-2}^{(j)}(0) \\
&+\int\limits_0^x A_{i+k-2}^{(k)}(x-y)f(ny)dy\Bigg) \\
&=\sum_{i=0}^{n-k}(-1)^i(i+k-1)\binom{n-k}{i}\Bigg(\sum_{j=1}^{k}\binom{k}{j} f^{(k-j)}(x)A_{i+k-2}^{(j-1)}(0) \\
&+f^{(k)}(x)\int\limits_0^x A_{i+k-2}^{(k)}(y)dy\Bigg).
\end{aligned}
$$

Заменом $x=0$ и елиминацијом чланова једнаких нули добијамо

$$
\begin{aligned}
&nf'(0)(k-2)!f^{k-1}(0)+f(0)A_{k-2}^{(k-1)}(0)-(n-k)(k-1)!f^{k+1}(0) \\
&=f(0)A_{k-2}^{(k-1)}(0)+f'(0)(k-2)!f^{k-1}(0)k-k(n-k)(k-2)!f^{k+1}(0),
\end{aligned}
$$

одакле следи да је $f'(0)=-f^2(0)$, што значи да (2.2) важи за $q=1$. Претпоставимо сада да (2.2) важи за свако $q\le r$. Доказаћемо да важи и за $q=r+1$.

Диференцирањем интегралне једначине (2.25) $k+r$ пута, заменом $x=0$ и елиминацијом чланова једнаких нули добијамо

$$
\begin{aligned}
&(k-1)\sum_{i=0}^{r+1}(-1)^i\binom{n-k}{i}\sum_{j=i+k-2}^{k+r-1} n^{k+r-1-j} f^{(k+r-1-j)}(0)A_{i+k-2}^{(j)}(0) \\
&=\sum_{i=0}^{r+1}(-1)^i(i+k-1)\binom{n-k}{i}\sum_{j=i+k-2}^{k+r-1}\binom{k+r}{j+1} f^{(k+r-1-j)}(0)A_{i+k-2}^{(j)}(0).
\end{aligned}
$$

Чланови за $i=0$ и $j=k+r-1$ поклапају се па се могу скратити. Раздвајањем суме на два дела: за $i=0$ и $i>0$ добијамо

$$
\begin{aligned}
&(k-1)\left(n^{r+1}f^{(r+1)}(0)A_{k-2}^{(k-2)}(0)+\sum_{j=k-1}^{k+r-2}n^{k+r-1-j}f^{(k+r-1-j)}(0)A_{k-2}^{(j)}\right)\\
&+(k-1)\sum_{i=1}^{r+1}\sum_{j=i+k-2}^{k+r-1}(-1)^i\binom{n-k}{i}n^{k+r-1-j}f^{(k+r-1-j)}(0)A_{i+k-2}^{(j)}(0)\\
&=(k-1)\binom{k+r}{k-1}f^{(r+1)}(0)A_{k-2}^{(k-2)}(0)\\
&+(k-1)\sum_{k-1}^{k+r-2}\binom{k+r}{j+1}f^{(k+r-1-j)}(0)A_{k-2}^{(j)}(0)\\
&+\sum_{i=1}^{r+1}\sum_{j=i+k-2}^{k+r-1}(-1)^i\binom{n-k}{i}(i+k-1)\binom{k+r}{j+1}f^{(k+r-1-j)}(0)A_{i+k-2}^{(j)}(0).
\end{aligned}
$$

Применом индуктивне хипотезе на изводе функције f, и затим, преко леме 2.4.1, на изводе функције A_{i+k-2} и груписањем сабирака имамо

$$
\begin{aligned}
&f^{(r+1)}(0)f^{k-1}(0)(k-1)!\left(n^{r+1}-\binom{k+r}{k-1}\right)\\
&=(-1)^{r+1}f^{k+r+1}(0)\Bigg((k-1)!\sum_{j=k-1}^{k-2}\left(\binom{k+r}{j+1}-n^{k+r-1-j}\right)\left\{{j+1 \atop k-1}\right\}\\
&+\sum_{i=1}^{r+1}\sum_{j=i+k-2}^{k+r-1}\binom{n-k}{i}\Bigg((i+k-1)\binom{k+r}{j+1}\\
&-(k-1)n^{k+r-1-j}\Bigg)(i+k-2)!\left\{{j+1 \atop i+k-1}\right\}\Bigg).
\end{aligned}
$$

Да бисмо доказали индуктивни корак остаје још показати

$$
\begin{aligned}
&(k-1)!\left(n^{r+1}-\binom{k+r}{k-1}+\sum_{j=k-1}^{k-2}\left(n^{k+r-1-j}-\binom{k+r}{j+1}\right)\left\{{j+1 \atop k-1}\right\}\right)\\
=&\sum_{i=1}^{r+1}\sum_{j=i+k-2}^{k+r-1}\binom{n-k}{i}\Big((i+k-1)\binom{k+r}{j+1}\\
-&(k-1)n^{k+r-1-j}\Big)(i+k-2)!\left\{{j+1 \atop i+k-1}\right\}.
\end{aligned}
$$

Спајајући делове за $i=0$ и $i>0$ поново у исту суму добијамо

$$
\sum_{i=0}^{r+1}\sum_{j=i+k-2}^{k+r-1}\binom{n-k}{i}\left((i+k-1)\binom{k+r}{j+1}-(k-1)n^{k+r-1-j}\right)\times
$$
$$
\times(i+k-2)!\left\{{j+1 \atop i+k-1}\right\}=0,
$$

односно,

$$
\begin{aligned}
&\sum_{i=0}^{r+1}\sum_{j=i+k-2}^{k+r-1}\binom{n-k}{i}(k-1)n^{k+r-1-j}(i+k-2)!\left\{{j+1 \atop i+k-1}\right\}\\
&=\sum_{i=0}^{r+1}\binom{n-k}{i}(i+k-1)!\sum_{j=i+k-2}^{k+r-1}\binom{k+r}{j+1}\left\{{j+1 \atop i+k-1}\right\}.
\end{aligned}
$$

Применом идентитета (2.11) и леме 2.3.1 доказ је завршен. □

Следећа теорема је уопштење теореме 2.2.2.

Теорема 2.4.2. *Нека је $X_1,\dots,X_n$ узорак из расподеле F која припада $\mathcal{F}$ и нека је X_0 случајна величина независна од узорка која има исту расподелу. Нека је k фиксирани број такав да је $1<k\le n$. Ако важи*

$$
X_{(k-1;n)}+\frac{1}{n-k+1}X_0\overset{d}{=}X_{(k;n)} \tag{2.26}
$$

онда је узорак из експоненцијалне расподеле с неким $\lambda>0$.

Доказ изостављамо јер је потпуно аналоган доказу теореме 2.4.1, наравно уз примену леме 2.3.2 у последњем кораку.

Последња у низу теорема, која је уопштење теореме 2.2.1, даје нам репрезентацију статистике поретка реда k преко пондерисане суме независних случајних величина.

Теорема 2.4.3. *Нека је $X_1, \ldots, X_n$ узорак из расподеле F која припада $\mathcal{F}$. Нека је k фиксирани број такав да је $1 \leq k \leq n$. Ако важи*

$$\frac{1}{n}X_1 + \frac{1}{n-1}X_2 + \cdots + \frac{1}{n-k+1}X_k \overset{d}{=} X_{(k;n)} \tag{2.27}$$

онда је узорак из експоненцијалне расподеле с неким $\lambda > 0$.

Доказ. Нека је $k \geq 2$. Изједначавањем густина као у претходним доказима добијамо

$$\begin{aligned}
&\int_0^x nf(n(x-y_2)) \int_0^{y_2} (n-1)f((n-1)(y_2-y_3)) \cdots \\
&\quad \cdots \int_0^{y_{k-1}} (n-k-2)f((n-k+2)(y_{k-1}-y_k))f((n-k+1)y_k)dy_2\cdots dy_k \\
&\quad = \frac{n!}{(k-1)!(n-k)!}F^{k-1}(x)(1-F(x))^{n-k}f(x),
\end{aligned}$$

односно,

$$\begin{aligned}
&\int_0^x f(n(x-y_2)) \cdots \int_0^{y_{k-1}} f((n-k+2)(y_{k-1}-y_k))f((n-k+1)y_k)dy_2\cdots dy_k \\
&\qquad = \frac{1}{(k-1)!}f(x)\sum_{i=0}^{n-k}(-1)^i\binom{n-k}{i}\int_0^x A_{k-2+i}(y)dy.
\end{aligned} \tag{2.28}$$

Означимо леву страну једнакости (2.28) са $J_{k,n}(x)$. Она се очигледно може представити као

$$J_{k,n}(x) = \int_0^x f(n(x-y_2))J_{k-1,n-1}(y_2)dy_2,$$
$$J_{1,1}(x) = f((n-k+1)x).$$

Извод реда $(k+r)$ израза $J_{k,n}$ је

$$\begin{aligned} J_{k,n}^{(k+r)}(x) &= \sum_{j=0}^{k+r-1} n^j f^{(j)}(0) J_{k-1,n-1}^{(k+r-j-1)}(x) \\ &+ \int_0^x f^{(k+1)}(n(x-y_2)) n^{k+r} J_{k-1,n-1}^{(r+1)}(y_2)dy_2. \end{aligned}$$

Стављајући $x=0$ добијамо

$$\begin{aligned} J_{k,n}^{(k+r)}(0) &= \sum_{j=0}^{k+r-1} n^j f^{(j)}(0) J_{k-1,n-1}^{(k+r-j-1)}(0), \\ J_{1,1}^{(s)}(0) &= (n-k+1)^s f^{(s)}(0), \text{ за свако } s \geq 0. \end{aligned} \tag{2.29}$$

Применом рекурентне формуле (2.29) $k-1$ пута добијамо

$$\begin{aligned} J_{k,n}^{(k+1)}(0) = \sum_{j_1=0}^{k+r-1} n^{j_1} f^{(j_1)}(0) \sum_{j_2=0}^{k+r-2-j_1} (n-1)^{j_2} f^{(j_2)}(0) \cdots \\ \cdots \sum_{j_{k-1}=0}^{r+1-\sum_{l=1}^{k-2} j_l} (n-k+2)^{j_{k-1}} f^{(j_{k-1})} (n-k+1)^{r+1-\sum_{l=1}^{k-1} j_l} f^{(r+1-\sum_{l=1}^{k-1} j_l)}(0). \end{aligned}$$

Извод реда $(k+r)$ леве стране израза (2.28) постаје

$$\sum_{\substack{j_1,\ldots,j_k \geq 0 \\ j_1+\cdots+j_k=r+1}} n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k} f^{(j_1)}(0) f^{(j_2)}(0)\cdots f^{(j_k)}(0).$$

Као и у претходним случајевима, доказаћемо индукцијом да (2.2) важи за свако q. За $r = 0$, извод реда k израза (2.28) за $x = 0$ је

$$\begin{aligned}(n + n - 1 + \cdots + n - k + 1)f'(0)f^{k-1}(0)\\ = \frac{1}{(k-2)!}f(0)A_{k-2}^{(k-1)}(0) + f'(0)f^{k-1}(0)k - k(n-k)f^{k+1}(0).\end{aligned} \tag{2.30}$$

Из (2.23) следи

$$A_{k-2}^{(k-1)}(0) = f^{k-2}(0)f'(0)(k-1)! + (k-2)f'(0)f^{k-2}(0)\frac{(k-1)!}{2}.$$

Заменом овог израза у (2.30) добијамо $f'(0) = -f^2(0)$ што значи да (2.2) важи за $q = 1$. Претпоставимо сада да је (2.2) задовољено за $q \leq r$. Показаћемо да важи и за $q = r + 1$. Извод реда $(k + r)$ израза (2.28) за $x = 0$ је

$$\begin{aligned}\sum_{\substack{j_1,\dots,j_k \geq 0 \\ j_1+\cdots+j_k=r+1}} n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k}f^{(j_1)}(0)f^{(j_2)}(0)\cdots f^{(j_k)}(0)\\ = \sum_{i=0}^{r+1}(-1)^i\frac{(i+k-1)!}{(k-1)!}\binom{n-k}{i}\sum_{j=i+k-2}^{k+r-1}\binom{k+r}{j+1}f^{(k+r-1-j)}(0)A_{i+k-2}^{(j)}(0).\end{aligned} \tag{2.31}$$

Применом индуктивне хипотезе лева страна израза (2.31) постаје

$$\begin{aligned}&f^{k-1}(0)f^{(r+1)}(n^{r+1} + \cdots + (n-k+1)^{r+1})\\ &\quad + \sum_{\substack{0 \leq j_1,\dots,j_k < r+1 \\ j_1+\cdots+j_k=r+1}} n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k}(-1)^{r+1}f^{r+1+k}(0),\end{aligned}$$

док се његова десна страна може изразити као

$$
\begin{aligned}
\sum_{i=1}^{r+1}\binom{n-k}{i}\sum_{j=i+k-2}^{k+r-1}&\binom{k+r}{j+1}(-1)^{r+1}f^{k+r+1}\frac{(i+k-1)!}{(k-1)!}\begin{Bmatrix} j+1\\ i+k-1\end{Bmatrix}\\
&+\sum_{j=i+k-1}^{k+r-2}f^{k+r+1}(0)\frac{(i+k-2)!}{(k-2)!}(-1)^{r+1}\begin{Bmatrix} j+1\\ k-1\end{Bmatrix}\\
&+\binom{k+r}{k-1}f^{(r+1)}(0)\frac{(i+k-2)!}{(k-2)!}+\frac{1}{(k-2)!}f(0)A_{k-2}^{(k+r-1)}(0).
\end{aligned}
$$

Израз $A_{k-2}^{(k+r-1)}(0)$ може се израчунати користећи (2.22) и (2.23), и једнак је

$$
\begin{aligned}
A_{k-2}^{k+r-1}(0) &= \frac{(k+r-1)!}{(r+1)!}f^{k-2}(0)f^{(r+1)}(0)\\
&+ \frac{(k+r-1)!}{(r+2)!}(k-2)f^{k-2}(0)f^{(r+1)}(0)\\
&+ \sum_{\substack{1\le j_1,\dots,j_{k-2}<r+2\\ 0\le j_{k-1}<r+1\\ j_1+\cdots+j_k=k+r-1}}(-1)^{r+1}f^{r+k}(0)\frac{(k+r-1)!}{j_1!\cdots j_{k-1}!}.
\end{aligned}
$$

Након горе наведених трансформација и груписања сабирака, израз

(2.31) своди се на

$$
\begin{aligned}
f^{(r+1)}(0)&\left(n^{r+1}+\cdots+(n-k+1)^{r+1}-\binom{k+r}{k-1}-\binom{k+r-1}{k-2}-\binom{k+r-1}{k-3}\right)\\
&=(-1)^{r+1}f^{r+2}(0)\left(\sum_{j=k-1}^{k+r-2}\binom{k+r}{j+1}\left\{{j+1 \atop k-1}\right\}\right.\\
&+\sum_{i=1}^{r+1}\binom{n-k}{i}\sum_{j=i+k-2}^{k+r-1}\binom{k+r}{j+1}\frac{(i+k-1)!}{(k-1)!}\left\{{j+1 \atop i+k-1}\right\}\\
&+\frac{1}{(k-2)!}\sum_{\substack{1\le j_1,\dots,j_{k-2}<r+2\\ 0\le j_{k-1}<r+1\\ j_1+\cdots+j_k=k+r-1}}\frac{(k+r-1)!}{j_1!\cdots j_{k-1}!}\\
&\left.-\sum_{\substack{0\le j_1,\dots,j_k<r+1\\ j_1+\cdots+j_k=r+1}}n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k}\right).
\end{aligned}
$$

Да бисмо доказали индуктивни корак остаје још да покажемо

$$
\begin{aligned}
&\left(n^{r+1}+\cdots+(n-k+1)^{r+1}-\binom{k+r}{k-1}-\binom{k+r-1}{k-2}-\binom{k+r-1}{k-3}\right)\\
&=\sum_{j=k-1}^{k+r-2}\binom{k+r}{j+1}\left\{{j+1 \atop k-1}\right\}+\frac{1}{(k-2)!}\sum_{\substack{r+2>j_1,\dots,j_{k-2}\ge 1\\ r+1>j_{k-1}\ge 0\\ j_1+\cdots+j_k=k+r-1}}\frac{(k+r-1)!}{j_1!\cdots j_{k-1}!}\\
&+\sum_{i=1}^{r+1}\binom{n-k}{i}\sum_{j=i+k-2}^{k+r-1}\binom{k+r}{j+1}\frac{(i+k-1)!}{(k-1)!}\left\{{j+1 \atop i+k-1}\right\}\\
&-\sum_{\substack{r+1>j_1,\dots,j_k\ge 0\\ j_1+\cdots+j_k=r+1}}n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k}.
\end{aligned}
$$

Враћањем сабирака назад у одговарајуће суме и применом идентитета (2.11) добијамо

$$\sum_{\substack{j_1,\ldots,j_k\geq 0\\ j_1+\cdots+j_k=r+1}} n^{j_1}(n-1)^{j_2}\cdots(n-k+1)^{j_k} = \sum_{i=0}^{r+1}\binom{n-k}{i}\frac{(i+k-1)!}{(k-1)!}\left\{{i+r+1 \atop i+k}\right\},$$

што следи из леме 2.3.4, па је доказ за $k \geq 2$ завршен.

Случај $k = 1$ доказује се на аналоган, али доста једноставнији начин, па га овде не наводимо. $\square$

Следећа директна последица теореме 2.4.3, хипотеза је коју су поставили Арнолд и Виљасењор [9].

Последица 2.4.1. *Нека је $X_1, \ldots, X_n$ случајни узорак из расподеле F која припада $\mathcal{F}$. Ако важи*

$$X_1 + \frac{1}{2}X_2 + \cdots + \frac{1}{n}X_n \overset{d}{=} X_{(n;n)},$$

онда је узорак из експоненцијалне расподеле с неким $\lambda > 0$.

Поглавље 3

Тестови сагласности на основу карактеризација

Карактеризације расподела погодне су за формирање тестова сагласности с одређеном расподелом. Идеја је потекла од Линика [48]. Данас конструкција тестова на основу карактеризација постаје један од главних праваца развоја теорије тестова сагласности.

Према [35] прво проучавање таквог типа тестова јавља се 1976. у раду Вашичека [79] у којем је предложен тест заснован на карактеризацији нормалне расподеле на основу максималне ентропије. Затим следе тестови експоненцијалности засновани на особини одсуства меморије ([6],[1],[45],[46],[58]). Експоненцијална расподела вероватно има примат у броју тестова заснованих на карактеризацијама који су посвећени управо њој.

Поред наведених, ту су још тестови из радова [15], [28], [29], [30] и многи други. Проучавање тестова сагласности на основу карактеризација неких непрекидних расподела налазимо у раду [54]. Тест сагласности с униформном расподелом на основу карактеризације Папатанасијуа (теорема 1.2.10) разматран је у раду [27].

Случајевима у којима је карактеризација на основу једнако расподељених статистика посвећено је такође неколико радова.

Занимљиво је да су у раду [63] представљени тестови сагласности с експоненцијалном расподелом на основу својства да $X/(X+Y)$ има униформну расподелу што није карактеризациона теорема (контрапример може се наћи у [43]).

У раду Никитина и Волкове [62] обрађени су тестови на основу карактеризације Ахсанулаха (теорема 1.1.3), а у радовима Волкове ([80],[81]) на основу карактеризације Розберга (теорема 1.1.2) и Јанева и Чакрабортија заснованој на рекурзивној репрезентацији узастопних максимума.

Тестови сагласности с нормалном расподелом на основу Пољине карактеризације обрађени су у радовима [55] и [49]. Тестови сагласности на основу Шеповог својства (теорема 1.1.9) разматрани су у раду [83].

Тест сагласности са степеном расподелом на основу карактеризације Пурија и Рубина налазимо у раду [84].

3.1 U-статистике и V-статистике

С обзиром на то да су тест статистике тестова који су предмет ове дисертације засноване на U-статистикама и V-статистикама, остатак поглавља биће посвећен њима.

Теорију U-статистика први је развио Хефдинг [31]. Детаљније проучавање ове проблематике може се наћи у књигама [42] и [75].

Нека је $X_1, \dots, X_n$ узорак из расподеле F и нека је $\vartheta = \vartheta(F)$ параметар који има непристрасну оцену на основу датог узорка, тј. постоји $m \geq 1$ и мерљива функција $\Phi(x_1, \dots, x_m)$ таква да је

$$E(\Phi(X_1, \dots, X_m)) = \vartheta.$$

Најмање m за које постоји оцена називамо *редом* параметра ϑ. Може се без губљења општости претпоставити да је функција Φ симетрична по својим аргументима јер ако није може се заменити симетричном функцијом

$$\frac{1}{m!}\sum_{\pi}\Phi(x_{i_1},\ldots,x_{i_m})$$

где је сума по свих $m!$ пермутација скупа $i_1,\ldots,i_m$.

Једна од непристрасних оцена параметра ϑ је статистика дата следећом дефиницијом.

Дефиниција 3.1.1. *Нека је $X_1, X_2, \ldots, X_n$ узорак из расподеле F и нека је $\Phi : \mathbb{R}^m \mapsto \mathbb{R}$ мерљива функција. Статистика*

$$U_n = \frac{(n-m)!}{n!}\sum_{\pi_{m,n}}\Phi(X_{i_1},\ldots,X_{i_m}), \tag{3.1}$$

где је $\pi_{m,n}$ скуп свих m-пермутација $i_1,\ldots,i_m$ скупа природних бројева од 1 до n назива се U-статистика, а функција Φ њено језгро. Број m назива се редом U-статистике.

Назив U-статистика потиче од првог слова енглеске речи за непристрасну оцену (енгл. *unbiased*). У случају да је језгро U-статистике Φ симетрична функција по својим аргументима, горња дефиниција своди се на

$$U_n = \binom{n}{m}^{-1}\sum_{1\le i_1\le\cdots\le i_m\le n}\Phi(X_{i_1},\ldots,X_{i_m}). \tag{3.2}$$

Примери U-статистика су узорачка средина ($\Phi(X_1) = X_1$), поправљена узорачка дисперзија ($\Phi(X_1, X_2) = (X_1 - X_2)^2/2$) итд.

Сродне U-статистикама су и такозване V-статистике (или фон-Мизесови функционали) који се дефинишу на следећи начин.

Дефиниција 3.1.2. *Нека је $X_1, X_2, \ldots, X_n$ узорак из расподеле F и нека је $\Phi : \mathbb{R}^m \mapsto \mathbb{R}$ мерљива функција. Статистика*

$$V_n = n^{-m}\sum_{i_1,\ldots,i_m=1}^{n}\Phi(X_{i_1},\ldots,X_{i_m}), \tag{3.3}$$

назива се V-статистика. Функција Φ је њено језгро, а m је ред ове статистике.

3.2 Асимптотска својства U-статистика

Приметимо да U-статистике и V-статистике имају једноставну структуру и представљају збир једнако расподељених случајних величина, али, осим у случају реда 1, сабирци нису независни. Међутим, постоји начин на који се тај проблем може заобићи, тј. U-статистике се могу с одређеним степеном тачности апроксимирати збиром независних једнако расподељених случајних величина. То се постиже методом “пројекције” U-статистике на своје аргументе.

Означимо са Φ_c следеће условно математичко очекивање

$$\begin{aligned}\Phi_c(x_1,\dots,x_c) &= E\Phi(x_1,\dots,x_c,X_{c+1},\dots,X_m)\\ &= E(\Phi(X_1,\dots,X_m)|X_1=x_1,\dots,X_c=x_c),\ \ 0\le c\le m.\end{aligned}$$

Нека је $\tilde{\Phi}=\Phi-\vartheta(F)$ и $\tilde{\Phi}_c=\Phi_c-\vartheta(F)$. Дефинишемо низ функција на следећи начин

$$\begin{aligned}&g_1(x_1)=\tilde{\Phi}_1(x_1)\\ &g_2(x_1,x_2)=\tilde{\Phi}_2(x_1,x_2)-g_1(x_1)-g_1(x_2)\\ &\dots\\ &g_m(x_1,\dots,x_m)=\tilde{\Phi}_m(x_1,\dots,x_m)-\sum_{i=1}^{m}g_1(x_i)-\sum_{1\le i_1<i_2\le m}g_2(x_{i_1},x_{i_2})\\ &-\dots-\sum_{1\le i_1<\dots<i_{m-1}\le m}g_{m-1}(x_{i_1},\dots,x_{i_{m-1}}).\end{aligned}$$

Нека је $r\le 1$ најмањи природан број за који је $g_1=\dots=g_{r-1}=0$ и $g_r\ne 0$. Број r тада се назива *рангом* језгра (или U-статистике). Ако је ранг језгра једнак један, тада језгро називамо *недегенерисаним*. Функције g_i називамо канонским функцијама, а $\tilde{\Phi}_c(x_1,\dots,x_c)$ пројекцијама реда c језгра Φ.

Дисперзија недегенерисане U-статистике може се израчунати као

$$\sigma_{\Phi}^2=m^2Eg_1^2n^{-1}.$$

Недегенерисаност фамилије U-статистика дефинишемо на следећи начин (в. [59]).

Дефиниција 3.2.1. *За фамилију U-статистика $\{U_n(t), t \in [a,b]\}$ с језгрима $\Phi(x_1, \ldots, x_m; t)$ и првим пројекцијама $\phi(s_1; t)$ на X_1 кажемо да је недегенерисана ако је функција дисперзија $\sigma^2_\phi(t) = E\phi^2(X_1; t)$ једнака нули само на крајевима интервала $[a,b]$ и у највише коначно много тачака из унутрашњости тог интервала.*

Хефдинг је у свом раду [31] показао да за U-статистике ранга r с језгром из L^1 простора важи следећа декомпозиција.

$$U_n = \sum_{c=r}^{m} \binom{m}{c} U_{nc},$$

где је

$$U_{nc} = \frac{1}{nc} \sum_{1 \le i_1 < \cdots < i_c \le n} g_c(X_{i_1}, \ldots, X_{i_c}).$$

Приметимо да је U_{nc} такође U-статистика, а да је U_{n1}, с обзиром да јој је ранг 1, збир независних једнако расподељених случајних величина.

Следеће две теореме представљају закон великих бројева и централну граничну теорему за U-статистике.

Теорема 3.2.1. *Нека је U_n U-статистика с језгром за које важи $E|\Phi| < \infty$. Тада за U_n важи јаки закон великих бројева, тј. низ U_n скоро сигурно конвергира параметру ϑ.*

Доказ се може наћи у [42] или [75].

Теорема 3.2.2 (Хефдинг, 1948)**.** *Нека је U_n U-статистика с недегенерисаним језгром за које важи $E(\Phi^2) < \infty$. Тада*

$$\frac{1}{\sqrt{DU_n}}(U_n - \vartheta) \xrightarrow{d} \mathcal{N}(0,1).$$

У случају недегенерисаних језгара $V-$статистике имају исту асимптотику као и U-статистике истог реда, па се приликом проучавања асимптотских својстава могу проучавати одговарајуће U-статистике.

3.3 U-емпиријске фунцкије расподеле и типови тест статистика

Познато је да емпиријска функција расподеле представља добру оцену теоријске функције расподеле. Како нам се у карактеризационим теоремама јављају расподеле неких статистика, оне се могу добро оценити такозваним U-емпиријским функцијама расподеле.

Дефиниција 3.3.1. *Нека је $X_1, X_2, \ldots, X_n$ узорак из расподеле F и нека је $h : \mathbb{R}^m \mapsto \mathbb{R}$ мерљива функција. Функција расподеле*

$$H_n(t) = \binom{n}{m}^{-1} \sum_{1 \le i_1 \le \cdots \le i_m \le n} \mathrm{I}\{h(X_{i_1}, \ldots, X_{i_m}) \le t\}, \ \ t \in \mathbb{R} \tag{3.4}$$

назива се U-емпиријска функција расподеле.

U-емпиријска функција расподеле $H_n(t)$ за фиксирано t је једна U-статистика чија је средња вредност

$$H_F(t) = P\{h(X_{i_1}, \ldots, X_{i_m}) \le t\}, \ \ t \in \mathbb{R}.$$

У случају $m = 1$ функција $H_n(t)$ поклапа се са стандардном емпиријском функцијом расподеле.

Аналогно U-емпиријским могу се дефинисати и V-емпиријске функције расподеле. Више о овим функцијама може се наћи у [37] и [42].

Следећа теорема (в. [28]) је теорема Гливенко-Кантелијевог типа о униформној конвергенцији U-емпиријских функција расподеле.

Теорема 3.3.1 (Хелмерс, Јансен, Серфлинг, 1988)**.** *За U-емпиријску функцију расподеле $H_n(t)$ и $\varepsilon > 0$ постоји позитивна константа C која не зависи од F тако да важи*

$$P\big\{\sup_t |H_n(t) - H_F(t)| > \varepsilon\big\} \le \big(1 + 4C\sqrt{[n/m]}\varepsilon\big)e^{-2[n/m]\varepsilon^2}.$$

Важност ове теореме огледа се у томе да су U-емпиријске (и V-емпиријске) функције расподеле с језгрима дефинисаним на основу карактеризације за велико n довољно блиске па се на основу њих може конструисати тест статистика за тест сагласности.

У научној литератури постоји неколико типова тест статистика које су засноване на некој врсти "растојања" међу статистикама. У даљем тексту пажња ће бити посвећена двама типовима статистика, интегралном и типу Колмогоров-Смирнова.

Претпоставимо да имамо карактеризацију расподеле F облика $g(X_1, \ldots, X_m) \overset{d}{=} h(X_1, \ldots, X_m)$. Нека је $F_n(t)$ емпиријска функција расподеле на основу узорка обима n и нека су $G_n(t)$ и $H_n(t)$ U-емпиријске (или V-емпиријске) функције расподеле на основу истог узорка с језгрима добијеним симетризацијом функција g и h које фигуришу у карактеризацији.

Тада тест статистику интегралног типа можемо дефинисати са

$$I_n = \int_{-\infty}^{\infty} (G_n(t) - H_n(t)) dF_n(t), \tag{3.5}$$

док се тест статистика типа Колмогоров-Смирнова дефинише са

$$D_n = \sup_t |G_n(t) - H_n(t)| \tag{3.6}$$

С обзиром да је у питању тест сагласности, нулта хипотеза је да је узорак из расподеле F за коју важи наведена карактеризација.

Није тешко видети да је интегрална статистика асимптотски еквивалентна U-статистици с језгром $\Phi(X_1, \ldots, X_{m+1})$ које је добијено симетризацијом функције

$$\mathrm{I}\{g(X_1, \ldots, X_m) \leq X_{m+1}\} - \mathrm{I}\{h(X_1, \ldots, X_m) \leq X_{m+1}\}.$$

Уколико је ово језгро недегенерисано, а показује се да је то случај у већини тестова заснованим на карактеризацијама ове врсте, тада тест статистика интегралног типа има, на основу Хефдингове теореме 3.2.2

асимптотски нормалну расподелу

$$\sqrt{n}I_n \xrightarrow{d} \mathcal{N}(0, m^2\sigma_\Phi^2),$$

где је σ_Φ^2 дисперзија пројекције језгра Φ на један од својих аргумената.

С друге стране израз $|G_n(t) - H_n(t)|$ за фиксирано t такође је U-статистика с језгром $\Xi(X_1, \ldots, X_m; t)$ које је добијено симетризацијом функције

$$\mathrm{I}\{g(X_1, \ldots, X_m) \leq t\} - \mathrm{I}\{h(X_1, \ldots, X_m) \leq t\}.$$

У случају да је језгро Ξ недегенерисано у складу с дефиницијом 3.2.1 може се показати на основу резултата Силвермана [76] да U-емпиријски случајни процес

$$\eta_n(t) = \sqrt{n}(G_n(t) - H_n(t))$$

слабо конвергира у $D(-\infty, \infty)$ кад $n \to \infty$ неком центрираном Гаусовом процесу $\eta(t)$ с нултим математичким очекивањем и коваријацијом која се може израчунати на основу Силверманове теореме. Међутим, статистика D_n конвергира у расподели случајној величини $\sup_t |\eta(t)|$ чију расподелу не можемо експлицитно одредити.

Без губљења општости можемо сматрати велике вредности статистика интегралног Колмогоровљевог типа значајним за одбацивање нулте хипотезе. Низ статистика I_n неће бити постојан против сваке алтернативе, међутим, у пракси он јесте постојан против великог броја стандардних алтернатива па је стога применљив. Низ статистика D_n наравно је постојан против произвољне алтернативе.

Поред наведених постоје и други могући типови тест статистика. На пример, један од њих је и ω^2 тип статистике,

$$W_n = \int_0^\infty (G_n(t) - H_n(t))^2 \, dF_n(t).$$

Међутим, за овај, а и за неке друге типове тест статистика, асимптотска теорија није још увек довољно развијена.

Поглавље 4

Бахадурова ефикасност

4.1 Асимптотска релативна ефикасност

Избор одговарајућег теста у некој практичној ситуацији један је од најзначајнијих задатака статистичара. Сходно томе неопходно је дефинисати неке критеријуме на основу којих се може извршити упоређивање тестова.

Да бисмо дефинисали неке од тих критеријума увешћемо формално неколико појмова теорије ефикасности тестова. Опширније о овоме може се наћи у монографији Никитина [57].

Нека је случајна величина X дефинисана на параметарском простору вероватноћа $(\Omega, \mathcal{A}, P_\theta)$, $\theta \in \Theta$ с одговарајућим узорачким простором. Нека је $\mathbf{X} = X_1, \dots, X_n$ узорак. Нулта хипотеза је $H_0(\theta \in \Theta_0 \subset \Theta)$ а алтернативна $H_1(\theta \in \Theta_1 = \Theta \setminus \Theta_0)$.

Нека је T_n низ статистика које ћемо користити у тестирању. Претпостављамо без губљења општости да је критична област облика $W = \{T_n \geq c\}$, где је c позитивна реална константа. Функција моћи теста дефинисана је са $M(\theta) = P_\theta\{T_n \geq c\}$, а моћ теста са $\sup_{\theta \in \Theta_1} P_\theta\{T_n \geq c\}$. За свако $\beta \in (0,1)$ и $\theta \in \Theta_1$ дефинишемо низ константи $c_n(\beta, \theta)$ такав да важи

$$P_\theta\{T_n > c_n\} \leq \beta \leq P_\theta\{T_n \geq c_n\}.$$

Нека је

$$\alpha_n(\beta, \theta) = \sup_{\theta' \in \Theta_0} P_{\theta'}\{T_n \geq c_n\}$$

минимални ниво значајности теста с низом тест статистика T_n за који је $M(\theta) \geq \beta$. Нека је $N_T(\alpha, \beta, \theta)$ минимални обим узорка за који ће тест с низом тест статистика T_n, за фиксирани ниво значајности α имати моћ теста у тачки θ не мању од β, тј.

$$N_T(\alpha, \beta, \theta) = \min\{n : \alpha_m(\beta, \theta) \geq \alpha, \text{за свако } m \geq n\},$$

где је α ниво значајности теста.

Сада можемо дефинисати критеријум поређења два теста. За две тест статистике T_n и V_n које тестирају исту нулту против исте алтернативне хипотезе дефинишемо релативну ефикасност $e_{V,T}$ као количник минималних обима узорака потребних да се достигне одговарајућа моћ теста, односно

$$e_{V,T}(\alpha, \beta, \theta) = \frac{N_T(\alpha, \beta, \theta)}{N_V(\alpha, \beta, \theta)}.$$

Уколико је релативна ефикасност мања од 1 закључујемо да ако користимо статистику T_n биће нам потребан мањи узорак да достигнемо одговарајућу моћ теста па је тест заснован на овој статистици бољи него онај заснован на V_n.

Проблем с овим критеријумом је тај што је израчунавање вредности $e_{V,T}$ у пракси исувише компликовано или чак немогуће. Из овог разлога о квалитету тестова заснованим на поменутим статистикама доносимо закључке на основу такозване асимптотске релативне ефикасности (АРЕ).

У литератури срећемо три основна типа асимптотске релативне ефикасности у зависности од тога по ком од параметара α, β или θ имамо асимптотско понашање. То су:

1. Бахадурова APE која мери релативну ефикасност тестова кад праг значајности тежи нули, тј.

$$e^{B}_{V,T}(\beta,\theta)=\lim_{\alpha\to 0} e_{V,T}(\alpha,\beta,\theta),\ \ \beta\in(0,1),\theta\in\Theta_1;$$

2. Хаџис-Леманова APE која мери релативну ефикасност тестова кад моћ теста тежи јединици, тј.

$$e^{HL}_{V,T}(\alpha,\theta)=\lim_{\beta\to 1} e_{V,T}(\alpha,\beta,\theta),\ \ \alpha\in(0,1),\theta\in\Theta_1;$$

3. Питманова APE која мери релативну ефикасност локално блиских алтернатива, тј.

$$e^{P}_{V,T}(\alpha,\beta)=\lim_{\theta\to\theta_0} e_{V,T}(\alpha,\beta,\theta),\ \ 0<\alpha<\beta<1,\theta_0\in\partial\Theta_0.$$

У сва три случаја подразумева се једнакост ако гранична вредност постоји.

У овом поглављу биће детаљније приказана основна теорија Бахадурове асимптотске ефикасности. Разлог из којег је изабран баш Бахадуров метод је због тога што је за разлику од Питмановог применљив и у случајевима када асимптотска расподела под нултом хипотезом није нормална (нпр. у случају тест статистика Колмогоровљевог типа). У случају интегралног типа статистика локалне Бахадурове и Питманове ефикасности се поклапају ([11], [86]).

4.2 Бахадурова ефикасност и теорија великих одступања

Понашање тест статистике под нултом хипотезом разматрали смо у претходном поглављу. Проучавање понашања под блиским алтернативама могуће је захваљујући развоју теорије великих одступања[1]. Пред-

[1] енгл. large deviation theory

мет њеног изучавања (в. [13]) су граничне вредности облика

$$n^{-1} \ln P_n \to -f,$$

где је P_n низ вероватноћа који експоненцијално тежи нули.

Нека је F_{T_n} функција расподеле тест статистике T_n, тј.

$$F_{T_n}(t, \theta) = P_\theta(T_n(\mathbf{X}) < t)$$

и нека је

$$G_{T_n}(t) = \inf_{\theta \in \Theta_0} \{F_{T_n}(t, \theta)\}.$$

Тада је p-вредност теста

$$L_{T_n}(\mathbf{X}) = 1 - G_{T_n}(T_n(\mathbf{X})).$$

Може се показати да под нултом хипотезом L_{T_n} има приближно униформну расподелу на сегменту $[0, 1]$ и да је $P_\theta\{L_n \leq u\} \leq u$ за $u \in [0, 1]$. Уколико имамо непрекидну расподелу тада важи једнакост.

Ако важи следећа скоро сигурна конвергенција (тј. постоји гранична вредност)

$$\frac{1}{n} \ln L_{T_n} \overset{\text{с.с.}}{\to} -\frac{1}{2} c_T(\theta), \ \theta \in \Theta_1 \ \ n \to \infty \tag{4.1}$$

где је $c_T(\theta)$ функција параметра θ, онда се $c_T(\theta)$ назива Бахадуров тачан нагиб[2]. Следећа теорема ([12], [24]) даје нам асимптотско понашање обима узорка када ниво значајности тежи нули.

Теорема 4.2.1 (Бахадур, 1967; Гренебом и Шорак, 1981)**.** *Претпоставимо да за низ тест статистика* $\{T_n\}$ *важи* (4.1) *и да је* $c_T(\theta) > 0$ *за* $\theta \in \Theta_1$. *Тада је*

$$N_T(\alpha, \beta, \theta) \sim -\frac{2 \ln(\alpha)}{c_T(\theta)}, \quad \alpha \to 0.$$

Директна последице ове теореме је да ако важи (4.1) Бахадурову

[2]енгл. Bahadur exact slope

АРЕ низова тест статистика $\{T_n\}$ и $\{V_n\}$ можемо одредити као количник Бахадурових тачних нагиба тј.

$$e^B_{V,T} = \frac{c_V(\theta)}{c_T(\theta)}.$$

Теорема 4.2.2 (Бахадур, 1967; 1971)**.** *Претпоставимо да T_n конвергира у вероватноћи*

$$T_n \xrightarrow{p} b(\theta), \ \ \theta \in \Theta_1, \ \ |b(\theta)| < \infty$$

и да је

$$\lim_{n\to\infty} \frac{\ln(1 - G_{T_n}(t))}{n} = -f(t), \ \ t \in I,$$

где је I неки интервал на коме је f непрекидна и $\{b(\theta), \ \theta \in \Theta_1\} \subset I$. Тада важи (4.1) *и*

$$c_T(\theta) = 2f(b(\theta)). \tag{4.2}$$

Докази ових теорема могу се наћи у [57].

Следи неколико тврђења која нам омогућавају да одредимо функцију f из теореме 4.2.2 за случај тест статистика дефинисаних у претходном поглављу.

За случај статистике интегралног типа (3.5) наводимо следећу теорему наведену у [61] (сродне резултате имамо и у [19] и [59]):

Теорема 4.2.3 (Никитин и Поникаров, 1999)**.** *Нека је језгро Φ U-статистике*

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} \Phi(X_{i_1}, \ldots, X_{i_m})$$

ограничена функција на $[0,1]^m$. Нека је поред тога $E\Phi = 0$ и Φ има ранг 1, тј.

$$\sigma^2_\Phi = E\phi^2(X_1) > 0,$$

где је $\phi(s_1) = E(\Phi(X_1, \ldots, X_m)|X_1 = s_1)$ пројекција језгра Φ на X_1. Тада за сваки низ реалних бројева γ_n који тежи нули кад $n \to \infty$ имамо

$$\lim_{n\to\infty}\frac{1}{n}\ln P\{U_n \geq a+\gamma_n\} = \sum_{j=2}^{\infty} b_j a^j,$$

где ред на десној страни једнакости конвергира за довољно мало $a > 0$. *Поред тога,* $b_2 = -1/(2m^2\sigma_\Phi^2)$.

За мале вредности a функција f из теореме 4.2.2 може се представити као

$$f(a) = \frac{1}{2m^2\sigma^2}a^2 + o(a^2). \tag{4.3}$$

Теорема која нам даје велика одступања у случају статистике Колмогоровљевог типа (3.6) може се наћи у [59]

Теорема 4.2.4 (Никитин, 2010)**.** *Претпоставимо да је недегенерисана фамилија статистика* $\{U_n(t)\}, t \in [a,b]$ *с ограниченим и центрираним језгрима* $\Phi(x_1,\dots,x_m;t)$ *и пројекцијама* $\phi(s_1;t) = E(\Phi(X_1,\dots,X_m;t)|X_1 = s_1)$ *за свако* t *и да задовољава услов монотоности по параметру. Тада важи*

$$\lim_{n\to\infty} n^{-1}\ln P\{\sup_t U_n(t) > \varepsilon\} = -g_T(\varepsilon,\Phi),$$

где је

$$g_T(\varepsilon) = \frac{\varepsilon^2}{2m^2\sigma_0^2} + O(\varepsilon^3), \ \ \varepsilon \to 0,$$

а $\sigma_0^2 = \sup_t E\phi^2(X_1;t)$.

У даљем тексту биће нам потребно да дефинишемо Кулбак-Лајблерово растојање између две расподеле.

$$K(\theta,\theta') := K(P_\theta, P_{\theta'}) = \begin{cases} \int \ln \frac{dP_\theta}{dP_{\theta'}} dP_\theta, & \text{ако је } P_\theta \ll P_{\theta'}; \\ \infty, & \text{иначе.} \end{cases}$$

Може се показати да је $K(\theta,\theta') \geq 0$, као и да је $K(\theta,\theta') = 0$ ако и само ако се мере P_θ и $P_{\theta'}$ поклапају на одговарајућој σ-алгебри. Означимо са

$K(\theta, \Theta_0)$ растојање расподеле P_θ од фамилије расподела $\{P_{\theta_0}, \theta_0 \in \Theta_0\}$, тј. нека је $K(\theta, \Theta_0) = \inf_{\theta_0 \in \Theta_0} K(\theta, \theta_0)$. Важи следећа теорема ([70], [13]).

Теорема 4.2.5 (Рагавачари, 1970; Бахадур, 1971)**.** *За свако* $\theta \in \Theta_1$ *скоро сигурно по* P_θ

$$\varliminf_{n \to \infty} \frac{1}{n} \ln L_n(X^{(n)}) \geq -K(\theta, \Theta_0).$$

Уколико за низ статистика $\{T_n\}$ важи (4.1), коришћењем ове теореме добијамо горњу границу Бахадуровог нагиба, тј.

$$c_T(\theta) \leq 2K(\theta, \Theta_0). \tag{4.4}$$

Двоструко Кулбак-Лајблерово растојање због ове неједнакости назива се горња граница Бахадурових тачних нагиба и у неком смислу има улогу сличну доњој граници Рао-Крамера за дисперзије непристрасних оцена.

Сада је природно дефинисати (апсолутну) Бахадурову ефикасност теста за конкретну алтернативу као количник Бахадуровог нагиба и Кулбак-Лајблерове горње границе

$$e_T(\theta) = \frac{c_T(\theta)}{2K(\theta, \Theta_0)}.$$

У већини случајева Бахадурову ефикасност није могуће изрчунати за сваку вредност параметра θ. Међутим, могуће је израчунати граничну вредност Бахадурове ефикасности

$$e_T = \lim_{\theta \to \theta_0} \frac{c_T(\theta)}{2K(\theta, \Theta_0)}, \quad \theta_0 \in \partial\Theta_0, \tag{4.5}$$

коју називамо *локалном Бахадуровом ефикасношћу.*

4.3 Локална Бахадурова ефикасност код тестова сагласности

Код теста сагласности нулта хипотеза је да расподела припада некој класи $\mathcal{K}$, а алтернативна да јој не припада. Нека је $G(x; \theta)$, $x \in [a, b]$

фамилија расподела таква да је $G(x;0)$ припада класи $\mathcal{K}$ и $G(x;\theta) \notin \mathcal{K}$ за $\theta \neq 0$. Тада можемо нашу нулту хипотезу написати у облику $H_0 : \theta = 0$. За блиске алтернативе, локална асимптотска Бахадурова ефикасност тада постаје

$$e_T = \lim_{\theta \to 0} \frac{c_T(\theta)}{2K(\theta)}. \tag{4.6}$$

Нека је $\mathcal{G} = \{G(x;\theta)\}$ класа алтернатива које задовољавају следеће услове регуларности

- постоје парцијални изводи $g(x,\theta)$ и $G'_\theta(x,\theta)$;
- $g(x,\theta)$ је апсолутно непрекидна функција параметра θ скоро свуда и $g(x,0) > 0$;
- $G'_\theta(x,\theta)$ је апсолутно непрекидна функција за скоро свако $\theta \in \Theta$, $G'_\theta(x,0)$ није идентички једнако нули и

$$\sup_x \sup_\theta |G'_\theta(x,\theta)| \leq M,$$

$$\lim_{x \downarrow a} G'_\theta(x,\theta) = \lim_{x \uparrow b} G'_\theta(x,\theta) = 0$$

- за скоро свако $x \in [a,b]$

$$\frac{\partial^2 G(x,\theta)}{\partial x \partial \theta}|_{\theta=0} = \frac{\partial^2 G(x,\theta)}{\partial \theta \partial x}|_{\theta=0},$$

Нека је $g(x;\theta)$ густина расподеле која припада $\mathcal{G}$, и нека је $h(x) = g'_\theta(x;0)$. Лако се види да је $\int_a^b h(x)dx = 0$.

За тест статистику U_n која је U-статистика с недегенерисаним језгром $\Phi(X_1,\ldots,X_m)$ може се показати да се њен лимес у вероватноћи под алтернативном хипотезом $b_U(\theta)$ може представити као

$$b_U(\theta) = m\theta \int_a^b \phi(x)h(x)dx + o(\theta), \quad \theta \to 0. \tag{4.7}$$

У случајевима регуларних расподела код просте нулте хипотезе Кулбак-Лајблерово растојање $K(\theta)$ испуњава следећу асимптотску релацију ([13], [57])

$$K(\theta) = \frac{1}{2}I(g)\theta^2 + o(\theta^2), \ \ \theta \to 0, \tag{4.8}$$

где је $I(g) \in (0, \infty)$ информациона функција Фишера, тј.

$$I(g) = \int_a^b \frac{h^2(x)}{g(x,0)}dx.$$

Узимајући у обзир теорему 4.2.3 о великим одступањима недегерисане U-статистике, као и (4.7) и (4.8), локална Бахадурова ефикасност из (4.6), у случају недегенерисане U-статистике постаје

$$e_U = \frac{\left(\int_a^b \phi(x)h(x)dx\right)^2}{\sigma^2 I(g)}.$$

4.4 Проблем локално оптималих алтернатива

Један од важних задатака асимптотске ефикасности статистичких тестова је да се за дату нулту и алтернативну хипотезу одреди тест статистика чија је ефикасност оптимална. Асимптотски оптималне тест статистике у Бахадуровом смислу су алтернативе код којих је ефикасност једнака јединици. Међутим таквих статистика је јако мало.

Постоји и други аспект гледања на исти проблем. Наиме, за дату тест статистику треба одредити скуп алтернативних расподела за које је тест оптималан. Како нам је од велике важности да тест разликује блиске алтернативе, одређујемо локално оптималне алтернативе, тј. оне за које је локална Бахадурова ефикасност максимална, тј. да важи релација

$$c_T(\theta) \sim 2K(\theta), \ \ \theta \to 0. \tag{4.9}$$

Такве алтернативе чине такозвани домен локално асимптотских оптималних (ЛАО) алтернатива. Важност ове проблематике први је истакао Бахадур [12]. Детаљније проучавање започето је у раду Никитина [56] и даље развијено у [57].

Идеја начина његовог одређивања у општем случају представљена је у [57]. Означимо са $\mathbf{V}$ скуп реалних, непрекидних функција дефинисаних на $[0,1]$. Назваћемо га водећим скупом за низ тест статистика и фамилију $\mathcal{G}$ уколико је услов локалне оптималности испуњен ако и само ако је

$$G'_\theta(G^{-1}(x;0);0) \in \mathbf{V}. \tag{4.10}$$

Елементе скупа $\mathbf{V}$ назваћемо водећим функцијама и означити са v па је услов (4.10) еквивалентан са

$$G'_\theta(x;0) = v(G(x;0)).$$

Због услова регуларности $G(x;\theta)$ можемо представити у облику

$$G(x,\theta) = G(x,0) + v(G(x,0))\theta + o(\theta), \ \ \theta \to 0.$$

Сада се проблем налажења фамилије локално оптималних алтернатива у Бахадуровом смислу своди на решавање диференцијалне једначине

$$G'_\theta(x;\theta) = v(G(x,0)), \ \ v \in \mathbf{V}.$$

Вратимо се сада случају када је тест статистика недегенерисана U-статистика за који постоји једноставнији начин решавања овог проблема. Нека је $g(x;\theta)$ густина из претходно дефинисане класе $\mathcal{G}$ и нека важи (4.8). Тада се (4.9) може записати као

$$\Big(\int\limits_a^b \phi(x)h(x)dx\Big)^2 = \sigma^2 I(g),$$

односно

$$\Big(\int_a^b \phi(x)h(x)dx\Big)^2 = \int_a^b \phi^2(x)g(x;0)dx \int_a^b \frac{h^2(x)}{g(x,0)}dx.$$

Из неједнакости Коши-Шварца видимо да се ова једнакост достиже уколико је $h(x) = C \cdot \phi(x)g(x,0)$ за неку позитивну константу C, па се решавањем диференцијалне једначине

$$\frac{d\ln g(x,\theta)}{d\theta}|_{\theta=0} = C\phi(x)$$

одређују ЛАО алтернативе. Може се показати да је у овом случају (в. [60]) водећи скуп једнак $v_U(x) = C\int_0^x \phi(G^{-1}(u))du$.

Поглавље 5

Тестови сагласности с Паретовом расподелом

У овом, као и у следећем поглављу, биће представљени резултати аутора на пољу тестова сагласности заснованим на карактеризацијама и њихове Бахадурове ефикасности. Изложени су у радовима [66], [64] и [39].

Иако због важности Паретове расподеле имамо велики број тестова сагласности (в. нпр. [26],[72]), напомињемо да су у раду [66] предложени први тестови сагласности с Паретовом расподелом засновани на карактеризацијама једнако расподељених функција који се појављују у литератури. Тестови ове врсте затим су проучавани и у [82].

Представићемо шест тестова сагласности с Паретовом расподелом. Они су засновани на карактеризацијама наведеним у теоремама 1.1.6, 1.1.7 и 1.1.8. За сваку од карактеризација формирамо и проучавамо два теста, по један тест интегралног и Колмогоровљевог типа.

Нека је $X_1, X_2, \dots, X_n$ узорак из ненегативне непрекидне расподеле F. Пошто ниједна од три карактеризације не зависи од параметра α, можемо тестирати сложену нулту хипотезу да је узорак из фамилије Паретових расподела (1.4), тј. $H_0 : F \in \mathcal{P}$, против опште алтернативе $H_1 : F \notin \mathcal{P}$ с истим носачем $[1, \infty)$.

Прва два теста која ћемо нешто опширније размотрити засновани су на карактеризационој теореми 1.1.6 (у даљем тексту карактеризација $\mathcal{A}$).

Тест статистике које ћемо користити су

$$I_n^{[\mathcal{A}]} = \int_1^\infty (M_n(t) - F_n(t)) dF_n(t)$$

и

$$D_n^{[\mathcal{A}]} = \sup_{t \geq 1} |M_n(t) - F_n(t)|,$$

где је $F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}\{X_i \leq t\}$ емпиријска функција расподеле узорка $X_1, X_2, \ldots, X_n$ а $M_n(t)$ је U-емпиријска функција расподеле заснована на карактеризационој теореми 1.1.6 дефинисана са

$$M_n(t) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{I}\left\{\max\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) \leq t\right\}, \quad t \geq 1.$$

Наредни тестови засновани су на специјалним случајевима теорема 1.1.7 и 1.1.8 за узорак обима три (у даљем тексту карактеризације $\mathcal{B}$ и $\mathcal{C}$). Прегледности ради дајемо исказ теорема и у овим случајевима.

Теорема 5.0.1. *Нека је X_1, X_2, X_3 узорак из ненегативне расподеле. Ако статистике $X_{(2;3)}/X_{(1;3)}$ и $X_{(1;2)}$ имају исту расподелу, тада X_1 има Паретову расподелу.*

Теорема 5.0.2. *Нека су X_1, X_2 и X_3 једнако расподељене ненегативне случајне величине са строго монотоном функцијом расподеле и монотоно растућом или опадајућом хазардном функцијом. Тада, $X_{(3;3)}/X_{(2;3)}$ и $(X_{(2;3)}/X_{(1;3)})^2$ имају исту расподелу ако и само ако X_1 има Паретову расподелу.*

Разлог због ког су узети баш ови специјални случајеви је тај што су најједноставнији и стога најпогоднији за практичну примену тестова.

Наведене две карактеризације веома су сличне у смислу да су обе засноване на неким количницима статистика поретка у узорку обима три. Због тога ћемо их разматрати заједно.

На основу карактеризације $\mathcal{B}$ уводимо следеће V-емпиријске функције расподеле:

$$G_n(t) = n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} I\{X_{(2),X_i,X_j,X_k}/X_{(1),X_i,X_j,X_k} \leq t\}, t \geq 1$$

и

$$H_n(t) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} I\{\min\{X_i, X_j\} \leq t\}, t \geq 1,$$

где су $X_{(l),X_a,X_b,X_c}$, $l = 1, 2$ статистике поретка реда l у узорку (X_a, X_b, X_c).

Сада уводимо тест статистике

$$I_n^{[\mathcal{B}]} = \int_1^{\infty} (G_n(t) - H_n(t))dF_n(t) \tag{5.1}$$

и

$$D_n^{[\mathcal{B}]} = \sup_{t\geq 1} |G_n(t) - H_n(t)|. \tag{5.2}$$

На основу карактеризације $\mathcal{C}$ аналогно уводимо следеће V-емпиријске функције расподеле

$$J_n(t) = n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} I\{X_{(3),X_i,X_j,X_k}/X_{(2),X_i,X_j,X_k} \leq t\}, t \geq 1$$

и

$$K_n(t) = n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} I\{(X_{(2),X_i,X_j,X_k}/X_{(1),X_i,X_j,X_k})^2 \leq t\}, t \geq 1,$$

као и одговарајуће тест статистике:

$$I_n^{[\mathcal{C}]} = \int_1^{\infty} (J_n(t) - K_n(t))dF_n(t) \tag{5.3}$$

$$D_n^{[\mathcal{C}]} = \sup_{t\geq 1} |J_n(t) - K_n(t)|. \tag{5.4}$$

Без губљења општости можемо претпоставити да су велике вредности тест статистика значајне.

5.1 Статистика интегралног типа $I_n^{[\mathcal{A}]}$

Прелазимо на асимптотске особине статистике интегралног типа $I_n^{[\mathcal{A}]}$. Можемо претпоставити без губљења општости да је параметар Паретове расподеле $\alpha = 1$. Користећи асимптотску еквивалентност $I_n^{[\mathcal{A}]}$ и U-статистике с језгром

$$\begin{aligned}\Upsilon_{\mathcal{A}}(X,Y,Z) \;&=\; \frac{1}{3}\,\mathrm{I}\left\{\max\left(\frac{X}{Y},\frac{Y}{X}\right)\leq Z\right\} + \frac{1}{3}\,\mathrm{I}\left\{\max\left(\frac{X}{Z},\frac{Z}{X}\right)\leq Y\right\}\\ &+\; \frac{1}{3}\,\mathrm{I}\left\{\max\left(\frac{Y}{Z},\frac{Z}{Y}\right)\leq X\right\} - \frac{1}{2},\end{aligned}$$

на основу закона великих бројева за U-статистике (теорема 3.2.1) добијамо

$$I_n^{[\mathcal{A}]} \xrightarrow[n\to\infty]{p} P\left\{\max\left(\frac{X}{Y},\frac{Y}{X}\right)\leq Z\right\} - \frac{1}{2}. \tag{5.5}$$

Сада ћемо испитати асимптотско понашање $I_n^{[\mathcal{A}]}$ под нултом хипотезом користећи горе дефинисане U-статистике.

Нека је $\upsilon_{\mathcal{A}}(X)$ пројекција $\Upsilon_{\mathcal{A}}(X,Y,Z)$ на X. Тада је

$$\begin{aligned}
\upsilon_{\mathcal{A}}(s) &= E(\Upsilon_{\mathcal{A}}(X,Y,Z)|X=s)\\
&= \frac{2}{3}P\left\{\max\left(\frac{s}{Y},\frac{Y}{s}\right)\leq Z\right\}+\frac{1}{3}P\left\{\max\left(\frac{Y}{Z},\frac{Z}{Y}\right)\leq s\right\}-\frac{1}{2}\\
&= \frac{2}{3}(\int\limits_{1}^{s}dz\int\limits_{\frac{s}{z}}^{\infty}z^{-2}y^{-2}dy+\int\limits_{s}^{\infty}dz\int\limits_{\frac{z}{s}}^{\infty}z^{-2}y^{-2}dy)+\frac{1}{3}P\left\{Y\leq s\right\}-\frac{1}{2}\\
&= \frac{2}{3}\frac{2\ln s+1}{2s}+\frac{1}{3}(1-s^{-1})-\frac{1}{2}\\
&= \frac{2}{3}\frac{\ln s}{s}-\frac{1}{6}.
\end{aligned}$$

Лако је показати да је $E(\upsilon_{\mathcal{A}}(X))=0$, па је дисперзија ове пројекције

$$\sigma^2_{\mathcal{A}}=\mathrm{D}(\upsilon_{\mathcal{A}}(X))=\int\limits_{1}^{\infty}(\upsilon_{\mathcal{A}})^2(s)s^{-2}ds=\frac{5}{972}. \tag{5.6}$$

Пошто је дисперзија ове пројекције позитивна, језгро $\Upsilon_{\mathcal{A}}(X,Y,Z)$ је недегенерисано, па можемо применити Хефдингову теорему 3.2.2 за U-статистике с недегенерисаним језгрима. С обзиром да је степен ове U-статистике три, асимптотска дисперзија је $3^2\sigma^2_{\mathcal{A}}$, па следи да је гранична расподела тест статистике $I_n^{[\mathcal{A}]}$:

$$\sqrt{n}I_n^{[\mathcal{A}]}\overset{d}{\to}\mathcal{N}\left(0,\frac{5}{108}\right). \tag{5.7}$$

5.1.1 Локална Бахадурова ефикасност

У даљем раду ћемо израчунати локалну асимптотску Бахадурову ефикасност за неке алтернативе и наћи локално оптималне алтернативе. Нека је $\mathcal{G}=\{G(x;\theta)\}$ класа алтернатива које задовољавају услове регуларности из одељка 4.3 таква да је $G(x;0)=1-x^{-2},\ x\leq 1$. Нека је $g(x;\theta)$ густина расподеле која припада $\mathcal{G}$, и нека је $h(x)=g'_\theta(x;0)$.

Сада ћемо израчунати Бахадуров тачан нагиб тест статистике $I_n^{[\mathcal{A}]}$. Функције $f(t)$ и $b_T(\theta)$ одредићемо помоћу следећих лема.

Лема 5.1.1. *Нека је $t > 0$. За статистику $I_n^{[\mathcal{A}]}$ функција $f_{I^{[\mathcal{A}]}}(t)$ је аналитичка за довољно мало $t > 0$ и важи*

$$f_{I^{[\mathcal{A}]}}(t) = \frac{54}{5}t^2 + o(t^2), \ \ t \to 0.$$

Доказ. Пошто је језгро $\Upsilon_{\mathcal{A}}$ ограничено, центрирано и недегенерисано, из теореме 4.2.3 о великим одступањима за недегенерисане U-статистике следи тврђење леме. □

Лема 5.1.2. *За дату алтернативну густину $g(x;\theta)$ чија расподела припада $\mathcal{G}$ важи*

$$b_{I^{[\mathcal{A}]}}(\theta) = 3\theta \int\limits_1^\infty \upsilon_{\mathcal{A}}(x)h(x)dx + o(\theta), \ \theta \to 0. \tag{5.8}$$

Доказ: Из (5.5) следи да је

$$\begin{aligned}
b_{I^{[\mathcal{A}]}}(\theta) &= P\left\{\max\left(\frac{X}{Y},\frac{Y}{X}\right) \leq Z\right\} - \frac{1}{2} \\
&= 1 - P\left\{\frac{X}{Y} > Z, X > Y\right\} - P\left\{\frac{Y}{X} > Z, Y > X\right\} - \frac{1}{2} \\
&= \frac{1}{2} - 2P\left\{\frac{X}{Y} > Z\right\} \\
&= \frac{1}{2} - 2\int\limits_1^\infty dz \int\limits_1^\infty dy \int\limits_{yz}^\infty g(x;\theta)g(y;\theta)g(z;\theta)dx \\
&= \frac{1}{2} - 2\int\limits_1^\infty dz \int\limits_1^\infty (1 - G(yz;\theta))g(y;\theta)g(z;\theta)dy.
\end{aligned}$$

Први извод је

$$\begin{aligned}
(b_{I^{[\mathcal{A}]}})'(\theta) &= 2\int\limits_1^\infty dz \int\limits_1^\infty dy \int\limits_1^{yz} g'_\theta(x;\theta)g(y;\theta)g(z;\theta)dx \\
&- 4\int\limits_1^\infty dz \int\limits_1^\infty (1 - G(yz;\theta))g'_\theta(y;\theta)g(z;\theta)dy.
\end{aligned}$$

Стављајући $\theta = 0$ добијамо

$$\begin{aligned}
(b_{I[\mathcal{A}]})'(0) &= 2\int_1^{\infty} dz \int_1^{\infty} dy \int_1^{yz} h(x) y^{-2} z^{-2} dx \\
&- 4\int_1^{\infty} dz \int_1^{\infty} (yz)^{-1} h(y) z^{-2} dy \\
&= -2\int_1^{\infty} dz \int_1^{\infty} dy \int_{yz}^{\infty} h(x) y^{-2} z^{-2} dx \\
&- 4\int_1^{\infty} dz \int_1^{\infty} (yz)^{-1} h(y) z^{-2} dy \\
&= -2\int_1^{\infty} h(x) dx \int_1^{x} z^{-2} dz \int_1^{\frac{x}{z}} y^{-2} dy \\
&- 4\int_1^{\infty} h(y) y^{-1} dy \int_1^{\infty} z^{-3} dz \\
&= 2\int_1^{\infty} x^{-1} h(x) \ln x dx = 3\int_1^{\infty} \upsilon(x) h(x) dx.
\end{aligned}$$

Како је $b_{I[\mathcal{A}]}(0) = 0$, развијајући $b_{I[\mathcal{A}]}(\theta)$ у Маклоренов ред добијамо (5.8).

$\square$

У претходном поглављу речено је да је Кулбак-Лајблерово растојање блиских расподела за које важе услови регуларности приближно једнако информационој функцији Фишера. Међутим, како ће нам због сложености нулте хипотезе бити потребно растојање алтернативе од целе класе нултих расподела, овде не можемо користити израз (4.8), већ налазимо Кулбак-Лајблерову горњу границу применом следеће леме.

Лема 5.1.3. *За дату густину $g(x;\theta)$ нека је Кулбак-Лајблерово растојање*

$$K(\theta) = \inf_{\lambda>0} \int_1^{\infty} \ln \frac{g(x;\theta)}{\lambda x^{-\lambda-1}} g(x;\theta) dx \tag{5.9}$$

добро дефинисано. Тада када $\theta \to 0$

$$2K(\theta) = \theta^2 \Bigg(\int\limits_1^\infty x^2 h^2(x) dx - \Bigg(\int\limits_1^\infty h(x) \ln x dx \Bigg)^2 \Bigg) + o(\theta^2). \tag{5.10}$$

Доказ. Инфимум у (5.9) достиже се за $\lambda = \left(\int\limits_1^\infty g(x;\theta) \ln x dx \right)^{-1}$. Тада је

$$\begin{aligned} K(\theta) &= \int\limits_1^\infty g(x;\theta) \ln g(x;\theta) dx + \left(\frac{1}{\int\limits_1^\infty g(x;\theta) \ln x dx} + 1 \right) \int\limits_1^\infty g(x;\theta) \ln x dx \\ &+ \ln \int\limits_1^\infty g(x;\theta) \ln x dx \int\limits_1^\infty g(x;\theta) dx \\ &= \int\limits_1^\infty g(x;\theta) \ln g(x;\theta) dx + 1 + \int\limits_1^\infty g(x;\theta) \ln x dx + \ln \int\limits_1^\infty g(x;\theta) \ln x dx. \end{aligned}$$

Из дефиниције је очигледно да је $K(0) = 0$. Када диференцирамо $K(\theta)$ добијамо

$$K'(\theta) = \int\limits_1^\infty g'_\theta(x;\theta) \ln g(x;\theta) dx + \left(1 + \frac{1}{\int\limits_1^\infty g(x;\theta) \ln x dx} \right) \int\limits_1^\infty g'_\theta(x;\theta) \ln x dx.$$

Стављајући $\theta = 0$, добијамо

$$K'(0) = \int\limits_1^\infty h(x) \ln(x^{-2}) dx + \left(1 + \frac{1}{\int\limits_1^\infty x^{-2} \ln x dx} \right) \int\limits_1^\infty h(x) \ln x dx = 0.$$

Други извод функције $K(\theta)$ је

$$\begin{aligned}
K''(\theta) &= \int\limits_1^\infty g''_{\theta^2}(x;\theta)\ln g(x;\theta)dx + \int\limits_1^\infty (g'_\theta(x;\theta))^2(g(x;\theta))^{-1}dx \\
&+ \left(1 + \frac{1}{\int\limits_1^\infty g(x;\theta)\ln x dx}\right)\int\limits_1^\infty g''_{\theta^2}(x;\theta)\ln x dx \\
&- \frac{\left(\int\limits_1^\infty g'_\theta(x;\theta)\ln x dx\right)^2}{\left(\int\limits_1^\infty g(x;\theta)\ln x dx\right)^2}.
\end{aligned}$$

Стављајући $\theta = 0$ добијамо

$$\begin{aligned}
K''(0) &= \int\limits_1^\infty g''_{\theta^2}(x;0)\ln(x^{-2})dx + \int\limits_1^\infty h^2(x)x^{-2})^{-1}dx \\
&+ \left(1 + \frac{1}{\int\limits_1^\infty x^{-2}\ln x dx}\right)\int\limits_1^\infty g''_{\theta^2}(x;0)\ln x dx \\
&- \frac{\left(\int\limits_1^\infty h(x)\ln x dx\right)^2}{\left(\int\limits_1^\infty x^{-2}\ln x dx\right)^2} \\
&= \int\limits_1^\infty x^2h^2(x)dx - \left(\int\limits_1^\infty h(x)\ln x dx\right)^2.
\end{aligned}$$

Из Маклореновог развоја функције $K(\theta)$ следи (5.10). □

На основу наведених теорема можемо израчунати локалне Бахадурове ефикасности у случају блиских алтернатива. Алтернативне расподеле које ће бити разматране су следеће:

- лог-Вејбулова расподела с густином

$$g_1(x;\theta) = (1+\theta)x^{-1}(\ln x)^{\theta}e^{-(\ln x)^{1+\theta}}, \ \ x \geq 1, \theta \in (0,1), \tag{5.11}$$

- лог-гама расподела с густином

$$g_2(x;\theta) = \frac{(\ln x)^{\theta}}{x^2\Gamma(1+\theta)}, \ \ x \geq 1, \theta \in (0,1), \tag{5.12}$$

- расподела с густином

$$g_3(x;\beta,\theta) = \frac{1}{x^2}(e^{-\theta(\ln x)^{\beta}} + \theta\beta(\ln x)^{\beta-1}e^{-\theta(\ln x)^{\beta}}), \ \ x \geq 1, \ \theta \in (0,1) \tag{5.13}$$

  за вредности параметра $\beta = 1.5$ и $\beta = 2$,

- инверзна бета расподела с густином

$$g_4(x;\theta) = \frac{1+\theta}{x^2}\Big(1-\frac{1}{x}\Big)^{\theta}, \ \ x \geq 1, \theta \in (0,1), \tag{5.14}$$

- Паретова расподела с такозваним "тилт" параметром (в. [52]) с густином

$$g_5(x;\theta) = \frac{1+\theta}{(x+\theta)^2}, \ \ x \geq 1, \theta \in (0,1). \tag{5.15}$$

Пример 5.1.1. *Нека је алтернативна расподела лог-Вејбулова* (5.11). *Први извод њене густине по θ за $\theta = 0$ је*

$$h(x) = \frac{1}{x^2}(-\ln x \ln\ln x + \ln\ln x + 1).$$

Користећи лему 5.1.3 добијамо

$$\begin{aligned} 2K_{g_1}(\theta) &= \theta^2\Big(\int\limits_1^{\infty}\frac{1}{x^2}((1-\ln x)\ln\ln x + 1)^2 dx \\ &- (\int\limits_1^{\infty}\frac{1}{x^2}((1-\ln x)\ln\ln x + 1)\ln x dx)^2\Big) + o(\theta^2), \theta \to 0, \end{aligned}$$

а применом леме 5.1.2 следи да је

$$b_{I[\mathcal{A}]}(\theta) = 2\theta \int\limits_1^\infty \frac{\ln x}{x^3}((1-\ln x)\ln\ln x + 1)dx + o(\theta), \theta \to 0.$$

Након израчунавања ових интеграла преко математичих очекивања логаритама гама расподеле и коришћењем особине дигама функције $\psi(x+1) = \psi(x) + \frac{1}{x}$, *добијамо*

$$2K_{g_1}(\theta) = \theta^2 \psi'(1) \tag{5.16}$$

и

$$b_{I[\mathcal{A}]}(\theta) = \frac{\theta}{4}.$$

Из (4.2) *и* (4.5), *применом леме 5.1.1, добијамо да је локална асимптотска Бахадурова ефикасност*

$$e_{I[\mathcal{A}]} = \frac{27}{20\psi'(1)} \approx 0.821.$$

Пример 5.1.2. *Нека је друга алтернатива с густином расподеле* g_3. *Пошто је*

$$h(x) = \frac{1}{x^2}(\beta \ln^{\beta-1} x - \ln^\beta x),$$

Кулбак-Лајблерова граница постаје

$$\begin{aligned} 2K_{g_3}(\theta) &= \theta^2 \Bigg(\int\limits_1^\infty \frac{1}{x^2}(\beta \ln^{\beta-1} x - \ln^\beta x)^2 dx \\ &- (\int\limits_1^\infty \ln x \frac{1}{x^2}(\beta \ln^{\beta-1} x - \ln^\beta x) dx)^2 \Bigg) + o(\theta^2) \\ &= \theta^2 \beta^2 \Gamma(2\beta - 1) - \Gamma^2(\beta+1) + o(\theta^2), \theta \to 0, \end{aligned}$$

и

$$\begin{aligned} b_{I^{[\mathcal{A}]}}(\theta) &= 2\theta\int_1^{\infty}\frac{\ln x}{x^3}(\beta\ln^{\beta-1}x-\ln^{\beta}x)dx+o(\theta)\\ &= \theta\frac{\Gamma(\beta+1)(\beta-1)}{2^{\beta+1}}+o(\theta),\theta\to 0. \end{aligned}$$

Локална асимптотска Бахадурова ефикасност је

$$e_{I^{[\mathcal{A}]}}=\frac{108}{5}\frac{\Gamma^2(\beta+1)(\beta-1)^2}{2^{2\beta+2}(\beta^2\Gamma(2\beta-1)-\Gamma^2(\beta+1))}.$$

За $\beta=2$ *ефикасност је* $\frac{27}{80}\approx 0.34$. *За* $\beta=1.5$ *добијамо* $\frac{27\pi}{160(4-\pi)}\approx 0.62$, *а за* $\beta=2$ *вредност је* 0.34. *Највећа ефикасност достиже се када* β *тежи јединици и гранична вредност је* $\frac{27}{20\xi'(1)}\approx 0.82$.

За остале алтернативе рачун је аналоган. Вредности Бахадурове ефикасности приказане су у табели 5.1.

алт.	g_1	g_2	$g_3(1.5)$	$g_3(2)$	g_4	g_5
$e_{I^{\mathcal{A}}}$	0.821	0.788	0.618	0.338	0.777	0.800

Табела 5.1: Бахадурова ефикасност за статистику $I_n^{[\mathcal{A}]}$

5.1.2 Локално оптималне алтернативе

Како нам због сложености нулте хипотезе не важи (4.8), не можемо директно искористити резултат из одељка 4.4, већ ћемо их применом сличног поступка одредити помоћу следеће теореме.

Теорема 5.1.1. *Нека је* $g(x;\theta)$ *густина из* $\mathcal{G}$ *која уз то задовољава и услов*

$$\int_1^{\infty}x^2h^2(x)dx<\infty.$$

Алтернативне густине

$$g(x;\theta) = \frac{1}{x^2} + \theta(C\frac{\upsilon_{\mathcal{A}}(x)}{x^2} + D\frac{\ln x - 1}{x^2}),\ x \geq 1,\ C > 0,\ D \in \mathbb{R},$$

за мало θ су асимптотски оптималне за тест чија је тест статистика $I_n^{[\mathcal{A}]}$.

Доказ: Означимо функцију

$$h_0(x) = h(x) - \frac{\ln x - 1}{x^2} \int_1^\infty h(s) \ln s ds. \tag{5.17}$$

Може се показати да ова функција задовољава следеће једнакости

$$\int_1^\infty h_0^2(x) x^2 dx = \int_1^\infty x^2 h^2(x) dx - \left(\int_1^\infty h(x) \ln x dx \right)^2 \tag{5.18}$$

$$\int_1^\infty \upsilon_{\mathcal{A}}(x) h_0(x) dx = \int_1^\infty \upsilon_{\mathcal{A}}(x) h(x) dx. \tag{5.19}$$

Из лема 5.1.1 и 5.1.2, користећи (5.6), добијамо да је локална асимптотска ефикасност

$$
\begin{aligned}
e_{I[\mathcal{A}]} &= \lim_{\theta\to 0}\frac{c_{I[\mathcal{A}]}(\theta)}{2K(\theta)} = \lim_{\theta\to 0}\frac{2f(b_{I[\mathcal{A}]}(\theta))}{2K(\theta)} = \lim_{\theta\to 0}\frac{2\cdot\frac{54}{5}b^2_{I[\mathcal{A}]}(\theta)}{2K(\theta)} = \lim_{\theta\to 0}\frac{b^2_{I[\mathcal{A}]}(\theta)}{9\sigma^2_{\mathcal{A}}2K(\theta)}\\
&= \lim_{\theta\to 0}\frac{9\theta^2\Big(\int\limits_1^\infty \upsilon_{\mathcal{A}}(x)h(x)dx\Big)^2}{9\int\limits_1^\infty \upsilon^2_{\mathcal{A}}(x)x^{-2}dx\Big(\theta^2\Big(\int\limits_1^\infty x^2h^2(x)dx-\big(\int\limits_1^\infty h(x)\ln x dx\big)^2\Big)\Big)}\\
&= \frac{\Big(\int\limits_1^\infty \upsilon_{\mathcal{A}}(x)h(x)dx\Big)^2}{\int\limits_1^\infty \upsilon^2_{\mathcal{A}}(x)x^{-2}dx\Big(\int\limits_1^\infty x^2h^2(x)dx-\big(\int\limits_1^\infty h(x)\ln x dx\big)^2\Big)}\\
&= \frac{\Big(\int\limits_1^\infty \upsilon_{\mathcal{A}}(x)h_0(x)dx\Big)^2}{\int\limits_1^\infty \upsilon^2_{\mathcal{A}}(x)x^{-2}dx\int\limits_1^\infty h_0^2(x)x^2dx}.
\end{aligned}
$$

Из неједнакости Коши-Шварца следи да је $e_{I[\mathcal{A}]} = 1$ ако и само ако $h_0(x) = C\upsilon_{\mathcal{A}}(x)x^{-2}$. Заменом ове једнакости у (5.17) добијамо израз за $h(x)$. Пошто је $h(x)$ за наше алтернативе овог облика, тврђење је доказано. $\square$

5.2 Статистика Колмогоровљевог типа $D_n^{[\mathcal{A}]}$

Сада ћемо прећи на испитивање асимптотских својстава статистике Колмогоровљевог типа $D_n^{[\mathcal{A}]}$ под нултом хипотезом. Овде такође претпостављамо да је параметар Паретове расподеле $\alpha = 1$. За фиксирано $t \in [1,\infty)$ израз $M_n(t) - F_n(t)$ је U-статистика с језгром

$$
\Xi_{\mathcal{A}}(X,Y;t) = I\left\{\max\left(\frac{X}{Y},\frac{Y}{X}\right)\leq t\right\} - \frac{1}{2}I\left\{X\leq t\right\} - \frac{1}{2}I\left\{Y\leq t\right\}.
$$

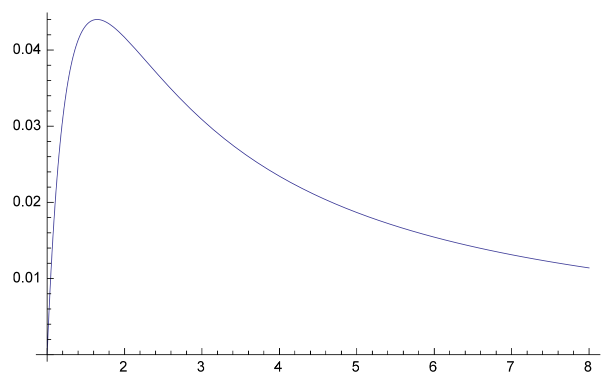
Нека је $\xi_{\mathcal{A}}(X;t)$ пројекција $\Xi_{\mathcal{A}}(X,Y;t)$ на X. Тада је

$$\begin{aligned}\xi_{\mathcal{A}}(s;t) &= E(\Xi_{\mathcal{A}}(X,Y;t)|X=s)\\ &= P\left\{\max\left(\frac{s}{Y},\frac{Y}{s}\right)\le t\right\}-\frac{1}{2}I\{s\le t\}-\frac{1}{2}P\left\{Y\le t\right\}\\ &= \frac{t}{s}-\frac{1}{st}-\frac{1}{2}+\frac{1}{2t}+I\{s\le t\}\left(\frac{1}{2}-\frac{t}{s}\right).\end{aligned}$$

Лако је показати да је математичко очекивање $\xi_{\mathcal{A}}(X;t)$ једнако нули. Дисперзија ове пројекције за фиксирано t је

$$\sigma^2_{\mathcal{A}}(t) = \mathrm{D}(\xi_{\mathcal{A}}(X;t)) = \frac{1}{12t^3}(t^2+t-2).$$

Стога је наша фамилија језгара $\Xi_{\mathcal{A}}(X,Y;t)$, недегенерисана у складу с дефиницијом 3.2.1.



Слика 5.1: График функције $\sigma^2_{\mathcal{A}}(t)$

Функција $\sigma^2(t)$ (слика 5.1) достиже максимум у тачки $t_{\mathcal{A}} = (\sqrt{7}-1)$, и тај максимум једнак је $\frac{7\sqrt{7}+10}{648} \approx 0.044$. Може се показати коришћењем метода из [76] да U-емпиријски случајни процес $\rho = \sqrt{n}(M_n(t) - F_n(t))$, $t \ge 1$, конвергира у расподели неком Гаусовом процесу. Није једноставно израчунати коваријацију овог процеса, а асимптотска расподела статистике $D_n^{[\mathcal{A}]}$ није позната.

5.2.1 Локална Бахадурова ефикасност

Сада ћемо израчунати Бахадурову ефикасност аналогно претходној тест статистици. Овде се функција $f(t)$ из (4.2) за статистику $D_n^{[\mathcal{A}]}$ одређује применом следеће теореме.

Теорема 5.2.1. *Нека је $t > 0$. Тада је $f_{D^{\mathcal{A}}}(t)$ аналитичка за мале вредности $t > 0$ и важи*

$$f_{D^{[\mathcal{A}]}}(t) = \frac{t^2}{8\sigma^2(t_{\mathcal{A}})} + o(t^2) \sim 2.84t^2, \ \ t \to 0.$$

Доказ ове теореме следи из теореме 4.2.4.

Следећом лемом налазимо $b_{D^{[\mathcal{A}]}}(\theta)$, лимес у вероватноћи статистике $D_n^{[\mathcal{A}]}$.

Лема 5.2.1. *Задату алтернативну густину $g(x;\theta)$ чија расподела припада $\mathcal{G}$ важи*

$$b_{D^{\mathcal{A}}}(\theta) = 2\theta \sup_{t\geq 1}\left|\int_1^{\infty} \xi_{\mathcal{A}}(x;t)h(x)dx\right| + o(\theta), \ \theta \to 0. \tag{5.20}$$

Доказ. Коришћењем теореме 3.3.1, добијамо

$$\begin{aligned} b_{D^{\mathcal{A}}}(\theta) &= \sup_{t\geq 1}\left|P\left\{\max\left(\frac{X}{Y}, \frac{Y}{X}\right) \leq t\right\} - G(t;\theta)\right| \\ &= \sup_{t\geq 1}\left|2\int_1^{\infty} dx \int_x^{xt} g(y;\theta)g(x;\theta)dy - \int_1^{t} g(x;\theta)dx\right|. \end{aligned} \tag{5.21}$$

Означимо

$$a_{D^{[\mathcal{A}]}}(\theta;t) = 2\int_1^{\infty} dx \int_x^{xt} g(y;\theta)g(x;\theta)dy - \int_1^{t} g(x;\theta)dx.$$

Лако се показује да је $a_{D^{[\mathcal{A}]}}(0;t) = 0$. Први извод функције $a_{D^{[\mathcal{A}]}}(\theta)$ по θ

је

$$
\begin{aligned}
a_V'(\theta;t) &= 2\int_1^\infty dx \int_x^{xt} g_\theta'(y;\theta)g(x;\theta)dy + 2\int_1^\infty dx \int_x^{xt} g(y;\theta)g_\theta'(x;\theta)dy \\
&- \int_1^t g_\theta'(x;\theta)dx.
\end{aligned}
$$

Први извод у тачки $\theta = 0$ је

$$
\begin{aligned}
a_{D[\mathcal{A}]}'(t) &= 2\int_1^\infty dx \int_x^{xt} h(y)\frac{1}{x^2}dy + 2\int_1^\infty dx \int_x^{xt} \frac{1}{y^2}h(x)dy - \int_1^t h(x)dx \\
&= \int_1^t h(x)dx + 2t\int_t^\infty h(x)x^{-1}dx - 2t^{-1}\int_1^\infty h(x)x^{-1}dx \\
&= \int_1^\infty \left[I\left\{x \le t\right\} + 2tx^{-1}\left(1 - I\left\{x \le t\right\}\right) - 2t^{-1}x^{-1}\right] h(x)dx.
\end{aligned}
$$

Применом Маклореновог развоја функције $a_{D[\mathcal{A}]}(\theta;t)$ и (5.21) добијамо (5.20). $\square$

Као и у случају интегралне статистике израчунаћемо локалне Бахадурове ефикасности за стандардне алтернативе. Наведеном списку алтернатива додаћемо и следећу:

- мешавина две Паретове расподеле с густином

$$
g_6(x;\beta,\theta) = \frac{1-\theta}{x^2} + \frac{\beta\theta}{x^{\beta+1}}, \ \ x \ge 1, \ \ \beta > 1, \ \theta \in (0,1). \tag{5.22}
$$

Разлог због ког ову алтернативу нисмо могли узети у обзир код статистике интегралног типа је тај што тест није постојан против ње, тј. лимес у вероватноћи $b_{I[\mathcal{A}]}(\theta)$, под алтернативом (5.22), је негативан.

Пример 5.2.1. *Нека је алтернативна хипотеза мешавина две Паретове расподеле* (5.22). *Први извод њене густине по* θ *у тачки* $\theta = 0$

је

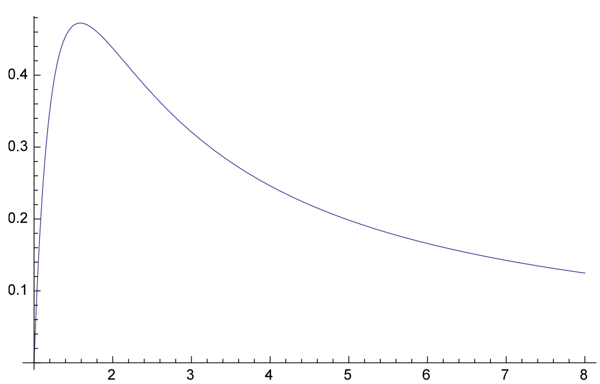
$$h(x) = -\frac{1}{x^2} + \frac{\beta}{x^{\beta+1}}.$$

Применом леме 5.1.3 добијамо

$$\begin{aligned}2K(\theta) &= \theta^2\Bigg(\Big(-1+\frac{\beta x^2}{x^{\beta+1}}\Big)^2 dx - (\int_1^\infty(-\frac{1}{x^2}+\frac{\beta}{x^{\beta+1}})\ln x dx)^2\Bigg) + o(\theta^2)\\ &= \theta^2\frac{\beta}{2} + o(\theta^2), \theta \to 0,\end{aligned}$$

а из леме 5.2.1 имамо

$$\begin{aligned}b_{D^{[\mathcal{A}]}}(\theta) &= 2\theta\sup_{t\geq 1}\left|\int_1^\infty\Big(I\{x\leq t\}(\frac{1}{2}-\frac{t}{x}) + x^{-1}(t-t^{-1})\Big)(-\frac{1}{x^2}+\frac{\beta}{x^{\beta+1}})dx\right|\\ &= \theta\sup_{t\geq 1}\left|(t^{-1}-t^{-\beta})\right| + o(\theta)\\ &= \theta\frac{(\beta-1)^2}{(1+\beta)(\beta)^{\frac{\beta}{\beta-1}}} + o(\theta),\ \theta\to 0.\end{aligned}$$



Слика 5.2: График функције $a'_{D^{\mathcal{A}}}(t)$ за мешавину расподела (5.22), $\beta = 4$

Користећи (4.2), (4.5) *и теорему 5.2.1 рачунамо локалну асимптотску Бахадурову ефикасност. Добијамо да је*

$$e_{D^{\mathcal{A}}} = \lim_{\theta\to 0}\frac{c_{D^{\mathcal{A}}}(\theta)}{2K(\theta)} = 5.68\frac{(2\beta-1)}{(1+\beta)^2(\beta)^{\frac{2}{\beta-1}}}. \tag{5.23}$$

За вредности $\beta = 1.5,\ 4$ *и* 8 *ефикасности су редом* $0.36,\ 0.63$ *и* 0.58. *Израз* (5.23) *достиже максимум за* $\beta = 4.646$ *и тада је* $e_{D^{\mathcal{A}}} \approx 0.636$.

Бахадурове ефикасности у случају осталих алтернативних расподела приказане су у табели 5.2

алт.	g_1	g_2	$g_3(1.5)$	$g_3(2)$	g_4	g_5	$g_6(1.5)$	$g_6(4)$	$g_6(8)$
$e_{D^{[\mathcal{A}]}}$	0.437	0.448	0.331	0.192	0.455	0.473	0.359	0.631	0.581

Табела 5.2: Бахадурова ефикасност за статистику $D_n^{[\mathcal{A}]}$

Видимо да су ефикасности у овом случају задовољавајуће али ипак приметно ниже него код статистике интегралног типа. То је уобичајна појава, статистике интегралног типа показује се да су ефикасније за стандардне алтернативе. Међутим, можемо запазити да постоје алтернативе против којих интегрална статистика није постојана, што даје додатни значај тесту Колмогоровљевог типа.

5.2.2 Локално оптималне алтернативе

Наравно, постоје и алтернативе за које је управо ова тест статистика оптимална. Неке од њих налазимо следећом теоремом.

Теорема 5.2.2. *Нека је* $g(x;\theta)$ *густина из* $\mathcal{G}$ *која задовољава и услов*

$$\int_1^\infty x^2 h^2(x)dx < \infty.$$

Алтернативне густине

$$g(x;\theta) = \frac{1}{x^2} + \theta\Big(C\frac{\xi_{\mathcal{A}}(x;t_{\mathcal{A}})}{x^2} + D\frac{\ln x - 1}{x^2}\Big),\ x \geq 1,\ C > 0,\ D \in \mathbb{R},$$

где је $t_{\mathcal{A}} = (\sqrt{7} - 1)$, *за мале вредности* θ *асимптотски су оптимлне за тест чија је тест статистика* $D_n^{[\mathcal{A}]}$.

Доказ: Нека је h_0 функција дефинисана у (5.17). Може се показати да поред (5.18), она такође задовољава и следећу једнакост

$$\int_1^\infty \xi_{\mathcal{A}}(x;t)h_0(x)dx = \int_1^\infty \xi_{\mathcal{A}}(x;t)h(x)dx. \tag{5.24}$$

Из теореме 5.2.1 и леме 5.2.1, добијамо да је асимптотска ефикасност једнака

$$\begin{aligned}
e_{D[\mathcal{A}]} &= \lim_{\theta\to 0}\frac{c_{D[\mathcal{A}]}(\theta)}{2K(\theta)} = \lim_{\theta\to 0}\frac{2f(b_{D[\mathcal{A}]}(\theta))}{2K(\theta)} = \lim_{\theta\to 0}\frac{b^2_{D[\mathcal{A}]}(\theta)}{4\sigma^2_{\mathcal{A}}(t_{\mathcal{A}})2K(\theta)}\\
&= \lim_{\theta\to 0}\frac{4\theta^2\sup\limits_{t\geq 1}\left(\int\limits_1^\infty \xi_{\mathcal{A}}(x;t)h(x)dx\right)^2}{4\sup\limits_{t\geq 1}\int\limits_1^\infty \xi^2_{\mathcal{A}}(x;t)x^{-2}dx\left(\theta^2\left(\int\limits_1^\infty x^2h^2(x)dx - \left(\int\limits_1^\infty h(x)\ln x dx\right)^2\right)\right)}\\
&= \frac{\sup\limits_{t\geq 1}\left(\int\limits_1^\infty \xi_{\mathcal{A}}(x;t)h(x)dx\right)^2}{\sup\limits_{t\geq 1}\int\limits_1^\infty \xi^2_{\mathcal{A}}(x;t)x^{-2}dx\left(\int\limits_1^\infty x^2h^2(x)dx - \left(\int\limits_1^\infty h(x)\ln x dx\right)^2\right)}\\
&= \frac{\sup\limits_{t\geq 1}\left(\int\limits_1^\infty \xi_{\mathcal{A}}(x;t)h_0(x)dx\right)^2}{\sup\limits_{t\geq 1}\int\limits_1^\infty \xi^2_{\mathcal{A}}(x;t)x^{-2}dx\int\limits_1^\infty h_0^2(x)x^2dx}.
\end{aligned}$$

Из Коши-Шварцове неједнакости добијамо да је $e_{D[\mathcal{A}]} = 1$ ако и само ако $h_0(x) = C\xi_{\mathcal{A}}(x;t_{\mathcal{A}})x^{-2}$. Заменом ове једнакости у (5.17) добијамо израз за $h(x)$. Пошто је $h(x)$ наших алтернатива тог облика, тврђење је доказано.

□

5.3 Статистике $I_n^{[\mathcal{B}]}$ и $I_n^{[\mathcal{C}]}$

Прелазимо на проучавање статистика интегралног типа датих израза (5.1) и (5.3). Као и у претходним случајевима можемо претпоставити без губљења општости да је $\alpha = 1$. Статистика $I_n^{[\mathcal{B}]}$ је V-статистика са следећим језгром:

$$\begin{aligned}\Upsilon_{\mathcal{B}}(X_1,X_2,X_3,X_4) &= \frac{1}{4}\sum_{\pi(j_1,j_2,j_3,j_4)} I\{X_{(2);X_{j_1},X_{j_2},X_{j_3}}/X_{(1);X_{j_1},X_{j_2},X_{j_3}} \le X_{j_4}\}\\ &- \frac{1}{12}\sum_{\pi(j_1,j_2,j_3)} I\{\min\{X_{j_1},X_{j_2}\} \le X_{j_3}\}\end{aligned}$$

где је $\pi(j_1,j_2,j_3,j_4)$ скуп свих пермутација скупа $\{1,2,3,4\}$ а $\pi(j_1,j_2,j_3)$ скуп свих 3-пермутација истог скупа.

Пројекција овог језгра под нултом хипотезом је

$$\begin{aligned}\upsilon_{\mathcal{B}}(s) &= E(\Upsilon_{\mathcal{B}}(X_1,X_2,X_3,X_4)|X_1=s)\\ &= \frac{1}{4}P\{X_{(2);X_2,X_3,X_4}/X_{(1);X_2,X_3,X_4} \le s\}\\ &+ \frac{3}{4}P\{X_{(2);s,X_2,X_3}/X_{(1);s,X_2,X_3} \le X_4\} - \frac{1}{4}P\{\min\{X_2,X_3\} \le X_4\}\\ &- \frac{1}{4}P\{\min\{X_2,X_3\} \le s\} - \frac{1}{2}P\{\min\{s,X_2\} \le X_3\}.\end{aligned}$$

Први и четврти члан израза поклапају се на основу теореме 5.0.1 па добијамо

$$\begin{aligned}&\upsilon_{\mathcal{B}}(s) = \frac{3}{4}(2P\{s \le X_2 \le X_3, \frac{X_2}{s} \le X_4\} + 2P\{X_2 \le s \le X_3, \frac{s}{X_2} \le X_4\}\\ &+ 2P\{X_2 \le X_3 \le s, \frac{X_3}{X_2} \le X_4\}) - \frac{1}{2}(1 - P\{s > X_2 > X_3\}\\ &- P\{X_2 > s > X_3\}) - \frac{1}{4}\cdot\frac{2}{3}\\ &= \frac{1}{4}\frac{3\ln s - \frac{1}{2}}{s^2} - \frac{1}{24}.\end{aligned}$$

Математичко очекивање ове пројекције једнако је нули, а дисперзија је

$$\sigma^2_{\mathcal{B}} = D(\upsilon_{\mathcal{B}}(X)) = \int_1^\infty (\upsilon_{\mathcal{B}}(s))^2\frac{1}{s^2}ds = \frac{13}{4500}. \qquad (5.25)$$

Како је дисперзија ове пројекције позитивна, језгро $\Upsilon_{\mathcal{B}}(X_1, X_2, X_3, X_4)$ је недегенерисано па применом Хефдингове теореме 3.2.2 добијамо следећу асимптотску расподелу

$$\sqrt{n} I_n^{[\mathcal{B}]} \xrightarrow{d} \mathcal{N}\left(0, \frac{52}{1125}\right).$$

Статистика $I^{[\mathcal{C}]}$ је V-статистика с језгром

$$\begin{aligned}\Upsilon_{\mathcal{C}}(X_1, X_2, X_3, X_4) = \frac{1}{4} &\sum_{\pi(j_1,j_2,j_3,j_4)} \mathrm{I}\{X_{(3);X_{j_1},X_{j_2},X_{j_3}}/X_{(2);X_{j_1},X_{j_2},X_{j_3}} \leq X_{j_4}\} \\ - \frac{1}{4} &\sum_{\pi(j_1,j_2,j_3,j_4)} \mathrm{I}\{(X_{(2);X_{j_1},X_{j_2},X_{j_3}}/X_{(1);X_{j_1},X_{j_2},X_{j_3}})^2 \leq X_{j_4}\}.\end{aligned}$$

Њена пројекција под нултом хипотезом је

$$\begin{aligned} \upsilon^{[\mathcal{C}]}(s) &= E(\Upsilon_{\mathcal{C}}(X_1, X_2, X_3, X_4)|X_1 = s) \\ &= \frac{1}{4}P\{X_{(3);X_2,X_3,X_4}/X_{(2);X_2,X_3,X_4} \leq s\} \\ &+ \frac{3}{4}P\{X_{(3);s,X_2,X_3}/X_{(2);s,X_2,X_3} \leq X_4\} \\ &- \frac{1}{4}P\{(X_{(2);X_2,X_3,X_4}/X_{(1);X_2,X_3,X_4})^2 \leq s\} \\ &- \frac{3}{4}P\{(X_{(2);s,X_2,X_3}/X_{(1);s,X_2,X_3})^2 \leq X_4\}. \end{aligned}$$

Први и трећи члан поклапају се на основу теореме 5.0.2 па се могу скратити. Добијамо

$$\begin{aligned}\upsilon^{[\mathcal{C}]}(s) &= \frac{3}{4}(2P\{s \le X_2 \le X_3, \frac{X_3}{X_2} \le X_4\} + 2P\{X_2 \le s \le X_3, \frac{X_3}{s} \le X_4\} \\ &+ 2P\{X_2 \le X_3 \le s, \frac{s}{X_3} \le X_4\}) - \frac{3}{4}(2P\{s \le X_2 \le X_3, \left(\frac{X_2}{s}\right)^2 \le X_4\} \\ &+ 2P\{X_2 \le s \le X_3, \left(\frac{s}{X_2}\right)^2 \le X_4\} + 2P\{X_2 \le X_3 \le s, \left(\frac{X_3}{X_2}\right)^2 \le X_4\}) \\ &= \frac{3}{4} \cdot \frac{2\ln s - 1}{s} + \frac{1}{s^3} - \frac{1}{4}.\end{aligned}$$

Математичко очекивање ове пројекције једнако је нули, а дисперзија је

$$\sigma^2_{\mathcal{C}} = \int\limits_1^\infty (\upsilon^{[\mathcal{C}]}(s))^2 \frac{1}{s^2} ds = \frac{19}{4200}.$$

Као и у претходном случају, језгро је недегенерисано и применом Хефдингове теореме добијамо

$$\sqrt{n} I_n^{[\mathcal{C}]} \xrightarrow{d} \mathcal{N}\left(0, \frac{38}{525}\right).$$

5.3.1 Локална Бахадурова ефикасност

Аналогно претходном, сада рачунамо локалну Бахадурову ефикасност статистика интегралног типа. За статистику $I_n^{[\mathcal{B}]}$ функције $f(t)$ и $b_T(\theta)$ налазимо применом следећих лема.

Лема 5.3.1. *Нека је $t > 0$. За статистику $I_n^{[\mathcal{B}]}$ функција $f_{I^{[\mathcal{B}]}}(t)$ је аналитичка за мало $t > 0$ и важи*

$$f_{I^{[\mathcal{B}]}}(t) = \frac{1125}{104} t^2 + o(t^2), \quad t \to 0.$$

Доказ. Пошто је језгро $\Upsilon_{\mathcal{B}}$ ограничено, центрирано и недегенерисано, применом теореме 4.2.3 добијамо тврђење леме. $\square$

Лема 5.3.2. *За алтернативну густину расподеле $g(x;\theta)$ која припада $\mathcal{G}$ важи*

$$b_{I^{[\mathcal{B}]}}(\theta) = 4\theta \int_1^\infty \upsilon_{\mathcal{B}}(x)h(x)dx + o(\theta),\ \theta \to 0. \tag{5.26}$$

Доказ. На основу закона великих бројева за U-статистике (теорема 3.2.1) добијамо

$$b_{I^{[\mathcal{B}]}}(\theta) = P_\theta\{X_{(2;3)}/X_{(1;3)} \leq X_4\} - P_\theta\{\min\{X_1, X_2\} \leq X_3\}.$$

Тада је

$$\begin{aligned}
b_{I^{[\mathcal{B}]}}(\theta) &= \int_1^\infty \int_1^\infty \int_y^{yz} 6(1-G(x;\theta))g(x;\theta)g(y;\theta)g(z;\theta)dxdydz \\
&- \int_1^\infty \int_x^\infty 2(1-G(x;\theta))g(x;\theta)g(y;\theta)dydx \\
&= 3\Big(1 - \int_1^\infty \int_1^\infty (1-G(yz;\theta))^2)g(y;\theta)g(z;\theta)dydz\Big) \\
&- 2\int_1^\infty (1-G(x;\theta))^2 g(x;\theta)dx.
\end{aligned}$$

Пошто је други члан константа, његов први извод једнак је нули, па је први извод $b_{I^{[\mathcal{B}]}}(\theta)$ за $\theta = 0$

$$\begin{aligned}
b'_{I^{[\mathcal{B}]}}(0) &= 6\int_1^\infty \int_1^\infty \frac{H(yz)}{y^3z^3}dydz - 6\int_1^\infty \int_1^\infty \frac{h(y)}{y^2z^4}dydz \\
&= 6\Big(\int_1^\infty \int_1^x \int_{\frac{x}{y}}^\infty \frac{h(x)}{y^3z^3}dxdydz + \int_1^\infty \int_x^\infty \int_1^\infty \frac{h(x)}{y^3z^3}dxdydz - \frac{1}{3}\int_1^\infty \frac{h(y)}{y^2}dy\Big) \\
&= \int_1^\infty \frac{h(x)}{x^2}(3\ln x - \frac{1}{2})dx = 4\int_1^\infty \upsilon_{\mathcal{B}}(x)h(x)dx.
\end{aligned}$$

Пошто је $b_{I^{[\mathcal{B}]}}(0) = 0$ користећи Маклоренов развој добијамо (5.26). $\square$

За израчунавање Бахадурове ефикасности у конкретном случају наводимо следећи пример.

Пример 5.3.1. *Нека је алтернативна расподела лог-Вејбулова* (5.11). *Њен први извод по θ дат је изразом* (5.1.1).

Из (5.26) *и* (5.1.1) *добијамо*

$$b_{I^{[\mathcal{B}]}}(\theta) = 4\theta \int\limits_1^\infty \upsilon_{\mathcal{B}}(x) \frac{\alpha}{x^{\alpha+1}}(-\alpha \ln x \ln \ln x + \ln \ln x + 1)dx = \frac{2}{9}\theta.$$

Кулбак-Лајблерова горња граница дата је изразом (5.16) *па је Бахадурова ефикасност*

$$e_{I^{[\mathcal{B}]}} = \frac{c_{I^{[\mathcal{B}]}}(\theta)}{2K_{g_1}(\theta)} = \frac{125}{117\psi'(1)} \approx 0.649.$$

Израчунавање у случају осталих алтернатива (5.12-5.15) је аналогно. Вредности локалних Бахадурових ефикасности приказани су у табели 5.3.

Прелазимо на рачунање Бахадуровиг тачног нагиба у случају статистике $I_n^{[\mathcal{C}]}$. Функције $f(t)$ и $b_T(\theta)$ налазимо коришћењем следећих лема.

Лема 5.3.3. *Нека је $t > 0$. За статистику $I_n^{[\mathcal{B}]}$ функција $f_{I^{[\mathcal{C}]}}(t)$ је аналитичка за довољно мало $t > 0$ и важи*

$$f_{I^{[\mathcal{C}]}}(t) = \frac{525}{76}t^2 + o(t^2), \ \ t \to 0.$$

Лема 5.3.4. *За алтернативну расподелу $g(x;\theta)$ која припада $\mathcal{G}$ важи*

$$b_{I^{[\mathcal{C}]}}(\theta) = 4\theta \int\limits_1^\infty \upsilon^{[\mathcal{C}]}(x)h(x)dx + o(\theta), \ \theta \to 0. \tag{5.27}$$

Докази ових лема аналогни су доказима лема 5.3.1 и 5.3.2 па их овде изостављамо.

Пример 5.3.2. *Нека је алтернативна расподела лог-Вејбулова* (5.11). *Из* (5.27) *и* (5.1.1) *добијамо*

$$b_{I^{[\mathcal{C}]}}(\theta) = 4\theta \int_1^\infty \upsilon^{[\mathcal{C}]}(x)\frac{\alpha}{x^{\alpha+1}}(-\alpha \ln x \ln\ln x + \ln\ln x + 1)dx = \frac{3}{4}(1-\ln 2)\theta,$$

а користећи (5.16) *добијамо да је Бахадурова ефикасност једнака*

$$e_{I^{[\mathcal{C}]}} = \frac{c_{I^{[\mathcal{C}]}}(\theta)}{2K_{g_1}(\theta)} = \frac{4725(1-\ln 2)^2}{608\psi'(1)} \approx 0.445.$$

За остале алтернативе вредности локалне Бахадурове ефикасности приказане су у табели 5.3.

алт.	g_1	g_2	$g_3(1.5)$	$g_3(2)$	g_4	g_5
$e_{I^{[\mathcal{B}]}}$	0.649	0.757	0.326	0.119	0.778	0.451
$e_{I^{[\mathcal{C}]}}$	0.445	0.277	0.610	0.486	0.244	0.737

Табела 5.3: Бахадурова ефикасност статистика $I_n^{[\mathcal{B}]}$ и $I_n^{[\mathcal{C}]}$

5.3.2 Локално оптималне алтернативе

Неке од локално оптималних алтернатива налазимо следећом теоремом.

Теорема 5.3.1. *Нека је $g(x;\theta)$ густина расподеле из $\mathcal{G}$ која такође задовољава услов*

$$\int_1^\infty x^2 h^2(x)dx < \infty.$$

Алтернативне густине расподеле

$$g(x;\theta) = \frac{1}{x^2} + \theta\Big(C\frac{\upsilon^{[l]}(x)}{x^2} + D\frac{\ln x - 1}{x^2}\Big),\ x \geq 1,\ C > 0,\ D \in \mathbb{R},$$

за мало θ асимптотски су оптималне за тест заснован на статистици $I_n^{[l]}$, $l = \mathcal{B}, \mathcal{C}$.

Доказ. Доказаћемо теорему за прву статистику ($l = \mathcal{B}$), а за другу је потпуно еквивалентан. Нека је

$$h_0(x) = h(x) - \frac{(\ln x - 1)}{x^2} \int\limits_1^\infty h(s) \ln s ds. \tag{5.28}$$

Може се показати да ова функција задовољава следеће једнакости:

$$\int\limits_1^\infty h_0^2(x) x^2 dx = \int\limits_1^\infty x^2 h^2(x) dx - \Bigg(\int\limits_1^\infty h(x) \ln x dx \Bigg)^2$$

$$\int\limits_1^\infty \upsilon_{\mathcal{B}}(x) h_0(x) dx = \int\limits_1^\infty \upsilon_{\mathcal{B}}(x) h(x) dx.$$

Из лема 5.3.1 и 5.3.2, користећи (5.25), добијамо да је локална асимптотска ефикасност једнака

$$\begin{aligned} e_{I[\mathcal{B}]} &= \lim_{\theta \to 0} \frac{c_{I[\mathcal{B}]}(\theta)}{2K(\theta)} = \lim_{\theta \to 0} \frac{2 \cdot \frac{1125}{108} b^2_{I[\mathcal{B}]}(\theta)}{2K(\theta)} = \lim_{\theta \to 0} \frac{b^2_{I[\mathcal{B}]}(\theta)}{9\sigma^2_{\mathcal{B}} 2K(\theta)} \\ &= \lim_{\theta \to 0} \frac{9\theta^2 \Bigg(\int\limits_1^\infty \upsilon_{\mathcal{B}}(x) h(x) dx \Bigg)^2 + o(\theta^2)}{9 \int\limits_1^\infty (\upsilon_{\mathcal{B}})^2(x) x^2 dx \Bigg(\theta^2 \Big(\int\limits_1^\infty x^2 h^2(x) dx - \big(\int\limits_1^\infty h(x) \ln x dx \big)^2 \Big) + o(\theta^2) \Bigg)} \\ &= \frac{\Bigg(\int\limits_1^\infty \upsilon_{\mathcal{B}}(x) h_0(x) dx \Bigg)^2}{\int\limits_1^\infty (\upsilon_{\mathcal{B}})^2(x) x^{-2} dx \int\limits_1^\infty h_0^2(x) x^2 dx}. \end{aligned}$$

Применом неједнакости Коши-Шварца добијамо да је $e_{I[\mathcal{B}]} = 1$ ако и само ако је $h_0(x) = C\upsilon_{\mathcal{B}}(x)\alpha x^{-2}$. Заменом ове једнакости у (5.28) добијамо израз за $h(x)$. Пошто је $h(x)$ за алтернативе из теореме овог облика, доказ је завршен. □

5.4 Статистике $D_n^{[\mathcal{B}]}$ и $D_n^{[\mathcal{C}]}$

Сада ћемо анализирати асимптотска својства статистика Колмогоровљевог типа датих изразима (5.2) и (5.4). Као и раније, можемо претпоставити да је $\alpha = 1$. За фиксирано $t \geq 1$, израз $G_n(t) - H_n(t)$ је V-статистика с језгром

$$\begin{aligned}\Xi_{\mathcal{B}}(X_1, X_2, X_3, t) &= I\{X_{(2);X_1,X_2,X_3}/X_{(1);X_1,X_2,X_3} \leq t\} \\ &- \frac{1}{3}(I\{\min\{X_1, X_2\} \leq t\} + I\{\min\{X_2, X_3\} \leq t\} \\ &+ I\{\min\{X_1, X_3\} \leq t\}).\end{aligned}$$

Пројекција овог језгра је

$$\begin{aligned}\xi_{\mathcal{B}}(s,t) &= E(\Xi_{\mathcal{B}}(X_1, X_2, X_3, t)|X_1 = s) \\ &= P\{X_{(2);s,X_2,X_3}/X_{(1);s,X_2,X_3} \leq t\} \\ &- \frac{2}{3}P\{\min\{s, X_2\} \leq t\} - \frac{1}{3}P\{\min\{X_2, X_3\} \leq t\} \\ &= 2P\{X_2 \leq st, s \leq X_2 \leq X_3\} + 2P\{s \leq X_2 t, X_2 \leq s \leq X_3\} \\ &+ 2P\{X_3 \leq tX_2, X_2 \leq X_3 \leq s\} - \frac{2}{3}(1 - I\{s > t\}P\{X_2 > t\}) \\ &- \frac{1}{3}(1 - (P\{X_2 > t\})^2).\end{aligned}$$

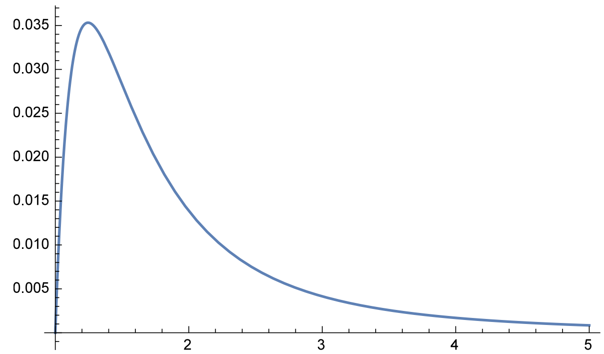
После извесног рачуна добијамо

$$\xi_{\mathcal{B}}(s,t) = \frac{t}{s^2} - \frac{1}{s^2 t^2} + \frac{1}{3t^2} - \frac{1}{3t} + I\{s \leq t\}\Big(\frac{1}{3t} - \frac{t}{s^2}\Big).$$

Лако је показати да је математичко очекивање ове пројекције једнако нули. Њена дисперзија за фиксирано t једнака је

$$\sigma^2_{\mathcal{B}}(t) = \int_1^{\infty}(\xi_{\mathcal{B}}(s,t))^2 s^{-2} ds = \frac{4}{45}(t^{-3} + t^{-4} - 2t^{-6}).$$

Функција $\sigma^2_{\mathcal{B}}(t)$ (слика 5.3) достиже свој максимум за $t_{\mathcal{B}} = 1.245$ и $\sigma^2_{\mathcal{B}}(t_{\mathcal{B}}) = 0.0353$. Закључујемо да је фамилија језгара $\Xi_{\mathcal{B}}(X_1, X_2, X_3, t)$



Слика 5.3: График функције $\sigma^2_{\mathcal{B}}(t)$

недегенерисана, док V-емпиријски процес

$$\rho_n^{[\mathcal{B}]}(t) = \sqrt{n}(G_n(t) - H_n(t)),\ t \geq 1$$

конвергира неком Гаусовом процесу чија се коваријација може израчунати на основу резултата из [76]. Расподела статистике $D_n^{[\mathcal{B}]}$ није позната.

Слично, за $t \geq 1$, израз $J_n(t) - K_n(t)$ је V-статистика с језгром

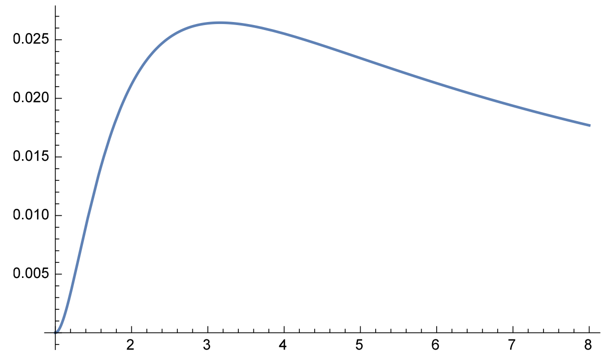
$$\begin{aligned}\Xi_{\mathcal{C}}(X_1, X_2, X_3, t) &= I\{X_{(3);X_1,X_2,X_3}/X_{(2);X_1,X_2,X_3} \leq t\} \\ &- I\{(X_{(2);X_1,X_2,X_3}/X_{(1);X_1,X_2,X_3})^2 \leq t\}.\end{aligned}$$

Аналогним рачуном као у претходном случају можемо извести да је пројекција језгра

$$\begin{aligned}\xi_{\mathcal{C}}(s,t) &= \frac{1}{s^2 t}(2 - 2s + s^2 t^{\frac{1}{2}} - s^2 t - t^{\frac{3}{2}} + 2st^2 - t^3 + I\{s \leq t^{\frac{1}{2}}\}t^{\frac{1}{2}}(t - s^2) \\ &+ I\{s \leq t\}t(t-s)^2).\end{aligned}$$

Њена дисперзија је

$$\sigma^2_{\mathcal{C}}(t) = \frac{1}{15}(3t^{-1} - 2t^{-\frac{3}{2}} + 2t^{-2} - 14t^{-\frac{5}{2}} + 13t^{-3} - 2t^{-4})$$



Слика 5.4: График функције $\sigma^2_{\mathcal{C}}(t)$

која има максимум 0.0265 за $t_{\mathcal{C}} = 3.160$ (слика 5.4). Ова фамилија је такође недегенерисана и сви закључци о асимптотској расподели статистике $D_n^{[\mathcal{C}]}$ су еквивалентни оним за статистику $D_n^{[\mathcal{B}]}$.

5.4.1 Локална Бахадурова ефикасност

Прелазимо на рачунање локалне Бахадурове ефикасности за статистике $D_n^{[\mathcal{B}]}$ и $D_n^{[\mathcal{C}]}$. Функције f из (4.2) одредићемо следећом теоремом.

Теорема 5.4.1. *Нека је $a \geq 0$. Тада су $f_{D^{[\mathcal{B}]}}(a)$ и $f_{D^{[\mathcal{C}]}}(a)$ аналитичке за довољно мало a и важи*

$$f_{D^{[\mathcal{B}]}}(a) = \frac{a^2}{18}\sigma^2_{\mathcal{B}}(t_{\mathcal{B}}) + o(a^2) \sim 1.58a^2,\ a \to 0$$
$$f_{D^{[\mathcal{C}]}}(a) = \frac{a^2}{18}\sigma^2_{\mathcal{C}}(t_{\mathcal{C}}) + o(a^2) \sim 2.10a^2,\ a \to 0$$

Доказ следи из теореме 4.2.4. Лимес у вероватноћи статистике $D_n^{[l]}$, $l = \mathcal{B}, \mathcal{C}$ под алтернативном хипотезом одредићемо следећом лемом.

Лема 5.4.1. *За дату алтернативну расподелу $g(x;\theta)$ из $\mathcal{G}$ важи*

$$b_{D^{[l]}}(\theta) = 3\theta \sup_{t\geq 1}\left|\int_1^{\infty} \xi^{[l]}(x,t)h(x)dx\right| + o(\theta),\ \theta \to 0. \tag{5.29}$$

Доказ. Лему ћемо доказати за прву статистику ($l = \mathcal{B}$). Други случај је еквивалентан. Користећи теорему 3.3.1 добијамо

$$b_{D^{[\mathcal{B}]}}(\theta) = \sup_{t \geq 1} \Big| P_\theta\{X_{(2;3)}/X_{(1;3)} \leq t\} - P_\theta\{\min\{X_1, X_2\} \leq t\} \Big|.$$

Нека је $a_{D^{[\mathcal{B}]}}(\theta; t) = P_\theta\{X_{(2;3)}/X_{(1;3)} \leq t\} - P_\theta\{\min\{X_1, X_2\} \leq t\}$. Тада имамо

$$\begin{aligned} 2a_{D^{[\mathcal{B}]}}(\theta; t) &= \int_1^\infty \int_1^{tx} 6(1 - G(x;\theta))g(x;\theta)g(y;\theta)dxdy - (1 - (P_\theta\{X_1 > t\})^2) \\ &= 3\int_1^\infty (1 - G(tx;\theta))^2 g(x;\theta)dx - (1 - (1 - G(t;\theta))^2). \end{aligned}$$

Први извод функције $a_{D^{[\mathcal{B}]}}(\theta; t)$ по θ у нули, $a'_{D^{[\mathcal{B}]}}(t)$, је

$$\begin{aligned} a'_{D^{[\mathcal{B}]}}(t) &= 6\int_1^\infty \frac{H(tx)}{tx^3}dx - 3\int_1^\infty \frac{h(x)}{t^2x^2}dx + \frac{2}{t}H(t) \\ &= \frac{6}{t}\int_1^t \int_1^\infty \frac{h(y)}{x^3}dxdy + \frac{6}{t}\int_t^\infty \int_{\frac{y}{t}}^\infty \frac{h(y)}{x^3}dxdy - 3\int_1^\infty \frac{h(x)}{t^2x^2}dx + \frac{2}{t}\int_1^t h(y)dy \\ &= \int_1^\infty h(y)(-\frac{3}{t}I\{y \leq t\} + \frac{3t}{y^2}I\{y > t\} - \frac{3}{t^2y^2} + \frac{2}{t}I\{y \leq t\})dy \\ &= 3\int_1^\infty \xi_{\mathcal{B}}(y;t)h(y)dy. \end{aligned}$$

Применом Маклореновог развоја функције $a_{D^{[\mathcal{B}]}}(\theta; t)$ добијамо тврђење леме. $\square$

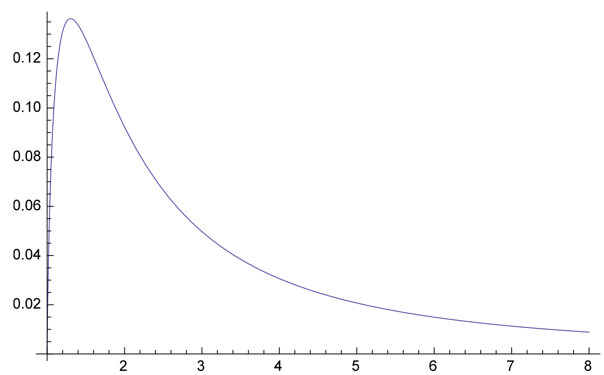
Као и у претходним примерима израчунаћемо Бахадурову ефикасност за лог-Вејбулову алтернативу.

Пример 5.4.1. *За статистику* $D_n^{[\mathcal{B}]}$ *из* (5.29) *и* (5.1.1) *добијамо*

$$b_{D^{[\mathcal{B}]}}(\theta) = 3\theta \sup_{t\geq 1} \left| \int_1^\infty \xi_{\mathcal{B}}(x) \frac{1}{x^2}(-\ln x \ln\ln x + \ln\ln x + 1)dx \right| = 0.4088\,\theta,$$

а користећи (5.16) *добијамо да је Бахадурова ефикасност*

$$e_{D^{[\mathcal{B}]}} = \frac{c_{D^{[\mathcal{B}]}}(\theta)}{2K_{g_1}(\theta)} \approx 0.320.$$



Слика 5.5: График функције $a'_{D^{[\mathcal{B}]}}(t)$ за лог-Вејбулову расподелу

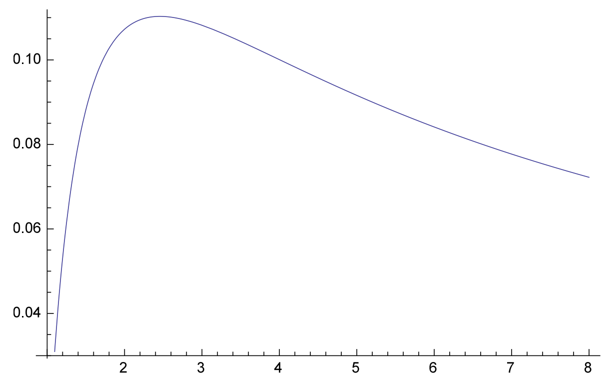
Слично, за статистику $D_n^{[\mathcal{C}]}$ *имамо*

$$b_{D^{[\mathcal{C}]}}(\theta) = 3\theta \sup_{t\geq 1} \left| \int_1^\infty \xi_{\mathcal{C}}(x) \frac{1}{x^2}(-\ln x \ln\ln x + \ln\ln x + 1)dx \right| = 0.3309\,\theta,$$

и

$$e_{D^{[\mathcal{C}]}} = \frac{c_{D^{[\mathcal{C}]}}(\theta)}{2K_{g_1}(\theta)} \approx 0.280.$$

За остале алтернативе рачунање је аналогно. Вредности локалних Бахадурових ефикасности приказани су у табели 5.4.



Слика 5.6: График функције $a'_{D[\mathcal{C}]}(t)$ за лог-Вејбулову расподелу

алт.	g_1	g_2	$g_3(1.5)$	$g_3(2)$	g_4	g_5	$g_6(1.5)$	$g_6(4)$	$g_6(8)$
$e_{D[\mathcal{B}]}$	0.320	0.414	0.141	0.047	0.436	0.207	0.124	0.462	0.632
$e_{D[\mathcal{C}]}$	0.280	0.174	0.410	0.362	0.156	0.513	0.500	0.323	0.116

Табела 5.4: Бахадурова ефикасност за статистике Колмогоровљевог типа

5.4.2 Локално оптималне алтернативе

Као и у претходном случају, одредићемо неке од локално оптималних алтернатива следећом теоремом.

Теорема 5.4.2. *Нека је $g(x;\theta)$ густина из $\mathcal{G}$ која такође задовољава услов*

$$\int\limits_1^\infty x^2 h^2(x)dx < \infty.$$

Алтернативне густине расподела

$$g(x;\theta) = \frac{1}{x^2} + \theta\Big(C\frac{\xi^{[l]}(x;t_l)}{x^2} + D\frac{\ln x - 1}{x^2}\Big),\ x \geq 1,\ C > 0,\ D \in \mathbb{R},$$

где је $t_{\mathcal{B}} = 1.245$, $t_{\mathcal{C}} = 3.16$, за мале вредности θ асимптотски су оптималне за тест заснован на статистици $D_n^{[l]}$, $l = \mathcal{B}, \mathcal{C}$.

Доказ је аналоган доказу теореме 5.3.1, па га овде не наводимо.

5.5 Упоређивање асимптотских ефикасности

Сада ћемо упоредити Бахадурове ефикасности наших тестова користећи алтернативне расподеле (5.11-5.15) и (5.22).

Бахадурове ефикасности наведених тестова поново су ради прегледности приказане су у табели 5.5 за статистике интегралног типа а у табели 5.6 за тест статистике Колмогоровљевог типа.

Можемо приметити да је опште правило да су интегралне статистике ефикасније него статистике Колмогоровљевог типа. Међутим, као што је раније речено, мана тестова заснованим на интегралним статистикама је што нису постојани против сваке алтернативе. Наведен је пример алтернативе (5.22) против које ниједан од интегралних тестова није постојан.

Можемо приметити да сваки од тестова има барем једну алтернативу против које је задовољавајуће ефикасан. Упоређивањем статистика заснованим на сродним карактеризацијама преко количника статистика поретка на узорку обима три, занимљиво је приметити да им се ефикасности прилично разликују. Лог-Вејбулова, лог-гама и инверзна бета расподела дају боље резултате за тест статистику $I_n^{[\mathcal{B}]}$, а остале алтернативе за $I_n^{[\mathcal{C}]}$ па не можемо закључити да је један од њих бољи од другог.

Тест заснован на статистици $I_n^{[\mathcal{A}]}$ изгледа најбољи у глобалу, међутим сваки од тестова има бар једну стандардну алтернативу против које је ефикаснији од своја два конкурента.

Сличне закључке можемо извести и у случају статистика Колмогоровљевог типа, али редослед ефикасности по алтернативама није идентичан. Видимо да је тест заснован на статистици $D_n^{[\mathcal{C}]}$ сада најефикаснији за "тилт" алтернативу g_5 као и g_3 за обе вредности параметра. Алтернатива мешавине Паретових расподела g_6 доноси нам занимљиву појаву. У зависности од вредности параметра β, редослед тестова по ефикасности се мења и сваки је најефикаснији за неку вредност. Ово нам указује да ако сумњамо да имамо контаминирани узорак, треба да изаберемо погодни тест у зависности од претпостављног количника параметара чланова мешавине. Ако је тај количник висок, бирамо $D_n^{[\mathcal{B}]}$, ако је низак, $D_n^{[\mathcal{C}]}$, а ако је "умерен" онда $D_n^{[\mathcal{A}]}$.

	Алтернативна расподела					
Тест	g_1	g_2	$g_3(1.5)$	$g_3(2)$	g_4	g_5
$I_n^{[\mathcal{A}]}$	0.821	0.788	0.618	0.338	0.777	0.800
$I_n^{[\mathcal{B}]}$	0.649	0.757	0.326	0.119	0.778	0.451
$I_n^{[\mathcal{C}]}$	0.445	0.277	0.610	0.486	0.244	0.737

Табела 5.5: Бахадурова ефикасност за статистике интегралног типа

	Алтернативна расподела								
Тест	g_1	g_2	$g_3(1.5)$	$g_3(2)$	g_4	g_5	$g_6(1.5)$	$g_6(4)$	$g_6(8)$
$D_n^{[\mathcal{A}]}$	0.437	0.448	0.331	0.192	0.455	0.473	0.359	0.631	0.581
$D_n^{[\mathcal{B}]}$	0.320	0.414	0.141	0.047	0.436	0.207	0.124	0.462	0.632
$D_n^{[\mathcal{C}]}$	0.280	0.174	0.410	0.362	0.156	0.513	0.500	0.323	0.116

Табела 5.6: Бахадурова ефикасност за статистике Колмогоровљевог типа

5.6 Критичне вредности тестова

Сада ћемо израчунати критичне вредности тестова за узорке малог обима. С обзиром да немамо тачну расподелу за мале вредности n, критичне вредности рачунамо Монте Карло методом.

Статистика $I_n^{[\mathcal{A}]}$ може се изразити као

$$\begin{aligned} I_n^{[\mathcal{A}]} &= \int_1^{\infty} (M_n(t) - F_n(t)) dF_n(t) = \frac{1}{n}\sum_{j=1}^{n}(M_n(x_j) - F_n(x_j)) \\ &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{r_j - j}{N} - \frac{j}{n}\right) \\ &= (2nN)^{-1}\left(2\sum_{j=1}^{n} r_j - (n+1)(n+N)\right), \end{aligned}$$

где је $N = \binom{n}{2}$, а r_j је ранг x_j у обједињеном узорку x_k, $1 \le k \le n$ и $\max\{\frac{x_j}{x_k}, \frac{x_k}{x_j}\}$, $1 \le j < k \le n$.

На сличан начин могу се изразити и остале интегралне статистике. Статистике Колмогоровљевог типа није потребно трансформисати јер им је облик погодан за програмирање.

Критичне вредности за свих шест тестова, израчунате из 10000 понављања у програмском језику R, дате су у табели 5.7.

n	$I_n^{[\mathcal{A}]}$	$I_n^{[\mathcal{B}]}$	$I_n^{[\mathcal{C}]}$	$D_n^{[\mathcal{A}]}$	$D_n^{[\mathcal{B}]}$	$D_n^{[\mathcal{C}]}$
10	0.13	0.11	0.24	0.41	0.57	0.47
20	0.09	0.08	0.14	0.28	0.37	0.29
30	0.07	0.06	0.11	0.22	0.29	0.22
40	0.06	0.05	0.09	0.19	0.25	0.19
50	0.06	0.05	0.08	0.16	0.22	0.17
100	0.04	0.03	0.05	0.12	0.15	0.11

Табела 5.7: Критичне вредности тестова с нивоом значајности 0.05

Поглавље 6

Тестови сагласности с експоненцијалном расподелом

Тестове које ћемо овде приказати оформљени су на основу карактеризације Арнолда и Виљасењора (последица 2.4.1) изложених у [39]. У даљем тексту дефинисаћемо тест статистике, одредити њихову асимптотску расподелу, израчунати локалну Бахадурову ефикасност против неких стандардних блиских алтернатива и пронаћи неке алтернативе из ЛАО класе.

6.1 Тест статистике

Нека је $X_1, X_2, \ldots, X_n$ узорак из расподеле F која припада класи $\mathcal{F}$.

Тестирамо сложену нулту хипотезу да је F експоненцијална расподела (1.1) где је $\lambda > 0$ непознати параметар.

Нека је $F_n(t)$ емпиријска функција расподеле. У складу с наведеном карактеризацијом дефинишемо V-емпиријску функцију расподеле

$$
\begin{aligned}
H_n^{[k]}(t) &= \frac{1}{n^k}\sum_{i_1,i_2,\ldots,i_k=1}^{n} \mathrm{I}\Big\{\max(X_{i_1},X_{i_2},\ldots,X_{i_k}) \le t\Big\},\\
G_n^{[k]}(t) &= \frac{1}{n^k k!}\sum_{i_1,\ldots,i_k=1}^{n}\Bigg[\sum_{\pi(j_1,\ldots,j_k)} \mathrm{I}\Big\{\frac{X_{i_1}}{j_1}+\frac{X_{i_2}}{j_2}+\ldots+\frac{X_{i_k}}{j_k}\le t\Big\}\Bigg],
\end{aligned}
$$

где $\pi(j_1,\ldots,j_k)$ представља скуп свих $k!$ пермутација природних бројева $1,2,\ldots,k,\ k\ge 2$, док $\pi(j_1,\ldots,j_i^*,\ldots,j_k)$, који се јавља доле, означава скуп свих $(k-1)!$ пермутација природних бројева $1,2,\ldots,k$ без i.

Уводимо два низа (две класе) статистика (у зависности од природног броја $k>1$) чија је расподела инваријантна у односу на параметар λ експоненцијалне расподеле. Прва статистика је интегралног, а друга Колмогоровљевог типа дефинисаних у трећем поглављу.

$$
\begin{aligned}
&I_n^{[k]} = \int_0^\infty \big(H_n^{[k]}(t)-G_n^{[k]}(t)\big)\,dF_n(t),\\
&D_n^{[k]} = \sup_{t\ge 0}\mid H_n^{[k]}(t)-G_n^{[k]}(t)\mid,
\end{aligned}
$$

Без губљења општости можемо претпоставити да су велике вредности тест статистика значајне за одбацивање нулте хипотезе.

6.2 Интегрална статистика $I_n^{[k]}$

Без губљења општости можемо претпоставити да је параметар експоненцијалне расподеле $\lambda=1$. Статистика $I_n^{[k]}$ асимптотски је еквивалентна V-статистици реда $(k+1)$ с центрираним језгром

$$\begin{aligned}
&\Upsilon_k(X_1, X_2, \ldots, X_{k+1})\\
&\quad= \frac{1}{k+1}\Big[\sum_{i=1}^{k+1} \mathrm{I}\big\{\max(X_1, \ldots, X_{i-1}, X_{i+1} \ldots, X_{k+1}) \leq X_i\big\}\\
&\quad - \frac{1}{k!}\sum_{i=1}^{k+1}\sum_{\pi(j_1,\ldots,j_i^*,\ldots,j_{k+1})} \mathrm{I}\Big\{\frac{X_1}{j_1} + \ldots + \frac{X_{i-1}}{j_{i-1}} + \frac{X_{i+1}}{j_{i+1}} + \ldots + \frac{X_{k+1}}{j_{k+1}} \leq X_i\Big\}\Big].
\end{aligned}$$

Да бисмо показали да језгро $\Upsilon_k(X_1, X_2, \ldots, X_{k+1})$ није дегенерисано, израчунаћемо његову пројекцију $\upsilon_k(s)$ под нултом хипотезом. За фиксирано $X_{k+1} = s$ ова пројекција је облика

$$\begin{aligned}
\upsilon_k(s) &= E\big(\Upsilon_k(X_1, X_2, \ldots, X_{k+1})|X_{k+1} = s\big)\\
&= \frac{1}{k+1}P\big\{\max\{X_1, \ldots, X_k\} \leq s\big\}\\
&+ \frac{k}{k+1}P\big\{\max\{s, X_2, \ldots, X_k\} \leq X_1\big\}\\
&- \frac{1}{(k+1)!}\sum_{\pi(j_1,\ldots,j_k)} P\Big\{\frac{X_1}{j_1} + \ldots + \frac{X_k}{j_k} \leq s\Big\}\\
&- \frac{k}{(k+1)!}\sum_{\pi(j_1,\ldots,j_k)} P\Big\{\frac{s}{j_1} + \frac{X_2}{j_2} + \ldots + \frac{X_k}{j_k} \leq X_1\Big\}.
\end{aligned}$$

На основу карактеризације први и трећи члан горњег израза поклапају се па се могу скратити. Други члан горњег израза једнак је

$$\begin{aligned}
&\frac{k}{k+1}P\big\{\max\{s, X_2, \ldots, X_k\} \leq X_1\big\}\\
&\qquad= \frac{k}{k+1}\int_0^\infty \mathrm{I}\{s \leq t\}P\{X_2 \leq t, \ldots, X_k \leq t\}dF(t)\\
&\qquad= \frac{k}{k+1}\int_s^\infty F^{k-1}(s)dF(s) = \frac{1}{k+1}\left(1 - F^k(s)\right),
\end{aligned}$$

где је $F(x) = 1 - e^{-x}$. Остаје да се израчуна последњи члан.

$$
\begin{aligned}
P\Big\{\frac{s}{j_1} + \frac{X_2}{j_2} + \ldots + \frac{X_k}{j_k} \le X_1\Big\} \\
&= \int_0^{\infty} e^{-x_2} dx_2 \ldots \int_0^{\infty} e^{-x_k} dx_k \int\limits_{\frac{s}{j_1}+\frac{x_2}{j_2}+\ldots+\frac{x_k}{j_k}}^{\infty} e^{-x_1} dx_1 \\
&= \frac{1}{(k+1)}\Big(1 + \frac{1}{j_1}\Big) e^{-s/j_1},
\end{aligned}
$$

после сумирања по свим пермутацијама индексā $j_1, j_2, \ldots, j_k$ и још мало додатног рачуна добијамо да је четврти члан једнак $\frac{1}{(k+1)^2}\sum_{r=1}^{k}(1 + \frac{1}{r})e^{-s/r}$, па је пројекција υ_k језгра Υ_k:

$$
\upsilon_k(s) = \frac{1 - (1 - e^{-s})^k}{k+1} - \frac{1}{(k+1)^2} \sum_{r=1}^{k} \Big(1 + \frac{1}{r}\Big) e^{-s/r}. \tag{6.1}
$$

Лако се показује да је $E(\upsilon_k(X_1)) = 0$. Након израчунавања добијамо да је дисперзија пројекције једнака

$$
\begin{aligned}
\sigma_k^2 = \int\limits_0^{\infty} \upsilon_k^2(s) e^{-s} ds = \frac{1}{(k+1)^3}\Bigg[\frac{-12k^4 - 38k^3 - 35k^2 - 11k}{4(k+1)^2(k+2)(2k+1)} \\
+ 2k! \sum_{r=1}^{k} \frac{1}{(k+1+\frac{1}{r})(k+\frac{1}{r})\cdots(2+\frac{1}{r})} + \frac{2}{k+1} \sum_{1\le i<j\le k} \frac{1}{i+j+ij}\Bigg].
\end{aligned} \tag{6.2}
$$

Јасно је из израза (6.2) да је Υ_k недегенерисано за свако k.

У случају недегенерисаних језгара можемо уместо V-статистике $I_n^{[k]}$ користити одговарајућу U-статистику с истим језгром која има исте асимптотске особине али је прилично једноставнија за израчунавања.

Како су од практичног значаја само статистике за мале вредности k, сада ћемо посебно размотрити случајеве када је $k = 2$ и $k = 3$.

6.2.1 Статистика $I_n^{[2]}$

У случају $k = 2$ из израза (6.1) и (6.2) добијамо да је пројекција језгра $\Upsilon_2(X, Y, Z)$ једнака

$$\upsilon_2(s) = \frac{4}{9}e^{-s} - \frac{1}{3}e^{-2s} - \frac{1}{6}e^{-s/2}, \tag{6.3}$$

док је његова дисперзија

$$\sigma_2^2 = \int_0^\infty \upsilon_2^2(s)e^{-s}ds = \frac{5}{13608} \approx 0.000367.$$

Применом Хефдингове теореме 3.2.2 за U-статистике с недегенерисаним језгрима, кад $n \to \infty$, добијамо граничну расподелу

$$\sqrt{n}I_n^{[2]} \xrightarrow{d} \mathcal{N}\Big(0, \frac{5}{1512}\Big).$$

Сада налазимо логаритамску асимптотику великих одступања низа статистика $I_n^{[2]}$ под нултом хипотезом. Језгро Υ_2 је центрирамо, недегенерисано и ограничено. Применом теореме 4.2.3 за велика одступања недегенерисаних U-статистика добијамо следећу теорему

Теорема 6.2.1. *За $a > 0$ важи*

$$\lim_{n\to\infty} n^{-1}\ln P_{H_0}\{I_n^{[2]} > a\} = -f_{I^{[2]}}(a),$$

где је функција $f_{I^{[2]}}$ аналитичка за довољно мало $a > 0$, и поред тога је

$$f_{I^{[2]}}(a) = \frac{a^2}{18\sigma_2^2} + o(a^2) \sim \frac{756}{5}a^2 = 151.2a^2, \quad a \to 0. \tag{6.4}$$

На основу закона великих бројева за U-статистике (в. теорему 3.2.1) под алтернативом H_1 статистика $I_n^{[2]}$ конвергира у вероватноћи ка

$$b_I^{[2]}(\theta) = P_\theta\big(\max(X, Y) \le Z\big) - P_\theta\Big(X + \frac{Y}{2} \le Z\Big).$$

Нека је класа алтернатива $\mathcal{G}$ дефинисана аналогно оној у одељку 5.1.1 с том разликом што је сада $G(x;0) = 1 - e^{-x}$. Лако се може показати да је за $G \in \mathcal{G}$

$$b_{I^{[2]}}(\theta) = 3\theta \int_0^\infty \upsilon_2(s)h(s)ds + o(\theta), \;\; \theta \to 0, \tag{6.5}$$

где је $h(x) = g'_\theta(x;0)$, а $\upsilon_2(s)$ пројекција из (6.3).

Навешћемо сада скуп стандардних алтернативних расподела блиских експоненцијалној које ће у даљем тексту бити разматране.

1. Макехамова расподела с густином

$$g_1(x,\theta) = \big(1 + \theta(1 - e^{-x})\big)e^{-x-\theta(e^{-x}-1+x)}, \theta > 0, x \geq 0;$$

2. Вејбулова расподела с густином

$$g_2(x,\theta) = (1+\theta)x^\theta e^{-x^{1+\theta}}, \theta > 0, x \geq 0;$$

3. гама расподела с густином

$$g_3(x,\theta) = \frac{x^\theta}{\Gamma(\theta+1)}e^{-x}, \theta > 0, x \geq 0;$$

4. мешавина експоненцијалних расподела с негативним тежинама (EMNW(β)) (в. [38]) с густином

$$g_4(x) = (1+\theta)e^{-x} - \theta\beta e^{-\beta x}, x \geq 0, \theta \in \Big(0, \frac{1}{\beta - 1}\Big].$$

Израчунаћемо локалну Бахадурову ефикасност против наведених алтернатива.

Кулбак-Лајблерово растојање блиских алтернатива од класе експоненцијалних расподела дајемо следећом лемом.

Лема 6.2.1. *За дату густину $g(x;\theta)$ нека је Кулбак-Лајблерово расто-*

јање

$$K(\theta) = \inf_{\lambda>0} \int\limits_1^{\infty} \ln \frac{g(x;\theta)}{\lambda e^{-\lambda x}} g(x;\theta)dx \tag{6.6}$$

добро дефинисано. Тада када $\theta \to 0$

$$2K(\theta) = \theta^2 \Bigg(\int_0^{\infty} h^2(x) e^x dx - \Big(\int_0^{\infty} x h(x) dx \Big)^2 \Bigg) + o(\theta^2). \tag{6.7}$$

Доказ је аналоган доказу теореме 5.1.3 па га овде изостављамо.

За Макехамову алтернативу из (6.5) следи да је

$$\begin{aligned} b_{I^{[2]}}(\theta) &= 3\theta \int_0^{\infty} \Big(\frac{4}{9} e^{-s} - \frac{1}{3} e^{-2s} - \frac{1}{6} e^{-s/2}\Big) e^{-s}(2 - 2e^{-s} - s) ds \\ &= \frac{\theta}{90} \approx 0.011\,\theta + o(\theta),\ \ \theta \to 0. \end{aligned}$$

Локални тачни Бахадуров нагиб низа $I_n^{[2]}$ кад $\theta \to 0$ једнак је

$$c_{I^{[2]}}(\theta) = \big(b_{I^{[2]}}(\theta)\big)^2 / (9\sigma_2^2) = 0.037\theta^2 + o(\theta^2).$$

Из леме 6.6 Кулбак-Лајблерово растојање Макехамове расподеле понаша се кад $\theta \to 0$ као

$$K_{g_1}(\theta) = \frac{\theta^2}{24} + o(\theta^2) \tag{6.8}$$

па је локална Бахадурова ефикасност

$$e_{I^{[2]}} = \lim_{\theta \to 0} \frac{c_{I^{[2]}}(\theta)}{2K_{g_1}(\theta)} = 0.448.$$

Рачунање ефикасности у случају осталих алтернатива је аналогно. Вредности Бахадурових ефикасности приказане су у табели 6.1.

6.2.2 Статистика $I_n^{[3]}$

За $k = 3$ из (6.1) и (5.6) добијамо да је пројекција језгра $\Upsilon_3(X, Y, Z, W)$ једнака

$$\upsilon_3(s) = \frac{5}{8}e^{-s} - \frac{3}{4}e^{-2s} + \frac{1}{4}e^{-3s} - \frac{3}{32}e^{-s/2} - \frac{1}{12}e^{-s/3}, \qquad (6.9)$$

а њена дисперзија

$$\sigma_3^2 = \int_0^\infty \upsilon_3^2(s)e^{-s}ds = \frac{14591}{30750720} \approx 0.000474.$$

Као и у претходном случају, на основу Хефдингове теореме, кад $n \to \infty$, важи следећа конвергенција у расподели

$$\sqrt{n}I_n^{[3]} \xrightarrow{d} \mathcal{N}\Big(0, \frac{14591}{1921920}\Big).$$

Асимптотику великих одступања низа статистика $I_n^{[3]}$ под нултом хипотезом налазимо на аналогни начин као и у претходном случају

Теорема 6.2.2. *За $a > 0$ важи*

$$\lim_{n\to\infty} n^{-1} \ln P_{H_0}\{I_n^{[3]} > a\} = -f_{I^{[3]}}(a),$$

где је функција $f_{I^{[3]}}$ аналитичка за довољно мало $a > 0$, и поред тога је

$$f_{I^{[3]}}(a) = \frac{a^2}{32\sigma_3^2} + o(a^2) \sim \frac{960960}{14591}a^2 = 65.86a^2, \;\; a \to 0. \qquad (6.10)$$

Под алтернативом H_1 статистика $I_n^{[3]}$ конвергира у вероватноћи ка

$$b_{I^{[3]}}(\theta) = P_\theta\big\{\max(X, Y, Z) \le W\big\} - P_\theta\Big\{X + \frac{Y}{2} + \frac{Z}{3} \le W\Big\}.$$

Лако је показати користећи (4.7) да је

$$b_{I^{[3]}}(\theta) = 4\theta \int_0^\infty \upsilon_3(s)h(s)ds + o(\theta),$$

где је $\upsilon_3(s)$ пројекција из (6.9).

За Макехамову алтернативу имамо

$$\begin{aligned} b_{I^{[3]}}(\theta) &= 4\theta\int_0^\infty \Big(\frac{5}{8}e^{-s}-\frac{3}{4}e^{-2s}+\frac{1}{4}e^{-3s}-\frac{3}{32}e^{-s/2}-\frac{1}{12}e^{-s/3}\Big)e^{-s}(2-2e^{-s}-s)ds \\ &= \frac{2}{105}\theta \approx 0.019\,\theta + o(\theta),\ \ \theta\to 0, \end{aligned}$$

и за локални тачни нагиб низа $I_n^{[3]}$ кад $\theta \to 0$ важи

$$c_{I^{[3]}}(\theta) = \big(b_{I^{[3]}}(\theta)\big)^2/(16\sigma_3^2) = 0.048\theta^2 + o(\theta^2).$$

Како је Кулбак-Лајблерово растојање дато изразом (6.8), локална Бахадурова ефикасност једнака је

$$e_{I^{[3]}} = \lim_{\theta\to 0}\frac{c_{I^{[3]}}(\theta)}{2K_{g_1}(\theta)} \approx 0.573.$$

За остале алтернативе рачун је аналоган. У табели 6.1 приказане су локалне Бахадурове ефикасности за наше четири алтернативе за $k=2$ и $k=3$, као и максималне вредности (по k).

Табела 6.1: Упоредна табела локалних Бахадурових ефикасности за $I_n^{[k]}$

Алтернатива	$k=2$	$k=3$	$\max_k$
Макехамова	0.448	0.573	0.875 за $k=14$
Вејбулова	0.621	0.664	0.710 за $k=8$
гама	0.723	0.708	0.723 за $k=2$
EMNW(3)	0.694	0.799	0.885 за $k=6$

6.3 Статистика Колмогоровљевог типа $D_n^{[k]}$

Сада ћемо анализирати статистику Коломогоровљевог типа (6.1). За фиксирано $t > 0$ израз $H_n^{[k]}(t) - G_n^{[k]}(t)$ је V-статистика са следећим језгром:

$$\begin{aligned}\Xi_k(X_1, X_2, \ldots, X_k; t) &= \mathrm{I}\Big\{ \max(X_1, X_2, \ldots, X_k) \leq t\Big\} \\ &- \frac{1}{k!}\sum_{\pi(j_1,\ldots,j_k)} \mathrm{I}\Big\{\frac{X_1}{j_1} + \frac{X_2}{j_2} + \ldots + \frac{X_k}{j_k} \leq t\Big\}.\end{aligned}$$

Нека је $\upsilon_k(X_1; t)$ пројекција језгра $\Xi_k(X_1, X_2, \ldots, X_k; t)$ на X_1. Тада је

$$\begin{aligned}\upsilon_k(s; t) &= E\big(\Xi_k(X_1, X_2, \ldots, X_k; t)|X_1 = s\big) \\ &= P\Big\{ \max(s, X_2, \ldots, X_k) \leq t\Big\} \\ &- \frac{1}{k!}\sum_{\pi(j_1,\ldots,j_k)} P\Big\{\frac{s}{j_1} + \frac{X_2}{j_2} + \ldots + \frac{X_k}{j_k} \leq t\Big\} \\ &= \mathrm{I}\{s \leq t\}(F(t))^{k-1} \\ &- \frac{1}{k}\sum_{j=1}^{k}\Bigg[\mathrm{I}\{s \leq jt\}\Bigg(1 - \sum_{\substack{i=1\\ i\neq j}}^{k}\Bigg(e^{-i(t-\frac{s}{j})}\prod_{\substack{h=1\\ h\neq i,j}}^{k}\frac{h}{h-i}\Bigg)\Bigg)\Bigg],\end{aligned} \tag{6.11}$$

где је $F(t)$ функција расподеле експонцијалне расподеле. Рачунање дисперзије ове пројекције у општем случају је исувише сложено, па ћемо је рачунати само за конкретне вредности k.

6.3.1 Статистика $D_n^{[2]}$

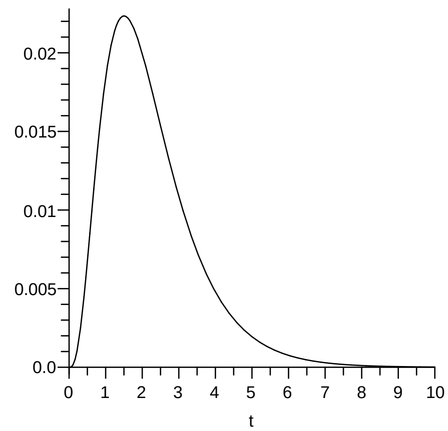
Фор $k = 2$ из израза (6.11) добијамо да је пројекција фамилије језгара $\Xi_2(X, Y; t)$ једнака

$$\xi_2(s; t) = \mathrm{I}\{s \leq t\}F(t) - \frac{1}{2}\,\mathrm{I}\{s \leq t\}F(2(t-s)) - \frac{1}{2}\,\mathrm{I}\{s \leq 2t\}F(t - s/2). \tag{6.12}$$

Дисперзије ових проејкција $\sigma_2^2(t)$ под H_0. дате су функцијом

$$\sigma_2^2(t) = \frac{1}{3}e^{-t} - \frac{5}{4}e^{-2t} - \frac{1}{3}e^{-3t} - \frac{1}{12}e^{-4t} - \frac{2}{3}e^{-3t/2} + 2e^{-5t/2} + \frac{1}{2}te^{-2t},$$

чији је график приказан на слици 6.1.



Слика 6.1: График функције $\sigma^2(t)$, $k = 2$.

Стога је наша фамилија језгара $\Xi_2(X, Y; t)$ недегенерисана (теорема 4.2.4) и важи

$$\sigma_2^2 = \sup_{t \geq 0} \sigma_2^2(t) = 0.02234.$$

Гранична расподела статистике $D_n^{[2]}$ није позната. Користећи Силвerманов резултат може се показати да

U-емпиријски случајни процес

$$\eta_n^{[2]}(t) = \sqrt{n}\left(H_n^{[2]}(t) - G_n^{[2]}(t)\right),\ t \geq 0,$$

слабо конвергира у $D(0, \infty)$ кад $n \to \infty$ неком центрираном Гаусовом процесу $\eta^{[2]}(t)$ чија се коваријација може израчунати на основу теореме. Тада низ статистика $\sqrt{n}D_n^{[2]}$ конвергира у расподели случајној величини $\sup_{t \geq 0} |\eta^{[2]}(t)|$ чију расподелу не можемо екплицитно наћи. Стога критичне вредности овог теста налазимо коришћењем симулација.

Фамилија језгара $\{\Xi_2(X, Y; t)\}, t \geq 0$, је центрирана и ограничена и недегенерисана. Применом теореме 4.2.4 о великом одступању супремума фамилије недегерисаних U-статистика (и V-статистика), добијамо следећи резултат.

Теорема 6.3.1. *За $a > 0$ важи*

$$\lim_{n\to\infty} n^{-1} \ln P_{H_0}\{D_n^{[2]} > a\} = -f_{D^{[2]}}(a),$$

где је функција $f_{D^{[2]}}$ непрекидна за довољно мало $a > 0$, и поред тога је

$$f_{D^{[2]}}(a) = (8\sigma_2^2)^{-1}a^2 + o(a^2) \sim 5.595a^2,\ a \to 0.$$

Локална Бахадурова ефикасност статистике $D_n^{[2]}$

На основу теореме 3.3.1 лимес у вероватноћи низа статистика $D_n^{[2]}$ под алтернативном хипотезом је

$$b_{D^{[2]}}(\theta) = \sup_{t\geq 0} |b_{D^{[2]}}(t,\theta)| = \sup_{t\geq 0} \Big| P_\theta\big\{ \max(X,Y) \leq t\big\} - P_\theta\Big\{X + \frac{Y}{2} \leq t\Big\}\Big|.$$

Уз претпоставку о регуларности алтернативне функције расподеле добијамо

$$b_{D^{[2]}}(t,\theta) = 2\theta \int_0^\infty \xi_2(s;t)h(s)ds + o(\theta),\ \ \theta \to 0, \tag{6.13}$$

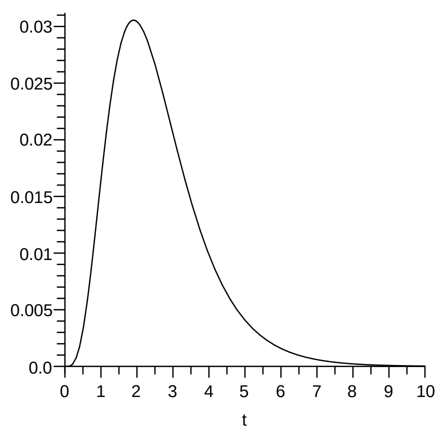
где је пројекција из (6.12).

Сада прелазимо на израчунавање локалних Бахадурових ефикасности за наше четири алтернативе.

За Макехамову алтернативу из (6.13) добијамо

$$\begin{aligned} b_{D^{[2]}}(t,\theta) &= \theta\Big(2\int_0^t F(t)e^{-s}(2-2e^{-s}-s)ds \\ &- \int_0^t F\big(2(t-s)\big)e^{-s}(2-2e^{-s}-s)ds \\ &- \int_0^{2t} F(t-s/2)e^{-s}(2-2e^{-s}-s)ds\Big) + o(\theta) \\ &= \theta\Big(\frac{2}{3}e^{-t} + (1-2t)e^{-2t} - 2e^{-3t} + \frac{1}{3}e^{-4t}\Big) + o(\theta),\ \theta \to 0, \end{aligned}$$

а график ове функције за мале вредности θ дат је на слици 6.2.



Слика 6.2: График функције $b_D^{[2]}(t,\theta)$, Макехамова алтернатива

Одавде следи да је

$$\sup_{t>0} b_{D^{[2]}}(t,\theta) = b_{D^{[2]}}(1.908,\theta) + o(\theta) = 0.03055\,\theta + o(\theta),\ \theta \to 0.$$

Локални тачни нагиб статистике $D_n^{[2]}$ кад $\theta \to 0$ задовољава

$$c_{D^{[2]}}(\theta) = \big(b_{D^{[2]}}(\theta)\big)^2/(4\sigma_2^2) = 0.0104\,\theta^2 + o(\theta^2).$$

Користећи $K_{g_1}(\theta)$ из (6.8), добијамо да је локална Бахадурова ефикасност једнака

$$e_{D^{[2]}} = \lim_{\theta\to 0}\frac{c_{D^{[2]}}(\theta)}{2K_{g_1}(\theta)} \approx 0.125.$$

За остале алтернативе рачун је аналоган. Вредности Бахадурових ефикасности за све четири алтернативе приказани су у табели 6.2.

Табела 6.2: Локалне Бахадурове ефикасности за статистику $D_n^{[2]}$.

Алтернатива	Ефикасност
Макехамова	0.125
Вејбулова	0.092
гама	0.093
EMNW(3)	0.149

Можемо уочити да су ефикасности врло мале, значајно мање него у случају других тестова експоненцијалности заснованих на карактеризацијама изузев [58].

6.3.2 Статистика $D_n^{[3]}$

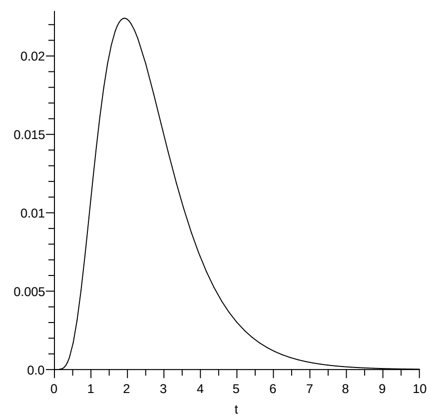
У случају $k = 3$ из (6.11) добијамо да је пројекција фамилије језгара $\Xi_3(X, Y, Z; t)$ једнака

$$\begin{aligned}
\xi_3(s;t) \quad &= \quad \mathrm{I}\{x \le t\}\Big[F^2(t) - F(2(t-x)) + \frac{2}{3}F(3(t-x))\Big] \\
&- \quad \mathrm{I}\{x \le 2t\}\Big[\frac{1}{2}F(t-x/2) - \frac{1}{6}F(3(t-x/2))\Big] \\
&- \quad \mathrm{I}\{x \le 3t\}\Big[\frac{2}{3}F(t-x/3) - \frac{1}{3}F(2(t-x/3))\Big]. \qquad (6.14)
\end{aligned}$$

Сада рачунамо дисперзије пројекција $\sigma_3^2(t)$ под H_0. Добијамо

$$\begin{aligned}
\sigma_3^2(t) \quad &= \quad \frac{8}{15}e^{-t} + \Big(\frac{1}{2}t - \frac{1}{24}\Big)e^{-2t} + \Big(\frac{41}{9} - \frac{4}{3}t\Big)e^{-3t} - \frac{179}{210}e^{-4t} + \frac{113}{210}e^{-5t} \\
&- \quad \frac{419}{2520}e^{-6t} - \frac{14}{15}e^{-3t/2} + \frac{122}{35}e^{-5t/2} - \frac{2}{3}e^{-7t/2} - \frac{2}{3}e^{-9t/2} - \frac{5}{7}e^{-5t/3} \\
&- \quad \frac{5}{2}e^{-7t/3} + \frac{10}{7}e^{-8t/3} - 4e^{-10t/3} - 2e^{-11t/3} + 2e^{-13t/3},
\end{aligned}$$

и график ове функције приказан је на слици 6.3.



Слика 6.3: График функције $\sigma_3^2(t)$.

Дакле, фамилија језгара $\Xi_3(X, Y, Z; t)$ је недегенерисана и важи

$$\sigma_3^2 = \sup_{t \geq 0} \sigma_3^2(t) = 0.02241.$$

Користећи исти начин закључивања као и у случају $D_n^{[2]}$ закључујемо да је немогуће наћи експлицитно граничну расподелу статистике $D_n^{[3]}$. Фамилија језгара $\{\Xi_3(X, Y, Z; t)\}$, $t \geq 0$, је центрирана и ограничена и применом теореме 4.2.4 добијамо следеће тврђење

Теорема 6.3.2. *За $a > 0$ важи*

$$\lim_{n \to \infty} n^{-1} \ln P_{H_0}(D_n^{[3]} > a) = -f_{D^{[3]}}(a),$$

где је функција $f_{D^{[3]}}$ непрекидна за довољно мало $a > 0$, и поред тога је

$$f_{D^{[3]}}(a) = (18\sigma_3^2)^{-1} a^2 (1 + o(1)) \sim 2.479 a^2, \ a \to 0.$$

Локална Бахадурова ефикасност статистике $D_n^{[3]}$

У овом случају лимес у вероватноћи под алтернативном хипотезом према теореми 3.3.1 једнак је

$$b_{D^{[3]}}(\theta) = \sup_{t\geq 0}|b_{D^{[3]}}(t,\theta)| = \sup_{t\geq 0}\Big|P_\theta\big\{\max(X,Y,Z)\leq t\big\} - P_\theta\Big\{X+\frac{Y}{2}+\frac{Z}{3}\leq t\Big\}\Big|.$$

Није тешко показати да $b_{D^{[3]}}(t,\theta)$ за регуларне алтернативе задовољава релацију

$$b_{D^{[3]}}(t,\theta) = 3\theta\int_0^\infty \xi_3(s;t)h(s)ds + o(\theta), \quad \theta\to 0. \tag{6.15}$$

Као и у претходним случајевима прво ћемо израчунати локалну Бахадурову ефикасност за Макехамову алтернативу. Из (6.15) добијамо да је

$$\begin{aligned}
b_{D^{[3]}}(t,\theta) = \theta\bigg(&\int_0^t\Big[F^2(t) - F(2(t-s)) + \frac{2}{3}F(3(t-s))\Big]e^{-s}(2-2e^{-s}-s)ds\\
&-\int_0^{2t}\Big[\frac{1}{2}F(t-s/2) - \frac{1}{6}F(3(t-s/2))\Big]e^{-s}(2-2e^{-s}-s)ds\\
&-\int_0^{3t}\Big[\frac{2}{3}F(t-s/3) - \frac{1}{3}F(2(t-s/3))\Big]e^{-s}(2-2e^{-s}-s)ds\bigg)\\
=\theta\Big(&\frac{8}{5}e^{-t} + \Big(\frac{9}{2}-6t\Big)e^{-2t} - 8e^{-3t} + 2e^{-4t} - \frac{1}{10}e^{-6t}\Big) + o(\theta), \theta\to 0,
\end{aligned}$$

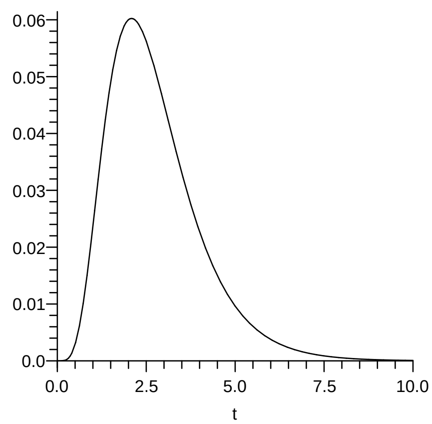
а график ове функције за мале вредности θ приказан је на слици 6.4. Дакле добијамо да је

$$\sup_{t>0} b_{D^{[3]}}(t,\theta) = b_{D^{[3]}}(2.087,\theta) = 0.0602\,\theta + o(\theta), \quad \theta\to 0.$$

Локални тачни нагиб низа $D_n^{[3]}$ кад $\theta\to 0$ задовољава релацију

$$c_{D^{[3]}}(\theta) = \big(b_{D^{[3]}}(\theta)\big)^2/(9\sigma_3^2) = 0.018\,\theta^2 + o(\theta^2),$$

а локална Бахадурова ефикасност једнака је $e_{D^{[3]}} = 0.216$.



Слика 6.4: График функције $b_D^{[3]}(t,\theta),$ Макехамова алтернатива

У табели 6.3 приказане су вредности локалне Бахадурове ефикасности за наше алтернативе.

Табела 6.3: Локалне Бахадурове ефикасности за статистику $D_n^{[3]}$

Алтернатива	Ефикасност
Макехамова	0.216
Вејбулова	0.152
Гама	0.138
EMNW(3)	0.230

Можемо приметити да су ефикасности веће него у случају претходне тест статистике, али још увек прилично ниске.

6.4 Услови локалне оптималности

Ефикасности наших тестова су веома мале у поређењу с максималним. Ипак, постоје посебне алтернативне расподеле за које су наши низови статистика $I_n^{[k]}$ и $D_n^{[k]}$ локално асимптотски оптимални (ЛАО) у Бахадуровом смислу.

Претпоставимо да за функције расподеле G из $\mathcal{G}$ поред услова регуларности из 4.3 важи и

$$\int_0^\infty h^2(x)e^x dx < \infty.$$

Под овим условима за статистику $I_n^{[k]}$ важи

$$b_{I^{[k]}}(\theta) = (k+1)\theta \int_0^\infty \xi_k(x)h(x)dx + o(\theta),\ \ \theta \to 0.$$

Уведимо помоћну функцију

$$h_0(x) = h(x) - (x-1)\exp(-x)\int_0^\infty uh(u)du. \tag{6.16}$$

Одавде директно следи

$$\begin{aligned} \int_0^\infty h^2(x)e^x dx - \Big(\int_0^\infty xh(x)dx\Big)^2 &= \int_0^\infty h_0^2(x)e^x dx, \\ \int_0^\infty \xi_k(x)h(x)dx &= \int_0^\infty \xi_k(x)h_0(x)dx. \end{aligned} \tag{6.17}$$

Израз за локалну Бахадурову ефикасност је

$$\begin{aligned} e_{I^{[k]}} &= \lim_{\theta\to 0}\frac{\big(b_{I^{[k]}}(\theta)\big)^2}{2(k+1)^2\sigma_k^2 K(\theta)} \\ &= \Big(\int_0^\infty \xi_k(x)h_0(x)dx\Big)^2 \Big/ \Big(\int_0^\infty \xi_k^2(x)e^{-x}dx \cdot \int_0^\infty h_0^2(x)e^x dx\Big). \end{aligned}$$

Из неједнакости Коши-Шварца добијамо да је ефикасност једнака један ако је $h_0(x) = C_1 e^{-x}\xi(x)$ за неку константу $C_1 > 0$, тако да је $h(x) = e^{-x}(C_1\xi(x) + C_2(x-1))$ за неке константе $C_1 > 0$ и C_2. Ове расподеле припадају домену ЛАО у оквиру класе $\mathcal{G}$.

Најједноставније такве алтернативе $g(x,\theta)$, за мало $\theta > 0$ су, за $k = 2$,

$$g(x,\theta) = e^{-x}\big(1 + \frac{\theta}{3}(\frac{4}{3}e^{-x} - e^{-2x} - \frac{1}{2}e^{-x/2})\big),$$

а за $k = 3$,

$$g(x,\theta) = e^{-x}\big(1 + \frac{\theta}{4}(\frac{5}{2}e^{-x} - 3e^{-2x} + e^{-3x} - \frac{3}{8}e^{-x/2} - \frac{1}{3}e^{-x/3})\big).$$

Размотримо сада статистику Колмогоровљевог типа (6.1). Може се показати да је

$$b_{D^{[k]}}(\theta) = k\theta \int_0^\infty \xi_k(x;t)h(x)dx + o(\theta),\ \ \theta \to 0.$$

За $h_0(x)$ дефинисано у (6.16), поред (6.17) важи и

$$\int_0^\infty \xi_k(x;t)h(x)dx = \int_0^\infty \xi_k(x;t)h_0(x)dx.$$

У овом случају ефикасност је једнака

$$\begin{aligned} e_{D^{[k]}} &= \lim_{\theta\to 0} \frac{\big(b_{D^{[k]}}(\theta)\big)^2}{\sup_{t\geq 0}\big(2k^2\sigma_k^2(t)K(\theta)\big)} \\ &= \frac{\sup_{t\geq 0}\Big(\int_0^\infty \xi_k(x;t)h_0(x)dx\Big)^2}{\sup_{t\geq 0}\Big(\int_0^\infty \xi_k^2(x;t)e^{-x}dx \cdot \int_0^\infty h_0^2 e^x dx\Big)}. \end{aligned}$$

Из неједнакости Коши-Шварца добијамо да је ефикасност једнака јединици ако $h(x) = e^{-x}\big(C_1\xi_k(x;t_0) + C_2(x-1)\big)$ за $t_0 = \mathrm{argmax}_{t\geq 0}\sigma_k^2(t)$ и неке константе $C_1 > 0$ и C_2. Алтернативне густине које имају такве функције $h(x)$ формирају домен ЛАО у класи $\mathcal{G}$.

Најједноставнији примери дати су, за $k = 2$,

$$
\begin{aligned}
g(x,\theta) = e^{-x}\Big[1 + \theta\cdot\ \mathrm{I}\{x \le t_2\}\big(1 - e^{-t_2}\big) - \frac{1}{2}\theta\cdot\Big(\ \mathrm{I}\{x \le t_2\}\big(1 - e^{-2(t_2-x)}\big) \\
+\ \mathrm{I}\{x \le 2t_2\}\big(1 - e^{-(t_2-x/2)}\big)\Big)\Big],
\end{aligned}
$$

где је

$$
\begin{aligned}
t_2 &= \operatorname*{argmax}_{t\ge 0}\Big(\frac{1}{3}e^{-t} - \frac{5}{4}e^{-2t} - \frac{1}{3}e^{-3t} - \frac{1}{12}e^{-4t} - \frac{2}{3}e^{-3t/2} + 2e^{-5t/2} + \frac{1}{2}te^{-2t}\Big)\\
&\approx 1.502,
\end{aligned}
$$

а у случају $k = 3$,

$$
\begin{aligned}
g(x,\theta) = e^{-x}\Big[1 + \theta\cdot\ \mathrm{I}\{x \le t_3\}\Big(\big(1 - e^{-t_3}\big)^2 + e^{-2(t_3-x)} - \frac{2}{3}e^{-3(t_3-x)} - \frac{1}{3}\Big)\\
- \frac{1}{3}\theta\cdot\ \mathrm{I}\{x \le 2t_3\}\Big(1 - \frac{3}{2}e^{-(t_3-x/2)} + \frac{1}{2}e^{-3(t_3-x/2)}\Big)\\
- \frac{1}{3}\theta\cdot\ \mathrm{I}\{x \le 3t_3\}\Big(1 - 2e^{-(t_3-x/3)} + e^{-2(t_3-x/3)}\Big)\Big],
\end{aligned}
$$

где је

$$
\begin{aligned}
t_3 &= \operatorname*{argmax}_{t\ge 0}\Big[\frac{8}{15}e^{-t} + \Big(\frac{1}{2}t - \frac{1}{24}\Big)e^{-2t} + \Big(\frac{41}{9} - \frac{4}{3}t\Big)e^{-3t} - \frac{179}{210}e^{-4t}\\
&+ \frac{113}{210}e^{-5t} - \frac{419}{2520}e^{-6t} - \frac{14}{15}e^{-3t/2} + \frac{122}{35}e^{-5t/2} - \frac{2}{3}e^{-7t/2} - \frac{2}{3}e^{-9t/2}\\
&- \frac{5}{7}e^{-5t/3} - \frac{5}{2}e^{-7t/3} + \frac{10}{7}e^{-8t/3} - 4e^{-10t/3} - 2e^{-11t/3} + 2e^{-13t/3}\Big]\\
&\approx 1.919.
\end{aligned}
$$

6.5 Анализа моћи тестова

Поред асимптотске ефикасности још један стандардни начин испитивања квалитета тестова и њихових упоређивања је и анализа моћи тестова. Иако су асимптотске ефикасности тестова интегралног типа задовољавајуће, оне за тестове Колмогоровљевог типа, барем у случају стандардних алтернатива, прилично су ниске, па је стога извршена и анализа моћи тестова да би се боље сагледала потенцијална примена ових тестова у пракси.

Коришћењем Монте Карло метода с 10000 понављања оцењене су моћи теста за обим узорка $n = 100$ и различите вредности параметра θ за исте алтернативе за које је рачуната и ефикасност. Резултати су приказани у табели 6.4 за тестове интегралног типа, а у табели 6.5 за тестове Колмогоровљевог типа, за $k = 2, 3,$ и 4, за нивое значајности $\alpha = 0.05$ и 0.025.

Теоријски гледано, концепт моћи теста ближе је повезан с Хаџис-Леманововом ефикасношћу (в. [57]) и редослед тестова по Бахадуровој ефикасности обично се не поклапа с редоследом тестова по моћима. Упркос томе, можемо видети да се редослед тестова слаже скоро у потпуности гледано по k, а прилично и гледано по алтернативама, с изузетком Вејбулове расподеле која предњачи по моћи а ефикасност јој није највећа.

Анализом, како Бахадурове ефикасности, тако и моћи тестова, закључујемо да можемо препоручити тестове интегралног типа за примену у пракси. Тестови Колмогоровљевог типа приказују веома ниске ефикасности, док су им моћи нешто прихватљивије. Као такви они у начелу нису препоручљиви за употребу, међутим у неким случајевима се и на њих може рачунати, нарочито ако су потенцијалне алтернативне расподеле блиске локално оптималним.

Табела 6.4: Симулиране моћи тестова са статистиком $I_n^{[k]}$.

Алтернатива	θ	k	$\alpha = 0.05$	$\alpha = 0.025$
Макехамова	1.5	2	0.51	0.39
	1.5	3	0.59	0.48
	1.5	4	0.63	0.51
	0.5	2	0.19	0.12
	0.5	3	0.22	0.13
	0.5	4	0.22	0.14
	0.25	2	0.11	0.06
	0.25	3	0.12	0.07
	0.25	4	0.13	0.08
Вејбулова	0.5	2	1.00	0.99
	0.5	3	1.00	0.99
	0.5	4	1.00	1.00
	0.25	2	0.74	0.61
	0.25	3	0.77	0.66
	0.25	4	0.78	0.68
гама	0.5	2	0.86	0.76
	0.5	3	0.85	0.76
	0.5	4	0.84	0.75
	0.25	2	0.42	0.30
	0.25	3	0.42	0.30
	0.25	4	0.42	0.30
EMNW(3)	0.5	2	0.98	0.97
	0.5	3	0.98	0.97
	0.5	4	0.98	0.97
	0.25	2	0.45	0.33
	0.25	3	0.47	0.34
	0.25	4	0.47	0.34

Табела 6.5: Симулиране моћи тестова са статистиком $D_n^{[k]}$

Алтернатива	θ	k	$\alpha = 0.05$	$\alpha = 0.025$
Макехамова	1.5	2	0.34	0.24
	1.5	3	0.41	0.29
	1.5	4	0.44	0.34
	0.5	2	0.13	0.08
	0.5	3	0.15	0.10
	0.5	4	0.15	0.10
	0.25	2	0.09	0.05
	0.25	3	0.09	0.05
	0.25	4	0.10	0.06
Вејбулова	0.5	2	0.88	0.80
	0.5	3	0.93	0.89
	0.5	4	0.95	0.91
	0.25	2	0.40	0.29
	0.25	3	0.47	0.35
	0.25	4	0.50	0.38
гама	0.5	2	0.45	0.32
	0.5	3	0.48	0.36
	0.5	4	0.48	0.36
	0.25	2	0.19	0.12
	0.25	3	0.20	0.13
	0.25	4	0.20	0.13
EMNW(3)	0.5	2	0.67	0.54
	0.5	3	0.68	0.55
	0.5	4	0.68	0.55
	0.25	2	0.22	0.14
	0.25	3	0.23	0.16
	0.25	4	0.23	0.16

Закључак

У дисертацији је наведено и доказано осам нових карактеризација, од којих су две за Паретову и шест за експоненцијалну расподелу. Карактеризације Паретове расподеле добијене су трансформацијом постојећих карактеризација, док карактеризације експоненцијалне расподеле представљају уопштења досадашњих резултата. Додатно је развијен метод доказивања у случају када се функција густине може развити у Маклоренов ред. Поред теорема карактеризације доказано је и неколико индентитета са Стирлинговим бројевима друге врсте. Ти резултати могу имати и самостални значај.

У делу посвећеном тестовима, предложено је шест тестова сагласности с Паретовом и две класе тестова сагласности с експоненцијалном расподелом. Коришћењем теорије развијене последњих двадесетак година, проучена је асимптотика тест статистика, велика одступања, и израчуната је локална Бахадурова ефикасност против различитих блиских алтернатива. Вредности ефикасности указују на то да новодобијени тестови углавном поседују квалитет да буду коришћени у пракси.

Правац даљег истраживања наравно укључује стварање нових карактеризација. Постоји велики број својстава расподела за које је потребно одредити оптималне услове под којима је одређено својство карактеризација. Ти услови могу бити различите врсте, од оних који дефинишу класу у оквиру које се карактерише дата расподела, до оних у којима се услов поставља на само својство, нпр. да важи за неки конкретан или за сваки обим узорка.

Код тестова сагласности, очигледан правац даљег истраживања је формирање нових тестова коришћењем модела приказаних у дисертацији за неке нове карактеризације. Друга, занимљивија могућност је коришћење неких других типова тест статистика које нису нужно засноване на U-емпиријским функцијама расподеле, већ на неким U-статистикама, U-емпиријским карактеристичним функцијама, итд.

Литература

[1] I. Ahmad, I. Alwasel, A goodness-of-fit test for exponentiality based on the memoryless property, J. R. Stat. Soc. Ser. B Stat. Methodol. vol.61(3) (1999) 681–689.

[2] M. Ahsanullah, On a characterization of the exponential distribution by spacings, Ann. Inst. Statist. Math. vol.30(1) (1978) 163–166.

[3] M. Ahsanullah, I. Bairamov, A characterization of uniform distribution, Istatistik, vol.2(3) (1999) 145–151.

[4] M. Ahsanullah, G.G. Hamedani, *Exponential distribution: Theory and Methods*, NOVA Science, New York, 2010.

[5] M. Ahsanullah, M. Rahman, A Characterization of the Exponential Distribution, J. Appl. Probab. vol.9(2) (1972) 457–461.

[6] J.E. Angus, Goodness-of-fit Test for Exponentiality Based on Loss of Memory Type Functional Equation, J. Statist. Plann. Inference vol.6(3) (1982) 241–251.

[7] B.C. Arnold, Some Characterizations of Cauchy Distribution, Aust. J. Statist. vol.21(2) (1979) 166–169.

[8] B.C. Arnold, N. Balakrishnan, H.N. Nagaraja, *A First Course in Order Statistics*, SIAM, Philadelphia, 2008.

[9] B.C. Arnold, J.A. Villasenor, Exponential characterizations motivated by the structure of order statistics in samples of size two, Statist. Probab. Lett. vol.83(2) (2013) 596–601.

[10] G. Arslan, A new characterization of the power distribution, J. Comput. Appl. Math. vol.260 (2014) 99–102.

[11] R.R. Bahadur, Stochastic comparison of tests, Ann. Math. Statist. vol.31(2) (1960) 276–295.

[12] R.R. Bahadur, Rates of Convergence of Estimates and Test Statistics, Ann. Math. Statist. vol.38(2) (1967) 303–324.

[13] R.R. Bahadur, *Some Limit Theorems in Statistics*, SIAM, Philadelphia, 1971.

[14] N. Balakrishnan, C.R. Rao, *Order Statistics, Theory & Methods*, Elsevier, Amsterdam, 1998.

[15] L. Baringhaus, N. Henze, Test of fit for exponentiality based on a characterization via the mean residual life function, Statist. Papers 41(2) (2000) 225–236.

[16] S.N. Bernstein, On the property characteristic of the normal law, Trudy Leningrad. Polytechn. Inst. (1941) 21–22; Collected papers, Vol IV, Nauka, Moscow (1964) 314–315.

[17] A. Castaño-Martinez, F. López-Blázquez, B. Salamanca-Miño, Random translations, contractions and dilations of order statistics and records, Statistics vol.46(1) (2012) 57–67.

[18] S. Chakraborty, G.P. Yanev, Characterization of exponential distribution through equidistribution conditions for consecutive maxima, J. Stat. Appl. Probab. vol.2(3) (2013) 237–242.

[19] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, Springer, New York, 2008.

[20] L.-Y. Deng, E. O. George, Some characterizations of the uniform distribution with applications to random number generation, Ann. Inst. Statist. Math. vol.44(2) (1992) 379–385.

[21] M.M. Desu, A characterization of the exponential distribution by order statistics, Ann. Math. Statist. vol.42(2) (1971) 837–838.

[22] J. Galambos, S. Kotz, *Characterizations of Probability Distributions*, Springer-Verlag, Berlin-Heidelberg-New York, 1978.

[23] J. Galambos, I. Simonelli, Comments on a recent limit theorem of Quine, Statist. Probab. Lett. vol.63(1) (2003) 89–95.

[24] P. Groeneboom, G. Shorack, Large deviation of goodness-of-fit statistics and linear combination of order statistics, Ann. Probab. vol.9(6) (1981) 971–987.

[25] J.L. Gross, *Combinatorial Methods with Computer Applications*, Chapman & Hall, 2008.

[26] S. Gulati, S. Shapiro, Goodness of Fit Tests for the Pareto Distribution, Statistical Models and Methods for Biomedical and Technical Systems, Birkhauser, Boston, (Vonta, F., Nikulin, M., Limnios, N., Huber, C., eds) (2008) 263–277.

[27] T. Hashimoto, S. Shirahata, A Goodness of Fit Test Based on a Characterization of Uniform Distribution, J. Japan Statist. Soc. vol.23(2) (1993) 123–130.

[28] R. Helmers, P. Janssen, R. Serfling, Glivenko-Cantelli properties of some generalized empirical DF's and strong convergence of generalized L-statistics, Probab. Theory Related Fields vol.79(1) (1988) 75–93.

[29] N. Henze, S.G. Meintanis, Goodness-of-fit tests based on a new characterization of the exponential distribution, Comm. Statist. Theory Methods vol.31(9) (2002) 1479–1497.

[30] N. Henze, S.G. Meintanis, Recent and classical tests for exponentiality: a partial review with comparisons, Metrika vol.61(1) (2005) 29–45.

[31] W. Hoeffding, A Class of Statistics with Asymptotically Normal Distribution, Ann. Math. Statist. vol.19(3) (1948) 293–395.

[32] W. Hoeffding, On the distribution of the expected values of the order statistics, Ann. Math. Statist, vol.24(1) (1953) 93–100.

[33] J.S. Huang, On a Theorem of Ahsanullah and Rahman, J. Appl. Probab. vol.11(1) (1974) 216–218.

[34] W.J. Huang, L.S. Chen, Note on a characterization of gamma distributions, Statist. Probab. Lett. vol.8(5) (1989) 485–487.

[35] C. Huber-Carol, N. Balakrishnan, M.S. Nikulin, M. Mesbah, Goodness-of-fit test and model validity, Birkhauser, Boston, 2002.

[36] H.M. Jansen van Rensburg, J.W.H. Swanepoel, A class of goodness-of-fit tests based on a new characterization of the exponential distribution, J. Nonparametr. Stat. vol.20(6) (2008) 539–551.

[37] P. L. Janssen, *Generalized empirical distribution functions with statistical applications*, Diepenbeek, Limburgs Universitair Centrum, 1988.

[38] V. Jevremović, A note on mixed exponential distribution with negative weights, Statist. Probab. Lett. vol.11(3) (1991) 259–265.

[39] M. Jovanović, B. Milošević, Ya. Yu. Nikitin, M. Obradović, K. Yu. Volkova, Tests of exponentiality based on Arnold-Villasenor characterization, and their efficiencies, arXiv preprint (2014) arXiv:1407.5014.

[40] A. M. Kagan, Yu. V. Linnik, C. R. Rao, *Characterization Problems in Mathematical Statistics* (translated from Russian), John Wiley and Sons, New York, 1973.

[41] L.B. Klebanov, A characterization of the normal distribution by a property of order statistics, Math. Notes vol.13(1) (1973) 71–73.

[42] V.S. Korolyuk, Yu.V. Borovskikh, *Theory of U-statistics*, Kluwer, Dordrecht, 1994.

[43] I. Kotlarski, On characterizing the gamma and the normal distribution, Pacific J. Math. vol.20(1) (1967) 69–76.

[44] S. Kotz, F.W. Steutel, Note on a characterization of exponential distributions, Statist. Probab. Lett. vol.6(3) (1988) 201–203.

[45] H.L. Koul, A test for new better than used, Comm. Statist. Theory Methods vol.6(6) (1977) 563–574.

[46] H.L. Koul, Testing for new is better than used in expectation, Comm. Statist. Theory Methods vol.7(7) (1978) 685–701.

[47] V.M. Kruglov, A Characterization of the Gaussian Distribution, Stoch. Anal. Appl. vol.31(5) (2013) 872–875.

[48] Yu. V. Linnik, Linear forms and statistical criteria (I,II) Ukrainian Math. J. vol.5(2) (1953) 207–243; vol.5(3) (1953) 247–290.

[49] V. V. Litvinova, Ya. Yu. Nikitin, Two families of normality tests based on Polya-type characterization and their efficiencies, J. Math. Sci. (N.Y.) vol.139(3) (2006) 6582–6588.

[50] E. Lukacs, A characterization of the normal distribution, Ann. Math. Statist. vol.13(1) (1942) 91–93.

[51] E. Lukacs, A characterization of the gamma distribution, Ann. Math. Statist. vol.26(2) (1955) 319–324.

[52] A. W. Marshall, I. Olkin, *Life distributions: Structure of nonparametric, semiparametric, and parametric families*, Springer, New York, 2007.

[53] B. Milošević, M. Obradović, Some characterizations of exponential distributions based on order statistics, arXiv preprint (2014) arXiv:1412.5019.

[54] K. Morris, D. Szynal, Goodness of Fit Tests Based on Characterizations of Continuous Distributions, Appl. Math. vol.27(4) (2000) 475–488.

[55] P. Muliere, Ya. Yu. Nikitin, Scale-invariant test of normality based on Polyas characterization, Metron vol.60(1-2) (2002) 21–33.

[56] Ya.Yu. Nikitin, Local Asymptotic Bahadur Optimality and Characterization Problems, Theory Probab. Appl. vol.29(1) (1984) 79–92.

[57] Ya.Yu. Nikitin, *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, New York, 1995.

[58] Ya.Yu. Nikitin, Bahadur Efficiency of Test of Exponentiality Based on Loss of Memory Type Functional Equation, J. Nonparametr. Stat. vol.6(1) (1996) 13–26.

[59] Ya.Yu. Nikitin, Large Deviations of U-empirical Kolmogorov-Smirnov Test, and Their Efficiency, J. Nonparametr. Stat. vol.22(5) (2010) 649–668.

[60] Ya.Yu. Nikitin, I. Peaucelle, Efficiency and Local Optimality of Nonparametric Tests Based on U- and V-statistics, Metron vol.62(2) (2004) 185–200.

[61] Ya.Yu. Nikitin, E.V. Ponikarov, Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics, Proceedings of Saint-Petersburg Mathematical Society, vol.7 (1999) 124–167; English translation in AMS Translations, ser. 2, vol.203 (2001) 107–146.

[62] Ya.Yu. Nikitin, K.Yu. Volkova, Asymptotic efficiency of exponentiality tests based on order statistics characterization, Georgian Math. J. vol.17(4) (2010) 749–763.

[63] H.A. Noughabi and N.R. Arghami, Testing Exponentiality Based on Characterizations of the Exponential Distribution, J. Stat. Comput. Simul. vol.81(11) (2011) 1641–1651.

[64] M. Obradović, On Asymptotic Efficiency of Goodness of Fit Tests for Pareto Distribution Based on Characterizations, Filomat (2015) accepted for publication.

[65] M. Obradović, Three Characterizations of Exponential Distribution Involving the Median of Sample of Size Three, arXiv preprint (2014) arXiv:1412.2563.

[66] M. Obradović, M. Jovanović, B. Milošević, Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics, Statistics (2014) published online 1–16. DOI: 10.1080/02331888.2014.919297

[67] V. Papathanasiou, Some characterizations of distributions based on order statistics, Statist. Probab. Lett. vol.9(2) (1990) 145–147.

[68] G. Pólya, Herleitung des Gauschen Fehlergesetzes aus einer Funktionalgleichung, Math. Z. vol.18 (1923) 96–108.

[69] P.S. Puri, H. Rubin, A Characterization Based on Absolute Difference of Two I.I.D. Random Variables, Ann. Math. Statist. vol.41(6) (1970) 2113–2122.

[70] M. Raghavachari, On theorem of Bahadur on the rate of convergence of test statistics, Ann. Math. Statist. vol.41(5) (1970) 1695–1699.

[71] M. Riedel, H.-J. Rossberg, Characterization of the exponential distribution function by properties of the difference $X_{k+s:n} - X_{k:n}$ of order statistics, Metrika vol.41(1) (1994) 1–19.

[72] M.L. Rizzo, New Goodness-of-fit Tests for Pareto Distribution, Astin Bull. vol.39(2) (2009) 691–715.

[73] H.-J. Rossberg, Characterization of the exponential and the Pareto distributions by means of some properties of the distributions which the differences and quotients of order statistics are subject to, Statistics vol.3(3) (1972) 207–216.

[74] D. Roy, A Generalization of the Lack of Memory Property and Related Characterization Results, Comm. Statist. Theory Methods vol.33(12) (2004) 3145–3158.

[75] R.J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York, 2002.

[76] B.W. Silverman, Convergence of a class of empirical distribution functions of dependent random variables, Ann. Probab. vol.11(3) (1983) 745–751.

[77] V.P. Skitovich, Linear forms of independent random variables and the normal distribution law, Izv. Akad. Nauk SSSR Ser. Mat. vol.18(2) (1954) 185–200.

[78] Y.-H. Too, G.D. Lin, Characterizations of uniform and exponential distributions, Statist. Probab. Lett. vol.7(5) (1989) 357–359.

[79] O. Vasicek, A test for normality based on sample entropy, J. R. Stat. Soc. Ser. B Stat. Methodol. vol.38(1) (1976) 54–59.

[80] K.Yu. Volkova, On asymptotic efficiency of exponentiality tests based on Rossbergs characterization, J. Math. Sci. (N.Y.) vol.167(4) (2010) 486–494.

[81] K.Yu. Volkova, Tests of exponentiality based on Yanev-Chakraborty characterization, and their efficiency, arXiv preprint (2014) arXiv:1405.7210.

[82] K.Yu. Volkova, On asymptotic efficiency of goodness-of-fit tests for the Pareto distribution based on its characterization, arXiv preprint (2014) arXiv:1408.4527.

[83] K.Yu. Volkova, Ya. Yu. Nikitin, On the asymptotic efficiency of normality tests based on the Shepp property, Vestnik St. Petersburg Univ. Math. vol.42(4) (2009) 256–261.

[84] K.Yu. Volkova, Ya Yu Nikitin, Goodness-of-Fit Tests for the Power Function Distribution Based on the Puri-Rubin Characterization and Their Efficiences, J. Math. Sci. (N.Y.) vol.199(2) (2014) 130–138.

[85] J. Weselowski, M. Ahsanullah, Switching order statistics through random power contractions, Aust. N. Z. J. Stat. vol.46(2) (2004) 297–303.

[86] H.S. Wieand, A condition under which the Pitman and Bahadur approaches to efficiency coincide, Ann. Statist. vol.4(5) (1976) 1003–1011.

[87] J.L. Xu, On a Characterization of the Normal Distribution by a Property of Order Statistics, Sankhyā Ser. A vol.60(1) (1998) 145–149.

[88] G.P. Yanev, S. Chakraborty, Characterizations of exponential distribution based on sample of size three, Pliska Stud. Math. Bulgar. vol.23 (2013) 237–244.

[89] G.F. Yeo, R.K. Milne, On characterization of exponential distributions, Statist. Probab. Lett. vol.7(4) (1989) 303–305.

Биографија аутора

Марко Обрадовић рођен је 24. децембра 1978. године у Крушевцу, где је завршио основну школу и два разреда гимназије. Међународну средњу школу *Collegio del mondo unito dell'Adriatico* завршио 1997. године у Дуину (Италија). Математички факултет у Београду, смер Вероватноћа и статистика, завршио је 2002. године с просечном оценом 9.85, а магистарске студије на истом смеру завршио 2007. године с просечном оценом 10 и одбрањеном тезом под насловом *Вероватноће разарања у процесима ризика с Гама маргиналним расподелама.*

На Математичком факултету запослен је од 2003. године, најпре као асистент приправник, а од 2009. године као асистент. Држао је вежбе из следећих предмета: *Вероватноћа и статистика, Увод у вероватноћу, Увод у статистику, Математичка статистика, Теорија вероватноћа, Линеарни статистички модели, Елементи актуарске математике, Статистички практикум 2, Теорија информације, Теорија узорака, Одабрана поглавља случајних процеса - стохастичка анализа, Биостатистика, Статистика у метеорологији, Математика (за студенте биохемије).* Ангажован је у комисијама за мастер радове, од којих је до сада двадесет четири одбрањено.

Учествовао је на пројектима Министарства за науку Републике Србије: *Стохастички процеси, екстремне вредности и примене у анализи временских серија, Геометрија, образовање и визуализација* и *Безбедност и заштита организовања и функционисања васпитно образовног система у Републици Србији,* те на Темпус пројекту *Master in Applied Statistics.*

Марко Обрадовић бави се научним истраживањем из области математичке статистике и теорије вероватноће, а заинтересован је и за примене статистике у другим наукама. Говори енглески, италијански и француски, а служи се и руским језиком.

До сада је објавио следеће научне радове:

1. M. Unkašević, I. Tošić, M. Obradović, Spectral analysis of the Koshava wind, Theoretical and applied climatology (ISSN: 0177-798X) vol.89(3-4) (2007) 239–244. **IF: 1.674**

2. V. Jevremović, M. Obradović, Bertrand's paradox – is there anything else?, Quantity & Quality (ISSN: 0033-5177) vol.46(6) (2012) 1709–1714. **IF: 0.768**

3. M. Obradović, M. Jovanović, B. Milošević, Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics, Statistics, DOI: 10.1080/02331888.2014.919297. (2014) 1–16. **IF: 1.594**

4. M. Obradović, M. Jovanović, B. Milošević, V. Jevremović, Estimation of $P\{X \leq Y\}$ for Geometric-Poisson Model, Hacettepe Journal of Mathematics and Statistics, DOI: 10.15672/HJMS.2014267477 (2014) 1–16. **IF: 0.433**

5. M. Obradović, M. Jovanović, B. Milošević, Optimal unbiased estimates of $P\{X < Y\}$ for some families of distributions, Metodološki zvezki - Advances in Methodology and Statistics (ISSN 1854-0031) vol.11(1) (2014) 21–29.

6. M. Obradović, On Asymptotic Efficiency of Goodness of Fit Tests for Pareto Distribution Based on Characterizations, Filomat (accepted for publication) (2015) 1–14. **IF: 0.753**

Имао је следећа саопштења на конференцијама:

1. M. Obradović, Ruin probabilities for risk processes with gamma distributed claim sizes, XII Srpski matematički kongres, 28. avgust - 2. septembar 2008, Novi Sad.

2. V. Jevremović, M. Obradović, Da li je Bertranov paradoks paradoks?, Treći simpozijum "Matematika i primene", 28-29. maj 2012, Matematički fakultet, Beograd.

3. M. Obradović, M. Jovanović, B. Milošević, Optimal unbiased estimates of $P\{X < Y\}$, Applied statistics, 22-25. septembar 2013, Ribno(Bled), Slovenija.

4. M. Obradović, On Asymptotic Efficiency of Goodness of Fit Tests for Pareto Distribution Based on Characterizations, XIII Srpski matematički kongres, 22-25. maj 2014, Vrnjačka Banja.

5. V. Božin, V. Lekić, B. Milošević, M. Obradović, Testiranje Markovljevog svojstva na konkretnim podacima korišćenjem funkcije transformacije i analiza konvolucionog ponašanja distribucije verovatnoće, XLI SYM-OP-IS, Divčibare, 16-19. septembar 2014, 648–652.

6. M. Minić, M. Obradović, Estimating parameters using Ranked Set Sampling Applied statistics, 22-25. septembar 2014, Ribno(Bled), Slovenija.

7. Z. Vidović, B. Milošević, M. Obradović, K. Ilijević, Tests of normality and their sensitivity against particular alternatives, Applied statistics, 21-24. septembar 2014, Ribno(Bled), Slovenija.

8. V. Jevremović, B. Milošević, M. Obradović, Karakterizacije raspodela verovatnoća s posebnim osvrtom na eksponencijalnu raspodelu, Peti simpozijum "Matematika i primene", 17-18. oktobar 2014, Matematički fakultet, Beograd.

Прилог 1.

Изјава о ауторству

Потписани Марко Обрадовић

број уписа ______________________________

Изјављујем

да је докторска дисертација под насловом

Карактеризације неких расподела и Бахадурова асимптотска ефикасност тестова сагласности

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанда

У Београду, 12. фебруара 2015.

МОбрадовић

Прилог 2.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Марко Обрадовић

Број уписа ______________________________

Студијски програм Вероватноћа и математичка статистика

Наслов рада Карактеризације неких расподела и Бахадурова асимптотска ефикасност тестова сагласности

Ментор проф. др Слободанка Јанковић

Потписани Марко Обрадовић

изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанда

У Београду, 12. фебруара 2015.

М Обрадовић

Прилог 3.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Карактеризације неких расподела и Бахадурова асимптотска ефикасност тестова сагласности

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство

2. Ауторство - некомерцијално

3. Ауторство – некомерцијално – без прераде

4. Ауторство – некомерцијално – делити под истим условима

5. Ауторство – без прераде

6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанда

У Београду, 12. фебруара 2015.

МОбрадовић

1. Ауторство - Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. Ауторство – некомерцијално. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. Ауторство - некомерцијално – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. Ауторство - некомерцијално – делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. Ауторство – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. Ауторство - делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.