УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

Бојана Милошевић

АСИМПТОТСКА СВОЈСТВА НЕПАРАМЕТАРСКИХ ТЕСТОВА ЗАСНОВАНИХ НА U-СТАТИСТИКАМА И V-СТАТИСТИКАМА СА НЕДЕГЕНЕРИСАНИМ И СЛАБО ДЕГЕНЕРИСАНИМ ЈЕЗГРОМ

— Докторска дисертација —

Београд, 2016.

UNIVERSITY OF BELGRADE
FACULTY OF MATHEMATICS

Bojana Milošević

ASYMPTOTIC PROPERTIES OF NON-PARAMETRIC TESTS BASED ON U-STATISTICS AND V- STATISTICS WITH NON-DEGENERATE AND WEAKLY DEGENERATE KERNEL

— Doctoral Dissertation —

Belgrade, 2016.

Ментор:

др Павле Младеновић, редовни професор
Математичког факултета Универзитета у Београду

Чланови комисије:

др Слободанка Јанковић, редовни професор
Математичког факултета Универзитета у Београду

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

Датум одбране:

Наслов докторске дисертације: *Асимптотска својства непараметарских тестова заснованих на U-статистикама и V-статистикама са недегенерисаним и слабо дегенерисаним језгром*

Резиме: Тестови сагласности са расподелом и тестови симетрије заузимају значајно место у непараметарској статистици. Велики број класичних тестова сагласности са расподелом је заснован на посматрању растојања између претпостављене функције расподеле и њене постојане оцене, емпиријске функције расподеле. Аналогно су направљени и тестови симетрије у којима се посматра разлика између оцена функције расподеле и репа расподеле. Један од новијих приступа који је посебно атрактиван последњих година подразумева прављење тестова на основу карактеризација расподела различитог типа. У таквим тестовима се јављају уопштења емпиријских функција расподеле U-емпиријске функције расподеле, U-емпиријске трансформације расподела (нпр. Лапласова трансформација, карактеристична функција итд.) и U-емпиријски моменти расподеле. Главна предност оваквих тестова је то што су често слободни од неког параметра расподеле па је њима омогућено тестирање сложених нултих хипотеза. За поређење тестова све више се користи Бахадурова ефикасност, за чије одређивање није неопходна асимптотска нормалност тест-статистике. Још једна предност ове врсте ефикасности је што се локално често поклапа са Питмановом ефикасношћу. Испоставља се да је за одређивање Бахадурове ефикасности неопходно наћи функцију великих одступања уколико важи нулта хипотеза. Уколико то није могуће, најчешће се одређује приближна Бахадурова ефикасност за коју је довољно познавање граничне расподеле тест статистике и њене граничне вредности у вероватноћи уколико важи алтернативна хипотеза.

Циљеви ове дисертације су формирање нових тестова сагласности и симетрије заснованих на U-статистикама и V-статистикама, одређивање граничних расподела нових статистика, одговарајућих функција великих одступања и одређивање Бахадурове ефикасности теста или њене апроксимације. Дисертација се састоји од два дела.

Први део се састоји од три поглавља. Прво је посвећено теорији U-статистика и V-статистика, као и U-емпиријским и V-емпиријским функцијама расподеле и другим емпиријским функцијама. Друго поглавље је посвећено U-статистикама и V-статистикама са оцењеним параметрима. У трећем поглављу дефинисана је асимптотска ефикасност непараметарских тестова. Највише пажње је посвећено Бахадуровој асимптотској ефикасности. У истом поглављу нађена је и функција великих одступања за нову класу тест статистика. Резултат је

приказан у раду [69].

Други део, који почиње четвртим поглављем, посвећен је новим тестовима заснованим на U-статистикама и V-статистикама. У четвртом поглављу су представљени типови карактеризација на основу којих су конструисани тестови у наредним поглављима и то карактеризације расподела на основу једнако расподељених статистика у којима су посебно издвојене карактеризације симетричних расподела, на основу функционалних једначина које задовољава функција расподеле, на основу независности статистика и на основу момената. У овом поглављу су показане и две нове карактеризације симетричних расподела. У петом поглављу су представљени нови тестови сагласности са расподелом и то четири теста сагласности са фамилијом експоненцијалних расподела, два теста сагласности са фамилијом Паретових расподела, два теста сагласности са фамилијом логистичких расподела и једна нова класа тестова сагласности са униформном $\mathcal{U}[0,1]$ расподелом којим је могуће тестирати сагласност са било којом унапред задатом непрекидном расподелом (в. [66], [69], [67]). Шесто поглавље посвећено је новим тестовима симетрије. Представљено је пет нових тестова симетрије. У новим тестовима поред U-статистика са недегенерисаним језгром јављају се и статистике са слабо дегенерисаним језгром. У седмом поглављу су представљени тестови сагласности са фамилијом експоненцијалних расподела заснованих на U-статистикама и V-статистикама са оцењеним параметрима. Посебно су издвојени тестови у којима се јављају U-емпиријске Лапласове трансформације. Поред ових представљени су и тестови засновани на U-емпиријским моментима. За тестове је поред Бахадурове, одређена и Питманова асимптотска ефикасност. На крају дисертације, у осмом поглављу приказан је кратак осврт на неке примене представљених тестова у анализи временских серија.

Кључне речи: U-статистике, асимптотска ефикасност, теорија великих одступања, тестови сагласности, тестови симетрије

Научна област: Математика

Ужа научна област: Вероватноћа и статистика

УДК: 519.224+519.234.3(0.433)

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

Doctoral Dissertation Title: *Asymptotic properties of non-parametric tests based on U-statistics and V-statistics with non-degenerate and weakly degenerate kernel*

Abstract: Goodness of fit and symmetry tests occupy a significant part of nonparametric statistic. Most of classical tests are based on the distance between the assumed distribution function and its consistent estimate, empirical distribution function. The symmetry tests are analogously constructed. A new approach that is especially attractive in recent years is making tests based on characterizations of different types. Those tests use U-empirical distribution functions (generalized empirical ones), U-empirical transforms (eg. Laplace transform, characteristic functions etc.) and U-empirical moments of distributions. The main advantage of these tests is that they are often free of some distribution parameters. Therefore they are suitable for testing composite hypothesis.

For purpose of comparison of tests the Bahadur efficiency has become very popular. One of the reasons is that it does not require the asymptotic normality of test statistics. In addition, Bahadur and Pitman efficiencies very often locally coincide. It turns out that for determining Bahadur efficiency it is necessary to find large deviations function under null hypothesis. If that is not possible, usually the approximate Bahadur efficiency is used. It requires the existence of asymptotic distribution and the asymptotic behavior of its tail under null hypothesis, and the limit in probability under alternative distribution.

The goals of the thesis are the construction of new goodness of fit and symmetry tests based on U-statistics and V-statistics, deriving the asymptotic distribution of proposed statistics, their large deviation functions and Bahadur efficiencies or their approximations. The thesis is divided into two parts.

The first part consists of three chapters. In the first chapter the theory of U-statistics and V-statistics as well as U-empirical and V-empirical distribution function and some other empirical transforms is presented. The second chapter is devoted to the U-statistics and V-statistics with estimated parameters. The third chapter deals with asymptotic efficiency of nonparametric tests. Most of the chapter is devoted to Bahadur efficiency. In the same chapter the large deviation function for a new class of tests statistics is derived. This result is presented in [69].

The second part, which starts with the fourth chapter, is dedicated to new tests based on U-statistics and V-statistics. In the fourth chapter some type of characterizations which are used within the next chapters for construction of tests are presented. They include characterizations based on equidistribution of some statistics among which the characterizations of symmetric distributions stand out, then those based on functional equations that the distribution function satisfies, those based on the independence of statistics and those based on moments. Two new characterizations of symmetric distributions are also presented. The fifth chapter deals with new goodness of fit tests. There are four new exponentiality tests, two new goodness of fit tests for a family of Pareto distribution, as well as two new goodness of fit tests for logistic distribution (see [66], [69]). Also one new class of uniformity tests which can be used as goodness of fit test for any predetermined continuous distribution is proposed (see [67]). The sixth chapter is devoted to new symmetry tests. Five new symmetry tests are proposed. In the seventh chapter there are new exponentiality tests based on U-statistics and V-statistics with estimated parameters. A special attention is given to new tests based on U-empirical Laplace transforms. Beside those, some tests based on U-empirical moments are also presented. For new presented tests local Bahadur and/or Pitman efficiency is calculated. In the final chapter, a brief review of some applications in time series analysis is shown.

Keywords: U-statistics, asymptotic efficiency,large deviation theory, goodness-of-fit tests, symmetry tests

Scietific Area: Mathematics

Scientific Sub-area: Probability and Statistics

UDC: 519.224+519.234.3(0.433)

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

Овој Дисертацији претходило је пуно степеница и препрека које је требало превазићи уз осмех на лицу. Зато осећам потребу да се захвалим неким људима који су ми у томе доста помогли.

Захваљујем се свим професорима који су ми предавали на пренесеном знању свих ових година. Сваки је на себи сопствен начин допринео развоју мог математичког мишљења а самим тим и квалитету Дисертације. Посебну захвалност дугујем професору Павлу Младеновићу који ми је први указао поверење запосливши ме на Математичком факултету и омогућивши да се мој таленат развија, као и на својим коментарима који су унапредили текст Дисертације. Захваљујем се професорки Слободанки Јанковић чији безрезервни савети су ми значајно помогли да до ове Дисертације дође и чија је брижна пажња једном нестрпљивом бићу стизала увек у право време. Захвалила бих се и професорки Весни Јевремовић на корисним сугестијама и интелектуалној подршци, као и професору Милану Мерклеу који ми је предавао математичку статистику и допринео да је заволим.

Посебну захвалност дугујем свом колеги и добром пријатељи Марку на сатима занимљивих разговора током којих су се рађале разне идеје које смо потом спроводили у дела. Надам се да ћемо се још дружити на такав начин. Захваљујем се и свом колеги Милану који је значајно допринео систематичности научног рада, као и професору Владимиру Божину за невероватну маштовитост.

Велику захвалност дугујем професору Јакову Никитину од кога сам много тога научила и који је све време од када смо почели сарадњу, увек био ту за мене.

Хвала свим пријатељима, највише Марку и Милани који су слушали све догодовштине током писања Дисертације и помогали ми својим саветима када је то било потребно.

Захвалила бих се свом момку Драгану на разумевању и стрпљењу који су били непоходни да би писање Дисертације било право уживање. Хвала брату Љубиши на топлим речима кад су биле најпотребније.

И на крају, највећу захвалност дугујем мами Елизабети и тати Бојану за све што су ми пружили, највише за љубав и веру да ће се све лепо завршити. Посебно хвала мами што се самном у најранијем детињству играла математике и допринела да је јако заволим.

Дисертацију посвећујем родитељима и свом деда Жарку који би сигурно био јако поносан да је доживео да је види.

Панчево, јануар 2016.

Бојана Милошевић

Садржај

Први део

Теорија U-статистика и V-статистика

Поглавље 1

U-статистике и V-статистике

У теорији оцењивања као природне оцене веома често се јављају U-статистике и V-статистике. U-статистике су симетричне и непристрасне оцене неких параметара па је њихова употреба посебно значајна у непараметарској статистици где се непознати параметри другачије интерпретирају а расподела елемената узорка припада некој класи расподела. Главна заслуга за развијање теорије U-статистика приписује се Хефдингу (в. [43]). Део његовог доприноса и доприноса његових следбеника биће приказан у овом поглављу а више се може наћи у књизи [53].

Нека је $X_1, \ldots, X_n$ 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 $\Phi(x_1, \ldots, x_m)$ таква да је

$$\vartheta(F) = E(\Phi(X_1, \ldots, X_m)) = \int .... \int \Phi(x_1, \ldots, x_m) dF(x_1) \cdots dF(x_m). \tag{1.1}$$

Најмање m за које постоји оцена називамо *редом* параметра ϑ.

Уколико функција Φ није симетрична по својим аргументима можемо је заменити симетричном функцијом

$$\Phi_0(x_1, \ldots, x_m) = \frac{1}{m!} \sum_{\pi(m)} \Phi(x_{\pi_1}, \ldots, x_{\pi_m})$$

где $\pi(m)$ представља скуп свих пермутација скупа $\{1, 2, ..., m\}$. Због тога, у даљем тексту, можемо претпоставити симетричност функције Φ по својим аргументима.

Дефиниција 1.0.1. *Нека је $X_1, X_2, \ldots, X_n$ прост случајан узорак из расподеле F и нека је $\Phi : \mathbb{R}^m \mapsto \mathbb{R}$ мерљива функција. U-статистика* *реда m са језгром Φ је статистика*

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1\le i_1<\cdots\le i_m\le n} \Phi(X_{i_1},\ldots,X_{i_m}), \tag{1.2}$$

Приметимо да је U-статистика, на овај начин дефинисана, непристрасна оцена параметра ϑ.

Најједноставнији пример U-статистике реда 1 је узорачка средина. Познати примери U-статистика су и непристрасне оцене централних момената расподеле, тј. параметра

$$\vartheta = E(X_1 - EX_1)^m, \;\; m \ge 2.$$

То су U- статистике са језгром

$$\Phi(x_1,...,x_m) = \frac{1}{m!}\sum_{\pi(m)}\Big(\sum_{k=1}^{m}(-1)^{k-1}\binom{m}{k-1}x_{\pi_1}^{m-k+1}x_{\pi_2}\cdots x_{\pi_k}+(-1)^m x_{\pi_1}\cdots x_{\pi_m}\Big).$$

Специјално, за $m=2$ добијамо U-статистику са језгром

$$\Phi(x_1,x_2) = \frac{1}{2}(x_1^2+x_2^2-2x_1x_2) = \frac{1}{2}(x_1-x_2)^2. \tag{1.3}$$

Трансформацијом израза (1.2) добија се да је U-статистика са језгром (1.3) поправљена узорачка дисперзија.

Асимптотска својства U-статистика зависе од њеног реда, али и од особина њеног језгра, тј. од ранга U-статистике дефинисаног на следећи начин.

Претпоставимо да је $E|\Phi(X_1,....,X_m)|<\infty$. Означимо са Φ_c математичко очекивање

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

Нека је

$$\tilde{\Phi}(x_1,....,x_m) = \Phi(x_1,...,X_m) - \vartheta(F)$$
$$\tilde{\Phi}_c(x_1,...,x_c) = \Phi_c(x_1,...,x_c) - \vartheta(F)$$

Даље, нека је са

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

дефинисан низ функција које имају својство комплетне дегенерисаности тј. $E(g_c(x_1,...,x_{c-1},X_c)) = 0$ за $c = 1,2,...,m$. Ово својство се веома често користи у извођењу особина U-статистика које ћемо у даљем тексту навести.

Нека је $r \geq 1$ најмањи природан број за који је $g_1 = \cdots = g_{r-1} = 0$ и $g_r \neq 0$. Број r тада се назива *рангом* језгра, односно U-статистике. Ако је ранг језгра једнак један, тада језгро називамо *недегенерисаним*, у супротном их називамо *дегенерисаним*. Ако је ранг језгра једнак два, тада језгро називамо *слабо дегенерисаним* док га у случајну ранга m називамо *комплетно дегенерисаним*. Степен дегенерисаности у највећој мери утиче на тип граничне расподеле статистике.

Свака U-статистика ранга r се може представити у облику збира комплетно дегенерисаних U-статистика односно важи Хефдингова репрезентација ([43]).

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

где је $U_{nc} = \frac{1}{\binom{n}{c}} \sum\limits_{1\le i_1<\cdots<i_c\le n} g_c(X_{i_1},...,X_{i_c})$.

Коришћењем ове репрезентације и мартингалских својстава U-статистика која нећемо наводити (в. [53]) може се показати да се дисперзија U-статистике чије језгро припада неком L^2 простору и има ранг r, може представити на два начина. Важи

$$
\sigma^2(U_n) = \frac{1}{\binom{n}{m}} \sum_{c=r}^{m} \binom{m}{c}\binom{n-m}{m-c}\eta_c,
$$

где је $\eta_c = E\tilde{\Phi}_c^2(X_1,...,X_c)$. Други начин да се представи дисперзија је

$$
\sigma^2(U_n) = \sum_{c=r}^{m} \frac{\binom{m}{c}^2}{\binom{n}{c}}\delta_c,
$$

где је $\delta_c = Eg_c^2(X_1, ..., X_c)$.

Из ове две репрезентације дисперзије, може се извести веза између η_c и δ_c, али и извести следећа асимптотска релација

$$\sigma^2(U_n) = r!\binom{m}{r}^2 \frac{1}{n^r}\eta_c + O\Big(\frac{1}{n^{r-1}}\Big).$$

Приметимо да су U-статистике збирови једнако расподељених случајних величина. Уколико је $m = 1$ ради се о збиру независних једнако расподељених случајних величина па, уз услов да језгро припада L^2 простору и да је недегенерисано, можемо применити централну граничну теорему и добити да $\sqrt{n}(U_n - \vartheta)$ има асимптотски нормалну $\mathcal{N}(0, \eta_1)$ расподелу. Можемо применити и закон великих бројева и добити да U_n скоро сигурно конвергира ка θ. Поставља се питање да ли гранична вредност и гранична расподела U-статистике постоје и који су у случају реда већег од 1. Одговор на ово питање даћемо у следећем одељку.

1.1 Асимптотска својства U-статистика

Следећа теорема представља закон великих бројева за U-статистике. Доказ се може наћи у [53].

Теорема 1.1.1. *Нека је U_n U-статистика са језгром Φ за које важи $E|\Phi(X_1, ..., X_m)| < \infty$. Тада за U_n важи јаки закон великих бројева, тј. низ U_n скоро сигурно конвергира параметру ϑ.*

Најпознатија формулација централне граничне теореме за U-статистике је следећа.

Теорема 1.1.2 (Хефдинг, 1948)**.** *Нека је U_n U-статистика са недегенерисаним језгром Φ за које важи*

$$E(\Phi^2(X_1, ..., X_m)) < \infty. \tag{1.4}$$

Тада

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

Може се показати да тврђење теореме 1.1.2 важи и уколико се услов (1.4) замени слабијим условом

$$E|g_c(X_1, ..., X_c)|^{\gamma_c} < \infty, \ c = 1, 2, ..., m \tag{1.5}$$

где је $\gamma_c = \frac{2c}{2c-1}$. Овај услов је испуњен уколико је $E|\Phi(X_1, ..., X_m)|^{\frac{4}{3}} < \infty$.

У случају U-статистика са дегенерисаним језгром гранична расподела није нормална. Из методолошких разлога прво ћемо навести теорему за U-статистике другог реда са дегенерисаним језгром, а затим произвољног реда са слабо дегенерисаним језгром, и на крају, произвољног реда са рангом r. Докази ових теорема се могу наћи у [53].

Теорема 1.1.3. *Претпоставимо да је $r = 2$ и да је $E\Phi^2(X_1, X_2) < \infty$. Тада*

$$nU_n \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j(\tau_j^2 - 1),$$

где је $\{\tau_j\}$ низ независних случајних величина са нормалном $\mathcal{N}(0,1)$ расподелом, $\{\lambda_j\}$ су сопствене вредности интегралног оператора $S : f \mapsto E(\Phi(X_1, x)f(X_1))$.

У случау U-статистика са слабо дегенерисаним језгром важи следећа теорема.

Теорема 1.1.4. *Претпоставимо да је $r = 2$ и да је $E\Phi^2(X_1, ..., X_m) < \infty$. Тада*

$$nU_n \xrightarrow{d} \binom{m}{2} \sum_{j=1}^{\infty} \lambda_j(\tau_j^2 - 1),$$

где је $\{\tau_j\}$ низ независних случајних величина са нормалном $\mathcal{N}(0,1)$ расподелом, $\{\lambda_j\}$ су сопствене вредности интегралног оператора $S : f \mapsto E(g_2(X_1, X_2)f(X_2)|X1)$.

Да бисмо приказали уопштење ове теореме потребно је да дефинишемо *Викове полиноме.* Нека је $f_1, f_2, ...f_m$ ортонормиран скуп функција у неком L^2 простору и $W(\cdot)$ Гаусова случајна мера. Тада је са

$$: \tau_{j_1} ... \tau_{j_n} := \int \cdots \int f_{j_1}(x_1) \cdots f_{j_n}(x_n) W(dx_1) \cdots W(dx_n)$$

дефинисан Виков полином.

Теорема 1.1.5. *Претпоставимо да је ранг језгра r и да је*

$$E|g_c(X_1, ..., X_c)|^{\gamma_{rc}} < \infty, \; c = r, r+1, ..., m, \tag{1.6}$$

где је $\gamma_{rc} = \frac{2c}{2c-r}$. *Тада*

$$n^{\frac{r}{2}} U_n \xrightarrow{d} \binom{m}{r} \sum_{(i_1,...,i_n)=\mathbf{1}}^{\infty} h_r(i_1, ..., i_r) : \tau_{i_1} \cdots \tau_{i_r},$$

где је

$$h_r(i_1, ..., i_r) = E(g_r(X_1, ..., X_r)\phi_{i_1}(X_1)\phi_{i_r}(X_r)),$$

а $\{\phi_j\}$ *ортонормирани систем функција у неком* L^2 *простору и* : $\tau_{i_1} \cdots \tau_{i_r}$ *је Виков полином.*

1.2 V-статистике

Нека је $X_1, ..., X_n$ прост случајан узорак. Функцију

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathrm{I}\{X_i \leq x\},\ x \in \mathbb{R},$$

називамо *емпиријском функцијом расподеле.* V-статистике, познате још и под називом фон-Мизесове статистике представљају функционале од емпиријских функција расподеле (статистичке функционале). Дефнисане су на следећи начин.

Дефиниција 1.2.1. *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.*

$$\begin{aligned} V_n = \vartheta(F_n) &= \int \cdots \int \Phi(x_1, ..., x_m) dF_n(x_1) \cdots dF_n(x_m) \\ &= \frac{1}{n^m} \sum_{1 \leq i_1, ..., i_m \leq n} \Phi(X_{i_1}, ..., X_{i_m}). \end{aligned}$$

назива се V*-статистика реда* m *са језгром* Φ.

Без умањења општости можемо претпоставити да је функција Φ симетрична по својим аргументима, јер ако није можемо је симетризовати на већ приказан начин.

Приметимо да се за $m = 1$ дефиниције U-статистика и V-статистика поклапају, док за $m \geq 2$ V-статистике представљају пристрасне оцене функционала $\vartheta(F)$ дефинисаног у (1.1). У даљем тексту видећемо да у неким ситуацијама U-статистике и V-статистике са истим језгром имају углавном иста асимптотска својства стога је за њихово проучавање довољно посматрати само једну од споменутих класа статистика.

Неки од примера V-статистика су узорачка средина и узорачка дисперзија. Веома често се природно јављају у конструкцији тест-статистика за тестове сагласности, посебно заснованих на карактеризацијама. Важно је и напоменути да је неретко имплементација V-статистика много једноставнија, а и временски мање захтевна код програма који су "векторски оријентисани" (R, $Matlab$).

1.2.1 Асимптотска својства V-статистика

Докази теорема које наводимо се могу наћи у [53].

Теорема 1.2.1. *Претпоставимо да је*

$$\max\{E\Phi(X_{i_1},...,X_{i_m})^2 : 1 \leq i_1 \leq \cdots \leq i_m \leq n\} < \infty,$$

Тада

$$\sqrt{n}(U_n - \vartheta(F_n)) \xrightarrow{d} 0,\ n \to \infty,$$

и ако постоје граничне расподеле случајних величина $\sqrt{n}(U_n - \vartheta(F))$ *и* $\sqrt{n}(\vartheta(F_n) - \vartheta(F))$ *онда се оне поклапају.*

Директна последица ове теореме је да у случају недегенерисаних језгара V-статистике имају исту асимптотику као и U-статистике истог реда и језгра.

Теорема 1.2.2. *Претпоставимо да је језгро* Φ V*-статистике* V_n *реда два и да је* $E|\Phi(X_1,X_1)| < \infty$, $E\Phi(X_1,X_2) = 0$ *и* $E\Phi^2(X_1,X_2) < \infty$. *Тада*

$$nV_n \xrightarrow{d} E\Phi(X_1,X_1) + \sum_{j=1}^{\infty} \lambda_j(\tau_j^2 - 1),\ \ n \to \infty, \tag{1.7}$$

где је $\{\tau_j\}$ *низ независних случајних величина са нормалном* $\mathcal{N}(0,1)$ *расподелом,* $\{\lambda_j\}$ *су сопствене вредности интегралног оператора* $S : f \mapsto E(g_2(X_1,X_2)f(X_2)|X1)$.

Уколико је оператор S нуклеаран, односно

$$\sum_{j=1}^{\infty} |\lambda_j| < \infty,$$

онда је

$$E\Phi(X_1,X_2) = \sum_{j=1}^{\infty} \lambda_j,$$

па се (1.7) своди на

$$nV_n \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \tau_j^2, \quad n \to \infty.$$

У доказу теореме се користи већ постојећи резулатат за U-статистику реда два са слабо дегенерисаним језгром, и репрезентација V-статистике преко одговарајуће U-статистике тј.

$$\begin{aligned} nV_n &= \frac{1}{n} \sum_{i=1}^{n} \Phi(X_i, X_i) + \frac{2}{n} \sum_{1 \le i < j \le n} \Phi(X_i, X_j) \\ &= \frac{1}{n} \sum_{i=1}^{n} \Phi(X_i, X_i) + (n-1) U_n. \end{aligned}$$

На сличан начин можемо добити граничне расподеле осталих V-статистика.

Пример 1.2.1. Нека су $X_1, ..., X_n$ независне и једнако расподељене случајне величине за које је $EX_1 = \mu$ и $\sigma^2(X_1) = \sigma^2$. Посматрајмо U-статистике и V-статистике са језгром $\Phi(x_1, x_2) = x_1 x_2$

$$\begin{aligned} U_n &= \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} X_i X_j \\ V_n &= \frac{1}{n^2} \sum_{1 \le i, j \le n} X_i X_j. \end{aligned}$$

Тада је $g_1(x_1) = E(X_1 X_2 | X_1 = x_1) = x_1 \mu - \mu^2$. Разликоваћемо два случаја.

1. Нека је $\mu \neq 0$. Сада је $E g_1(X_1)^2 = \mu^2 \sigma^2$. Применом теореме 1.1.2 закључујемо да

$$\sqrt{n}(U_n - \mu^2) \xrightarrow{d} \mathcal{N}(0, 4\mu^2\sigma^2).$$

Исто важи и за статистику V_n.

2. Нека је $\mu = 0$. У овом случају ради се о слабо дегенерисаном језгру јер је $g_1(X_1) = 0$ и $g_2(x_1, x_2) = x_1 x_2$. Треба да нађемо λ за које је

$$\lambda f(x) = x E(Y f(Y)).$$

Диференцирањем добијамо да је

$$f(x) = cx + d; \ c, d \text{ су константе.}$$

Заменом у полазној једначини добијамо да је $\lambda = 1$, $d = 0$ па применом теореме 1.1.4 добијамо

$$nU_n \xrightarrow{d} \sigma^2(Z^2 - 1),$$

где је Z случајна величина са стандардном нормалном расподелом. Како је $E\Phi(X_1, X_1) = EX_1^2 = \sigma^2$ добијамо да важи

$$nV_n \xrightarrow{d} \sigma^2 Z^2.$$

До истог резултата можемо доћи и на следећи начин. Приметимо да је

$$\frac{n-1}{\sigma^2}U_n = \Big(\frac{\sum_{i=1}^n X_i}{\sqrt{n\sigma}}\Big)^2 - \sum_{i=1}^n \frac{X_i^2}{n\sigma^2}.$$

Применом централне граничне теореме добијамо да први сабирак у расподели тежи случајној величини са нормалном $\mathcal{N}(0,1)$ расподелом. На други сабирак можемо применити закон велиих бројева и добити да он у вероватноћи тежи 1, па применом теореме Слуцког добијамо тврђење.

Напомена: *Теорема Слуцког коју на овом месту користимо односи се на збирове низова случајних величина. Наиме, уколико су $\{A_n\}$ и $\{B_n\}$ низови случајних величина за које важи*

$$A_n \xrightarrow{d} A, \quad n \to \infty,$$
$$B_n \xrightarrow{p} b, \quad n \to \infty,$$

где је b константа, тада

$$A_n + B_n \xrightarrow{d} A + b, \quad n \to \infty.$$

1.3 U-емпиријске функције расподеле

Дефиниција 1.3.1. *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. Функције расподеле*

$$U_n(t) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} \mathrm{I}\{h(X_{i_1}, \ldots, X_{i_m}) \le t\}, \quad t \in \mathbb{R} \tag{1.8}$$

$$V_n(t) = \frac{1}{n^m} \sum_{1 \le i_1 \ne \cdots \ne i_m \le n} \mathrm{I}\{h(X_{i_1}, \ldots, X_{i_m}) \le t\}, \quad t \in \mathbb{R} \tag{1.9}$$

називају се редом U-емпиријска и V-емпиријска функција расподеле.

За $m = 1$ $U_n(t)$ се поклапа са емпиријском функцијим расподеле па је природно очекивати да U-емпиријске и V-емпиријске функције добро апроксимирају функцију расподеле. О томе нам говори следећа теорема ([37]).

Теорема 1.3.1 (Хелмерс, Јансен, Серфлинг (1988))**.** *За U-емпиријску функцију расподеле $U_n(t)$ и $\varepsilon > 0$ постоји позитивна константа C која не зависи од F тако да важи*

$$P\big\{ \sup_t |U_n(t) - U(t)| > \varepsilon \big\} \leq \big(1 + 4C\sqrt{[n/m]}\varepsilon\big)e^{-2[n/m]\varepsilon^2}, \tag{1.10}$$

где је $U(t) = E(U_n(t)) = P\{h\{(X_1, ..., X_m) < t\}$.

Приметимо да је за свако фиксирано t $U_n(t)$ статистика са језгром $\Phi(x_1, ..., x_m; t) = \mathrm{I}\{h(x_1, ..., x_m) < t\}$. Више о фамилијама U-статистика се може наћи у књизи [47].

За фамилију U-статистика недегенерисаност дефинишемо на следећи начин (в. [77]).

Дефиниција 1.3.2. *За фамилију U-статистика $\{U_n(t), t \in [a, b]\}$ с језгрима $\Phi(X_{i_1}, \dots, X_{i_m}; t)$ и првим пројекцијама $\phi(s_1; t) = g_1(s_1; t)$ 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.*

За овакву фамилију статистика Силверман ([95]) је показао да важи следеће тврђење. Нека је растојање ρ_q

$$\rho_q(x, y) = \sup_{t \in R} \Big| \frac{1}{q(t)}(x(t) - y(t)) \Big|, \;\; x, y \in D(-\infty, \infty),$$

при чему је q тежинска функција дефинисана на $(-\infty, \infty)$.

Теорема 1.3.2 (Силверман (1983))**.** *Претпоставимо да је $q(s) = q(1-s)$ за $0 \leq s \leq \frac{1}{2}$ и нека је $v(t) = \frac{t^{\frac{1}{2}}}{q(t)}$. Претпоставимо још да важе следећи услови*

1. *q је неопадајућа и ненегативна функција на $[0, \frac{1}{2}]$;*
2. *v је растућа на $[0, \frac{1}{2}]$ и $\lim_{t \to 0^+} v(t) = 0$;*
3. *$\int_0^{\frac{1}{2}} (\log \frac{1}{t})^{\frac{1}{2}} dv(t) < \infty$.*

Нека је $\psi(x) = q(U(x))$ за свако x. Тада $\sqrt{n}(U_n(t) - U(t))$ конвергира слабо у ρ_ψ метрици на $D(-\infty, \infty)$ ка неком центрираном Гаусовом процесу који је непрекидан у свим тачкама непрекидности функције $U(t)$.

У раду [93] Рејмгарт и Зејлен су посматрали случајни процесс

$$\eta_n(t) = \sqrt{n}(U_n(t) - U(t)), t \in R^r, \tag{1.11}$$

где су ознаке исте као у (1.10) и показали тврђење аналогно тврђењу 1.3.2.

1.4 Остале U-емпиријске функције

У овом поглављу наводимо још неке U-емпиријске функције које су због својих својстава попут U-емпиријских функција расподеле, погодне за конструкцију непараметарских тестова.

Претпоставимо да позитивна случајна величина X има функцију расподеле F и Лапласову трансформацију

$$L(t) = E(e^{-tX}) = \int_0^\infty e^{-tx} dF(x),$$

за t за које дати интеграл конвергира. Тада је *емпиријска Лапласова трансформација*

$$L_n(t) = \frac{1}{n}\sum_{i=1}^{n} e^{-tX_i} = \int_0^\infty e^{-tx} dF_n(x), \tag{1.12}$$

где је F_n емпиријска функција расподеле.

Емпиријска Лапласова трансформација има многа лепа својства међу којима су и следећа која се могу наћи у раду [20]. Та својства говоре о граничној расподели емпиријске Лапласове трансформације и постојаности оцене.

Теорема 1.4.1. *Претпоставимо да $L(s)$ постоји на интервалу $I = [-\sigma, \infty)$, $\sigma > 0$. Ако је $J \subset I$ затворени интервал тада за сваки природан број k важи*

$$\sup_{s \in J} |L_n^{(k)}(s) - L^{(k)}(s)| \to 0, \ n \to \infty \quad s.s,$$

где су $L_n^{(k)}(s)$ и $L^{(k)}(s)$ k-ти изводи емпиријске Лапласове трансформације, односно Лапласове трансформације, редом.

Теорема 1.4.2. *Нека је*

$$W_{k,n}(s) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j^k e^{-sX_j} - E(X^k e^{-sX}).$$

Тада ако $2s \in I$

$$W_{k,n}(s) \xrightarrow{D} \mathcal{N}(0, \sigma_k^2(s)),$$

где је

$$\sigma_k^2(s) = \frac{1}{2^{2k}} L^{(2k)}(2s) - (L^{(k)}(s))^2.$$

Природно се намеће оцена Лапласове трансформације неке функције случајних величина $\Phi(X_1, ..., X_m)$, тј.

$$L_{n,\Phi}(t) = \frac{1}{\binom{n}{m}} \sum_{i_1 < \cdots < i_m} e^{-t\Phi(X_1,...,X_m)},$$

коју називамо *U-емпиријска Лапласова трансформација случајне величине* $\Phi(X_1, ..., X_m)$. Аналогно се дефинише V-емпиријска функција.

Уколико случајна величина X није позитивна, природно је посматрати њену емпиријску карактеристичну функцију, односно

$$\varphi_n(t) = \frac{1}{n} \sum_{i=1}^{n} e^{itX_i} = \int_{-\infty}^{\infty} e^{itx} dF_n(x), \tag{1.13}$$

или за случајну величину $\Phi^(X_1, ..., X_m)$, њену *U-емпиријску карактеристичну функцију*

$$\varphi_{n,\Phi}(t) = \frac{1}{\binom{n}{m}} \sum_{i1 < \cdots < i_m} e^{it\Phi(X_1,...,X_m)}.$$

У раду [27] су показана следећа лепа својства емпиријске карактеристичне функције.

Теорема 1.4.3. *За фиксирано* $T > 0$ *је*

$$P\{\lim_{n\to\infty} \sup_{|t|<T} |\varphi_n(t) - \varphi(t)| = 0\} = 1.$$

Теорема 1.4.4. *Нека је* $\{Y_n(t),\ t \in R\}$ *случајни процес дефинисан са*

$$Y_n(t) = \sqrt{n}(\varphi_n(t) - \varphi(t)).$$

Тада за свако $t \in [a, b]$, $-\infty < a < b < \infty$

$$Y_n(t) \xrightarrow{d} Y(t),$$

где је $\{Y(t),\ t \in R\}$ *центриран Гаусов процес са коваријационом функцијом*

$$K(t_1, t_2) = \varphi(t_1 + t_2) - \varphi(t_1)\varphi(t_2).$$

У литератури се не могу наћи резултати о слабој конвергенцији U-емпиријских Лапласових трансформација и карактеристичних функција. Међутим тест-статистике засноване на разлици U-емпиријских и теоријских функција се веома често своде на U- статистике или V-статистике па се о њиховим асимптотским својствима закључује на основу теорије U-статистика.

Интуитивно се очекује да случајни процес $U_n^L(t) = \sqrt{n}(L_{n,\Phi}(t) - L_n(t))$, ако функција Φ задовољава услов да је за свако t језгро $e^{-t\Phi(X_1,\ldots,X_m)}$ недегенерисано, конвергира неком центрираном Гаусовом процесу. До тог закључка смо дошли јер су тада све коначнодимензионе расподеле процеса $U_n^L(t)$ асимптотски нормалне (в. [53]).

Поглавље 2

U-статистике и V-статистике са оцењеним параметрима

Веома често се у тест-статистикама јављају оцене неког параметра, односно статистика T_n се може приказати у облику

$$T_n(\hat{\lambda}) = T_n(X_1, ..., X_n; \hat{\lambda}),$$

где је $\hat{\lambda}$ оцена параметра λ на основу узорка $X_1, ..., X_n$. Посматрајмо помоћну функцију

$$T_n(\gamma) = T_n(X_1, ..., X_n; \gamma).$$

Уколико $\{T_n(\gamma), \gamma \in \Gamma\}$ представља фамилију U-статистика, тада статистику $T_n(\hat{\lambda})$ називамо *U-статистика са оцењеним параметром.* Аналогна је дефиниција V-статистике са оцењеним параметром. У тексту који следи говорићемо о U-статистикама са оцењеним параметрима, а уколико неко својство не важи за V-статистике са оцењеним параметрима то ћемо нагласити.

У поглављу 1 дата је расподела U-статистика у зависности од особина њиховог језгра. У овом поглављу ћемо приказати резултате из литературе који нам омогућавају да нађемо граничну расподелу за неке класе U-статистика са оцењеним параметрима.

Први резултати на ову тему датирају из 1982. године када је Рандлес у свом раду [92] дао довољне услове да је гранична расподела тест-статистике са оцењеним парметром нормална.

Нека је $\{T_n(\gamma), \gamma \in \Gamma\}$ фамилија U-статистика са језгром Φ реда m и

$$\theta(\gamma) = \lim_{n\to\infty} E_\lambda \Phi(X_1, ..., X_m; \gamma).$$

Теорема 2.0.5. *Претпоставимо да је*

$$\sqrt{n}(\hat{\lambda}-\lambda)=O_p(1),\ \ n\to\infty. \tag{2.1}$$

Даље, претпоставимо да постоји околина λ*, у ознаци* $K(\lambda)$*, и позитивна константа* K_1 *таква да за свако* $\gamma\in K(\lambda)$ *и сферу са центром у* γ *полупречника* d *у ознаци* $D(\gamma,d)$ *тако да је* $D(\gamma,d)\subset K(\lambda)$ *и да су испуњена следећа два услова:*

$$E(\sup_{\gamma'\in D(\gamma,d)\subset K(\lambda)}|\phi(X_1,...,X_m;\gamma')-\phi(X_1,...,X_m;\gamma)|)\leq K_1 d; \tag{2.2}$$

$$\lim_{d\to 0}E(\sup_{\gamma'\in D(\gamma,d)\subset K(\lambda)}|\phi(X_1,...,X_m;\gamma')-\phi(X_1,...,X_m;\gamma)|^2)=0.$$

Тада важи

$$\sqrt{n}(T_n(\hat{\lambda})-\theta(\hat{\lambda})-T_n(\lambda)+\theta(\lambda))\xrightarrow{p}0,\ \ n\to\infty.$$

Следећа теорема је директна последица теореме 2.0.5.

Теорема 2.0.6. *Претпоставимо да је функција* $\theta(\gamma)$ *диференцијабилна у тачки* $\gamma=\lambda$ *и да важи услов* (2.1)*. Претпоставимо да*

$$\sqrt{n}(T_n(\lambda)-\theta(\lambda))\xrightarrow{d}\mathcal{N}(0,m^2\eta_1^2), \tag{2.3}$$

где је $\eta_1>0$ *дисперзија прве пројекције језгра* Φ *на* X_1*.*

Уколико је

$$\frac{\partial}{d\gamma}\theta(\gamma)|_{\gamma=\lambda}=0 \tag{2.4}$$

Тада важи

$$\sqrt{n}(T_n(\hat{\lambda})-\theta(\lambda))\xrightarrow{d}\mathcal{N}(0,m^2\eta_1^2).$$

Уколико је

$$\begin{aligned}&\frac{\partial}{d\gamma}\theta(\gamma)|_{\gamma=\lambda}\neq 0\\ &\sqrt{n}(T_n(\lambda)-\theta(\lambda),\hat{\lambda}-\lambda)\xrightarrow{d}\mathcal{N}_{p+1}(0,\Sigma),\ \ n\to\infty\end{aligned} \tag{2.5}$$

тада је испуњено

$$\sqrt{n}(T_n(\hat{\lambda})-\theta(\lambda))\xrightarrow{d}\mathcal{N}(0,\tau^2),$$

где је

$$\begin{aligned}\tau^2&=D^T\Sigma D>0\\ D^T&=\Big(1,\frac{\partial\theta(\cdot)}{\partial\gamma_1},...,\frac{\partial\theta(\cdot)}{\partial\gamma_p}\Big).\end{aligned}$$

Видимо да ако су $T_n(\gamma)$ за свако фиксирано γ U-статистике са недегенерисаним језгром, онда и одговарајућа U-статистика са оцењеним параметром има нормалну расподелу са параметром дисперзије који се може одредити. Овај резултат је доста значајан јер нам омогућује да нађемо расподелу тест-статистика у тестовима сагласности код којих се тестира сложена хипотеза, у којима фигурише оцена непознатог параметра како би се постигло да тестови буду слободни од параметара расподеле. Део ове тезе су управо такви тестови па ће сва наведена тврђења касније бити илустрована примерима.

Ајверсон и Рандлес су у свом раду [45] показали следећи закон великих бројева за U-статистике са оцењеним параметром.

Теорема 2.0.7. *Претпоставимо да је $E(\Phi(X_1, ..., X_m; \lambda)) < \infty$ и да постоји околина тачке λ $K(\lambda)$ таква да ако је $D(\lambda, d) \subset K(\lambda)$ сфера са центром у λ полупречника d важи*

$$\lim_{d\to 0} E(\sup_{\gamma\in D(\gamma,d)\subset K(\lambda)} |\phi(X_1, ..., X_m; \gamma) - \phi(X_1, ..., X_m; \lambda)|) = 0.$$

Ако су ови услови испуњени тада важе следеће импликације:

$$\hat{\lambda} \xrightarrow{p} \lambda \Rightarrow T_n(\hat{\lambda}) \xrightarrow{p} \theta(\lambda);$$
$$\hat{\lambda} \xrightarrow{s.s.} \lambda \Rightarrow T_n(\hat{\lambda}) \xrightarrow{s.s.} \theta(\lambda).$$

Поглавље 3

Асимптотска ефикасност непараметарских тестова

Избор адекватног теста у некој конкретној ситуацији је један од важних задатака статистичара. Зато је неопходно дефинисати неке критеријуме на основу којих ћемо извршити поређење одговарајућих тестова. У овом поглављу дефинисаћемо неке типове асимптотске релативне ефикасности, и изложити основна тврђења која ће нам бити потребна у даљем тексту. Више се може наћи у књизи Никитина (в. [75]) и Серфлинга (в. [94]).

Нека је случајна величина X дефинисана на параметарском простору вероватноћа $(\Omega, \mathcal{A}, P_\theta)$, $\theta \in \Theta$ са одговарајућим узорачким простором. Означимо са $X^{(n)}$ прост случајан узорак. Нулта и алтернативна хипотеза у тестирању су

$$H_0 : \theta \in \Theta_0 \subset \Theta \quad H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0. \tag{3.1}$$

Нека је $\{T_n\}$ низ тест статистика које користимо у тестирању. Претпоставићемо да је критична област за тестирање облика $W = \{T_n \geq c\}$, где је $c \in R$ позитивна константа. Овај облик критичне области је вероватно најчешће коришћени облик. Аналогно се разматрају други облици критичне области.
За тако дефинисану критичну област функција моћи теста је

$$M(\theta) = P_\theta\{T_n \geq c\}.$$

За свако $\beta \in (0,1)$ и $\theta \in \Theta_1$ дефинишемо низ константи $c_n := c_n(\beta, \theta)$ тако да је испуњено

$$P_\theta\{T_n > c_n\} \leq \beta \leq P_\theta\{T_n \geq c_n\}. \tag{3.2}$$

Нека је $\alpha_n(\beta, \theta)$ минимални ниво значајности теста са тест статистиком T_n за који је $M(\theta) \geq \beta$, и нека је $N_T(\alpha, \beta, \theta)$ минимални обим узорка за који ће тест за фиксиран ниво значајности теста α имати моћ теста у тачки θ не мању од β.

Вратимо се основној поставци проблема. За две тест статистике T_n и V_n којима тестирамо нулту хипотезу H_0 против исте алтернативе, дефинисаћемо релативну ефикасност (V_n у односу на T_n) као однос минималних обима узорака потребних да се достигне одговарајућа моћ теста, тј.

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

Уколико је релативна ефикасност мања од 1 закључујемо да ако користимо статистику T_n биће нам потребан мањи узорак да достигнемо одговарајућу моћ па је тест заснован на овој статистици бољи него тест заснован на тест статистици V_n. Међутим, израчунавање $e_{V,T}$ је веома компликовано или није уопште могуће. С друге стране разликовање алтернатива које нису блиске није толико значајно, као ни поређење тестова за високе нивое значајности или мале моћи. Зато о квалитету тестова заснованим на поменутим статистикама закључујемо на основу асимптотске релативне ефикасности (АРЕ). У зависности од тога да ли посматрамо релативне ефикасности за ниске нивое значајности, велике моћи или блиске алтернативе разликујемо три основне врсте АРЕ:

- Бахадурова АРЕ:

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

- Хаџис-Леманова АРЕ:

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

- Питманова АРЕ:

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

У сва три случаја се подразумева да гранична вредности постоји. Поред ових АРЕ постоје АРЕ које се добијају кад нека два од три посматрана параметра теже граничним вредностима (Чернофова, Хефдингова, Рубин-Сетураманова АРЕ).

У тези ћемо више пажње посветити Бахадуровој АРЕ као и Питмановој АРЕ које имају доста сличности локалног карактера.

3.1 Питманова асимптотска ефикасност

Употреба Питманове ефикасности као мере квалитета теста датира из 1949. Питманови резултати из тог периода су у великој мери допринели развоју непараметарске статистике. Ова мера ефикасности се најчешће користи кад је гранична расподела тест статистике нормална. Следећа теорема говори о томе како уз неке додатне претпоставке можемо одредити Питманову АРЕ.

Теорема 3.1.1. *Нека је* $\Theta = R^1$ *и* $\Theta_0 = (-\infty, \theta_0)$. *Претпоставимо да низ тест статистика* $\{T_n\}$ *задовољава следеће услове:*

1. *Постоје функције средње вредности* $\mu_n(\theta)$ *и функције стандардног одступања* $\sigma_n(\theta)$ *тако да је за* $\theta = \theta_0$ *и* $\theta_n = \theta_0 + \frac{k}{\sqrt{n}}, k \geq 0$

$$\lim_{n\to\infty} P_\theta\Big\{\frac{T_n - \mu_n(\theta)}{\sigma_n(\theta)} < z\Big\} = \Phi(z), z \in R.$$

2. *Постоји десни извод* $\mu_n'(\theta_0) > 0$.

3. $$\lim_{n\to\infty} \frac{\mu_n'(\theta_n)}{\mu_n'(\theta_0)} = 1, \quad \lim_{n\to\infty} \frac{\sigma_n(\theta_n)}{\sigma_n(\theta_0)} = 1, \lim_{n\to\infty} \frac{\mu_n'(\theta_0)}{\sqrt{n}\sigma_n(\theta)} = c > 0.$$

Нека је $\tilde{T}_n$ *други низ тест статистика који задовољава ове услове код којих су асимптотске средње вредности и стандардна одступање функције* $\tilde{\mu}_n$ *и* $\tilde{\sigma}_n$. *Тада Питманова АРЕ постоји за све* $0 < \alpha < \beta < 1$ *и може се израчунати коришћењем формуле*

$$e^{\mathcal{P}}_{T,\tilde{T}}(\alpha, \beta, \theta_0) = \lim_{n\to\infty} \Big(\frac{\mu_n'(\theta_0)\tilde{\sigma_n}(\theta_0)}{\sigma(\theta_0)_n\tilde{\mu}_n'(\theta_0)}\Big)^2 = \frac{c^2}{\tilde{c}^2}. \tag{3.3}$$

Нека је $I(\theta_0)$ информациона функција Фишера. У раду [91] је показано да уколико важе неки додатни услови постоји горња граница c^2, наиме важи

$$\frac{c^2}{I(\theta_0)} \leq 1.$$

Зато се природно уводи појам *апсолутне Питманове асимптотске ефикасности* која представља количник у горњој неједнакости.

3.2 Бахадурова асимптотска ефикасност

Да бисмо приказали теореме које се односе на одређивање Бахадурове АРЕ увешћемо следеће ознаке. За свако $\theta \in \Theta$, $t \in R$ и низ тест статистика $\{T_n\}$ означимо

$$F_n(t,\theta) = P_\theta\{T_n(X^{(n)}) < t\} \text{ и } G_n(t) = \inf_{\theta \in \Theta_0} \{F_n(t,\theta)\}.$$

Тада је p-вредност теста

$$L_n(X^{(n)}) = 1 - G_n(T_n(X^{(n)})).$$

Ако постоји

$$\lim_{n\to\infty} \frac{1}{n} \log L_n \stackrel{p}{=} -\frac{1}{2} c_T(\theta),\ \theta \in \Theta_1 \tag{3.4}$$

где је $c_T(\theta)$ функција параметра θ, онда се $c_T(\theta)$ назива Бахадуров тачан нагиб. Следећа теорема нам омогућава да ако важи (3.4) Бахадурову ARE низова тест статистика $\{T_n\}$ и $\{V_n\}$ можемо одредити као количник Бахадурових нагиба.

Теорема 3.2.1. *Претпоставимо да за низ тест статистика $\{T_n\}$ важи* (3.4) *и да је $c_T(\theta) > 0$ за $\theta \in \Theta_1$. Тада је*

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

Доказ ове теореме као и докази теорема које следе у овом поглављу се могу наћи у [75].

Теорема 3.2.2. *Претпоставимо да је за $\theta \in \Theta_1$*

$$\lim_{n\to\infty} T_n \stackrel{P_\theta}{=} b(\theta),$$

при чему је $|b(\theta)| < \infty$. Даље претпоставимо да је

$$\lim_{n\to\infty} \frac{\log(1 - G_n(t))}{n} = -f(t),\ t \in I, \tag{3.5}$$

где је I неки интервал на коме је f непрекидна и $\{b(\theta),\ \theta \in \Theta_1\} \subset I$. Тада важи (3.4) *и*

$$c_T(\theta) = 2f(b(\theta)).$$

Функција f из једнакости (3.5) назива се *функција великих одступања.*

Како бисмо дошли до природне дефиниције Бахадурове ефикасности дефинисаћемо Кулбак-Лајблерово растојање између две расподеле.

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

Познато је да је Кулбак-Лајблерово растојање ненегативна величина као и да је идентички једнако нули ако и само ако се ради о истим расподелама. 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 Рагавачарија и Бахадура.

Теорема 3.2.3 (Бахадур, Рагавачари 1970)**.** *За свако $\theta \in \Theta_1$ је*

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

са вероватноћом P_θ једнаком 1.

Уколико за низ статистика $\{T_n\}$ важи (3.4) коришћењем ове теореме добијамо горњу границу Бахадуровог нагиба, тј.

$$c_T(\theta) \leq 2K(\theta,\Theta_0). \tag{3.6}$$

Апсолутна Бахадурова ефикасност се сада природно дефинише на следећи начин:

$$e_T(\theta) = \frac{c_T(\theta)}{2K(\theta,\Theta_0)}. \tag{3.7}$$

Поменути количник није увек једноставно одредити. Са друге стране, од суштинског значаја је испитати ефикасност теста против блиских алтернатива па тако долазамо до појма *локалне Бахадурове ефикасности* дефинисане са

$$e_T = \lim_{\theta\to\theta_0} \frac{c_T(\theta)}{2K(\theta,\Theta_0)}, \quad \theta_0 \in \partial\Theta_0. \tag{3.8}$$

Уколико је за неку алтернативну рапсподелу испуњено да је $e_T = 1$ такву расподелу зовемо *локално оптималном алтернативном расподелом* за наш низ тест статистика $\{T_n\}$.

Следи неколико тврђења која нам омогућавају да одредимо функцију f из теореме 3.2.2.

3.2.1 Одређивање функције великих одступања

Нека је X реална случајна величина са функцијом расподеле F. Тада је њена момент-генераторна функција $M_X(t) = Ee^{tX}$, за $t \in \mathbb{R}$ (под претпоставком да је очекивање коначно). Нека је $\rho = \inf_{t\geq 0} M_X(t)$. Тада је $0 \leq \rho \leq 1$.

Теорема 3.2.4. *Нека је $\{X_n\}$ низ независних и једнако расподељених случајних величина. Тада је*

$$P\{X_1 + X_2 + \cdots + X_n \geq 0\} \leq \rho^n, \quad n \leq 1. \tag{3.9}$$

Теорема 3.2.5. *Нека је $\{u_n\}$ низ реалних бројева такав да је $\lim_{n\to\infty} u_n = u$, $|u| < \infty$ и $P\{X > u\} > 0$. Тада важи:*

$$\lim_{n\to\infty} \frac{1}{n} \log P\{X_1 + X_2 + \cdots + X_n \geq nu_n\} = -f(u), \tag{3.10}$$

где је f функција одређена условом

$$e^{-f(u)} = \inf_{t\geq 0} e^{-tu} M_X(t). \tag{3.11}$$

Теорему 3.2.5 је формулисао и доказао Чернов 1952. године (в. [18]). Како се многе тест статистике могу представити у облику збира независних и једнако расподељених случајних величина, његов резултат је нашао широку примену (нпр. в. [1], [34], [74], [88], [82], [97], [98]).

Ако претпоставимо да је $M_X(t) < \infty$ у околини нуле и $EX_1 = 0$ и $D(X_1) = \sigma^2 > 0$ развијањем функција $M_X(t)$ и e^{-tu} у Маклоренов ред, добијамо да је

$$f(u) = \frac{u^2}{2\sigma^2}(1 + o(1)), \ u \to 0. \tag{3.12}$$

Следеће теореме се односе на одређивање функције великих одступања за U-статистике и V-статистике. Резултати које ћемо представити и који су веома коришћени у радовима аутора налазе се у раду Никитина и Поникарова [79].

Теорема 3.2.6 (Никитин, Поникаров (1999))**.** *Нека је језгро Φ U-статистике*

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

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

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

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* $n \to \infty$ *је*

$$\lim_{n\to\infty} \frac{1}{n} \ln P\{U_n \geq \varepsilon + \gamma_n\} = \sum_{j=2}^{\infty} b_j \varepsilon^j := -g_\Phi(\varepsilon), \tag{3.13}$$

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

За мале вредности ε функција f из теореме 3.2.2 може се представити у облику

$$f(\varepsilon) = \frac{1}{2m^2\sigma^2}\varepsilon^2 + o(\varepsilon^2). \tag{3.14}$$

У случају U-статистика са слабо дегенерисаним језгром важи следећа теорема.

Теорема 3.2.7 (Никитин, Поникаров 1999)**.** *Нека је језгро* Φ *ограничена функција на* $[0,1]^m$ *и нека је* λ_0 *најмање* λ *које задовољава интегралну једначину*

$$x(s_1) = \lambda \int_0^1 \Phi^*(s_1, s_2)x(s_2)ds_2, \tag{3.15}$$

односно λ_0 *је проста карактеристична вредност линеарног интегралног оператора са језгром* Φ^* *са доменом* $L^2[0,1]$ *и кодоменом* $L^2[0,1]$*, где је* $\Phi^*(s_1, s_2) = E(\Phi(X_1, X_2..., X_m)|X_1 = s_1, X_2 = s_2)$*. Тада је*

$$\lim_{n\to} n^{-1} \log P\{T_n \geq \varepsilon\} = \sum_{j=2}^{n} c_j \varepsilon^{\frac{j}{2}},$$

при чему је ред са десне стране израза конвергентан за довољно мало $\varepsilon > 0$ *и*

$$c_2 = -\frac{\lambda_0}{m(m-1)}.$$

Обе теореме важе и за одговарајуће V-статистике. Важно је истаћи да су језгра Φ функције дефинисане на $[0,1]^m$ јер се без умањења општости може сматрати да је под претпоставком да важи H_0 $X^{(n)}$ прост случајан узорак из униформне $\mathcal{U}[0,1]$ расподеле.

Напомињемо да за U-статистике и V-статистике ранга већег од два, до сада нису одређене функције великих одступања.

Посматрајмо статистику

$$K_n = \sup_{\mathbf{t}\in T} \left|U_n(\mathbf{X}_m, \mathbf{t})\right|,$$

где је $\{U_n(\mathbf{t}),\ \ \mathbf{t} \in T\}$ фамилија U-статистика реда m са центрираним, ограниченим, недегенерисаним језгром Φ, а параметарски скуп је $T = [a_1, b_1] \times \cdots \times [a_p, b_p] \subset R^p$. Претпоставимо да је фамилија U-статистика $\{U_n(\mathbf{t})\}$ недегенерисана, односно њена функција дисперзије $\sigma^2_\varphi(\mathbf{t}) = E(\varphi^2(X_1, \mathbf{t}))$ је строго позитивна за свако $\mathbf{t}$ из унутрашњости скупа T осим у коначно много тачака. Означимо

$$\sigma_0^2 = \sup_{\mathbf{t}\in T} \sigma^2_\varphi(\mathbf{t}).$$

Додатно претпоставимо још и да фамилија $\{U_n(\mathbf{t})\}$ задовољава још један нови услов, такозвани *вишедимензиони услов монотоности по параметру* који је природно уопштење услова монотоности по параметру датог у раду Никитина [77] у коме је посматран случај $p = 1$.

Уколико постоји низ подела интервала T такав да је у чворовима поделе функција дисперзије позитивна и да за свако $k_i = 0, ..N - 1$, $i = 1, ..., p$ важи

$$\sup_{\mathbf{t}\in \prod_{i=1}^{p}[t_{k_i}, t_{k_i+1}]} U_n(\mathbf{t}) \le U_n(t_{k_1+1}, ..., t_{k_p+1}) + \Delta_n(N), \tag{3.16}$$

где низ $\Delta_n(N)$ конвергира ка нули брзо, односно постоји низ τ_N који конвергира ка нули кад $N \to \infty$ тако да је брзина конвергенције дата следећим изразом

$$\lim_{N\to\infty} \lim_{n\to\infty} \frac{1}{n} \log P\{\Delta_n(N) > \tau_N\} = -\infty. \tag{3.17}$$

кажемо да фамилија $\{U_n(\mathbf{t})\}$ задовољава вишедимензиони услов монотоности по параметру.

У раду Никитина [77] је показана следећа теорема.

Теорема 3.2.8 (Никитин (2010))**.** *Нека је $\{U_n(t), t \in [a, b] \subset R\}$ фамилија U-статистика са недегенерисаним језгрима $\Phi(\cdot, t)$ која су ограничена и центрирана за свако t и задовољава услов монотоности по параметру. Тада је за $\varepsilon > 0$ испуњено*

$$\lim_{n\to\infty} \frac{1}{n} \log P\{K_n \ge \varepsilon\} = - \inf_{t\in[a,b]} g_\phi(\varepsilon, t), \tag{3.18}$$

где је $g_\phi(\varepsilon, t)$ *функција дефинисана са* $g_\phi(\varepsilon, t) := -\sum_{j=2}^{\infty} b_j(t)\varepsilon^j$*,. За довољно мало* ε *ова гранична вредност се може приказати у облику*

$$g_K(\varepsilon) := \inf_{t\in[a,b]} g_\phi(\varepsilon, t) = \frac{\varepsilon^2}{2m^2\sigma_0^2} + o(\varepsilon^2),\ \varepsilon \to 0, \tag{3.19}$$

где је $\sigma_0^2 = \sup_{t\in[a,b]} \sigma_\varphi^2(t)$*.*

У раду аутора [69] је показано уопштење када је параметарски скуп T вишедимензионалан.

Теорема 3.2.9 (Милошевић, Обрадовић 2015)**.** *Нека је* $\{U_n(\mathbf{t}), \mathbf{t} \in T\}$ *фамилија* U*-статистика са недегенерисаним језгрима* $\Phi(\cdot, \mathbf{t})$ *која су ограничена и центрирана за свако* $\mathbf{t}$ *и задовољава вишедимензиони услов монотоности по параметру. Тада је за* $\varepsilon > 0$ *испуњено*

$$\lim_{n\to\infty} \frac{1}{n} \log P\{K_n \geq \varepsilon\} = -\inf_{\mathbf{t}\in T_0} g_\phi(\varepsilon, \mathbf{t}), \tag{3.20}$$

где је $g_\phi(\varepsilon, \mathbf{t})$ *функција дефинисана са* $g_\phi(\varepsilon, \mathbf{t}) := -\sum_{j=2}^{\infty} b_j(\mathbf{t})\varepsilon^j$*,. За довољно мало* ε *ова гранична вредност се може приказати у облику*

$$g_K(\varepsilon) := \inf_{\mathbf{t}\in T} g_\phi(\varepsilon, \mathbf{t}) = \frac{\varepsilon^2}{2m^2\sigma_0^2} + O(\varepsilon^3),\ \varepsilon \to 0. \tag{3.21}$$

где је $\sigma_0^2 = \sup_{\mathbf{t}\in T} \sigma_\phi^2(\mathbf{t})$*.*

Доказ. Показаћемо да важи неједнакости

$$\overline{\lim} \frac{1}{n} \log P\{\sup_{\mathbf{t}\in T} U_n(\mathbf{t}) > \varepsilon\} \leq g_K(\varepsilon) \tag{3.22}$$

$$\underline{\lim} \frac{1}{n} \log P\{\sup_{\mathbf{t}\in T} U_n(\mathbf{t}) < \varepsilon\} \geq g_K(\varepsilon). \tag{3.23}$$

У доказу неједнакости (3.22) кључну улогу играју вишедимензиони услов монотоности по параметру и следећа лема.

Лема 3.2.1. *Нека је* U_n U*-статистика са језгром* Φ_m*,* $|\Phi_m| < M$*, које је недегенерисано и* $z > 0$*. Тада је*

$$P\{|U_n| > z\} \leq 4e^{-\frac{nz^2}{2m^2\sigma^2 + Lz}},$$

где је $L = M\big(2^{m+3}m^m + \frac{2}{3m}\big)$*.*

Доказ ове леме се може наћи у [7]. Уведимо ознаку $T_0 = \{\mathbf{t} \in T : \sigma^2_\phi(\mathbf{t}) > \frac{\sigma_0^2}{2}\}$. Нека је T подељен тако да важи (3.16) и нека је τ_N низ такав да важи (3.17). Тада је

$$\begin{aligned} P\{\sup_{\mathbf{t}\in T} U_n(\mathbf{t}) > \varepsilon\} &\leq \sum_{k_1=1}^{N-1} \cdots \sum_{k_p=1}^{N-1} P\{\sup_{\mathbf{t}\in \prod_{i=1}^{p}[t_{k_i}, t_{k_i+1}]} U_n(\mathbf{t}) \geq \varepsilon\} \\ &\leq \sum_{k_1=1}^{N-1} \cdots \sum_{k_p=1}^{N-1} P\{U_n((t_{k_1+1}, ..., t_{k_p+1})) \geq \varepsilon\} \\ &+ \sum_{k_1=1}^{N-1} \cdots \sum_{k_p=1}^{N-1} P\{\Delta_n(N) \geq \tau_N\} \\ &\leq \sum_{k_1=1}^{N-1} \cdots \sum_{k_p=1}^{N-1} P\{U_n((t_{k_1+1}, ..., t_{k_p+1})) \geq \varepsilon\} \\ &+ (N-1)^p P\{\Delta_n(N) \geq \tau_N\}. \end{aligned} \tag{3.24}$$

Због брзине конвергенције низа τ_N ка нули други сабирак је занемарљив. Посматрајмо суму

$$\sum_{k_1=1}^{N-1} \cdots \sum_{k_p=1}^{N-1} P\{U_n((t_{k_1+1}, ..., t_{k_p+1})) \geq \varepsilon\}.$$

Можемо је поделити на две суме, једну за коју су чворови поделе у скупу T_0 и другу за коју су у скупу T_0^c. Уколико је $(t_{k_1+1}, ..., t_{k_p+1}) \in T_0$ тада се на статистику $U_n((t_{k_1+1}, ..., t_{k_p+1}))$ може применити теорема за велика одступања за недегенерисане U-статистике па је

$$\begin{aligned} \frac{1}{n} \log P\{U_n((t_{k_1+1}, ..., t_{k_p+1})) > \varepsilon - \tau_n\} &= g(\varepsilon - \tau_N, (t_{k_1+1}, ..., t_{k_p+1})) + o(1) \\ &\leq g_K(\varepsilon - \tau_N) + o(1),\ n \to \infty. \end{aligned}$$

За сабирке за које $(t_{k_1+1}, ..., t_{k_p+1}) \in T_0^c$ је

$$P\{U_n((t_{k_1+1}, ..., t_{k_p+1})) > \varepsilon - \tau_N\} \leq 4e^{-\frac{n(\varepsilon-\tau_N)^2}{m^2\sigma_0^2 + L(\varepsilon-\tau_N)}}.$$

Мајорирајући израз (3.24) коришћењем ових неједнакости добија се

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log P\{\sup_{\mathbf{t}\in T} U_n(\mathbf{t}) > \varepsilon\} \leq \min\Big(- g_K(\varepsilon - \tau_N), \frac{-(\varepsilon - \tau_N)^2}{m^2\sigma_0^2 + L(\varepsilon - \tau_N)}\Big) \tag{3.25}$$

Да бисмо довршили доказивање неједнакости 3.22 потребно је још да покажемо да је функција $g_K(\cdot)$ непрекидна.

Коришћењем теореме 3.2.6 је за $\mathbf{t} \in T_0$

$$g(\varepsilon, \mathbf{t}) = \sum_{j=2}^{\infty} b_j(\mathbf{t})\varepsilon^j, \tag{3.26}$$

одакле је за $\varepsilon_1 > \varepsilon_2$ за које ред (3.26) конвергира

$$|g_K(\varepsilon_1) - g_K(\varepsilon_2)| \le \sup_{\mathbf{t}\in T_0} |\sum_{j=2}^{\infty} b_j(\mathbf{t})(\varepsilon_1^j - \varepsilon_2^j)| \le |\varepsilon_1 - \varepsilon_2| \sum_{j=2}^{\infty} k||b_j||\varepsilon_1^{k-1}.$$

Ред са десне стране конвергира па је непрекидност показана. Вратимо се сад на неједнакост (3.25). Пуштајући да $N \to \infty$ добијамо да је

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log P\{\sup_{\mathbf{t}} U_n(\mathbf{t}) > \varepsilon\} \le \min\Big(- g_K(\varepsilon), \frac{-(\varepsilon)^2}{m^2\sigma_0^2 + L(\varepsilon)}\Big) = -g_K(\varepsilon). \tag{3.27}$$

Последња једнакост важи за довољно мало ε. Овим је први део теореме доказан. Сада показујемо неједнакост (3.23). Јасно је да важи

$$P\{\sup_{\mathbf{t}\in T} U_n(\mathbf{t}) > \varepsilon\} \ge \sup_{\mathbf{t}\in T_0} P\{U_n(\mathbf{t}) > \varepsilon\} \ge P\{U_n(\mathbf{t}) > \varepsilon\},\ \mathbf{t} \in T_0.$$

Одавде је

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log P\{\sup_{\mathbf{t}} U_n(\mathbf{t}) > \varepsilon\} \ge - \inf_{\mathbf{t}\in T_0} g(\varepsilon, \mathbf{t}) = -g_K(\varepsilon).$$

Потребно је још показати да важи (3.21).

$$g_K(\varepsilon) \le \inf_{\mathbf{t}\in T_0} \frac{\varepsilon^2}{2m^2\sigma_\phi^2(\mathbf{t})} + \sup_{\mathbf{t}\in T_0} |\sum_{k=3}^{\infty} b_k(\mathbf{t})\varepsilon^k| = \frac{\varepsilon^2}{2m^2\sigma_0^2} + \sup_{\mathbf{t}\in T_0} |\sum_{k=3}^{\infty} b_k(\mathbf{t})\varepsilon^k|.$$

$$g_K(\varepsilon) \ge \inf_{\mathbf{t}\in T_0} \frac{\varepsilon^2}{2m^2\sigma_\phi^2(\mathbf{t})} + \inf_{\mathbf{t}\in T_0} \sum_{k=3}^{\infty} b_k(\mathbf{t})\varepsilon^k \ge \frac{\varepsilon^2}{2m^2\sigma_0^2} - \sup_{\mathbf{t}\in T_0} |\sum_{k=3}^{\infty} b_k(\mathbf{t})\varepsilon^k|.$$

Одавде је

$$|g_K(\varepsilon) - \frac{\varepsilon^2}{2m^2\sigma_0^2}| \le \sum_{k=3}^{\infty} \|b_k\|\varepsilon^k = o(\varepsilon^2),\ \varepsilon \to 0.$$

Познато је да је функција великих одступања за статистике Колмогоровљевог типа иста у случају једностраних и двостраних статистика (в. [75]) те је овим доказ завршен. □

3.2.2 Одређивање Бахадурове ефикасности за посебну класу алтернатива

Нека је $\mathcal{G} = \{G(x,\theta),\ \ \theta \in \Theta\}$ класа алтернатива које задовољавају следеће услове регуларности (в. [78])

1. постоје парцијални изводи $\frac{\partial g(x,\theta)}{\partial \theta}$ и $\frac{\partial G(x,\theta)}{\partial \theta}$;

2. $g(x,\theta)$ је апсолутно непрекидна функција параметра θ скоро свуда и $g(x,0) > 0$;

3. $\frac{\partial G(x,\theta)}{\partial \theta}$ је апсолутно непрекидна функција за скоро свако $\theta \in \Theta$, $\frac{\partial G(x,\theta)}{\partial \theta}\big|_{\theta=0}$ није идентички једнако нули и

$$\sup_x \sup_\theta \left| \frac{\partial G(x,\theta)}{\partial \theta} \right| \leq M, \tag{3.28}$$

$$\lim_{x \downarrow a} \frac{\partial G(x,\theta)}{\partial \theta} = \lim_{x \uparrow b} \frac{\partial G(x,\theta)}{\partial \theta} = 0; \tag{3.29}$$

4. за скоро свако $x \in [a,b]$

$$\frac{\partial^2 G(x,\theta)}{\partial x \partial \theta}\Big|_{\theta=0} = \frac{\partial^2 G(x,\theta)}{\partial \theta \partial x}\Big|_{\theta=0}. \tag{3.30}$$

Напоменимо да су ови услови задовољени за већину познатих алтернативних функција расподеле.

Нека су $h(x)$ и $H(x)$ функције дефинисане са

$$h(x) = \frac{\partial g(x,\theta)}{\partial \theta}\Big|_{\theta=0},\ \ H(x) = \frac{\partial G(x,\theta)}{\partial \theta}\Big|_{\theta=0}.$$

У раду [78] су показане следеће теореме.

Теорема 3.2.10 (Никитин, Посел 2004)**.** *Нека је U_n U-статистика са недегенерисаним језгром $\Phi(X_1,\ldots,X_m)$ и алтернативна функција расподеле $G(x,\theta)$ задовољава претходно наведене услове регуларности. Тада за $b_U(\theta)$, граничну вредност у вероватноћи под алтернативном хипотезом важи*

$$b_U(\theta) = m \int\limits_a^b \phi(x)h(x)dx \cdot \theta + o(\theta),\ \ \theta \to 0. \tag{3.31}$$

Теорема 3.2.11 (Никитин, Посел 2004)**.** *Нека је U_n U-статистика са слабо дегенерисаним језгром $\Phi(X_1, \ldots, X_m)$, Претпоставимо да фамилија алтернатива $\mathcal{G}$ задовољава претходно наведене услове регуларности. Тада за $b_U(\theta)$, граничну вредност у вероватноћи под алтернативном хипотезом важи*

$$b_U(\theta) = \frac{m(m-1)}{2} \int\limits_{R^2} \Phi^*(x_1, x_2) g_\theta(x_1, 0) g_\theta(x_1, 0) dx_1 dx_2 \cdot \theta^2 + o(\theta^2), \quad \theta \to 0. \tag{3.32}$$

3.3 Приближна Бахадурова асимптотска ефикасност

Веома често функцију великих одступања није могуће одредити. Због тога се многи тестови пореде на основу приближне Бахадурове ефикасности дефинисане на следећи начин.

Претпоставимо да за низ тест статистика $\{T_n\}$ постоји гранична расподела, тј. да постоји недегенерисана функција расподеле $F(t)$ таква да је

$$\lim_{n\to\infty} P\{T_n \le t\} = F(t),$$

где је $F(t)$ нека недегенерисана функција расподеле. Даље, претпоставимо да је

$$\lim_{t\to\infty} \log(1 - F(t)) = -\frac{a_T t^2}{2}$$

и да за свако $\theta \in \Theta_1$ постоји гранична вредност у вероватноћи P_θ

$$\lim_{n\to\infty} \frac{T_n}{\sqrt{n}} = b_T(\theta) > 0.$$

Тада је *приближна релативна Бахадурова ефикасност* (ПРБЕ) теста заснованог на низу тест статистика $\{T_n\}$ у односу на тест заснован на низу тест статистика $\{V_n\}$ за који важе исти услови,

$$e^*_{T,V} = \frac{c^*_T}{c^*_V}, \tag{3.33}$$

где је

$$c^*_T = a_T b^2_T(\theta) \tag{3.34}$$

приближни Бахадуров нагиб статистике T_n.

Приметимо да је, уколико тест статистика има нормалну граничну расподелу коефицијент a_T реципрочна вредност њене асимптотске дисперзије.

Уколико посматрамо блиске алтернативе (за мало θ) говоримо о *локалној приближној релативној Бахадуровој ефикасности.* Локална ПРБЕ се веома често поклапа са Питмановом локалном АРЕ а неретко се доста једноставније рачуна. Довољни услови да би дошло до поклапања дати су у раду [105]. Услове није увек лако проверити да важе али је понекад то једноставније него рачунање Питманове локалне АРЕ па се онда она одређује индиректно, рачунањем локалне ПРБЕ. Због своје једноставности локална ПРБЕ је коришћена у многим радовима (в. нпр. [57], [52], [14], [65]).

Други део

Тестови засновани на U-статистикама и V-статистикама

Поглавље 4

Карактеризације расподела

Стална потрага за моћнијим и ефикаснијим тестовима довела је до проналажења великог броја карактеризација расподела, тј. особина који поседује само нека фамилија расподела. Један од најчешће помињаних примера у литератури је својство одсуства памћења експоненцијалне расподеле односно следеће тврђење.

Теорема 4.0.1. *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 $y \geq 0$. Тада је $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$, где је λ позитивна константа.

Наведена карактеризација је само један од многобројних примера карактеризација експоненцијалне расподеле (в. [29], [84]). Поред карактеризација експоненцијалне расподеле у литератури се може наћи мноштво примера карактеризација и за нормалну и униформну расподелу (в. [50], [29], [6], [4], [9]).

Класификацију карактеризација није једноставно направити и могуће је датом проблему приступити на више начина. Неки критеријуми за груписање карактеризација могу бити фамилије расподела које се карактеришу као и особине расподеле које фигуришу у карактеризацијама. Приказаћемо неке типове карактеризација који су коришћени у радовима аутора за конструкцију тестова сагласности са расподелом и тестова симетрије и две нове карактеризације симетричне расподеле.

Оргинални резултати односе се на нове карактеризације расподела симетричних око нуле и приказани су теоремама 4.1.6 и 4.1.7.

4.1 Карактеризације расподела засноване на једнако расподељеним статистикама

У овом одељку наводимо тип карактеризација који је веома погодан за конструкцију тестова сагласности, јер се тест-статистике могу правити и оцењивањем функције расподеле посматраног обележја U-емпиријским и V-емпиријским функцијама расподелама, али и оцењивањем Лапласове трансформације и карактеристичне функције посматраног обележја са одговарајућим U-емпиријским и V-емпиријским функцијама.

Следеће карактеризације се односе на експоненцијалну расподелу.

Теорема 4.1.1 (Пури, Рубин 1970). *Нека су X_1 и X_2 две независне копије случајне величине X са функцијом густине $f(x)$. Тада X и $|X_1 - X_2|$ имају исту расподелу ако и само ако је $f(x) = \lambda e^{-\lambda x}$ за неко $\lambda \geq 0$.*

У наредним карактеризацијама фигуришу статистике поретка.

Теорема 4.1.2 (Десу 1971). *Нека је $F(\cdot)$ функција расподеле случајне величине X. Нека је $X_1, X_2, ...X_n$ прост случајан узорак из расподеле F и $W = \min(X_1, ..., X_n)$. Ако је $F(\cdot)$ недегенерисана функција расподеле, онда за сваки природан број n, nW и X су једнако расподељене ако и само ако је $F(x) = 1 - e^{-\lambda x}$, за $x \geq 0$, где је λ нека позитивна константа.*

Теорема 4.1.3 (Милошевић, Обрадовић 2015). *Нека је $X_1, \ldots, X_n$ прост случајан узорак апсолутно непрекидних случајних величина чија густина $f(x)$ се може развити у Маклоренов ред за свако $x > 0$ и нека је X_0 случајна величина независна од узорка која има исту расподелу. Нека је k фиксирани број такав да је $1 < k \leq n$. Ако важи*

$$X_{(k-1;n)} + \frac{1}{n-k+1} X_0 \overset{d}{=} X_{(k;n)} \tag{4.1}$$

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

Од посебног значаја ће нам бити карактеризација за $n = 2, k = 2$ и $n = 3, k = 3$. Тада се једнакост 4.1 своди на

$$X_0 + \min(X_1, X_2) \overset{d}{=} \max(X_1, X_2) \tag{4.2}$$

$$X_0 + \mathrm{med}(X_1, X_2, X_3) \overset{d}{=} \max(X_1, X_2, X_3), \tag{4.3}$$

где $\mathrm{med}(\cdot)$ означава медијану одговарајућег низа. Докази теорема 4.1.2, 4.1.1 и 4.1.3 се могу наћи у радовима [22], [89], [68]. Специјалан случај (4.3) је показан у раду [83].

4.1.1 Карактеризације симетричних расподела

У овом одељку навешћемо неколико карактеризација расподела које су симетричне око нуле, тј. за функцију расподеле случајне величине X важи

$$F(x) = 1 - F(-x), \;\; x \in R.$$

Теорема 4.1.4 (Барингхаус, Хензе 1992)**.** *Нека су X и Y независне, једнако расподељене, апсолутно непрекидне случајне величине са функцијом расподеле F. Тада је расподела случајне величине X симетрична око нуле ако и само ако су случајне величине $|X|$ и $|\max(X,Y)|$ једнако расподељене.*

Наведена карактеризација се може наћи у [12]. У раду [3] је дата следећа карактеризација.

Теорема 4.1.5 (Ахсанулах 1992)**.** *Претпоставимо да су $X_1, ..., X_k$ независне апсолутно непрекидне једанко расподељене случајне величине са функцијом расподеле $F(x)$. Означимо $X_{1,k} = \min(X_1, ..., X_k)$ и $X_{k,k} = \max(X_1, ..., X_k)$. Тада су $|X_{1,k}|$ и $|X_{k,k}|$ једнако расподељене ако и само ако је расподела случајне величине X_1 симетрична око нуле.*

Означимо са $X_{(k)}$ k-ту статистику поретка простог случајног узорка $X_1, ..., X_n$. Тада важи следеће тврђење.

Теорема 4.1.6 (Обрадовић, Милошевић 2015)**.** *Претпоставимо да је $X_1, ..., X_n$, прост случајан узорак апсолутно непрекидних случајних величина са функцијом расподеле $F(x)$ и $k \leq \frac{n}{2}$. Тада су случајне величине $|X_{(k)}|$ и $|X_{(n-k+1)}|$ једнако расподељене ако и само ако је расподела случајне величине X_1 симетрична око нуле.*

Доказ. Једнака расподељеност случајних величина $|X_{(k)}|$ и $|X_{(n-k+1)}|$ се може приказати следећом једнакошћу:

$$\sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j} = \sum_{j=0}^{n-k} \binom{n}{j} F(-x)^j (1 - F(-x))^{n-j}. \tag{4.4}$$

Десна страна једнакости је једнака суми

$$\sum_{j=k}^{n} \binom{n}{j} F(-x)^{n-j} (1 - F(-x))^{j}. \tag{4.5}$$

Посматрајмо функцију

$$r(a) = \sum_{j=k}^{n} \binom{n}{j} a^j (1-a)^{n-j}.$$

Како бисмо испитали ток ове функције нађимо њен први извод.

$$\begin{aligned}
r'(a) &= n\sum_{j=k}^{n}\binom{n-1}{j-1}a^{j-1}(1-a)^{n-j} - n\sum_{j=k}^{n}\binom{n-1}{n-j-1}a^{j-1}(1-a)^{n-j-1}\\
&= n\sum_{j=k-1}^{n-1}\binom{n-1}{j}a^{j}(1-a)^{n-j-1} - n\sum_{j=k}^{n}\binom{n-1}{j}a^{j-1}(1-a)^{n-j-1}\\
&= n\binom{n-1}{k-1}a^{k-1}(1-a)^{n-k}.
\end{aligned}$$

Како је $r'(a) \geq 0$ за $a \in [0,1]$ закључујемо да је на том интервалу $r(a)$ неопадајућа функција.

Приметимо да је израз (4.4) након трансформације (4.5) еквивалентан са

$$r(a) = r(b),\ a = F(x),\ \ b = (1 - F(-x)).$$

Како је функција $r(x)$ монотона закључујемо да је $a = b$, односно

$$F(x) = 1 - F(-x), \tag{4.6}$$

што је и требало показати.

Уколико важи (4.6) једнакост (4.4) тривијално важи за свако $x \in R$. □

Теорема 4.1.7 (Милошевић, Обрадовић 2015). *Претпоставимо да су $X_1, ..., X_{2k+1}$ независне апсолутно непрекидне једнако расподељене случајне величине са функцијом расподеле $F(x)$. Тада је расподеле узорачке медијане симетрична око нуле ако и само ако је расподела случајне величине X_1 симетрична око нуле.*

Доказ је аналоган доказу теореме 4.1.6 па ћемо га изоставити.

Следећа карактеризација се може наћи у раду [23].

Теорема 4.1.8 (Донати-Мартин, Сонг, Јор 1994). *Ако су X и Y независне и једнако расподељене случајне величине са функцијом расподеле F тада су случајне величине $|X+Y|$ и $|X-Y|$ једнако расподељене ако и само ако је случајна величина X симетрична у односу на нулу.*

4.2 Карактеризације расподела засноване на функционалним једначинама

Овде ће бити речи о функционалним једначинама коју задовољава функција расподеле или нека њена трансформација. Једна од карактеризација овог типа је већ наведено својство одсуства памћења експоненцијалне расподеле,

Следеће три карактеризације се могу наћи у [29].

Теорема 4.2.1. *Нека је F функција расподеле ненегативне, непрекидне случајне величине X. Тада за $x, y > 1$ важи следећа једнакост*

$$1 - F(xy) = (1 - F(x))(1 - F(y))$$

ако и само ако X има Паретову расподелу са функцијом расподеле

$$F(x) = 1 - x^{-\lambda}, \;\; x > 1, \lambda > 0. \tag{4.7}$$

Теорема 4.2.2. *Нека је F функција расподеле непрекидне случајне величине X која је симетрична око нуле. Тада следећа једнакост важи*

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

ако и само ако X има логистичку расподелу са функцијом расподеле

$$F(x) = \frac{1}{1 + e^{-\lambda x}}, \; \mathbf{R}, \; \lambda > 0. \tag{4.8}$$

Наредна карактеризација говори о константном очекиваном преосталом животном веку експоненцијално расподељене случајне величине.

Теорема 4.2.3. *Нека је функција расподеле F ненегативне, непрекидне случајне величине X. Тада важи следећа једнакост*

$$\int_x^{\infty} (1 - F(y))dy = \frac{1}{\lambda}(1 - F(x)), \; x > 0, \lambda > 0$$

ако и само ако је $F(x) = 1 - e^{-\lambda x}, \; x \geq 0$.

У следећим карактеризацијама се јављају трансформације расподеле попут карактеристичне функције и момент-генераторне функције.

Теорема 4.2.4 (Лин, Ху 2007)**.** *Нека су X_1, X_2, X_3 независне и једнако расподељене случајне величине са функцијом расподеле F, $F(0) = \frac{1}{2}$ и каркатеристичном функцијом ϕ. Претпоставимо да је функција $t^2\phi(t)$ апсолутно интеграбилна на R. Тада је F функција расподеле случајне величине са стандардном логистичком расподелом односно*

$$F(x) = \frac{1}{1 + e^{-x}}, \; x \in \mathbf{R}$$

ако и само ако је

$$\phi_{(2)}(t) = (1 + t^2)\phi(t), \tag{4.9}$$

где је $\phi_{(2)}(t)$ *каркатеристична функција узорачке медијане* $\mathrm{med}(X_1, X_2, X_3)$.

Једнакост (4.9) се може интерпретирати и на следећи начин. Нека X_0 је независна од узорка X_1, X_2, X_3 и има исту расподелу тада је 4.9 еквивалентно са

$$X_0 = \mathrm{med}(X_1, X_2, X_3) + L,$$

где је L случајна величина са са стандардном Лапласовом расподелом са карактеристичном функцијом $\phi(t) = \frac{1}{(1+t^2)}$, $t \in R$. Поред ове теореме у раду [59] се може наћи и уопштење за узорачку медијану узорка обима $m > 3$, као и слична тврђења која карактеришу логистичку расподелу у којима фигуришу екстремне вредности узорка.

У раду [41] представљена је нова карактеризација експоненцијалне расподеле на основу својства које карактеристична функција $\phi(t) = u(t) + iv(t)$ задовољава.

Теорема 4.2.5 (Хензе, Меинтанис 2002). *Нека ненегативна случајна величина X има непрекидно диференцијабилну функцију густине f тако да постоји $\lim_{x\to 0^+} f(x)$ и $f'(x)$ је интеграбилна функција. Тада је*

$$v(t) - \frac{1}{\theta} t u(t) = 0,$$

ако и само X има експоненцијалну $\mathcal{E}(\theta)$ расподелу.

У раду [32] представљена је карактеризација нормалне расподеле.

Теорема 4.2.6 (Гош 1981). *Нека је X случајна величина чија је момент-генераторна функција $M_X(t)$. Тада X има нормалну $\mathcal{N}(0,1)$ расподелу ако и само ако*

$$M'_X(t) = t,\ t \in R.$$

4.3 Карактеризације расподела засноване на независности статистика

Карактеризација којом започињемо одељак једна је од првих карактеризационих теорема овог типа и односи се на нормалу расподелу (в. [13]).

Теорема 4.3.1 (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.*

Лукач је формулисао карактеризациону теорему за гама расподелу (в. [61]).

Теорема 4.3.2 (Лукач 1955). *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 случајне величине X и Y имају гама расподеле с истим параметром скалирања.*

Приметимо да се у теореми не поставља услов једнаке расподељености случајних величина X и Y. Међутим, приликом прављења одговарајућих тест-статистика тај услов се природно намеће уколико се ради о тестирању хипотезе на основу простог случајног узорка.

Још једна карактеризација гама расподеле овог типа дата је у раду [54].

Теорема 4.3.3 (Котларски, 1967). *Нека су X_1, X_2 и X_3 независне позитивне случајне величине и U_1, U_2*

$$U_1 = \frac{X_1}{X_1 + X_2}, \quad U_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3}.$$

Потребан и довољан услов да би случајне величина X_k имала гама расподелу $\Gamma(p_k, a)$, за $k = 1, 2, 3$, је да су случајне величине U_1 и U_2 независне са бета расподелама $\mathcal{B}(p_1, p_2)$ и $\mathcal{B}(p_1 + p_2, p_3)$, редом.

Карактеризација овог типа за експоненцијалну расподелу може се наћи у раду [28].

Теорема 4.3.4 (Фис 1958). *Нека су апсолутно непрекидне случајне величине X и Y независне и једнако расподељене. Случајне величине $\min\{X, Y\}$ и $|X - Y|$ су независне ако и само ако X и Y имају функцију расподеле $F(x) = 1 - e^{-\lambda x}$, $x > 0$, $\lambda > 0$.*

4.4 Карактеризације расподела засноване на моментима

Карактеризације у чијим формулацијама учествују моменти расподеле су веома честе. Навешћемо неке од њих.

Карактеризација степене расподеле преко рекурентне везе момената дата је у раду Туа и Лина (в. [99]) следећом теоремом.

Теорема 4.4.1 (Ту и Лин 1989). *Нека су $X_1, \dots X_n$ независне једнако расподељене случајне величине чији моменти одговарајућих редова постоје. Тада важи*

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

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

$$EX_1^2 - EX_{(2;2)} + \frac{2}{3} = 0.$$

Теорема 4.4.2 (Папатанасију 1990). *Нека су X_1 и X_2 независне и једнако расподељене случајне величине. Важи неједнакост*

$$Cov(\min(X_1, X_2), \max(X_1, X_2)) \leq \frac{1}{3} DX_1,$$

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

Напомињемо да је ово карактеризација произвољне униформне расподеле. Њена формулација и доказ се могу наћи у [87]. Још једна карактеризација у којој фигуришу неједнакости између момената статистика поретка може се наћи у [86].

Теорема 4.4.3 (Пападатос 1999). *Нека су X_1, X_2 и X_3 независне и једнако расподељене случајне величине. Важи неједнакост*

$$Cov(\min(X_1, X_2, X_3), \max(X_1, X_2, X_3)) \leq \frac{6}{a^2} DX_1,$$

где је $a \approx 5.96941$ јединствено решење једначине $\tanh \frac{a}{2} = \frac{a}{6}$, при чему се једнакост достиже ако и само X_1 има синус хиперболичку расподелу са густином

$$g(x, \mu, \sigma) = \frac{1}{a\sqrt{(x-\mu)^2 + \lambda^2\sigma^2}}, \quad |x - \mu| < a\sigma\sqrt{\frac{2}{a^2 - 24}},$$

где је $\lambda = \sqrt{\frac{2(36-a^2)}{a^2-24}} \approx 0.25089.$

Поглавље 5

Тестови сагласности

Тестови сагласности заузимају веома важно место у непараметарској статистици и зато се у последњих неколико деценија у литератури може наћи много примера таквих тестова. Највише је тестова сагласности са експоненцијалном, нормалном и Паретовом расподелом. Један од разлога за то је велика примена, а други је велики број карактеризација поменутих расподела које могу послужити као основа за конструкцију тест-статистика. Тестови засновани на карактеризацијама веома су актуелни последњих година нпр. у радовима [5], [2] и [55] предложени су тестови засновани на својству *одсуства меморије* експоненцијалне расподеле, док су радовима [56], [44] предложени тестови експоненцијалности засновани на карактеризацији о *константном преосталом животном веку*.

У тестовима заснованим на једнако расподељеним статистикама природно се јављају U-статистике, зато је теорија U-статистика заузела значајно место у одређивању граничне расподеле тест-статистика. Примери таквих тестова експоненцијалности могу се наћи у радовима [80], [101], [102]. Тестови нормалности овог типа могу се наћи у радовима [72], [104]. Од тестова сагласности са Паретовом расподелом заснованих на карактеризацијама једнако расподељених статистика наводимо [103], [49].

Тестова сагласности заснованих на карактеризацијама независности неких статистика није било све до 2015. године када је у раду [11] предложена једна класа ових тестова.

Тестова заснованих на карактеризацијама које укључују моменте има доста у литератури. Први тест сагласности са униформном расподелом овог типа био је предложен у раду [36] заснован на каратеризацији из рада [87]. Тест заснован на карактеризацији из рада [99] био је предложен у раду [70]. Исти аутори предложили су тестове засноване на карактеризацијама које садрже и моменте рекорда случајних величине (в. [71]).

У поменутим радовима као мера квалитета најчешће је коришћена Бахадурова асимптотска ефикасност. Један од главних разлога за то је што се она може рачунати и кад гранична расподела тест-статистике није нормална.

Организација овог поглавља је следећа. Прво ћемо представити нове тестове сагласности (са експоненцијалном расподелом) засноване на карактеризацијама једнако расподељених статистика. Ти резултати су садржани у раду аутора [66]. Следећи резултати односе се на тестове засноване на независности статистика узорка (са експоненцијалном расподелом) и тестове засноване на функционалним једначинама коју задовољава функција расподеле обележја (са Паретовом и логистичком расподелом). Нове тест-статистике интегралног типа су оригинални резултата аутора док су нове тест-статистике Колмогоровљевог типа представљене у коауторском раду [69]. На крају су представљени тестови засновани на карактеризацији у којој фигуришу моменти расподела и који за разлику од претходно наведених, имају просту нулту хипотезу односно да обележје има неку, унапред дефинисану функцију расподеле F_0. Резултати су садржани у раду аутора [99]. У свим предложеним тестовима подразумевамо да одбацујемо нулту хипотезу уколико је вредност тест-статистике велика. За сваки од тестова биће одређена Бахадурова асимптотска ефикасност против блиских алтренатива, и уколико је могуће, нађена класа локално оптималних алтернатива за сваки од представљених тестова.

5.1 Тестови засновани на карактеризацијама на основу једнако расподељених статистика

Нека је $X_1, X_2, ..., X_n$ прост случајан узорак. Конструисаћемо тест експоненцијалности заснован на карактеризацији из теореме 4.1.3 за случај који смо посебно издвојили (4.3).

Тест-статистике које предлажемо су:

$$I_n = \int_0^{\infty} (H_n(t) - G_n(t))dF_n(t), \tag{5.1}$$

$$K_n = \sup_{t \geq 0} \big| H_n(t) - G_n(t) \big|, \tag{5.2}$$

где су

$$G_n(t) = \frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathrm{I}\{X_i + \mathrm{med}(X_j, X_k, X_l) < t\},$$
$$H_n(t) = \frac{1}{n^3}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathrm{I}\{\max(X_j, X_k, X_l) < t\}.$$

Након интеграције добијамо да је

$$\begin{aligned} I_n = \frac{1}{n^5}\sum_{i_1,..,i_5}\Big(&\mathrm{I}\{\max(X_{i_2}, X_{i_3}, X_{i_4}) < X_{i_5}\} \\ &- \mathrm{I}\{X_{i_1} + \mathrm{med}(X_{i_2}, X_{i_3}, X_{i_4}) < X_{i_5}\}\Big). \end{aligned}$$

Приметимо да I_n и K_n не зависе од параметра скалирања λ. То је јако важна особина предложених тест-статистика која нам омогућава да тестирамо сложену нулту хипотезу.

Тест-статистика I_n асимптотски је еквивалентна V-статистици са симетричним језгром (в. [53])

$$\begin{aligned} \Psi(X_1, X_2, X_3, X_4, X_5) = \frac{1}{5!}\sum_{\pi(1:5)}\Big(&\mathrm{I}\{\max(X_{\pi_2}, X_{\pi_3}, X_{\pi_4}) < X_{\pi_5}\} \\ &- \mathrm{I}\{X_{\pi_1} + \mathrm{med}(X_{\pi_2}, X_{\pi_3}, X_{\pi_4}) < X_{\pi_5}\}\Big). \end{aligned}$$

Њена пројекција на X_1 ако важи H_0 је

$$\begin{aligned} \psi(s) &= E(\Psi(X_1, X_2, X_3, X_4, X_5)|X_1 = s) \\ &= \frac{1}{5}\Big(P\{\max(X_2, X_3, X_4) < X_5\} - P\{s + \mathrm{med}(X_2, X_3, X_4) < X_5\}\Big) \\ &+ \frac{3}{5}\Big(P\{\max(s, X_3, X_4) < X_5\} - P\{X_2 + \mathrm{med}(s, X_3, X_4) < X_5\}\Big) \\ &+ \frac{1}{5}\Big(P\{\max(X_3, X_4, X_5) < s\} - P\{X_2 + \mathrm{med}(X_3, X_4, X_5) < s\}\Big). \end{aligned}$$

С обзиром на то да тест-статистика не зависи од параметра λ можемо да претпоставимо да је $\lambda = 1$. Након израчунавања добијамо

$$\psi(s) = -\frac{1}{20} + \frac{2}{5}e^{-3s} - \frac{9}{10}e^{-2s} + \frac{1}{2}e^{-s}.$$

Лако је показати да је очекивана вредност пројекције нула а да за њену дисперзију важи

$$\sigma_I^2 = E(\psi^2(X_1)) = \frac{29}{42000}.$$

На основу овог закључујемо да је језгро Ψ недегенерисано. Користећи теорему 1.2.1 за V-статистике са недегенерисаним језгром добијамо да је гранична расподела наше тест-статистике нормална $\mathcal{N}(0, \frac{29}{1680})$.

Лема 5.1.1 (Милошевић 2015)**.** *За дату тест-статистику I_n функција великих одступања f_I из теореме 3.2.2 је аналитичка за довољно мало $\varepsilon > 0$ и важи*

$$f_I(\varepsilon) = \frac{840}{29}\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0.$$

Доказ. Језгро Ψ је ограничено, центрирано и недегенерисано. Сада можемо применити теорему 3.2.6 о великим одступањима за недегенерисане U-статистике и добијамо тврђење леме. □

Следећа лема директна је последица теореме 3.2.10.

Лема 5.1.2 (Милошевић 2015)**.** *За дату алтернативну расподелу $g(x;\theta)$ из фамилије $\mathcal{G}$ која је дефинисана у одељку 3.2.2 важи*

$$b(\theta) = 5\theta \int\limits_0^\infty \psi(x)h(x)dx + o(\theta),\quad \theta \to 0.$$

Функција $h(x)$ је овде и у тексту који следи дефинисана са $h(x) = g'_\theta(x, 0)$.

Кулбак-Лајблерово растојање алтернативних расподела са густином $g(x, \theta)$ из $\mathcal{G}$ од класе експоненцијалних расподела са густином $\{\lambda e^{-\lambda x},\ \ x \geq 0\}$, је

$$K(\theta) = \inf_{\lambda>0} \int_0^\infty \ln[g(x,\theta)/\lambda \exp(-\lambda x)]g(x,\theta)\ dx. \tag{5.3}$$

У раду ([82]) је показано да за мало θ израз (5.3) се може приказати у облику

$$2K(\theta) = \Big(\int\limits_0^\infty h^2(x)e^x dx - \Big(\int\limits_0^\infty xh(x)dx \Big)^2 \Big) \cdot \theta^2 + o(\theta^2). \tag{5.4}$$

У даљем тексту одредићемо локалну Бахадурову ефикасност наших тестова за следеће алтернативе:

- Вејбулова расподела с густином

$$g(x,\theta) = e^{-x^{1+\theta}}(1+\theta)x^\theta,\ \theta > 0,\ x \geq 0; \tag{5.5}$$

- Макехамова расподела с густином

$$g(x,\theta) = (1+\theta(1-e^{-x}))\exp(-x-\theta(e^{-x}-1+x)),\ \theta > 0,\ x \geq 0; \quad (5.6)$$

- гама расподела с густином

$$g(x,\theta) = \frac{x^\theta}{\Gamma(\theta+1)}e^{-x}, \theta > 0, x \geq 0; \quad (5.7)$$

- мешавина експоненцијалних расподела с негативним тежинама МЕНТ(β) (в. [48]) с густином

$$g(x,\theta) = (1+\theta)e^{-x} - \beta\theta e^{\beta x},\ \theta \in \left(0, \frac{1}{\beta-1}\right],\ x \geq 0; \quad (5.8)$$

- расподела линеарне стопе отказа (ЛСО) са густином

$$g(x,\theta) = (1+\theta x)e^{-x-\theta\frac{x^2}{2}}, \theta > 0, x \geq 0; \quad (5.9)$$

- генералисана експоненцијална расподела (ГЕ) ([73]) са густином

$$g(x,\theta) = e^{1-(1+x)^{1+\theta}}(1+\theta)(1+x)^\theta\ \theta > 0, x \geq 0; \quad (5.10)$$

- проширена експоненцијална расподела (ПЕ) ([30]) са густином

$$g(x,\theta) = \frac{1+\theta x}{1+\theta}e^{-x},\ \theta > 0,\ x \geq 0. \quad (5.11)$$

Следећим примерима илустроваћемо израчунавање локалне Бахадурове ефикасости.

Пример 5.1.1. Нека је алтернативна расподела Вејбулова са густином (5.5). Први извод по θ функције густине у тачки $\theta = 0$ је

$$h(x) = e^{-x} + e^{-x}\log x - e^{-x}x\log x.$$

Користећи (5.4) добијамо да је Кулбак-Лајблерово растојање

$$K(\theta) = \frac{\pi^2}{6}\theta^2 + o(\theta^2),\ \ \theta \to 0.$$

Коришћењем леме 5.1.2 добијамо

$$\begin{aligned} b_I(\theta) &= 5\theta\int_0^\infty \psi(x)(e^{-x} + e^{-x}\log x - e^{-x}x\log x)dx + o(\theta) \\ &= \log\left(\frac{3}{2^{11/8}}\right)\theta + o(\theta) \approx 0.146\theta + o(\theta),\ \ \theta \to 0. \end{aligned}$$

Сада можемо искористити тврђење леме 5.1.1 и (3.8) и добити да је локална Бахадурова ефикасност $e_B(I) = 0.746$.

Слично се могу добити и резултати за алтернативе (5.6)-(5.10). Њихове ефикасности су приказане у табели 5.1. Изузетак је алтернатива (5.11) за коју се лема 5.1.2 и израз (5.4) не могу примети. Следећим примером је приказано израчунавање у том случају.

Пример 5.1.2. Посматрајмо ПЕ алтернативу са функцијом густине (5.11). Њен први извод по θ у тачки $\theta = 0$ је

$$h(x) = -e^{-x} + e^{-x}x.$$

Изрази $\int_0^\infty h(x)\psi(x)dx$ и $\int_0^\infty h^2(x)e^x dx - \left(\int_0^\infty xh(x)dx\right)^2$ су једнаки нули па нам је потребан развој функција $b_I(\theta)$ $2K(\theta)$ до првог члана различитог од нуле. Гранична вредност статистике I_n кад $n \to \infty$ у вероватноћи $b_I(\theta)$ је

$$\begin{aligned} b_I(\theta) &= P\{\max(X_2, X_3, X_4) < X_5\} - P\{X_1 + \mathrm{med}(X_2, X_3, X_4) < X_5\} \\ &= \frac{1}{4} - 6\int\limits_0^\infty g(x,\theta)\int\limits_0^\infty G(y,\theta)(1 - G(y,\theta))g(y,\theta)\int\limits_{x+y}^\infty g(z,\theta)dzdydx \\ &= \frac{\theta^2(168 + 356\theta + 161\theta^2)}{2304(1+\theta)^4} = \frac{7}{96}\theta^2 + o(\theta^2), \;\; \theta \to 0. \end{aligned}$$

Двоструко Кулбак-Лајблерово растојање (5.3) алтернативе (5.11) класе експоненцијалних расподела је

$$2K(\theta) = 2\frac{(1+\theta)\log(1+2\theta) - e^{1/\theta}\theta Ei(-(1/\theta))}{(1+\theta)^2} = \theta^4 + o(\theta^4), \;\; \theta \to 0, \tag{5.12}$$

где је $Ei(z) = -\int_{-z}^\infty u^{-1}e^{-u}du$ експоненцијални интеграл. Применом леме 5.1.1 и израза (3.8) добијамо да је локална Бахадурова ефикасност $e_B(I) = 0.481$.

Можемо приметити да су у табели 5.1 све ефикасности релативно високе осим у случају (5.9) и (5.11) коју смо уврстили као пример како се може одредити локална Бахадурова ефикасност када се не може применити лема 5.1.2.

Сада ћемо испитати асимптотска својства статистике K_n дате изразом (5.2). За фиксирано $t > 0$ израз $H_n(t) - G_n(t)$ је $V-$ статистика са језгром

$$\begin{aligned} \Xi(X_1, X_2, X_3, X_4, t) = \frac{1}{4!}\sum_{\pi(1:4)} \Big(&\mathrm{I}\{\max(X_{\pi_2}, X_{\pi_3}, X_{\pi_4}) < t\} \\ &- \mathrm{I}\{X_{\pi_1} + \mathrm{med}(X_{\pi_2}, X_{\pi_3}, X_{\pi_4}) < t\}\Big). \end{aligned}$$

Табела 5.1: Локална Бахадурова ефикасност за тест-статистику I_n

Алт.	Ефикасност
Вејбул	0.746
Макехам	0.772
Гама	0.701
МЕНТ(3)	0.916
ЛСО	0.308
ГЕ	0.556
ПЕ	0.481

Пројекција ове фамилије језгара на X_1 ако важи H_0 је

$$\begin{aligned}\xi(s,t) &= E(\Xi(X_1,X_2,X_3,X_4,t)|X_1=s)\\ &= \frac{1}{4}\Big(P\{\max(X_2,X_3,X_4)<t\} - P\{s+\mathrm{med}(X_2,X_3,X_4)<t\}\Big)\\ &+ \frac{3}{4}\Big(P\{\max(s,X_3,X_4)<t\} - P\{X_2+\mathrm{med}(s,X_3,X_4)<t\}\Big)\\ &= \frac{1}{4}\mathrm{I}\{s<t\}e^{-s-3t}(-e^s-2e^{4s}+6e^{2t}+3e^{s+t}+3e^{3s+t}-e^{s+2t}(9-6s))\\ &+ \frac{1}{4}\mathrm{I}\{s\geq t\}e^{-3t}(-1+6e^t-2e^{3t}-e^{2t}(3-6t)).\end{aligned}$$

Дисперзије ових пројекција су

$$\sigma_K^2(t) = \frac{9}{80}e^{-6t} - \frac{3}{8}e^{-5t} - \frac{3}{8}e^{-4t} - \frac{9}{8}e^{-3t} + \frac{33}{16}e^{-2t} - \frac{3}{10}e^{-t}.$$

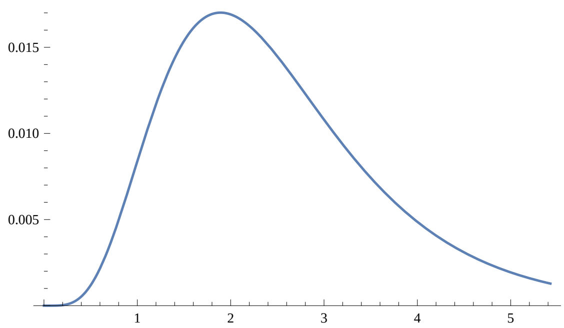
График ове функције приказан на слици 5.1. Проналазимо да важи

$$\sigma_K^2 = \sup_{t\geq 0}\sigma_K^2(t) = 0.017. \tag{5.13}$$

Супремум се достиже за $t_0 = 1.892$. Са графика 5.1 и (5.13) закључујемо да је фамилија језгара $\Xi(X_1,X_2,X_3,X_4,t)$ недегенерисана, па се за одређивање функције великих одступања може користити теорема из рада [77]. Коришћењем теореме 1.3.2 може се показати да V-емпиријски процес

$$\sqrt{n}\left(H_n(t) - G_n(t)\right),\ t\geq 0,$$

конвергира слабо у $D(0,\infty)$ кад $n\to\infty$ ка центрираном Гаусовом процесу $\nu(t)$ чија коваријациона функција се може одредити. На основу



Слика 5.1: График функције $\sigma_K^2(t)$,

тога закључујемо да $\sqrt{n}K_n$ конвергира у расподели ка случајној величини $\sup_{t\geq 0}|\nu(t)|$ чију расподелу не знамо.

У табели 5.2 су приказане критичне вредности за тест-статистику K_n добијене Монте Карло методом са 10000 понављања.

Табела 5.2: Критичне вредности за тест-статистику K_n

n	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
10	0.49	0.56	0.62	0.70
20	0.33	0.39	0.43	0.48
30	0.26	0.30	0.34	0.38
40	0.23	0.26	0.29	0.31
50	0.20	0.23	0.25	0.28
100	0.14	0.16	0.17	0.19

Лема 5.1.3 (Милошевић 2015)**.** *За статистику K_n функција f_K из теореме 3.2.2 је аналитичка за довољно мало $\varepsilon > 0$, и важи*

$$f_K(\varepsilon) = \frac{1}{32\sigma_K^2}\varepsilon^2 + o(\varepsilon^2) \approx 1.84\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0.$$

Доказ. Да бисмо применили теорему 3.2.8 потребно је да покажемо да статистика задовољава услов монотонсти по параметру. Нека су чворови поделе интеравала $[0,\infty)$ тачке $t_{k,N} = F^{-1}\left(\frac{k}{N}\right),\ k = 0, 1, ..., N$ Тада важи

$$\sup_{t_k \leq t < t_{k+1}} |H_n(t) - G_n(t)| \leq H_n(t_{k+1}) - G_n(t_{k+1}) + G_n(t_{k+1}) - G_n(t_k).$$

Нека је $\Delta_n = G_n(t_{k+1}) - G_n(t_k)$. Приметимо да је

$$\Delta_n = (F_n(t_{k+1}))^3 - (F_n(t_k))^3,$$

где је F_n емпиријска функција расподеле узорка $X_1, ..., X_n$. Тада је

$$\Delta_n < 2(F_n(t_{k+1}) - F_n(t_k))$$

Одатле закључујемо да важи

$$P\{n\Delta_n > \tau\} \leq P\Big\{n(F_n(t_{k+1}) - F_n(t_k)) > \frac{\tau}{2}\Big\}.$$

У раду [76] је показано да ако случајана величина S_n има биномну $\mathcal{B}\Big(n, \frac{1}{N}\Big)$ расподелу онда важи

$$P\{S_n > n\tau\} \leq 4^n e^{-n\tau \log N}. \tag{5.14}$$

Како случајна величина $n(F_n(t_{k+1}) - F_n(t_k))$ има биномну $\mathcal{B}(n, \frac{1}{N})$ расподелу добијамо да важи

$$P\Big\{n(F_n(t_{k+1}) - F_n(t_k)) > \frac{\tau}{2}\Big\} \leq 4^n e^{-\frac{n\tau \log N}{2}}.$$

За $\tau = \frac{2}{\sqrt{\log N}}$ добијамо да је услов монотоности по параметру задовољен. □

Гранична вредност у вероватноћи тест-статистике K_n кад $n \to \infty$ дата је следећом лемом.

Лема 5.1.4 (Милошевић 2015)**.** *За дату алтернативну расподелу $g(x;\theta)$ из $\mathcal{G}$ важи*

$$b_K(\theta) = 4\theta \sup_{t\geq 0} \Big| \int_0^\infty \xi(x;t)h(x)dx \Big| + o(\theta),\ \theta \to 0.$$

Доказ. Користећи Гливенко-Кантелијеву теорему за V-статистике, аналогну теореми 1.3.1, (в. [37]) добијамо

$$b_K(\theta) = \sup_{t\geq 0} \big| P\{\max(X_2, X_3, X_4) < t\} - P\{X_1 + \mathrm{med}(X_2, X_3, X_4) < t\} \big| \tag{5.15}$$

$$= \sup_{t\geq 0} \Big| G^3(t,\theta) - 6 \int\limits_0^t g(x,\theta) \int\limits_0^{t-x} G(y,\theta)(1 - G(y,\theta))g(y,\theta)dydx \Big|.$$

Нека је

$$a(t,\theta) = G^3(t,\theta) - 6\int\limits_0^t g(x,\theta)\int\limits_0^{t-x} G(y,\theta)(1-G(y,\theta))g(y,\theta)dydx. \quad (5.16)$$

После диференцирања унутар интеграла добијамо да важи

$$\begin{aligned} a'_\theta(t,0) &= \int\limits_0^t h(x)(4\xi(x,t) + e^{-3t}(1-6e^t+2e^{3t}+e^{2t}(3-6t))dx \\ &= 4\int\limits_0^\infty h(x)\xi(x,t)dx. \end{aligned}$$

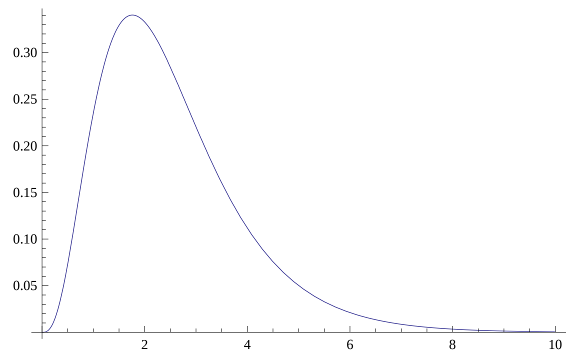
Развијајући $a(t,\theta)$ у Маклоренов ред и коришћењем (5.15) добијамо тврђење леме.

Следе примери израчунавања локалних Бахадурових ефикасности за алтернативе (5.5) и (5.11), док су резултати за све претходно поменуте алтернативе приказани у табели 5.3.

Пример 5.1.3. Нека је алтернативна расподела Вејбулова са функцијом густине (5.5). Користећи лему 5.1.4 добијамо

$$a(t,\theta) = 4\theta\int_0^\infty \xi(x,t)(e^{-x}+e^{-x}\log x - e^{-x}x\log x)dx + o(\theta), \quad \theta\to 0.$$

График функције $a'_\theta(t,0)$, је приказан на слици 5.2. Супремум функције



Слика 5.2: График функције $a'_\theta(t,0)$

$a(t,\theta)$ се достиже у тачки $t_1 \approx 1.761$, па је $b_K(\theta) \approx 0.34\theta + o(\theta), \ \ \theta\to 0$. Користећи леме 5.1.3, 5.1.4 и израз (3.8) добијамо да је локална Бахадурова ефикасност тест-статистике K_n приближно једнака 0.258.

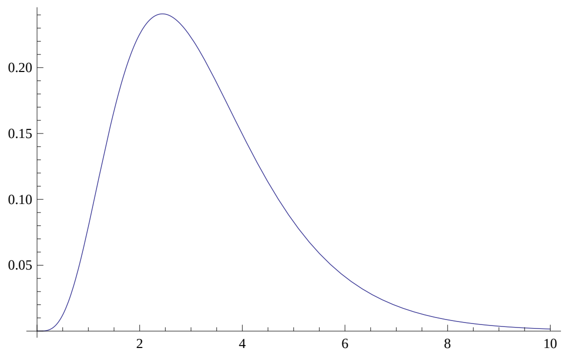
Пример 5.1.4. Нека је алтернативна расподела (5.11). Тада је функција $a(t,\theta)$ из (5.16)

$$\begin{aligned} a(t,\theta) &= \frac{3}{4(1+\theta)^4}e^{-3t}\theta^2\big(-2-8\theta-9\theta^2-2t^3\theta^2-4t^2\theta(1+2\theta) \\ &+8e^t(1+\theta)(2+t+6\theta+4t\theta+t^2\theta)-t(2+10\theta+13\theta^2) \\ &+e^{2t}(-14-56\theta-39\theta^2+t(6+18\theta+11\theta^2))\big) \\ &= \frac{3}{2}(-e^{-3t}+8e^{-2t}-7e^{-t}-te^{-3t}+4e^{-2t}t+3e^{-t}t)\theta^2+o(\theta^2). \end{aligned}$$

График функције $a_2(t)$, коефицијента уз θ^2, у претходном изразу приказан је на графику 5.3. Сада је

$$\sup_{t\geq 0}|a(t,\theta)| \approx 0.241\theta^2+o(\theta^2),\ \theta\to 0.$$

Двоструко Кулбак-Лајблерово растојање дато је изразом (5.12). Користећи леме 5.1.3, 5.1.4 и израз (3.8) добијамо да је локална Бахадурова ефикасност приближно 0.213.



Слика 5.3: График фукције $a_2(t)$

Можемо приметити да су ефикасности мање него за тест интегралног типа. Међутим у поређењу са другим тестовима истог типа заснованих на карактеризацијама који се могу наћи нпр. у [101], резултати су задовољавајући.

Следећом теоремом су приказане неке класе локално оптималних алтернатива.

Табела 5.3: Локална Бахадурова ефикасност тест-статистике K_n

Алтернатива	Ефикасност
Вејбул	0.258
Макехам	0.370
Гамма	0.212
МЕНТ(3)	0.364
ЛСО	0.213
ГЕ	0.298
ПЕ	0.213

Теорема 5.1.1 (Милошевић 2015)**.** *Нека је $g(x;\theta)$ функција густине из $\mathcal{G}$ за коју важи*

$$\int_0^\infty e^x h^2(x)dx < \infty.$$

Тада су, за мало $\theta > 0$, густине алтернативе

$$g(x;\theta) = e^{-x} + e^{-x}\theta(C_1\psi(x) + D_1(x-1)),\ x \geq 0,\ C_1 > 0,\ D_1 \in \mathbb{R} \quad (5.17)$$

$$g(x;\theta) = e^{-x} + e^{-x}\theta(C_2\xi(x,t_0) + D_2(x-1)),\ x \geq 0,\ C > 0,\ D_2 \in \mathbb{R}, \quad (5.18)$$

где је $t_0 \approx 1.892$, локално оптималне за тест заснован на статистици I_n и K_n, редом.

Доказ. Уведимо помоћну функцију

$$h_0(x) = h(x) - (x-1)e^{-x}\int_0^\infty h(s)sds, \quad (5.19)$$

која задовољава следеће услове:

$$\int_0^\infty h_0^2(x)e^x dx = \int_0^\infty e^x h^2(x)dx - \left(\int_0^\infty h(x)xdx\right)^2, \quad (5.20)$$

$$\int_0^\infty \psi(x)h_0(x)dx = \int_0^\infty \psi(x)h(x)dx$$

$$\int_0^\infty \xi(x)h_0(x)dx = \int_0^\infty \xi(x)h(x)dx. \quad (5.21)$$

Бахадурова локална ефикасност теста заснованог на тест-статистици I_n је

$$
\begin{aligned}
e_I^B &= \lim_{\theta\to 0}\frac{c_I(\theta)}{2K(\theta)} = \lim_{\theta\to 0}\frac{2f(b_I(\theta))}{2K(\theta)} = \lim_{\theta\to 0}\frac{2\cdot\frac{840}{29}b_I^2(\theta)}{2K(\theta)} = \lim_{\theta\to 0}\frac{b_I^2(\theta)}{25\sigma_I^2 2K(\theta)}\\
&= \lim_{\theta\to 0}\frac{25\theta^2\Bigg(\int\limits_0^\infty \psi(x)h(x)dx\Bigg)^2 + o(\theta^2)}{25\int\limits_0^\infty \psi^2(x)e^{-x}dx\Bigg(\theta^2(\int\limits_0^\infty e^x h^2(x)dx - \big(\int\limits_0^\infty h(x)xdx\big)^2) + o(\theta^2)\Bigg)}\\
&= \frac{\Bigg(\int\limits_0^\infty \psi(x)h(x)dx\Bigg)^2}{\int\limits_0^\infty \psi^2(x)e^{-x}dx\Bigg(\int\limits_0^\infty e^x h^2(x)dx - \big(\int\limits_0^\infty h(x)xdx\big)^2\Bigg)}\\
&= \frac{\Bigg(\int\limits_0^\infty \psi(x)h_0(x)dx\Bigg)^2}{\int\limits_0^\infty \psi^2(x)e^{-x}dx\int\limits_0^\infty h_0^2(x)e^x dx}.
\end{aligned}
$$

На основу неједнакости Коши-Шварца закључујемо да је $e_I^B = 1$ ако и само ако је $h_0(x) = C_1\psi(x)e^{-x}$. Када уврстимо добијени израз у (5.19) добијамо $h(x)$. Густине из тврђења теореме имају исту функцију $h(x)$,па је овим први део доказа завршен.

$$
\begin{aligned}
e_K &= \lim_{\theta\to 0}\frac{c_K(\theta)}{2K(\theta)} = \lim_{\theta\to 0}\frac{2f(b_K(\theta))}{2K(\theta)} == \lim_{\theta\to 0}\frac{b_K^2(\theta)}{16\sigma_K(t_0)^2 2K(\theta)}\\
&= \lim_{\theta\to 0}\frac{16\theta^2\sup\limits_{t\geq 0}\Bigg(\int\limits_0^\infty \xi(x,t)h(x)dx\Bigg)^2 + o(\theta^2)}{16\sup\limits_{t\geq 0}\int\limits_0^\infty \xi^2(x,t)e^{-x}dx\Bigg(\theta^2(\int\limits_0^\infty e^x h^2(x)dx - \big(\int\limits_0^\infty h(x)xdx\big)^2) + o(\theta^2)\Bigg)}\\
&= \frac{\sup\limits_{t\geq 0}\Bigg(\int\limits_0^\infty \xi(x,t)h(x)dx\Bigg)^2}{\sup\limits_{t\geq 0}\int\limits_0^\infty \xi^2(x,t)e^{-x}dx\Bigg(\int\limits_0^\infty e^x h^2(x)dx - \big(\int\limits_0^\infty h(x)xdx\big)^2\Bigg)}\\
&= \frac{\sup\limits_{t\geq 0}\Bigg(\int\limits_0^\infty \xi(x,t)h_0(x)dx\Bigg)^2}{\int\limits_0^\infty \xi^2(x,t)e^{-x}dx\int\limits_0^\infty h_0^2(x)e^x dx}.
\end{aligned}
$$

Услов једнакости у Коши-Шварцвов неједнакости, $e_K = 1$ се постиже ако и само ако је $h_0(x) = C\xi(x, t_0)e^{-x}$. Лако се показује да фамилија алтернатива (5.18) задовољава овај услов. □

5.1.1 Емпиријска моћ теста у случају малих узорака

Да бисмо испитали квалитет теста за узорке обима $n = 20$ и $n = 50$ против познатих алтернатива упоредили смо их са другим тестовима експоненцијалности који се могу наћи у раду [42]. Моћи тестова су приказане у табелама 5.4 и 5.5. Ознаке у табелама су преузете из рада [42]. Укратко ћемо описати сваки од тестова који су упоређени. Нека је $Y_1, ..., Y_n$ нормиран полазни узорак, односно $Y_j = \frac{X_j}{\bar{X}_n}$. Напомињемо да је сваки од тестова заснован на различитим својствима експоненцијалне расподеле и да се делимично и због тога баш ови тестови најчешће бирају за поређење са новим предложеним тестовима.

Епс и Пули су 1986. године предложили тест-статистику која је заснована на емпиријској карактеристичној функцији (в. [25])

$$EP_n = \sqrt{48n}\Big(\frac{1}{n}e^{-Y_j} - \frac{1}{2}\Big).$$

Барингхаус и Хензе су 2000. предложили следеће тест-статистике засноване на константном преосталом животном веку експоненцијално расподељене случајне величине (в. [38])

$$\overline{KS}_n = \sup_{t\geq 0}\Big|\frac{1}{n}\sum_{j=1}^{n}\min(Y_j, t) - \frac{1}{n}\sum_{j=1}^{n}\mathrm{I}\{Y_j \leq t\}\Big|,$$
$$\overline{CM}_n = \int_0^\infty \Big(\frac{1}{n}\sum_{j=1}^{n}\min(Y_j, t) - \frac{1}{n}\sum_{j=1}^{n}\mathrm{I}\{Y_j \leq t\}\Big)^2 e^{-t}dt.$$

Д'Агостино и Стивенс су 1986. предложили тест-статистику на основу узорачких размака (в. [21])

$$S_n = 2n - \frac{2}{n}\sum_{j=1}^{n} jY_{(j)}.$$

Оукс и Кокс су 1984. године предложили тест-статистику (в [19])

$$CO_n = n + \sum_{j=1}^{n}(1 - Y_j)\log Y_j.$$

Статистика ω^2 је стандардна статистика Крамера и фон Мизеса, а KS стандардна статистика Колмогорова и Смирнова. Алтернативе за које су поређени тестови су: Вејбулова расподела (W), гама расподела Γ, лог-нормална расподела (LN), халфнормална расподела (HN), униформна расподела на $[0,1]$ U, Ченова расподела (CH), расподела линеарне стопе отказа (LF), модификована расподела екстремних вредности (EV).

Подебљаним словима су означене моћи наших тестова које су веће од моћи свих конкурентских тестова. Може се приметити изванредно понашање статистике I_n, док се статистика K_n одлично показала за узорке обима $n = 20$. С обзиром на то да смо успели да оптимизујемо код и постигнемо изузетно брзо рачунање предложених тест-статистика препоручујемо га за тестирање експоненцијалности узорака малих и средњих обима.

Напомињемо да су моћи тестова одређене Монте Карло методом са $N = 10000$ понављања и да је ниво значајности теста $\alpha = 0.05$.

Табела 5.4: Проценат одбачених нултих хипотеза за различите тестове сагласности $n = 20$, $\alpha = 0.05$

Алтернатива	EP	$\overline{KS}$	$\overline{CM}$	ω^2	KS	S	CO	I	K
$W(1.4)$	36	35	35	34	28	35	37	**46**	32
$\Gamma(2)$	48	46	47	47	40	46	54	**59**	32
$LN(0.8)$	25	28	27	33	30	24	33	9	6
HN	21	24	22	21	18	21	19	**30**	**25**
U	66	72	70	66	52	70	50	**79**	**89**
$CH(0.5)$	63	47	61	61	56	63	80	23	20
$CH(1.0)$	15	18	16	14	13	15	13	**22**	**19**
$CH(1.5)$	84	79	83	79	67	84	81	22	20
$LF(2.0)$	28	32	30	28	24	29	25	**39**	**32**
$LF(4.0)$	42	44	43	41	34	42	37	**53**	**44**
$EV(0.5)$	15	18	16	14	13	15	13	**22**	**19**
$EV(1.5)$	45	48	47	43	35	46	37	**57**	**52**

Табела 5.5: Проценат одбачених нултих хипотеза за различите тестове сагласности $n = 50$, $\alpha = 0.05$

Алтернативе	EP	$\overline{KS}$	$\overline{CM}$	ω^2	KS	S	CO	I	K
$W(1.4)$	80	71	77	75	64	79	82	**82**	62
$\Gamma(2)$	91	86	90	90	83	90	96	94	72
$LN(0.8)$	45	62	60	76	71	47	66	14	7
HN	54	50	53	48	39	54	45	**58**	50
U	98	99	99	98	93	99	91	**99**	**100**
$CH(0.5)$	94	90	94	95	92	94	99	41	37
$CH(1.0)$	38	36	37	32	26	38	30	**41**	**38**
$CH(1.5)$	100	100	100	100	98	100	100	40	38
$LF(2.0)$	69	65	69	64	53	69	60	**73**	62
$LF(4.0)$	87	82	87	83	72	87	80	**88**	79
$EV(0.5)$	38	36	37	32	26	38	30	**41**	37
$EV(1.5)$	90	88	90	86	75	90	78	**90**	88

5.2 Тестови засновани на карактеризацијама на основу независности статистика

У овом поглављу ћемо представити статистике интегралног и Колмогоровљевог типа засноване на карактеризацији из теореме 4.3.4

$$I_n^{\mathcal{E}} = \int_0^\infty \int_0^\infty (H_n^{\mathcal{E}}(x,y) - G_n^{\mathcal{E}}(x,y))dF_n(x)dF_n(y),$$

$$K_n^{\mathcal{E}} = \sup_{t_1,t_2>0} |H_n^{\mathcal{E}}(t_1,t_2) - G_n^{\mathcal{E}}(t_1,t_2)|,$$

где су

$$G_n^{\mathcal{E}}(t_1,t_2) = \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} \mathrm{I}\{\min(X_i,X_j) \le t_1, |X_i - X_j| \le t_2\},$$

$$H_n^{\mathcal{E}}(t_1,t_2) = \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} \mathrm{I}\{\min(X_i,X_j) \le t_1\} \cdot \frac{1}{\binom{n}{2}} \sum_{1\le k<l\le n} \mathrm{I}\{|X_k - X_l| \le t_2\}.$$

U-емпиријске вишедимензионе расподеле.

Статистика I_n је асимптотски еквивалентна U-статистици са симетричним језгром (в. [53])

$$\Psi^{\mathcal{E}}(X_1,X_2,X_3,X_4)=\frac{1}{4!}\sum_{\pi(4)}\mathrm{I}\{\min(X_{\pi_1},X_{\pi_2})\le X_{\pi_3}\}(\mathrm{I}\{|X_{\pi_4}-X_{\pi_5}|\le X_{\pi_6}\}$$
$$-\,\mathrm{I}\{|X_{\pi_1}-X_{\pi_2}|\le X_{\pi_6}\})$$

Прва пројекција језгра на X_1 под претпоставком да важи нулта хипотеза је

$$\begin{aligned}\psi^{\mathcal{E}}(s)&=E(\Psi^{\mathcal{E}}(X_1,X_2,X_3,X_4)|X_1=s)=\frac{1}{3}P\{|s-X_2|\le X_3\}\\&-\frac{1}{2}P\{|s-X_2|\le X_3,\min(s,X_2)\le X_4\}+\frac{1}{6}P\{|X_3-X_2|\le s\}\\&-\frac{1}{4}P\{|X_3-X_2|\le s,\min(X_3,X_2)\le X_4\}+\frac{1}{6}P\{|X_4-X_2|\le X_3\}\\&-\frac{1}{4}P\{|X_4-X_2|\le X_3,\min(X_4,X_2)\le s\}\\&=\frac{1}{3}se^{-s}-\frac{1}{3}e^{-s}+\frac{3}{8}e^{-2s}-\frac{1}{24}.\end{aligned}$$

Очекивање пројекције је нула док је њена дисперзија

$$\sigma^2_{I^{\mathcal{E}}}=E(\psi^{\mathcal{E}}(X_1)^2)=\frac{7}{77760},$$

на основу чега закључујемо да се ради о U-статистици са недегенерисаним језгром. Као и у претходним поглављима, коришћењем теореме 1.1.2 закључујемо да је гранична расподела статистике $\sqrt{n}I_n^{\mathcal{E}}$ нормална $\mathcal{N}(0,\frac{112}{77760})$.

Функција великог одступања, потребна за рачунање локалне Бахадурове асимптотске ефикасности одређена је у следећој леми.

Напомињемо да су ова лема, као и остале леме у којима није наглашен аутор, оригинални резултати дисертације.

Лема 5.2.1. *Нека је $\varepsilon>0$. Функција великог одступања из теореме 3.2.2 аналитичка је за довољно мало ε и важи*

$$f_{I^{\mathcal{E}}}(\varepsilon)=\frac{77760}{224}\varepsilon^2+o(\varepsilon^2),\ \varepsilon\to 0.$$

Следећа лема нам даје граничну вредност у вероватноћи статистике $I_n^{\mathcal{E}}$ кад $n\to\infty$.

Лема 5.2.2. *За дату алтернативну функцију густине $g(x;\theta)$ из фамилије $\mathcal{G}$ дефинисане у одељку 3.2.2 важи*

$$b_{I^{\mathcal{E}}}(\theta) \sim 4\theta \int_0^\infty \psi^{\mathcal{E}}(x)h(x)dx \cdot \theta + o(\theta),\ \theta \to 0. \tag{5.22}$$

Пример 5.2.1. За Вејбулову расподелу је $h(x) = e^{-x} + e^{-x}\log x - e^{-x}x\log x$ и двоструко Кулбак-Лајблерово растојање је $2K(\theta) = \frac{\pi^2}{6}\theta^2 + o(\theta^2),\ \theta \to 0$. Применом леме 5.2.2 добијамо

$$b_{I^{\mathcal{E}}}(\theta) = \frac{1}{6}(1 + 2\log 2 - 2\log 3)\theta + o(\theta),\ \theta \to 0.$$

Користећи лему 5.2.1 добијамо да Бахадуров нагиб

$$c_{I^{\mathcal{E}}}(\theta) \approx 22.06\theta^2 + o(\theta^2),\ \theta \to 0,$$

одакле добијамо да је локална Бахадурова ефикасност приближно 0.419.

Локалне Бахадурове ефикасности за тест-статистику $I_n^{\mathcal{E}}$ за алтернативе (5.5)-(5.11) и (5.9) приказане су у табели 5.6. Како је израз (5.22) идентички једнак нули за одређивање локалног Бахадуровог нагиба не можемо искористити лему 5.2.2 већ граничну вредност у вероватноћи $b(\theta)$ рачунамо по дефиницији а затим добијени израз развијамо у Маклоренов ред закључно са квадратним чланом.

Табела 5.6: Локална Бахадурова ефикасност за тест-статистику $I_n^{\mathcal{E}}$

Алтернатива	Ефикасност
Вејбул	0.419
Макехам	0.714
Гамма	0.701
МЕНТ(3)	0.542
ЛСО	0.535
ГЕ	0.680
ПЕ	0.536

Сада ћемо испитати асимптотска својства статистике $K_n^{\mathcal{E}}$. За фиксирано $t_1 > 0$ и $t_2 > 0$ израз $H_n^{\mathcal{E}}(t_1,t_2) - G_n^{\mathcal{E}}(t_1,t_2)$ је U-статистика са језгром

$$\begin{aligned}\Xi^{\mathcal{E}}(X_1,X_2,X_3,X_4,t_1,t_2) = &\frac{1}{2}\mathrm{I}\{\min(X_1,X_2)\leq t_1\}\big(\mathrm{I}\{|X_1-X_2|\leq t_2\}\\ &-\mathrm{I}\{|X_3-X_4|\leq t_2\}\big)\\ &+\frac{1}{2}\mathrm{I}\{\min(X_2,X_3)\leq t_1\}\big(\mathrm{I}\{|X_2-X_3|\leq t_2\}\\ &-\mathrm{I}\{|X_1-X_4|\leq t_2\}\big).\end{aligned}$$

Пројекција ове фамилије језгра на X_1 ако важи H_0 је

$$\begin{aligned}\xi^{\mathcal{E}}(s,t_1,t_2) &= E(\Xi^{\mathcal{E}}(X_1,X_2,X_3,X_4,t_1,t_1)|X_1=s)\\ &=\frac{1}{2}P\{\min(X_2,s)\leq t_1,|X_2-s|\leq t_2\}\\ &-\frac{1}{2}P\{\min(X_2,s)\leq t_1,|X_3-X_4|\leq t_2\}\\ &+\frac{1}{2}P\{\min(X_2,X_3)\leq t_1,|X_2-X_3|\leq t_2\}\\ &-\frac{1}{2}P\{\min(X_2,X_3)\leq t_1,|s-X_4|\leq t_2\}\\ &=\frac{1}{2}e^{-s-2t_1-t_2}\big(-1+e^s+e^{t_2}(-e^s+e^{t_2})\mathrm{I}\{s>t_2\}\\ &-e^{t_1}\mathrm{I}\{s>t_1\}(e^s-e^{t_1}+e^{t_2}(-e^s+e^{t_1+t_2})\mathrm{I}\{s>t_1+t_2\})\big).\end{aligned}$$

Дисперзије пројекција под условом да важи нулта хипотеза H_0 су

$$\begin{aligned}\sigma^2_{\mathcal{E}}(t_1,t_2) = \frac{1}{12}e^{-5t_1-3t_2}\Big(&(-1+e^{t_1})^2(-1-e^{t_1}+e^{t_2}+e^{2t_2})\mathrm{I}\{t_1\leq t_2\}\\ &+(-1+e^{t_2})^2(-1+e^{t_1}-e^{t_2}+e^{2t_1})\mathrm{I}\{t_1>t_2\}\Big)\end{aligned}$$

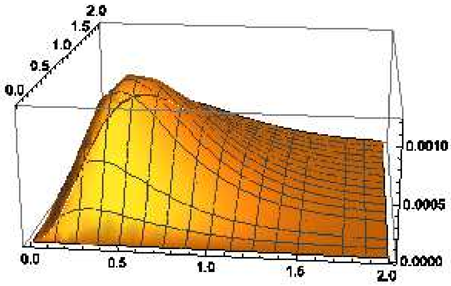
График је приказан на 5.4.

Добијамо да је

$$\sigma^2_{\mathcal{E}} = \sup_{t_1,t_2\geq 0}\sigma^2_{\mathcal{E}}(t_1,t_2)\approx\sigma^2_{\mathcal{E}}(0.452757,0.669169)\approx 0.02234,$$

на основу чега закључујемо да је наша фамилија језгара $\{\Xi^{\mathcal{E}}(X_1,X_2,X_3,X_4,t_1,t_2),\ t_1\geq 0, t_2\geq 0\}$ недегенерисана. Гранична расподела статистике $K_n^{\mathcal{E}}$ није позната, али се може показати да U-емпиријски процес

$$\nu_n^{\mathcal{E}}(t_1,t_2)=\sqrt{n}\left(H_n^{\mathcal{E}}(t_1,t_2)-G_n^{\mathcal{E}}(t_1,t_2)\right),\ t_1\,t_2>0$$



Слика 5.4: График функције $\sigma^2_{\mathcal{E}}(t_1, t_2)$,

слабо конвергира у $D((0,\infty)^2)$ кад $n \to \infty$ центрираном Гаусовом пољу $\nu^{\mathcal{E}}(t_1, t_2)$ (в. [93]). Тада низ тест-статистика $\sqrt{n}K_n^{\mathcal{E}}$ конвергира у расподели ка случајној величини $\sup_{t_1\geq 0, t_2\geq 0}|\nu^{\mathcal{E}}(t_1, t_2)|$ али је одређивање тачне расподеле ове случајне величине јако компликовано па се у пракси користе таблице критичних вредности добијене Монте-Карло симулацијама.

Да бисмо применили теорему 3.2.9 о функцији великих одступања за овај тип U-статистика потребно је да покажемо да статистика $K_n^{\mathcal{E}}$ задовољава вишедимензиони услов монотоности по параметру. Поделимо интервале $[0,\infty)$ и $[0,\infty]$ на N делова тако да су чворови у подели $t_{i,N}^k = \log\frac{N}{N-i}$, $k = 1, 2$, $i = 1, ..., N$. Означимо $T_{i,j,N} = [t_{i,N}^{(1)}, t_{i+1,N}^{(1)}] \times [t_{j,N}^{(2)}, t_{j+1,N}^{(2)}]$, $i, j = 1, 2, ..., N$ За фиксирано i, j

$$\begin{aligned}\sup_{(t_1,t_2)\in T_{i,j,N}} |H_n^{\mathcal{E}}(t_1,t_2) - G_n^{\mathcal{E}}(t_1,t_2)| \leq{}& G_n^{\mathcal{E}}(t_{i+1,N}^{(1)}, t_{j+1,N}^{(2)}) - H_n^{\mathcal{E}}(t_{i+1,N}^{(1)}, t_{j+1,N}^{(2)}) \\ &+ H_n^{\mathcal{E}}(t_{i+1,N}^{(1)}, t_{j+1,N}^{(2)}) - H_n^{\mathcal{E}}(t_{i,N}^{(1)}, t_{j,N}^{(2)}).\end{aligned}$$

Сада је

$$\Delta_n(N) = H_n^{\mathcal{E}}(t_{i+1,N}^{(1)}, t_{j+1,N}^{(2)}) - H_n^{\mathcal{E}}(t_{i,N}^{(1)}, t_{j,N}^{(2)}).$$

$$\begin{aligned}\Delta_n(N) &= \frac{1}{\binom{n}{2}} \sum_{1\leq r\leq s} \mathrm{I}\{\min(X_r, X_s) \in [t_{i,N}^{(1)}, t_{i+1,N}^{(1)}], |X_r - X_s| \in [t_{j,N}^{(2)}, t_{j+1,N}^{(2)}]\} \\ &\leq \frac{1}{n^2} \sum_{1\leq r\leq s} \mathrm{I}\{\min(X_r, X_s) \in [t_{i,N}^{(1)}, t_{i+1,N}^{(1)}]\}.\end{aligned}$$

Нека је τ_N низ реалних бројева који конвергира га нули. Означимо $\mathcal{T}_i =$

$[t_{i,N}^{(1)}, t_{i+1,N}^{(1)}]$. Тада важи:

$$\begin{aligned}
P\{\Delta_N > \tau_N\} &\leq P\Big\{\frac{1}{\binom{n}{2}}\sum_{1\leq r\leq s}\mathrm{I}\{\min(X_r, X_s)\in \mathcal{T}_i\}\geq \tau_N\}\\
&\leq P\Big\{\frac{1}{\binom{n}{2}}\sum_{1\leq r\leq s}\mathrm{I}\{X_r\in \mathcal{T}_i\}\geq \tau_N\Big\} + P\Big\{\frac{1}{\binom{n}{2}}\sum_{1\leq r\leq s}\mathrm{I}\{X_s\in \mathcal{T}_i\}\geq \tau_N\Big\}\\
&\leq P\Big\{\frac{2}{n-1}\sum_{1\leq r\leq s}\mathrm{I}\{X_r\in \mathcal{T}_i\}\geq \tau_N\Big\}\\
&= P\Big\{\sum_{r=1}^{n}\mathrm{I}\{X_r\in \mathcal{T}_i\}\geq \frac{n-1}{2}\tau_N\Big\}.
\end{aligned}$$

Приметимо да су сви сабирци независне и једнако расподељене Бернулијеве случајне величине па је њихов збир биномна случајна величина $\mathcal{B}(n, \frac{1}{N})$.

Примењујући неједнакост (5.14) и стављајући да је $\tau_N = (\log N)^{-\frac{1}{2}}$, за довољно велико N добијамо

$$P\{\Delta_N \geq \tau_N\} \leq e^{-(n-1)\sqrt{\log N}}.$$

Овим је услов монотоности по параметру показан. Применом теореме 3.2.9 добијамо да важи следеће тврђење.

Лема 5.2.3 (Милошевић, Обрадовић 2015)**.** *Нека је $\varepsilon > 0$. Функција великих одступања за тест-статистику $K_n^{\mathcal{E}}$ је аналитичка у нули и може се представити у облику:*

$$f_{K^{\mathcal{E}}}(\varepsilon) = 0.715\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0.$$

У следећој леми одредићемо лимес у вероватноћи тест-статистике $K_n^{\mathcal{E}}$ кад $n \to \infty$ под претпоставком да важи алтернативна хипотеза H_1.

Лема 5.2.4 (Милошевић, Обрадовић 2015)**.** *Претпоставимо да је алтернативна густина $g(x;\theta)$ која припада фамилији $\mathcal{G}$. Тада је гранична вредност у P_θ вероватноћи тест-статистике $K_n^{\mathcal{E}}$ кад $n \to \infty$*

$$b_{K^{\mathcal{E}}}(\theta) = 4\theta \sup_{t_1,t_2>0}\Big|\int_0^\infty \xi^{\mathcal{E}}(x;t_1,t_2)h(x)dx\Big| + o(\theta),\ \theta \to 0. \tag{5.23}$$

Доказ се изводи применом Гливенко-Кантелијеве леме за V статистике. Означимо са $a_{\mathcal{E}}(t_1,t_2,\theta)$ граничну вредност у вероватноћи статистике $H_n^{\mathcal{E}}(t_1,t_2) - G_n^{\mathcal{E}}(t_1,t_2)$ кад $n \to \infty$. Применом Гливенко-Кантелијеве теореме за V-статистике се добија функција $a^{\mathcal{E}}(t_1,t_2,\theta)$. Тада је

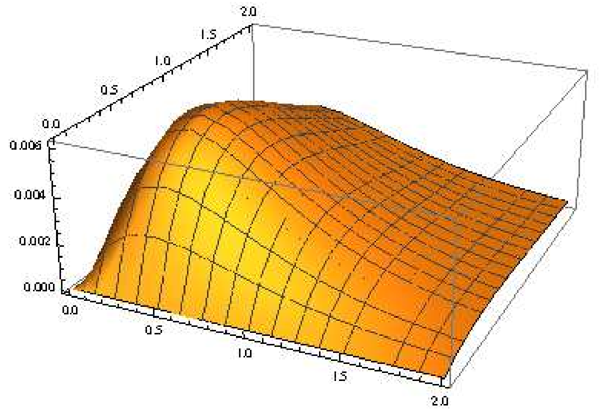
$$a_{\mathcal{E}}(t_1,t_2,\theta) = a'_{\mathcal{E}}(t_1,t_2,0)\theta + o(\theta),\ \theta \to 0,$$

при чему се добија да је

$$a'_{\mathcal{E}}(t_1,t_2,0)=\int\limits_0^{\infty}\xi^{\mathcal{E}}(x;t_1,t_2)h(x)dx.$$

Рачунање локалне Бахадурове ефикасности статистике $K_n^{\mathcal{E}}$ илустроваћемо примером.

Пример 5.2.2. Нека је алтернатива Макехамова расподела са густином (5.6). Тада је $h(x)=-2e^{-2x}+2e^{-x}-e^{-x}x$. Функција $a'_{\mathcal{E}}(t_1,t_2,0)$ приказана је на графику 5.5.



Слика 5.5: График функције $a'_{\mathcal{E}}(t_1,t_2,0)$ у случају Макехамове расподеле

Супремум се достиже у тачки $(t_0^{1\mathcal{E}},t_0^{2\mathcal{E}})\approx(0.405,0.693)$ и приближно износи 0.00617. Применом леме 5.2.3 добијамо да је локална Бахадурова ефикасност теста приближно 0.375.

Локалне Бахадурове ефикасности за остале алтернативе приказане су у табели 5.7. Неке класе локално оптималних алтернатива за тест-статистике $I_n^{\mathcal{E}}$ и $K_n^{\mathcal{E}}$ дате су следећом теоремом.

Теорема 5.2.1. *Нека је $g(x;\theta)$ функција густине из $\mathcal{G}$ за коју важи*

$$\int\limits_0^{\infty}e^{x}h^2(x)dx<\infty.$$

Тада су, за мало $\theta>0$, густине алтернативе

$$g(x;\theta)=e^{-x}+e^{-x}\theta(C_1\psi^{\mathcal{E}}(x)+D_1(x-1)),\ x\geq 0,\ C_1>0,\ D_1\in\mathbb{R}\quad(5.24)$$

$$g(x;\theta)=e^{-x}+e^{-x}\theta(C_2\xi^{\mathcal{E}}(x,t_{10},t_{20})+D_2(x-1)),\ x\geq 0,\ C>0,\ D_2\in\mathbb{R},\quad(5.25)$$

Табела 5.7: Локална Бахадурова ефикасност за тест-статистику $K_n^{\mathcal{E}}$

Алт.	Ефикасност
Вејбул	0.200
Макехам	0.375
Гамма	0.131
МЕНТ(3)	0.334
ЛСО	0.235
ГЕ	0.315
ПЕ	0.235

где је $(t_{10}, t_{20}) \approx (0.452757, 0.669169)$, *локално оптималне за тест-статистике* $I_n^{\mathcal{E}}$ *и* $K_n^{\mathcal{E}}$, *редом.*

Доказ је аналоган доказу теореме 5.1.1 па ћемо га изоставити.

5.3 Тестови засновани на карактеризацијама на основу функционалне једначине коју задовољава функција расподеле

У овом одељку ћемо приказати тестове сагласности са Паретовом и логистичком расподелом. Тестови сагласности са Паретовом расподелом су засновани на карактеризацији Паретове расподеле из теореме 4.2.1 а тестови сагласности са логистичком расподелом су засновани на основу карактеризације 4.2.2.

5.3.1 Тестови сагласности са Паретовом расподелом

Да бисмо тестирали сложену нулту хипотезу H_0 да је узорак из Паретове расподеле предлажемо тест-статистике интегралног и Колмогоровљевог типа које не зависе од параметра облика λ

$$I_n^{\mathcal{P}} = \int\limits_0^{\infty}\int\limits_0^{\infty} (H_n^{\mathcal{P}}(x,y) - G_n^{\mathcal{P}}(x,y))dF_n(x)dF_n(y)$$

$$K_n^{\mathcal{P}} = \sup_{t_1,t_2>1} |H_n^{\mathcal{P}}(t_1,t_2) - G_n^{\mathcal{P}}(t_1,t_2)|,$$

где су

$$H_n^{\mathcal{P}}(t_1,t_2) = \frac{1}{\binom{n}{2}} \sum_{1\le i_1<i_2\le n} \frac{1}{2!}\sum_{\pi(2)} \mathrm{I}\{X_{i_{\pi(1)}} > t_1\}\mathrm{I}\{X_{i_{\pi(2)}} > t_2\}$$

и

$$G_n^{\mathcal{P}}(t_1,t_2) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\{X_i > t_1t_2\},$$

U-емпиријске вишедимензионе функције расподеле. Статистика $I_n^{\mathcal{P}}$ је асимптотски еквивалентна U-статистици са симетричним језгром (в. [53])

$$\Psi^{\mathcal{P}}(X_1,X_2,X_3,X_4) = \frac{1}{4!}\sum_{\pi(4)}\Big(\mathrm{I}\{X_{\pi_1} > X_{\pi_3}\}\mathrm{I}\{X_{\pi_2} > X_{\pi_4}\} - \mathrm{I}\{X_{\pi_1} > X_{\pi_3}X_{\pi_4}\}\Big).$$

Прва пројекција језгра на X_1 под претпоставком да важи нулта хипотеза је

$$\begin{aligned}\psi^{\mathcal{P}}(s) = E(\Psi^{\mathcal{P}}(X_1,X_2,X_3,X_4)|X_1=s) = \frac{1}{2}P\{X_3>s\}P\{X_2>X_4\}\\ + \frac{1}{2}P\{s>X_3\}P\{X_2>X_4\} - \frac{1}{4}P\{s>X_3X_4\}\\ - \frac{1}{4}P\{X_2>X_3X_4\} - \frac{1}{2}P\{X_2>sX_3\}.\end{aligned}$$

Очекивање пројекције је нула док је њена дисперзија

$$\sigma^2_{I^{\mathcal{P}}} = E(\psi^{\mathcal{P}}(X_1)^2) = \frac{5}{6912},$$

на основу чега закључујемо да је гранична расподела тест-статистике $\sqrt{n}I_n^{\mathcal{P}}$ нормална $\mathcal{N}(0,\frac{5}{432})$.

За фиксиране $t_1>1$ и $t_2>1$ израз $H_n^{\mathcal{P}}(t_1,t_2) - G_n^{\mathcal{P}}(t_1,t_2)$ је U-статистика са симетричним језгром

$$\begin{aligned}\Xi^{\mathcal{P}}(X_1,X_2;t_1,t_2) =& \frac{1}{2}\mathrm{I}\{X_1>t_1\}\mathrm{I}\{X_2>t_2\} + \frac{1}{2}\mathrm{I}\{X_1>t_2\}\mathrm{I}\{X_2>t_1\}\\ &- \frac{1}{2}\mathrm{I}\{X_1>t_1t_2\} - \frac{1}{2}\mathrm{I}\{X_2>t_1t_2\}.\end{aligned}$$

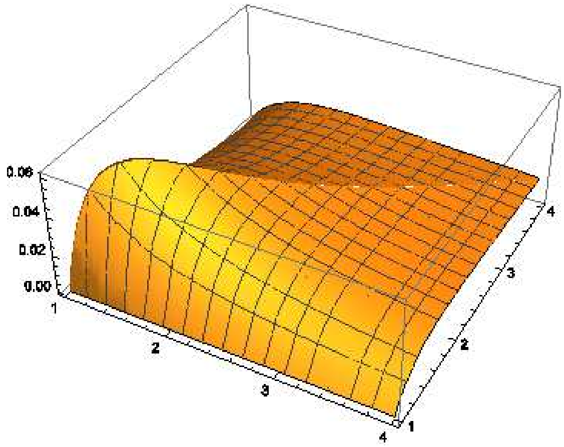
Пројекција језгра на X_1 уколико важи нулта хипотеза је

$$\begin{aligned}\xi^{\mathcal{P}}(s) =& E(\Xi^{\mathcal{P}}(X_1,X_2;t_1,t_2)|X_1=s) = -\frac{1}{2t_1t_2} + \frac{1}{2t_2}\mathrm{I}\{s>t_1\}\\ &+ \frac{1}{2t_1}\mathrm{I}\{s>t_2\} - \frac{1}{2}\mathrm{I}\{s>t_1t_2\}.\end{aligned}$$

Дисперзија ове пројекције је

$$\sigma^2_{\mathcal{P}}(t_1,t_2) = \frac{-1+t_1-t_2+t_1t_2}{4t_1^2t_2^2}\mathrm{I}\{t_1<t_2\} + \frac{-1-t_1+t_2+t_1t_2}{4t_1^2t_2^2}\mathrm{I}\{t_1\geq t_2\}.$$

График ове функције је приказан на слици 5.6.



Слика 5.6: График функције $\sigma^2_{\mathcal{P}}(t_1,t_2)$,

Добијамо да је

$$\sigma^2_{\mathcal{P}} = \sup_{t_1>1,t_2>1} \sigma^2_{\mathcal{P}}(t_1,t_2) = \sigma^2_{\mathcal{P}}(1.414, 1.414) = 0.0625,$$

на основу чега закључујемо да је фамилија језгара $\{\Xi^{\mathcal{P}}(X,Y,t_1,t_2),\ t_1 > 1, t_2 > 1\}$ недегенерисана.

Гранична расподела статистике $K_n^{\mathcal{P}}$ није позната, али се може показати да U-емпиријски процес

$$\nu_n^{\mathcal{P}}(t_1,t_2) = \sqrt{n}\left(H_n^{\mathcal{P}}(t_1,t_2) - G_n^{\mathcal{P}}(t_1,t_2)\right),\ t_1\,t_2 > 1$$

слабо конвергира у $D((1,\infty)^2)$ кад $n\to\infty$ центрираном Гаусовом пољу $\nu^{\mathcal{P}}(t_1,t_2)$ (в. [93]). Тада низ тест-статистика $\sqrt{n}K_n^{\mathcal{P}}$ конвергира у расподели ка случајној величини $\sup_{t_1\geq 0,t_2\geq 0}|\nu^{\mathcal{P}}(t_1,t_2)|$ али је одређивање тачне расподеле ове случајне величине јако компликовано па се у пракси користе таблице критичних вредности добијене Монте-Карло симулацијама.

Лако је показати да тест-статистика $K_n^{\mathcal{P}}$ задовољава услов монотоности по параметру па коришћењем теорема 3.2.6 и 3.2.9 добијамо да важи следеће тврђење.

Лема 5.3.1. *Нека је* $\varepsilon > 0$. *Функције великих одступања из теореме 3.2.2 за тест-статистике* $I_n^{\mathcal{P}}$ *и* $K_n^{\mathcal{P}}$ *су аналитичке у нули и могу се представити у облику:*

$$f_{I^{\mathcal{P}}}(\varepsilon) = \frac{216}{5}\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0,$$
$$f_{K^{\mathcal{P}}}(\varepsilon) = 2\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0.$$

Следећа лема даје нам граничну вредност у вероватноћи тест-статистика $I_n^{\mathcal{P}}$ и $K_n^{\mathcal{P}}$ кад $n \to \infty$ уколико важе блиске алтернативе.

Лема 5.3.2. *Претпоставимо да је алтернативна густина* $g(x;\theta)$ *која припада фамилији* $\mathcal{G}$ *дефинисаној у одељку 3.2.2. Тада су граничне вредности у вероватноћи* P_θ *за тест-статистик3* $I_n^{\mathcal{P}}$ *и* $K_n^{\mathcal{P}}$ *кад* $n \to \infty$ *редом једнаке*

$$b_{I^{\mathcal{P}}}(\theta) = 4\theta \int_0^\infty \psi^{\mathcal{P}}(x;t_1,t_2)h(x)dx + o(\theta),\ \theta \to 0,$$

$$b_{K^{\mathcal{P}}}(\theta) = 2\theta \sup_{t_1,t_2>0} \Big| \int_0^\infty \xi^{\mathcal{P}}(x;t_1,t_2)h(x)dx \Big| + o(\theta),\ \theta \to 0.$$

Делови лема 5.3.1 и 5.3.2 који се односе на статисику $K_n^{\mathcal{P}}$ приказани су у раду [69].

У раду [85] смо показали да се двоструко Кулбак-Лајблерово растојање блиске алтернативе од фамилије Паретових расподела може приказати у облику

$$2K^{\mathcal{P}}(\theta) = \theta^2 \Bigg(\int_1^\infty x^2 h^2(x)dx - \Big(\int_1^\infty h(x)\ln x dx \Big)^2 \Bigg) + o(\theta^2),\quad \theta \to 0. \quad (5.26)$$

Одабрали смо алтернативе које се веома често користе као алтернативе Паретовој расподели а за које важи да се смањивањем параметра θ приближавамо баш Паретовој расподели. То су:

- лог-Вејбулова расподела са густином
$$g_1(x;\theta) = (1+\theta)x^\theta e^{-x^{1+\theta}},\ x \ge 1,\ \theta \in (0,1), \quad (5.27)$$

- лог-гама расподела са густином
$$g_2(x;\beta,\theta) = \frac{x^\theta}{\Gamma(1+\theta)} e^{-x},\ x \ge 1,\ \theta \in (0,1), \quad (5.28)$$

- а Леј-Паиндавеине алтернатива са густином

$$g_3(x,\theta) = \frac{1}{x^2} - \frac{\pi\theta\cos\big(\pi(1-\frac{1}{x})\big)}{x^2},\ x \geq 1,\ \theta \in \big(0, \frac{1}{\pi}\big), \tag{5.29}$$

- инверзна бета расподела са густином

$$g_4(x;\theta) = \frac{1+\theta}{x^2}\Big(1-\frac{1}{x}\Big)^{\theta},\ x \geq 1,\ \theta \in (0,1), \tag{5.30}$$

- Паретова расподела са тилт параметром (в. [63]) са густином

$$g_5(x;\theta) = \frac{1+\theta}{(x+\theta)^2},\ x \geq 1,\ \theta \in (0,1), \tag{5.31}$$

- мешавина Паретових алтернатива са густином

$$g_6^{[\beta]}(x,\theta) = \frac{1-\theta}{x^2} + \frac{\beta\theta}{x^{\beta+1}}, x \geq 1,\ \theta \in (0,1). \tag{5.32}$$

Рачунање локалне Бахадурове ефикасности илустроваћемо примером. Резултати за остале алтернативе приказани су у табели 5.8. Можемо приметити да је тест интегралног типа доста ефикасан, посебно у случају лог-Вејбулове алтернативе и Паретове расподеле са тилт параметром док су ефикасности теста Колмогоровљевог типа доста мање.

Табела 5.8: Локална Бахадурова ефикасност за тест-статистике $I_n^{\mathcal{P}}$ и $K_n^{\mathcal{P}}$

Алт.	$I_n^{\mathcal{P}}$	$K_n^{\mathcal{P}}$
g_1	0.821	0.158
g_2	0.788	0.174
g_3	0.720	0.23
g_4	0.777	0.181
g_5	0.800	0.1875

Пример 5.3.1. Нека је алтернатива инверзна бета расподела са густином (5.30). Тада је

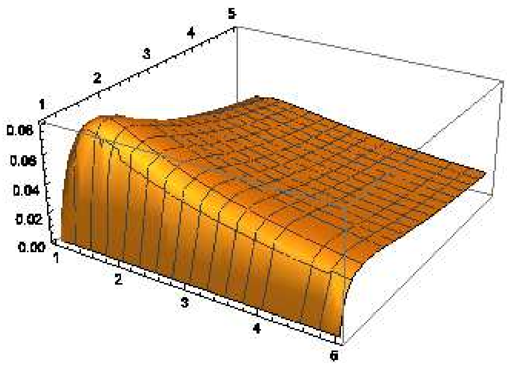
$$h(x) = \frac{1}{x^2} + \frac{\log\big(1-\frac{1}{x}\big)}{x^2}.$$

Гранична вредност у вероватноћи тест-статистике $I_n^{\mathcal{P}}$ кад $n \to \infty$ је

$$b_{I^{\mathcal{P}}}(\theta) = \frac{1}{12}(\pi^2 - 9)\theta + o(\theta),\ \theta \to 0.$$

Из израза (5.26) и леме 5.3.1 добијамо да локална Бахадурова ефикасност статистике $I_n^{\mathcal{P}}$ једнака 0.777.

Граничну вредност у вероватноћи тест-статистике $K_n^{\mathcal{P}}$ кад $n \to \infty$ рачунамо тражећи супремум функције $a_{\mathcal{P}}(t_1, t_2, \theta)$ дефинисане аналогно функцији $a_{\mathcal{E}}(t_1, t_2, \theta)$. График функције $a'_{\mathcal{P}}(t_1, t_2, 0)$ приказан је на слици 5.7.



Слика 5.7: График функције $a'_{\mathcal{P}}(t_1, t_2, 0)$ у случајну инверзне бета расподеле

Супремум се достиже у тачки $(t_0^{1\mathcal{P}}, t_0^{2\mathcal{P}}) \approx (1.447, 1.447)$ и износи приближно 0.0812. Користећи израз (5.26) и лему 5.3.1 добијамо да локална Бахадурова ефикасност статистике $K_n^{\mathcal{P}}$ приближно једнака 0.181.

Против алтернативе (5.32) интегрални тест није постојан. За тест-статистику $K_n^{\mathcal{P}}$ се добија да је локална Бахадурова ефикасност

$$e_{K^{\mathcal{P}}}(\beta) = \frac{8\beta - 4}{\beta^{2\frac{\beta+1}{\beta-1}}}.$$

Максимална ефикасност се постиже за $\beta \approx 6.114$ и износи приближно 0.292.

Неке локално оптималне алтернативе за тест-статистике $I_n^{\mathcal{P}}$ и $K_n^{\mathcal{P}}$ дате су следећом теоремом.

Теорема 5.3.1. *Нека је $g(x;\theta)$ функција густине из $\mathcal{G}$ за коју важи*

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

Тада су, за мало $\theta > 0$*, густине алтернативе*

$$g(x;\theta) = \frac{1}{x^2} + \frac{1}{x^2}\theta(C_1\psi^{\mathcal{P}}(x) + D_1(\log x - 1)),\ x \geq 0,\ C_1 > 0,\ D_1 \in \mathbb{R} \tag{5.33}$$

$$g(x;\theta) = e^{-x} + e^{-x}\theta(C_2\xi^{\mathcal{P}}(x, t_{10}, t_{20}) + D_2(\log x - 1)),\ x \geq 0,\ C > 0,\ D_2 \in \mathbb{R}, \tag{5.34}$$

где је $(t_{10}, t_{20}) = (1.414, 1.414)$*, локално оптималне за тест-статистике* $I_n^{\mathcal{P}}$ *и* $K_n^{\mathcal{P}}$*, редом.*

Доказ је аналоган доказу теореме 5.1.1 уз напомену да је сада помоћна функција

$$h_0(x) = h(x) - \frac{\log x - 1}{x^2}\int\limits_1^{\infty} h(s)sds.$$

5.3.2 Тест сагласности са логистичком расподелом

За тестирање сложене нулте хипотезе H_0 да је узорак $X_1, \ldots, X_n$ из логистичке расподеле са функцијом расподеле (4.8) предлажемо тест-статистике интегралног и Колмогоровљевог типа засновану на карактеризацији 4.2.2.

$$I_n^{\mathcal{L}} = \int\limits_0^{\infty}\int\limits_0^{\infty}(H_n^{\mathcal{L}}(x,y) - G_n^{\mathcal{L}}(x,y))dF_n(x)dF_n(y)$$

$$K_n^{\mathcal{L}} = \sup_{t_1,t_2\in\mathbb{R}}|H_n^{\mathcal{L}}(t_1,t_2) - G_n^{\mathcal{L}}(t_1,t_2)|,$$

где су

$$H_n^{\mathcal{L}}(t_1,t_2) = \frac{1}{\binom{n}{3}}\sum_{i_1<i_2<i_3}\frac{1}{3!}\sum_{\pi(1:3)}\mathrm{I}\{X_{i_{\pi(1)}} < t_1 + t_2\}\mathrm{I}\{X_{i_{\pi(2)}} > t_1\}\mathrm{I}\{X_{i_{\pi(3)}} > t_2\}$$

$$G_n^{\mathcal{L}}(t_1,t_2) = \frac{1}{\binom{n}{3}}\sum_{i_1<i_2<i_3}\frac{1}{3!}\sum_{\pi(1:3)}\mathrm{I}\{X_{i_{\pi(1)}} > t_1 + t_2\}\mathrm{I}\{X_{i_{\pi(2)}} < t_1\}\mathrm{I}\{X_{i_{\pi(3)}} < t_2\}.$$

Под претпоставком да важи нулта хипотеза H_0 тест-статистике $I_n^{\mathcal{L}}$ и $K_n^{\mathcal{L}}$ не зависе од параметра скалирања λ па у даљим израчунавањима можемо претпоставити да је $\lambda = 1$.

Статистика $I_n^{\mathcal{L}}$ је асимптотски еквивалентна U-статистици са симетричним језгром (в. [53])

$$\begin{aligned}
&\Psi^{\mathcal{L}}(X_1, X_2, X_3, X_4, X_5) = \\
&= \frac{1}{5!}\sum_{\pi(5)}\Big(\mathrm{I}\{X_{\pi_1} < X_{\pi_2} + X_{\pi_3}\}\mathrm{I}\{X_{\pi_4} > X_{\pi_2}\}\mathrm{I}\{X_{\pi_5} > X_{\pi_3}\} \\
&- \mathrm{I}\{X_{\pi_1} > X_{\pi_2} + X_{\pi_3}\}\mathrm{I}\{X_{\pi_4} < X_{\pi_2}\}\mathrm{I}\{X_{\pi_5} < X_{\pi_3}\}\Big).
\end{aligned}$$

Прва пројекција језгра на X_1 под претпоставком да важи нулта хипотеза је

$$\begin{aligned}
\psi^{\mathcal{L}}(s) &= E(\Psi^{\mathcal{L}}(X_1, X_2, X_3, X_4, X_5)|X_1 = s) \\
&= \frac{1}{5}P\{s < X_2 + X_3, X_4 > X_2, X_5 > X_3\} \\
&+ \frac{2}{5}P\{X_4 < X_2 + X_3, s > X_2, X_5 > X_3\} \\
&+ \frac{2}{5}P\{X_2 < s + X_3, X_4 > s, X_5 > X_3\} \\
&- \frac{1}{5}P\{s > X_2 + X_3, X_4 < X_2, X_5 < X_3\} \\
&- \frac{2}{5}P\{X_4 > X_2 + X_3, s < X_2, X_5 < X_3\} \\
&- \frac{2}{5}P\{X_2 > s + X_3, X_4 < s, X_5 < X_3\}.
\end{aligned}$$

Очекивање пројекције је нула а дисперзија

$$\sigma^2_{I^{\mathcal{L}}} = E(\psi^{\mathcal{L}}(X_1)^2) \approx 4.146 \cdot 10^{-4}.$$

на основу чега закључујемо да је гранична расподела тест-статистике $\sqrt{n}I_n^{\mathcal{L}}$ нормална $\mathcal{N}(0, 10.4 \cdot 10^{-4})$.

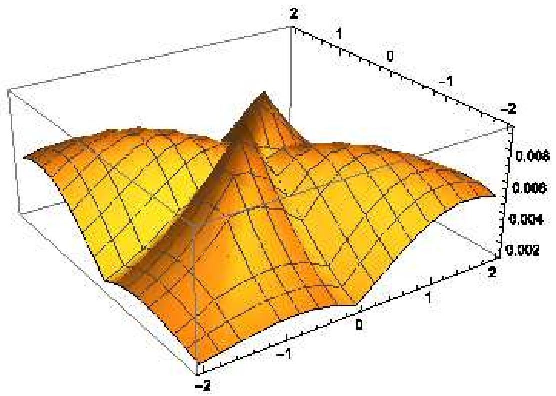
За фиксиране $t_1, t_2 \in R$ статистика $H_n^{\mathcal{L}}(t_1, t_2) - G_n^{\mathcal{L}}(t_1, t_2)$ је U-статистика са симетричним језгром

$$\begin{aligned}
\Xi^{\mathcal{L}}(X_1, X_2, X_3, t_1, t_2) = &\frac{1}{3!}\sum_{\pi(3)}\Big(\mathrm{I}\{X_{i_1} < t_1 + t_2\}\mathrm{I}\{X_{i_2} > t_1\}\mathrm{I}\{X_{i_3} > t_2\} \\
&-\mathrm{I}\{X_{i_1} > t_1 + t_2\}\mathrm{I}\{X_{i_2} < t_1\}\mathrm{I}\{X_{i_3} < t_2\}\Big).
\end{aligned}$$

Пројекција језгра на X_1 ако важи H_0 је

$$\begin{aligned}\xi^{\mathcal{L}}(s,t_1,t_2) =&E(\Xi^{\mathcal{L}}(X_1,X_2,X_3,t_1,t_2)|X_1=s) = \frac{-1+e^{t_1}+e^{t_2}+e^{t_1+t_2}}{3(1+e^{t_1})(1+e^{t_2})(1+e^{t_1+t_2})}\\ &-\frac{e^{t_2}(1+e^{t_1})}{3(1+e^{t_2})(1+e^{t_1+t_2})}\mathrm{I}\{s>t_1\} - \frac{e^{t_1}(1+e^{t_2})}{3(1+e^{t_1})(1+e^{t_1+t_2})}\mathrm{I}\{s>t_2\}\\ &+\frac{1+e^{t_1+t_2}}{3(1+e^{t_1})(1+e^{t_2})}\mathrm{I}\{s>t_1+t_2\}\end{aligned}$$

Израз за дисперзију $\sigma^2_{\mathcal{L}}$ је превише гломазан да би се приказао. График $\sigma^2_{\mathcal{L}}$ је приказан на 5.8. Супремум дисперзије је



Слика 5.8: График функције $\sigma^2_{\mathcal{L}}(t_1,t_2)$,

$$\sigma^2_{\mathcal{L}} = \sup_{t_1,t_2\in R}\sigma^2_{\mathcal{L}} \approx \sigma^2_{\mathcal{L}}(0.669169, 0.669169) \approx 0.00945498.$$

На основу супремума и графика закључујемо да је фамилија језгара $\{\Xi^{\mathcal{L}}(\mathbf{X},t_1,t_2)),\ t_1,t_2\in R\}$ недегенерисана и да U-емпиријско поље

$$\nu_n^{\mathcal{L}}(t_1,t_2) = \sqrt{n}\left(H_n^{\mathcal{L}}(t_1,t_2) - G_n^{\mathcal{L}}(t_1,t_2)\right),\ t_1\,t_2>1$$

конвергира у $D(-\infty,\infty)$ ка центрираном Гаусовом пољу $\nu^{\mathcal{L}}(t_1,t_2)$ (в. [93]). Тада низ тест-статистика $\sqrt{n}K_n^{\mathcal{L}}$ конвергира у расподели ка случајној величини $\sup_{t_1,t_2\in R}|\nu^{\mathcal{L}}(t_1,t_2)|$.

Применом теорема 3.2.6 и 3.2.9 добијамо да важи следеће тврђење.

Лема 5.3.3. *Нека је $\varepsilon>0$. Функције великих одступања из теореме 3.2.2 за тест-статистике $I_n^{\mathcal{L}}$ и $K_n^{\mathcal{L}}$ су аналитичке у нули и могу се представити у облику:*

$$f_{I^{\mathcal{L}}}(\varepsilon) \approx 48.25\varepsilon^2 + o(\varepsilon^2),\ \varepsilon\to 0,$$
$$f_{K^{\mathcal{L}}}(\varepsilon) =\approx 5.87\varepsilon^2 + o(\varepsilon^2),\ \varepsilon\to 0.$$

Следећа лема нам даје граничну вредност у вероватноћи наших тест-статистика под претпоставком да важи алтернативна хипотеза H_1.

Лема 5.3.4. *Претпоставимо да је алтернативна густина $g(x;\theta)$ која припада фамилији $\mathcal{G}$ дефинисаној у одељку 3.2.2. Тада су граничне вредности у вероватноћи P_θ тест-статистика $I_n^{\mathcal{L}}$ и $K_n^{\mathcal{L}}$ кад $n \to \infty$*

$$b_{I^{\mathcal{L}}}(\theta) = 5\theta \int\limits_{-\infty}^{\infty} \psi^{\mathcal{L}}(x)h(x)dx + o(\theta),\ \theta \to 0, \tag{5.35}$$

$$b_{K^{\mathcal{L}}}(\theta) = 3\theta \sup_{t_1,t_2\in\mathbb{R}} \Big| \int\limits_{-\infty}^{\infty} \xi^{\mathcal{L}}(x;t_1,t_2)h(x)dx \Big| + o(\theta),\ \theta \to 0. \tag{5.36}$$

Делови лема 5.3.3 и 5.3.4 који се односе на статисику $K_n^{\mathcal{L}}$ приказани су у раду [69].

За логистичку расподелу се у литератури не може наћи много тестова сагласности а самим тим не постоје стандарндне алтернативе против којих се мери ефикасност теста. Зато су алтернативе које ћемо посматрати:

- алтернатива транслације са густином

$$g_l(x,\theta) = \frac{e^{-(x-\theta)}}{(1+e^{-(x-\theta)})^2},\ \ x \in R,\ \theta \in (0,1), \tag{5.37}$$

- генерализована логистичка расподела (ГЛ) са густином

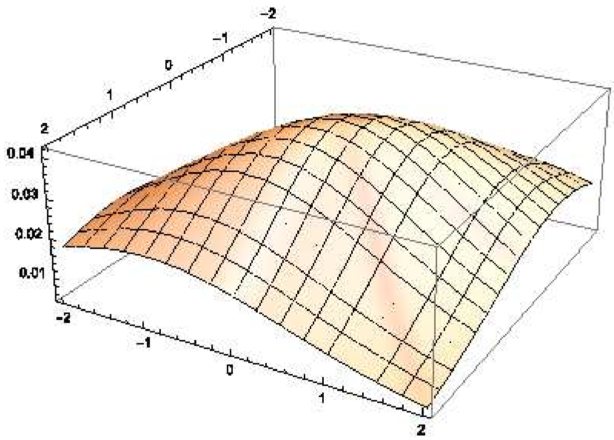
$$g_g(x,\theta) = \frac{(1+\theta)e^{-x}}{(1+e^{-x})^{2+\theta}},\ \ x \in R,\ \theta \in (0,1). \tag{5.38}$$

У случају логистичке расподеле није могуће наћи аналитички израз за инфимум (5.3). Зато не можемо извести израз за двоструко Кулбак-Лајблерово растојање сличан изразу (5.26) за Паретову расподелу. Зато ћемо за сваку алтернативу посебно одредити инфимум израза (5.3). При томе користимо теорему о имплицитним функцијама и чињеницу да су алтернативе блиске. За алтернативу (5.37) добијамо да је двоструко Кулбак-Лајблерово растојање постигнуто за $\widetilde{\lambda}_l(\theta) = 1 + o(\theta)$ и износи приближно $0.33\theta^2 + o(\theta^2)$, док је за алтернативу (5.38) једнак приближно $0.82\theta^2 + o(\theta^2)$ и постиже се за $\widetilde{\lambda}_g(\theta) \approx 1 - 0.35\theta + o(\theta),\ \ \theta \to 0$.

График функције $|a'_{lo}(t_1,t_2,0)|$ за алтернативу транслације приказан је на слици 5.9. Супермум апсолутне вредности функције се достиже у координатном почетку и износи 0.0417. График функције $|a'_{lo}(t_1,t_2,0)|$ за генерализоваy логистичку расподелу приказан је на слици 5.10. Супремум апсолутне вредности функције се достиже у координатном почетку и износи 0.0577623. У табели 5.9 су приказане локалне Бахадурове ефикасности за алтернативе (5.37) и (5.38).

Табела 5.9: Локална Бахадурова ефикасност за тест-статистике $I_n^{\mathcal{L}}$ и $K_n^{\mathcal{L}}$

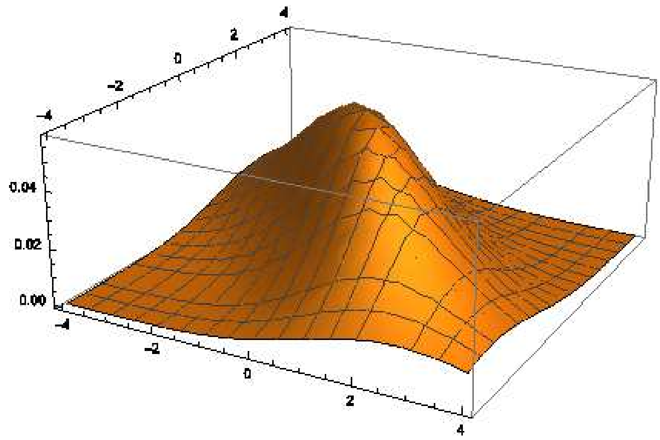
Алт.	$I_n^{\mathcal{L}}$	$K_n^{\mathcal{L}}$
g_l	0.988	0.551
g_g	0.845	0.430



Слика 5.9: График функције $|a'_{\mathcal{L}}(t_1, t_2, 0)|$ за алтернативу транслације

5.4 Тестови засновани на карактеризацијама на основу момената

У овом поглављу представљамо тестове сагласности засноване на карактеризацији униформне расподеле из теореме 4.4.1 за $k = 1$. Резултати су приказани у раду аутора [99]. Нека су $X_1, \dots, X_n$ независне и једнако расподељене случајне величине са непрекидном функцијом расподеле F_0. Без умањена општости можемо претпоставити да се ради о униформној $\mathcal{U}[0,1]$ расподели. Уколико желимо да тестирамо да је узорак из расподеле F_0 применићемо наш тест за униформну расподелу на трансформисан узорак $F_0(X_1), F_0(X_2), \dots, F_0(X_n)$.



Слика 5.10: График функције $|a'_{\mathcal{L}}(t_1,t_2,0)|$ за генерализовану логистичку расподелу

Да бисмо тестирали H_0 да је $F(x)=x,\ x\in(0,1)$ предлажемо следећу класу тест-статистика.

$$\begin{aligned}T_n^m=&\frac{1}{\binom{n}{m+1}}\sum_{i_1<\cdots<i_{m+1}}\Big(X^2_{(1),X_{i_1},\ldots,X_{i_m}}-\frac{2}{n+1}X_{(2),X_{i_1},\ldots,X_{i_{m+1}}}\\&+\frac{2}{(n+1)(n+2)}\Big)\end{aligned}\tag{5.39}$$

Ове статистике су асимптотски еквивалентне са U-статистикама са симетричним језгром

$$\begin{aligned}\Phi_m(X_1,..,X_{m+1})=&\frac{1}{(m+1)!}\sum_{\pi(m+1)}X^2_{(1),X_{\pi_1},\ldots,X_{\pi_m}}-\frac{2}{n+1}X_{(2),X_1,\ldots,X_{m+1}}\\&+\frac{2}{(n+1)(n+2)}.\end{aligned}$$

Прва пројекција језгра $\Phi_m(X_1,..,X_{m+1})$ на X_{m+1} ако важи H_0 је

$$\begin{aligned}\phi_m(s)=&E(\Phi_m(X_1,..,X_{m+1})|X_{m+1}=s)=\frac{m-1}{m}E(X_{(1),s,X_1,\ldots,X_{m-1}})\\&+\frac{1}{m}E(X_{(1),X_1,\ldots,X_m})-\frac{2}{n+1}E(X_{(2),s,X_1,\ldots,X_m})\\&+\frac{2}{(n+1)(n+2)},\ s\in[0,1].\end{aligned}$$

Лако се може показати да су прве пројекције идентички једнаке нули. Зато се, за разлику од претходно посматраних статистика у дисертацији, ради о U-статистикама са дегенерисаним језгром које, као што смо могли да видимо у поглављу о U-статистикама, имају другачија асимптотска својства од U- статистика са недегенерисаним језгром. Друга пројекција језгра $\Phi_m(X_1, .., X_{m+1})$ на (X_m, X_{m+1}) ако важи нулта хипотеза је

$$\begin{aligned}
\phi^*_m(s,t) &= E(\Phi_m(X_1,..,X_{m+1})|X_m = s_1, X_{m+1} = t)\\
&= \frac{m-1}{m+1}E(X_{(1),s,t,X_1,...,X_{m-3}}) + \frac{1}{m+1}E(X_{(1),s,X_1,...,X_{m-2}})\\
&+ \frac{1}{m+1}E(X_{(1),t,X_1,...,X_{m-2}}) - \frac{2}{n+1}E(X_{(2),s,t,X_1,...,X_{m-1}})\\
&+ \frac{2}{(n+1)(n+2)}, \quad s,t \in [0,1].
\end{aligned}$$

$$\begin{aligned}
E(X_{(1),s,t,X_1,...,X_{m-3}}) &= \mathrm{I}\{s<t\}\big(sP\{s < X_{(1),X_1,...,X_{m-3}}\}\\
&+ E(X_{(1),X_1,...,X_{m-3}}\mathrm{I}\{X_{(1),X_1,...,X_{m-3}} \geq s\})\big)\\
&+ \mathrm{I}\{s>t\}\big(tP\{t < X_{(1),X_1,...,X_{m-3}}\}\\
&+ E(X_{(1),X_1,...,X_{m-3}}\mathrm{I}\{X_{(1),X_1,...,X_{m-3}} \geq t\})\big)
\end{aligned}$$

$$\begin{aligned}
E(X_{(1),s,X_1,...,X_{m-2}}) &= sP\{s < X_{(1),X_1,...,X_{m-2}}\}\\
&+ E(X_{(1),X_1,...,X_{m-2}}\mathrm{I}\{s > X_{(1),X_1,...,X_{m-2}}\})
\end{aligned}$$

$$\begin{aligned}
E(X_{(1),t,X_1,...,X_{m-2}}) &= tP\{t < X_{(1),X_1,...,X_{m-2}}\}\\
&+ E(X_{(1),X_1,...,X_{m-2}}\mathrm{I}\{t > X_{(1),X_1,...,X_{m-2}}\})
\end{aligned}$$

$$\begin{aligned}
E(X_{(2),s,t,X_1,...,X_{m-1}}) &= \mathrm{I}\{s<t\}\big(tP\{X_{(1),X_2,...,X_{m-1}} > t\}\\
&+ sP\{X_{(1),X_2,...,X_{m-1}} < s < X_{(2),X_2,...,X_{m-1}}\\
&+ E(X_{(1),X_1,...,X_{m-1}}\mathrm{I}\{s < X_{(1),X_1,...,X_{m-1}} < t\})\\
&+ E(X_{(2),X_1,...,X_{m-1}}\mathrm{I}\{X_{(2),X_1,...,X_{m-1}} > s\})\}\big)\\
&+ \mathrm{I}\{s>t\}\big(sP\{X_{(1),X_2,...,X_{m-1}} > s\}\\
&+ tP\{X_{(1),X_2,...,X_{m-1}} < t < X_{(2),X_2,...,X_{m-1}}\\
&+ E(X_{(1),X_1,...,X_{m-1}}\mathrm{I}\{t < X_{(1),X_1,...,X_{m-1}} < s\})\\
&+ E(X_{(2),X_1,...,X_{m-1}}\mathrm{I}\{X_{(2),X_1,...,X_{m-1}} > t\})\}\big).
\end{aligned}$$

Након рачунања добијамо

$$\phi_m^*(s,t) = -\frac{2}{m(1+m)^2(2+m)}\Big(-2 + 2(1-t)^m + m^2\big(-(1-s)^{m+1} + (1-t)^m t\big) + m\big(-2(1-s)^{m+1} + (1-t)^m + 2(1-t)^m t\big)\Big) + \frac{2}{m(1+m)}\mathrm{I}\{s<t\}((1-t)^m - (1-s)^m).$$

За $m = 1$ и $m = 2$ добијају се следеће друге пројекције језгра

$$\phi_1^*(s_1,s_2) = \frac{s_1^2}{2} + \frac{s_2^2}{2} - \max(s_1,s_2) + \frac{1}{3}, \tag{5.40}$$

$$\phi_2^*(s_1,s_2) = \frac{1}{3}\mathrm{I}\{s<t\}(s^2 - 2t + t^2 - s^2) + \frac{1}{3}\mathrm{I}\{s>t\}(t^2 - 2s + s^2 - t^2\Big) + \frac{1}{6} + \frac{1}{3}s^2 + \frac{1}{3}t^2 - \frac{2}{9}s^3 - \frac{2}{9}t^3. \tag{5.41}$$

Очигледно је да је друга пројекција различита од нуле. Зато можемо закључити да је наша класа статистика класа U-статистика са слабо дегенерисаним језгром.

Користећи теорему 1.1.4 добијамо да важи

$$nT_n^m \xrightarrow{d} \binom{m+1}{2}\sum_{i=1}^{\infty}\nu_i(\tau_i^2 - 1),$$

где су ν_i сопствене вредности интегралног оператора S дефинисаног са

$$Sf(t) = \int\limits_0^1 \phi_m^*(s,t)f(s)ds.$$

Дакле, треба да решимо интегралну једначину

$$\nu f(t) = \int\limits_0^1 \phi_m^*(s,t)f(s)ds, \tag{5.42}$$

уз услов $\int\limits_0^1 f(s)ds = 0$.

Означимо са $y(t) = \int_0^t f(s)ds$. Након диференцирања израз (5.42) постаје

$$\nu y''(t) = -2\frac{(1-t)^{m-1}}{m+1}y(t), \quad y(0) = 0, \;\; y(1) = 0.$$

Да бисмо решили овај гранични проблем уведимо смену $t = 1 - s$ и $\lambda = \nu^{-1}$. Тада добијамо једначину

$$y''(s) + \frac{2}{m+1}\lambda y(s)s^{m-1} = 0,\ y(0) = 1,\ y(1) = 1. \tag{5.43}$$

У тексту који следи показаћемо како се може решити (5.43) у општем случају. За $m = 1$ и $m = 2$ одредићемо тачну или приближну вредност највеће карактеристичне вредности (реципрочне вредности најмање сопствене вредности) потребну за одређивање локалне Бахадурове ефикасности теста.

За $m = 1$ израз (5.43) постаје познати Штурм-Луивилов гранични проблем

$$y''(t) + \lambda y(t) = 0,\ y(0) = 0,\ y(1) = 0,$$

па се цео спектар оператора S може одредити. Добијамо да су карактеристичне вредности (реципрочне вредности сопствених вредности) $\lambda_k = k^2\pi^2,\ \ k = 1, 2...$ па је главна карактеристична вредност $\lambda_1 = \pi^2$.

Интресантно је да тест-статистика Крамера и фон Мизесова статистика има исто језгро као наша статистика T_n^1. Због тога ће оне имати слична асимптотска својства.

У случају (5.41), израз (5.43) постаје

$$y''(s) + \frac{2}{3}\lambda y(s)s = 0,\ y(0) = 1,\ y(1) = 1.$$

Користећи резултат из књиге [51] (2.162, стр.440) за $a = 0$, $c = 0$ и $b \neq 0$ добијамо да је решење једначине линеарна комбинација Беселових функција прве врсте

$$y(s) = 2^{\frac{1}{6}}\lambda^{\frac{1}{6}}\sqrt{s}\left(J_{-\frac{1}{3}}\Big((\frac{2}{3}s)^{\frac{3}{2}}\sqrt{\lambda}\Big)\Gamma(\frac{2}{3})C_1 + J_{\frac{1}{3}}\big((\frac{2}{3}s)^{\frac{3}{2}}\sqrt{\lambda}\big)\Gamma(\frac{4}{3})C_2\right),$$

при чему је

$$J_\nu(x) = \sum_{k=0}^{\infty}\frac{(-1)^k\left(\frac{x}{2}\right)^{\nu+2k}}{k!\Gamma(\nu+k+1)}.$$

Из граничних услова добијамо систем

$$\begin{aligned}
&\lim_{s\to 0} y(s) = 0 = C_1,\\
&y(1) = 0 = 2^{\frac{1}{6}}\lambda^{\frac{1}{6}}\frac{1}{\sqrt{3}}J_{\frac{1}{3}}\Big(\frac{2\sqrt{2\lambda}}{3^{\frac{3}{2}}}\Big)\Gamma\Big(\frac{4}{3}\Big)C_2,
\end{aligned}$$

који има решења уколико је λ нула Беселове функције $J_{\frac{1}{3}}\big((\frac{2}{3})^{\frac{3}{2}}\sqrt{\lambda}\big)$. Користећи нумеричке методе добијамо да је главна карактеристична вредност $\lambda_1 = 28.4344$.

У општем случају, да бисмо одредили λ за $m \geq 1$ треба да одредимо када следећи систем има решења:

$$\Big(1+\frac{1}{m}\Big)^{-\frac{1}{m+1}}\Big(\frac{m}{m+1}\Big)^{\frac{-1}{m+1}}C_1 = 0,$$
$$(2\lambda)^{\frac{1}{2+2m}}\big(1+\frac{1}{m}\big)^{-\frac{1}{m+1}}(m^2(m+1))^{\frac{1}{2+2m}}J_{\frac{1}{m+1}}\Big(\frac{2\sqrt{2\lambda}}{(1+m)^{\frac{3}{2}}}\Big)\Gamma\big(\frac{2+m}{1+m}\big)C_2 = 0.$$

Зато је проблем одређивања карактеристичних вредности λ еквивалентан проблему одређивања нула Беселових функција $J_{\frac{1}{m+1}}(x)$, који се може решити за свако $m \geq 1$.

У тексту који следи одредићемо локалну Бахадурову ефикасност тестова и упоредити их са неким конкурентским тестовима.

С обзиром на то да је нулта хипотеза проста двоструко Кулбак-Лајблерово растојање блиске алтернативе са функцијом густине $g(x,\theta)$ је

$$2K(\theta) = I(g)\theta^2 + o(\theta^2), \quad \theta \to 0, \tag{5.44}$$

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

Конкурентски тестови су

- тест Комогорова и Смирнова

$$D_n = \sup_{t\in(0,1)} |F_n(t) - t|; \tag{5.45}$$

- тест Андерсона и Дарлинга

$$A_n^2 = \int_0^1 \frac{(F_n(t)-t)^2}{t(1-t)}dt; \tag{5.46}$$

- тест Крамера и фон Мизеса

$$W_n^2 = \int_0^1 (F_n(t)-t)^2 dt; \tag{5.47}$$

- тест Фортиана и Гране заснован на максималним корелацијама [26]

$$Q^c = |Q_n - 1| = |\frac{6}{n^2}\sum_{i=1}^{n}(2i-n-1)X_{(i)} - 1|; \qquad (5.48)$$

- тест Хашимота и Ширахате ([36])

$$C_n = \binom{n}{4}^{-1} \sum_{i<j<k<l} h(X_i, X_j, X_k, X_l); \qquad (5.49)$$

где је

$$\begin{aligned}
&h(X_i, X_j, X_k, X_l) = \frac{1}{36}\Big((X_i - X_j)^2 + (X_i - X_k)^2 + (X_i - X_l)^2 \\
&+ (X_j - X_k)^2 + (X_j - X_l)^2 + (X_k - X_l)^2\Big) \\
&- \frac{1}{6}\Big((\max(X_i, X_j) - \max(X_k, X_l))(\min(X_i, X_j) - \min(X_k, X_l) \\
&+ (\max(X_i, X_k) - \max(X_j, X_l))(\min(X_i, X_k) - \min(X_j, X_l) \\
&+ (\max(X_i, X_l) - \max(X_j, X_k))(\min(X_i, X_l) - \min(X_j, X_k))\Big).
\end{aligned}$$

Алтернативе униформној против којих ћемо одредити локалне Бахадурове ефикасности су:

- степена расподела са функцијом густине

$$g_1(x, \theta) = (\theta + 1)x^{\theta},\ x \in (0, 1),\ \theta > 0; \qquad (5.50)$$

- расподела са густином

$$g_2(x, \theta) = 1 + \theta(2x - 1)\ x \in (0, 1),\ \theta > 0; \qquad (5.51)$$

- мешавина униформне и степене расподеле са густином

$$g_3(x, \theta) = 1 - \theta + \beta\theta x^{\beta-1},\ x \in (0, 1); \qquad (5.52)$$

- друга Леј-Паиндавеине алтернатива са густином

$$g_4(x, \theta) = 1 - \theta\pi \sin \pi x,\ x \in (0, 1)\ \theta \in [0, \pi^{-1}]. \qquad (5.53)$$

Може се показати да ове алтернативе задовољавају услове регуларности дефинисане у одељку 3.2.2 .

За неке од конкурентских тестова су познате фукције великих одступања неопходне за одређивање локалне Бахадурове ефикасности. Функције се могу наћи у [75] и [31] за тестове (5.45)-(5.48)

$$
\begin{aligned}
f_D(a) &= -2a^2 + o(a^2),\ a \to 0,\\
f_A(a) &= -a + o(a^2),\ a \to 0,\\
f_Q(a) &= -\frac{5}{2}a^2 + o(a^2),\ a \to 0.
\end{aligned}
$$

У тексту који следи ћемо одредити функцију великих одступања за Хашимото-Ширахата тест-статистику 5.49. Напомињемо да то у раду Хашимота и Ширахате није урађено.

Приметимо да је C_n U-статистика са језгром $h(X_1, X_2, X_3, X_4)$.

Није тешко показати да се ради о слабо дегенерисаном језгру. Пројекција језгра на (X_1, X_2) је

$$
\begin{aligned}
h^*(s_1, s_2) = E(h(X_1, X_2, X_3, X_4)|X_1 = s_1, X_2 = s_2) = \frac{s_1^2 s_2 + s_2^2 s_1}{6}\\
+\frac{\min(s_1, s_2)}{18} - \frac{2s_1 s_2}{9} - \frac{s_1^2 s_2^2}{6}
\end{aligned}
$$

Да бисмо применили теорему 3.2.7 потребно је да нађемо најмање λ која задовољава интегралну једначину

$$
x(s_1) = \lambda \int\limits_0^1 h^*(s_1, s_2)x(s_2)ds_2,\ \int\limits_0^1 x(s_2)ds_2 = 0,
$$

односно

$$
x'''(s_1) = -\frac{\lambda}{18}x'(s_1),\ x(0) = x(1) = 0,\ \int\limits_0^1 x(s_2)ds_2 = 0.
$$

Проналазимо да је тражено $\lambda = \lambda_0 = 72\pi^2$, па је функција великих одступања за статистику C_n

$$
f_C(a) = -6\pi^2 a + o(a),\ \ a \to 0.
$$

Следећим примером илустроваћемо одређивање локалне Бахдурове ефикасности за наше тест-статистике и Хашимото-Ширахата тест.

Пример 5.4.1. Нека је алтернативна расподела степена расподела са густином (5.50). Тада је

$$h(x) = 1 + \log x.$$

Лако се показује да степена расподела задовољава услове регуларности дефинисане у одељку 3.2.2 па је

$$2K(\theta) = 1 \cdot \theta^2 + o(\theta^2).\ \theta \to 0.$$

Коришћењем теореме 3.2.11 добијамо да је гранична вредност у вероватноћи тест-статистике C_n кад $n \to \infty$, под претпоставком да се ради о степеној расподели блиској униформној

$$b_C(\theta) = \frac{4 \cdot 3}{2 \cdot 1944}\theta + o(\theta),\ \theta \to 0.$$

Применом теореме 3.2.7 добијамо да је тражена локална Бахадурова ефикасност

$$e_C = 72\pi^2 \frac{1}{1944} = 0.37.$$

Слично за тест-статистике T_n^1 и T_n^2 добијамо

$$b_{T^1}(\theta) = \frac{2}{27}\theta + o(\theta), \theta \to 0,$$
$$b_{T^2}(\theta) = \frac{3 \cdot 37}{1296}\theta + o(\theta), \theta \to 0.$$

На основу чега закључујемо да су локалне Бахадурове ефикасности

$$e_{T^1} = \pi^2 \cdot \frac{2}{27} = 0.73,$$
$$e_{T^2} = 28.43 \cdot \frac{37}{1296} = 0.81.$$

Локалне Бахадурове ефикасности за остале алтернативе приказане су у табели 5.10. Видимо да су предложени тестови доста великих ефикасности, у поређењу са конкурентским тестовима.

Предложене тестове упоредићемо и као тестове сагласности са нормалном и Кошијевом расподелом против одговарајућих алтернатива транслације. Приказаћемо како се у овом случају, када се уместо нулте хипотезе да је X обележје са униформном $\mathcal{U}[0,1]$ расподелом, тестира нулта хипотеза да је X са функцијом расподеле F_0 за коју важи $0 < F_0(x) < 1$ за свако x. Нека је алтернативна функција расподеле $F(x,\theta) = F_0(x - \theta)$, одговарајућа густина $f(x,\theta) = f_0(x - \theta)$

Табела 5.10: Локална Бахадурова ефикасност

Алтернативе	A	D	$T^{(2)}$	Q^c	C	$T^{(1)}$
g_1	0.40	0.54	**0.81**	0.14	0.37	**0.73**
g_2	0.5	0.75	**0.95**	0	0.66	**0.98**
$g_3(3)$	0.48	0.74	**0.82**	0.06	0.63	**0.94**
g_4	0.49	0.81	**0.96**	0	0.76	**1**

и $h^*(x) = f'_\theta(x,0)$. Ако је нулта хипотеза тачна онда $F_0(X)$ има униформну $U[0,1]$ расподелу а алтернативна функција расподеле која одговара функцији расподеле $F_0(x-\theta)$ је

$$G(x,\theta) = P_\theta\{F_0(X) \leq x\} = P_\theta\{X \leq F_0^{-1}(x)\} = F_0(F_0^{-1}(x) - \theta),\ x \in (0,1).$$

Дакле, $G(x,\theta)$ је (блиска) алтернативна функција расподеле униформној функцији расподеле. Приметимо да је $G(x,0) = x$. Тада се наше тестирање може представити у облику

$$H_0 : U[0,1]\ (\theta = 0) \text{ против } H_1 :\ G(x,\theta)\ (\theta > 0).$$

Густина алтернативе је

$$g(x,\theta) = \frac{\partial G(x,\theta)}{\partial x} = \frac{f_0(F_0^{-1}(x),\theta)}{f_0(F_0^{-1}(x))} = \frac{f_0(F_0^{-1}(x)-\theta)}{f_0(F_0^{-1}(x))}.$$

Тада је

$$h(x) = g'_\theta(x,0) = \frac{h^*(F_0^{-1}(x))}{f_0(F_0^{-1}(x))},$$

одакле следи следећи низ једнакости

$$\int_0^1\int_0^1 h(x)h(y)\phi_m^*(x,y)dxdy = \int_0^1\int_0^1 \frac{h^*(F_0^{-1}(x))}{f_0(F_0^{-1}(x))}\frac{h^*(F_0^{-1}(y))}{f_0(F_0^{-1}(y))}\phi_m^*(x,y)dxdy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{h^*(u)}{f_0(u)}\frac{h^*(v)}{f_0(v)}\phi_m^*(F_0(u),F_0(v))f_0(u)f_0(v)dudv$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^*(u)h^*(v)\phi_m^*(F_0(u),F_0(v))dudv.$$

Како $G(x,\theta)$ и $G(x,0)$ имају исти носач $(0,1)$ можемо одредити двоструко Кулбак-Лајблерово растојање у функцији од Фишерове информационе функције, односно важи

$$I = \int_0^1 h^2(x)dx = \int_0^1 \Big(\frac{h^*(F_0^{-1}(x))}{f_0(F_0^{-1}(x))}\Big)^2 dx = \int_{-\infty}^{\infty}\Big(\frac{h^*(u)}{f_0(u)}\Big)^2 f_0(u)du$$
$$= \int_{-\infty}^{\infty}\frac{(h^*(u))^2}{f_0(u)}du.$$

Пример 5.4.2. Желимо да тестирамо нулту хипотезу да обележје има нормалну $\mathcal{N}(0,1)$ расподелу са функцијом расподеле $F(x)$ и густином $f(x)$, против алтернативе да има $\mathcal{N}(\theta,1)$, где је очекивање $\theta > 0$. Одавде добијамо да је

$$h^*(x) = \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}}.$$

Двоструко Кулбак-Лајблерово растојање је

$$2K(\theta) = 1 \cdot \theta^2 + o(\theta^2),\ \theta \to 0.$$

Тада је гранична вредност у вероватноћи статистике T_n^1, кад $n \to \infty$

$$b_{T^1}(\theta) = \int\limits_{-\infty}^{\infty} h^*(x)\phi_1^*(F(x))dx \cdot \theta + o(\theta) \approx 0.092\theta + o(\theta),\ \theta \to 0,$$

одакле добијамо да је локална Бахадурова ефикасност $0.092 \cdot \pi^2 = 0.91$.

Табела 5.11: Локална Бахадурова ефикасност тестова против алтернатива транслације

Стат.	Нормална р.	Кошијева р.
A_n^2	0.96	0.66
D_n	0.64	0.81
Q_n^c	0	0
C_n	0.49	1
T_n^1	0.91	0.76
T_n^2	0.87	0.72

Интресантно је приметити да је Хашимо-Ширахата локално оптималан тест за Кошијеву расподелу. Следећом теоремом дате су неке класе локално оптималних алтернатива за статистику T_n^m

Теорема 5.4.1 (Милошевић 2015)**.** *Нека је $g(x;\theta)$ функција густине из фамилије $\mathcal{G}$. Тада, за мало θ, алтернативе густине са функцијом $h(x)$*

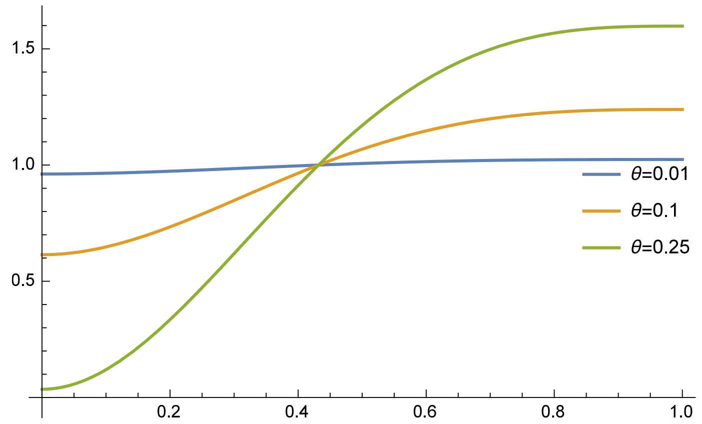
$$h(x) = Cf_{0,m}(x),\ x \in [0,1],\ C > 0,$$

где $f_{0,m}(x)$ је сопствена функција која одговара првој сопственој вредности интегралног оператора (5.42), су локално оптималне алтернативе за тест-статистику T_n^m.

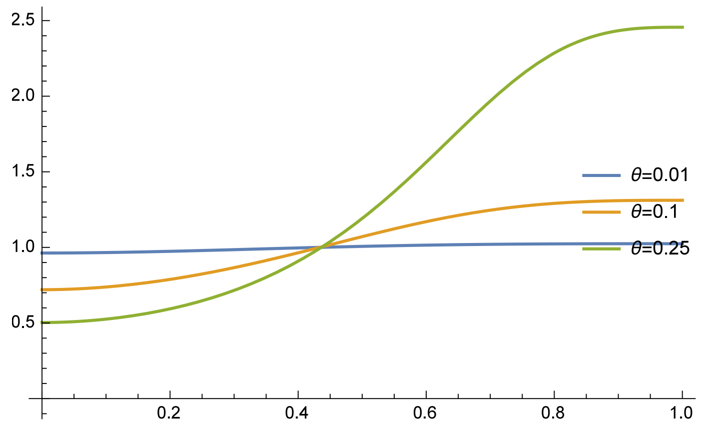
Примењујући ову теорему за $T_n^{(1)}$ добијамо да су неке класе локално оптималних алтернатива

$$g_{(1),1}(x,\theta) = 1 + \theta\cos\pi x, \quad x \in [0,1],$$
$$g_{(1),2}(x,\theta) = \frac{3(2-\theta)\theta}{2\arctan\sqrt{\frac{3\theta}{2-\theta}}}\frac{1}{1-\theta\cos\pi x}, x \in [0,1].$$

На сликама 5.11, 5.12 су приказане неке класе локално оптималних алтернатива за статистику T_n^2.



Слика 5.11: Локално оптималне алтернативе $g_{(2),1}(x,\theta)$



Слика 5.12: Локално оптималне алтернативе $g_{(2),2}(x,\theta)$

Поглавље 6

Тестови симетрије

Тестирање својства симетрије расподеле око нуле још је један од проблема непараметарске статистике и могао би се дефинисати на следећи начин. Нека је $X_1, ..., X_n$ прост случајан узорак обима n са непрекидном функцијом расподеле F. Желимо да тестирамо нулту хипотезу

$$H_0 : \quad F(x) = 1 - F(-x), \quad x \in R, \tag{6.1}$$

против алтернативе да је функција расподеле посматраног обележја $G(x, \theta)$ за коју једнакост (6.1) не важи осим у случају $\theta = 0$. Класу расподела за које важи (6.1) означићемо са $\mathcal{H}$. У монографији Никитина ([75]) представљене су тест-статистике које су аналог стандардним тест-статистикама Колмогорова и Смирнова, Крамера и фон Мизеса, Вотсона и Дарлинга за тестирање сагласности са расподелом. Испитана су њихова асимптотска својства и нађене функције великих одступања потребне за одређивање њихове Бахадурове ефикасности против алтернатива транслације. У радовима Бурђа и Никитина [16], [17] су предложени тестови симетрије засновани на линеарној комбинацији тест-статистике знакова и Вилкоксонове, односно Маезонове тест-статистике. Одређене су њихове Питманове ефикасности против алтернатива транслације. У раду Барингхауса и Хензеа су предложени тестови Крамер-фон Мизесовог и Колмогоровљевог типа заснованих на карактеризацији из теореме 4.1.4, док је у раду Литвинове ([60]) предложен тест интегралног типа на основу исте карактеризације. У оба рада су одређене Бахадурове ефикасности тестова.

Важно је напоменути да се Бахадурова ефикасност тестова симетрије рачуна на мало другачији начин од Бахадурове ефикасности тестова сагласности. За разлику од неједнакости (3.6) може се показати да за

Бахадуров нагиб низа тест-статистика $\{T_n\}$ и $g(x,\theta)$ важи

$$c_T(\theta) \leq 2\inf\Big\{\int\limits_{-\infty}^{\infty}\log\frac{g(x,\theta)}{g(x,0)}g(x,\theta)dx:\ g(x,0)\in\mathcal{H}\Big\}$$

$$= 2\int\limits_{-\infty}^{\infty} g(x,\theta)\log\frac{2g(x,\theta)}{g(x,\theta)+g(-x,\theta)}dx. \tag{6.2}$$

Уколико функција $g(x,\theta)$ припада фамили $\mathcal{G}$ дефинисаној у одељку 3.2.2 онда је десна страна неједнакости (6.2) асимптотски еквивалентна са

$$\frac{1}{4}I_1(g)\cdot\theta^2,$$

где је

$$I_1(g) = \int\limits_{-\infty}^{\infty}\frac{(g'_\theta(x)-g'_\theta(-x))^2}{g(x,0)}dx.$$

Зато је у овом случају природно дефинисати локалну Бахадурову ефикасност као количник леве и десне стране неједнакости (6.2)

$$e_T^B = \lim_{\theta\to 0}\frac{c_T(\theta)}{2\int\limits_{-\infty}^{\infty} g(x,\theta)\log\frac{2g(x,\theta)}{g(x,\theta)+g(-x,\theta)}dx}. \tag{6.3}$$

У овом поглављу предлажемо неколико нових тестова симетрије заснованих на U-емпиријским функцијама расподеле. Почећемо са тест-статистикама интегралног и Колмогоровљевог типа заснованих на карактеризацији из теореме 4.1.6. Нулту хипотезу одбацујемо ако је вредност тест-статистике велика. Применићемо карактеризацију за $n = 2k$, и посматрати централне две статистике поретка, односно расподела обележја X је симетрична ако и само ако је

$$|X_{(k),X_1,\dots,X_{2k}}| \overset{d}{=} |X_{(k+1),X_1,\dots,X_{2k}}|.$$

Нека је

$$I_n^{1,(k)} = \int_0^\infty (H_n^{1,(k)}(t) - G_n^{1,(k)}(t))dQ_n(t), \tag{6.4}$$

где су

$$H_n^{1,(k)}(t) = \frac{1}{\binom{n}{2k}}\sum_{\mathcal{I}_{2k}}\mathrm{I}\{|X_{(k),X_{i_1},\dots X_{i_{2k}}}| < t\},$$

$$G_n^{1,(k)}(t) = \frac{1}{\binom{n}{2k}}\sum_{\mathcal{I}_{2k}}\mathrm{I}\{|X_{(k+1),X_{i_1},\dots X_{i_{2k}}}| < t\},$$

U-емпиријске функције расподеле, Q_n емпиријска функција расподеле узорка $|X_1|, ..., |X_n|$, и $\mathcal{I}_m = \{(i_1, ..., i_m) : 1 \leq i_1 < i_2 < \cdots < i_m \leq n\}$.

Након интеграције добијамо да се статистика (6.4) може представити у облику

$$\frac{1}{n\binom{n}{2k}} \sum_{\mathcal{I}_{2k}} \sum_{i_{2k+1}=1}^{n} \mathrm{I}\{|X_{(k),X_{i_1},...X_{i_{2k}}}| < |X_{i_{2k+1}}|\} - \mathrm{I}\{|X_{(k+1),X_{i_1},...X_{i_{2k}}}| < |X_{i_{2k+1}}|\}.$$

Ради знатно једноставнијег рачуна важно је показати да овако дефинисана статистика не зависи од избора расподеле и да без умањења општости можемо претпоставити да се ради о узорку из униформне $U[-1, 1]$ расподеле.

Увешћемо и додатну претпоставку да је функција F строго растућа функција. Тада њен инверз F^{-1} постоји и такође је растућа функција. Нека је $y = F(t)$. Тада је

$$I_n^{1,(k)} = \int_{\frac{1}{2}}^{1} (H_n^{(k)}(F^{-1}(y)) - G_n^{(k)}(F^{-1}(y)))dQ_n(F^{-1}(y)). \tag{6.5}$$

За $y > \frac{1}{2}$ важи

$$\begin{aligned}
H_n^{1,(k)}(F^{-1}(y)) &= \frac{1}{\binom{n}{2k}} \sum_{\mathcal{I}_{2k}} \mathrm{I}\{|X_{(k),X_{i_1},...X_{i_{2k}}}| < F^{-1}(y)\} \\
&= \frac{1}{\binom{n}{2k}} \sum_{\mathcal{I}_{2k}} \mathrm{I}\{-F^{-1}(y) < X_{(k),X_{i_1},...X_{i_{2k}}} < F^{-1}(y)\} \\
&= \frac{1}{\binom{n}{2k}} \sum_{\mathcal{I}_{2k}} \mathrm{I}\{F^{-1}(1-y) < X_{(k),X_{i_1},...X_{i_{2k}}} < F^{-1}(y)\} \\
&= \frac{1}{\binom{n}{2k}} \sum_{\mathcal{I}_{2k}} \mathrm{I}\{1-y < F(X_{(k),X_{i_1},...X_{i_{2k}}}) < y\} \\
&= \frac{1}{\binom{n}{2k}} \sum_{\mathcal{I}_{2k}} \mathrm{I}\{1-y < F(X)_{(k),F(X_{i_1}),...F(X_{i_{2k}})} < y\} \\
&= \frac{1}{\binom{n}{2k}} \sum_{\mathcal{I}_{2k}} \mathrm{I}\{1-y < U_{(k),U_{i_1},...,U_{i_{2k}}} < y\},
\end{aligned}$$

где су $\{U_i\}$ независне случајне величине са униформном $U[0, 1]$ расподелом. Користили смо да уколико је F функција расподеле случајне величине симетричне у односу на нулу тада је за $y > \frac{1}{2}$

$$-F^{-1}(y) = F^{-1}(1-y).$$

Слично се показује да је

$$G_n^{1,(k)}(F^{-1}(y)) = \frac{1}{\binom{n}{2k}} \sum \mathrm{I}\{1-y < U_{(k+1),U_{i_1},\dots,U_{i_{2k}}} < y\}.$$

Заменом добијених израза за U-емпиријске функције у (6.5) наше тврђење је показано. У даљем тексту претпостављамо да је нулта хипотеза нашег теста да су $X_1, \dots X_n$ из $U[-1,1]$ расподеле.

Тест статистика $I_n^{1,(k)}$ је асимптотски еквивалентна U-статистици са симетричним језгром

$$\begin{aligned}\Phi^{1,(k)}(X_1, \dots . X_{2k+1}) = \frac{1}{(2k+1)!} \sum_{\pi(2k+1)} & \mathrm{I}\{|X_{(k),X_{\pi(i_1)},\dots,X_{\pi(i_{2k})}}| < |X_{\pi(i_{2k+1})}|\} \\ & - \mathrm{I}\{|X_{(k+1),X_{\pi(i_1)},\dots,X_{\pi(i_{2k})}}| < |X_{\pi(i_{2k+1})}|\}\end{aligned}$$

чија је пројекција на X_1, под претпоставком да важи H_0 и $s \in [-1,1]$

$$\begin{aligned}&\phi^{1,(k)}(s) = E(\Phi^{1,(k)}(X_1, \dots X_{2k+1})|X_1 = s) \\ &= \frac{1}{2k+1}\Big(P\{|X_{(k),X_2,\dots X_{2k+1}}| < |s|\} - P\{|X_{(k+1),X_2,\dots,X_{2k+1}}| < |s|\}\Big) \\ &+ \frac{2k}{2k+1}\Big(P\{|X_{(k),s,\dots X_{2k}}| < |X_{2k+1}|\} - P\{|X_{(k+1),s,\dots,X_{2k}}| < |X_{2k+1}|\Big).\end{aligned}$$

Приметимо да је разлика у првој загради једнака нули због карактеризације. Након рачунања друге две вероватноће добијамо да је

$$\phi^{1,(k)}(s) = \begin{cases} -\frac{1}{(2k+1)2^{2k-1}}((s^2-1)^k - (-1)^k), & s \in (-1,0) \\ \frac{1}{(2k+1)2^{2k-1}}((s^2-1)^k - (-1)^k), & s \in (0,1); \end{cases}$$

Лако се показује да је $E(\phi^{1,(k)}(X_1)) = 0$. Дисперзија пројекције је

$$\begin{aligned}\sigma^2_{I^{1,(k)}} &= E(\phi^{1,(k)}(X_1)^2) \\ &= \frac{\sqrt{\pi}\Gamma(k+\frac{3}{2})\Gamma(2k+1) - 2\Gamma(2k+\frac{3}{2})(\sqrt{\pi}\Gamma(k+1) - \Gamma(k+\frac{3}{2})))}{2^{4k-1}((2k+1)^2\Gamma(k+\frac{3}{2})\Gamma(\frac{3}{2}+2k)}\end{aligned}$$

Користећи теорему 1.1.2 добијамо да је гранична расподела статистике $\sqrt{n}I^{1,(k)}$ нормална $\mathcal{N}(0,(2k+1)^2\sigma^2_{I^{1,(k)}})$.

Нека је

$$K_n^{1,(k)} = \sup_{t>0}\Big|H_n^{1,(k)}(t) - G_n^{1,(k)}(t)\Big|.$$

Означимо са $K_n^{1,(k)*}(t) = H_n^{(k)}(t) - G_n^{(k)}(t)$.

За свако фиксирано $t \in (0,1)$ $K_n^{1,(k)*}(t)$ је U-статистика са језгром

$$\Xi^{1,(k)}(X_1,...,X_{2k},t) = \mathrm{I}\{|X_{(k),X_1,...X_{2k}}| < t\} - \mathrm{I}\{|X_{(k+1),X_1,...,X_{2k}}| < t\}.$$

Прва пројекција на X_1 под претпоставком да важи H_0 је

$$\begin{aligned}\xi^{1,(k)}(s,t) &= E(\Xi^{(k)}|X_1 = s)\\ &= \binom{2k-1}{k-1} t\Big(\frac{t+1}{2}\Big)^{k-1}\Big(\frac{1-t}{2}\Big)^{k-1}\Big(- \mathrm{I}\{s < -t\} + \mathrm{I}\{s > t\}\Big)\end{aligned}$$

У раду Ахсанулаха и Никитина ([81]) посматран је случај $k = 1$. За $k = 2$ добија се пројекција

$$\xi^{1,(2)}(s,t) = \frac{3}{4}\mathrm{I}\{s < -t\}(t^3 - t) + \frac{3}{4}\mathrm{I}\{s > t\}(t - t^3).$$

Њена дисперзија је

$$\sigma^2_{K^{1,(2)}}(t) = \frac{9}{16}(1-t)(t^3-t)^2.$$

График функције дисперзије за $k = 2$ приказан је на слици 6.1. Супремум се достиже у тачки $t_0 = 0.47$ и једнак је 0.040. На основу графика закључујемо да је фамилија језгара $\{\Xi^{1,(2)}(\cdot;t), t \in (0,1)\}$ недегенерисана. Користећи теорему Силвермана 1.3.2 закључујемо да уколико важи нулта хипотеза H_0 онда процес $\sqrt{n}K_n^{1,(2)*}(t)$ слабо конвергира центрираном Гаусовом процесу $\{\kappa^{1,(2)}(t),\ t \in (0,1)\}$ док статистика $\sqrt{n}K_n^{1,(2)}(t)$ конвергира у расподели случајној величини $\sup_t |\kappa^{1,(2)}(t)|$ чија је расподела непозната.

Дисперзија пројекције у општем случају је

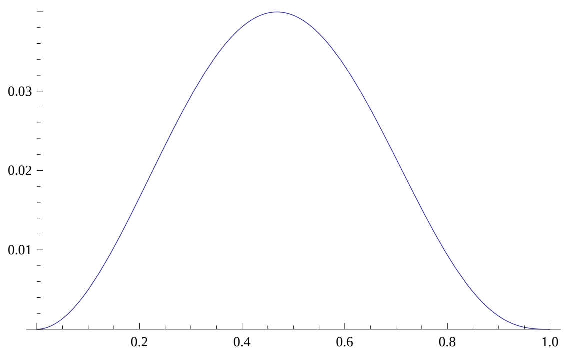
$$\sigma^2_{K^{1,(k)}}(t) = \binom{2k-1}{k-1}^2 2^{4-4k} t^2(1-t)(1-t^2)^{2k-2},\ \ t \in (0,1).$$

Супремум дисперзије пројекције у општем случају је

$$\sigma^2_{K^{1,(k)}} = \binom{2k-1}{k-1}^2 \frac{(\sqrt{32k-7}-1)^2(\frac{-1+\sqrt{32k-7}}{2(4k-1)} - 1)^{2k-1}(1 + \frac{-1+\sqrt{32k-7}}{2(4k-1)})^{2k-2}}{2^{4k-2}(4k-1)^2}.$$

Посматрајући функцију дисперзије за свако фиксирано $k \in N$, користећи теорему 1.3.2 долазимо до закључка да уколико важи нулта хипотеза H_0 онда процес $\sqrt{n}K_n^{1,(k)*}(t)$ слабо конвергира центрираном Гаусовом процесу $\{\kappa^{1,(k)}(t),\ t \in (0,1)\}$, односно статистика $\sqrt{n}K_n^{1,(k)}$ конвергира у расподели случајној величини $\sup_t |\kappa^{1,(k)}(t)|$,

Следећа лема нам даје функције великих одступања за тест статистике $I_n^{1,(k)}$ и $K_n^{1,(k)}$.



Слика 6.1: График функције $\sigma^2_{K^{1,(2)}}(t)$

Лема 6.0.1 (Милошевић, Обрадовић 2015)**.** *За статистике $I_n^{1,(k)}$ и $K_n^{1,(k)}$ под претпоставком да важи H_0 функција великих одступања из теореме 3.2.2 постоји, аналитичка је за довољно мало $\varepsilon > 0$ и важи*

$$f_{I^{1,(k)}}(\varepsilon) = \frac{1}{2(2k+1)^2\sigma^2_{I^{1,(k)}}}\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0,$$
$$f_{K^{1,(k)}}(\varepsilon) = \frac{1}{8k^2\sigma^2_{K^{1,(k)}}}\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0.$$

Доказ следи из теорема 3.2.6 и 3.2.9.

Лема 6.0.2 (Милошевић, Обрадовић 2015)**.** *За дату алтернативну расподелу $g(x;\theta)$ из $\mathcal{G}$ дефинисане у одељку 3.2.2 граничне вредности у вероватноћи P_θ тест-статистика $I_n^{1,(k)}$ и $K_n^{1,(k)}$ су*

$$b_{I^{1,(k)}}(\theta) = (2k+1)\sup\int_{-\infty}^{\infty}\psi^{1,(k)}(2G(x,0)-1)h(x)dx\cdot\theta + o(\theta),\quad \theta \to 0. \tag{6.6}$$

$$b_{K^{1,(k)}}(\theta) = 2k\sup_{t\in R}|\int_{-\infty}^{\infty}\xi^{1,(k)}(2G(x,0)-1;t)h(x)dx|\cdot\theta + o(\theta),\quad \theta \to 0. \tag{6.7}$$

Доказ. Први део теореме који се односи на интегралну статистику последица је тврђења 3.2.10. Други део се доказује коришћењем закона великих бројева за U-статистике.

$$\begin{aligned} K_n^{1,(k)*}(t) &\xrightarrow{P_\theta} a_{1,(k)}(\theta) = P\{|X_{(k),X_1,\dots X_{2k}}| < t\} - P\{|X_{(k+1),X_1,\dots,X_{2k}}| < t\} \\ &= \binom{2k}{k}G(x,\theta)^k(1-G(x,\theta))^k - \binom{2k}{k}G(-t,\theta)^k(1-G(-t,\theta))^k. \end{aligned}$$

Даље је

$$
\begin{aligned}
a'_{1,(k)}(0) &= -\binom{2k}{k} kG(t,0)^{k-1}(1-G(t,0))^{k-1}(2G(t,0)-1)(H(t)+H(-t)) \\
&= -\binom{2k}{k} kG(t,0)^{k-1}(1-G(t,0))^{k-1}(2G(t,0)-1) \\
&\cdot \left(\int_{-\infty}^{\infty} h(s)\mathrm{I}\{s<t\}ds + \int_{-\infty}^{\infty} h(s)\mathrm{I}\{s<-t\}ds \right) \\
&= 2k\binom{2k-1}{k-1} G(t,0)^{k-1}(1-G(t,0))^{k-1}(2G(t,0)-1) \\
&\cdot \left(\int_{-\infty}^{\infty} h(s)\mathrm{I}\{s>t\}ds - \int_{-\infty}^{\infty} h(s)\mathrm{I}\{s<-t\}ds \right) \\
&= 2k \int_{-\infty}^{\infty} \xi^{1,(k)}(2G(s,0)-1;2G(t,0)-1)h(s)ds.
\end{aligned}
$$

Развијајући функцију $a_{1,(k)}(\theta)$ у Маклоренов ред добијамо тврђење леме. □

У даљем тексту одредићемо локалне Бахадурове ефикасности предложених тестова против алтернатива транслације чија је функција расподеле $G(x,\theta) = G(x-\theta,0)$ и асиметричних алтернатива Азалинијевог типа са густином $g(x,\theta) = 2g(x,0)G(\theta x,0)$ (в. [8]), и то кад је $G(x,0)$ нормална, логистичка и Кошијева расподела.

У табелама су алтернативе са параметром транслације означене са индексом L а алтернативе Азалинијевог типа са индексом A. Резултати који се односе на тест-статистике $I_n^{1,(k)}$ и $K_n^{1,(k)}$ се налазе у табели 6.1. Генерално говорећи, тестови су прилично ефикасни, осим у случају Кошијеве расподеле. Међутим и у том случају, за алтернативе транслације, се постижу ефикасности веће него за друге тестове проучаване у литератури (в. [81]). За алтернативе Азалинијевог типа ефикасности нисмо нашли у литератури. Интресантно је још приметити да су предложени тестови истих ефикасности за алтернативу транслације за нормалну расподелу и Азалинијевог типа.

Следећа теорема нам даје неке класе локално оптималних алтернатива за тест-статистике $I_n^{1,(k)}$ и $K_n^{1,(k)}$.

Теорема 6.0.2 (Милошевић, Обрадовић 2015). *Нека је $g(x;\theta)$ функција*

Табела 6.1: Локална Бахадурова ефикасност тестова $I_n^{1,(k)}$ и $K_n^{1,(k)}$

Алт.	$I_n^{1,(1)}$	$I_n^{1,(2)}$	$I_n^{1,(3)}$	$I_n^{1,(4)}$	$K_n^{1,(1)}$	$K_n^{1,(2)}$	$K_n^{1,(3)}$	$K_n^{1,(4)}$
$\mathcal{N}_L$	0.977	0.957	0.932	0.911	0.764	0.810	0.804	0.794
$\mathcal{N}_A$	0.977	0.957	0.932	0.911	0.764	0.810	0.804	0.794
$\mathcal{L}_L$	0.937	0.981	0.989	0.986	0.750	0.865	0.885	0.889
$\mathcal{L}_A$	0.962	0.920	0.885	0.858	0.747	0.767	0.754	0.741
$\mathcal{C}_L$	0.358	0.497	0.587	0.649	0.376	0.570	0.662	0.716

густине из $\mathcal{G}$ за коју важи

$$\int_{-\infty}^{\infty} \frac{h^2(x)}{g(x,0)}dx < \infty.$$

Тада су, за мало $\theta > 0$, густине алтернативе

$$g(x;\theta) = g(x,0) + C\psi^{1,(k)}(x), C \in \mathbb{R},\ g(x,0) \in \mathcal{H} \tag{6.8}$$

$$g(x;\theta) = g(x,0) + C\xi^{1,(k)}(x,t_0), C \in \mathbb{R},\ g(x,0) \in \mathcal{H}, \tag{6.9}$$

где је t_0 тачка у којој дисперзије $\sigma^2_{K^{1,(k)}}(t)$ достиже максимум, локално оптималне за тест-статистике $I_n^{1,(k)}$ и $K_n^{1,(k)}$.

Доказ. Нека је $h_0(x) = \frac{h(x)-h(-x)}{2}$ помоћна функција и нека је $g(x,0) \in \mathcal{H}$. Тада је

$$\frac{1}{4}\int_{-\infty}^{\infty} \frac{(h(x)-h(-x))^2}{g(x,0)}dx = \int_{-\infty}^{\infty} \frac{h_0^2}{g(x,0)}dx.$$

Користећи непарност функције језгра $\phi^{1,(k)}$ добијамо да важи следећи низ једнакости

$$\begin{aligned}
\int_{-\infty}^{\infty} \phi^{1,(k)}(x)h_0(x)dx &= \frac{1}{2}\int_{-\infty}^{\infty} \phi^{1,(k)}(x)h(x)dx - \frac{1}{2}\int_{-\infty}^{\infty} \phi^{1,(k)}(x)h(-x)dx \\
&= \frac{1}{2}\int_{-\infty}^{\infty} \phi^{1,(k)}(x)h(x)dx + \frac{1}{2}\int_{-\infty}^{\infty} \phi^{1,(k)}(x)h(x)dx \qquad (6.10) \\
&= \int_{-\infty}^{\infty} \phi^{1,(k)}(x)h(x)dx.
\end{aligned}$$

На исти начин се показује да је

$$\int\limits_{-\infty}^{\infty} \xi^{1,(k)}(x,t)h_0(x)dx = \int\limits_{-\infty}^{\infty} \xi^{1,(k)}(x,t)h(x)dx. \tag{6.11}$$

За локалну Бахадурову ефикасност теста $I_n^{1,(k)}$ важи

$$e^B_{I^{1,(k)}} = \lim_{\theta\to 0} \frac{c_I(\theta)}{\frac{1}{4}I_1(g)\cdot\theta^2} = \frac{\left(\int\limits_{-\infty}^{\infty} \phi^{1,(k)}(x)h(x)dx\right)^2}{\frac{1}{4}\int\limits_{-\infty}^{\infty} \frac{(h(x)-h(-x))^2}{g(x,0)}dx \int\limits_{-\infty}^{\infty} (\phi^{1,(k)}(x))^2 g(x,0)dx}$$

$$= \frac{\left(\int\limits_{-\infty}^{\infty} \phi^{1,(k)}(x)h_0(x)dx\right)^2}{\int\limits_{-\infty}^{\infty} \frac{h_0^2(x)}{g(x,0)}dx \int\limits_{-\infty}^{\infty} (\phi^{1,(k)}(x))^2 g(x,0)dx} \leq 1.$$

Услов једнакости у примењеној Коши-Шварц неједнакости се постиже уколико је за неко $C \in R$

$$\frac{h_0(x)}{\sqrt{g(x,0)}} = C\phi^{1,(k)}(x)\sqrt{g(x,0)},$$

односно уколико је

$$h_0(x) = C\phi^{1,(k)}(x)g(x,0). \tag{6.12}$$

Како за алтернативе (6.8) важи (6.12) тврђење првог дела теореме је доказано.

$$e^B_{K^{1,(k)}} = \lim_{\theta\to 0} \frac{c_K(\theta)}{\frac{1}{4}I_1(g)\cdot\theta^2}$$

$$= \frac{\sup\limits_{t\in R}\left(\int\limits_{-\infty}^{\infty} \xi^{1,(k)}(x,t)h(x)dx\right)^2}{\int\limits_{-\infty}^{\infty} \frac{h_0^2(x)}{g(x,0)}dx \sup\limits_{t\in R}\int\limits_{0}^{\infty}(\xi^{1,(k)}(x,t))^2 g(x,\theta)dx} \leq 1$$

Из услова једнакости у примењеној Коши-Шварц неједнакости следи тврђење теореме. $\square$

Следеће две тест-статистике које предлажемо су засноване на карактеризацији из теореме 4.1.7.

Нека су

$$I_n^{2,(k)} = \int_{-\infty}^{\infty} (H_n^{2,(k)}(t) - G_n^{2,(k)}(t)) dF_n(t), \tag{6.13}$$

$$K_n^{2,(k)} = \sup_{t \in R} \Big| H_n^{2,(k)}(t) - G_n^{2,(k)}(t) \Big|. \tag{6.14}$$

где су

$$H_n^{2,(k)}(t) = \frac{1}{\binom{n}{2k+1}} \sum_{\mathcal{I}_{2k+1}} \mathrm{I}\{-X_{(k+1),X_{i_1},\dots X_{i_{2k+1}}} < t\},$$
$$G_n^{2,(k)}(t) = \frac{1}{\binom{n}{2k+1}} \sum_{\mathcal{I}_{2k+1}} \mathrm{I}\{X_{(k+1),X_{i_1},\dots X_{i_{2k+1}}} < t\},$$

U-емпиријске функције расподеле.

Након интеграције добијамо да се статистика (6.13) може представити у облику

$$\frac{1}{n\binom{n}{2k+1}} \sum_{\mathcal{I}_{2k+1}} \sum_{i_{2k+2}=1}^{n} \mathrm{I}\{X_{(k+1),X_{i_1},\dots X_{i_{2k+1}}} > -X_{i_{2k+2}}\} - \mathrm{I}\{X_{(k+1),X_{i_1},\dots X_{i_{2k}}} < X_{i_{2k+2}}\}.$$

На исти начин као и за статистику $I_n^{1,(k)}$ се показује да под претпоставком да важи нулта хипотеза симетрије, и тест-статистике $I_n^{2,(k)}$ и $K_n^{2,(k)}$ не зависе од претпостављене расподеле.

Тест статистика $I_n^{2,(k)}$ је асимптотски еквивалентна U-статистици са симетричним језгром

$$\begin{aligned}\Phi^{2,(k)}(X_1, \dots . X_{2k+2}) = \frac{1}{(2k+2)!} \sum_{\pi(2k+2)} & \mathrm{I}\{X_{(k+1),X_{\pi(i_1)},\dots,X_{\pi(i_{2k+1})}} > -X_{\pi(i_{2k+2})}\} \\ & - \mathrm{I}\{X_{(k+1),X_{\pi(i_1)},\dots,X_{\pi(i_{2k+1})}} < X_{\pi(i_{2k+2})}\}\end{aligned}$$

чија је пројекција на X_1, под претпоставком да важи H_0 и $s \in [-1, 1]$

$$\begin{aligned}\phi^{2,(k)}(s) &= E(\Phi^{2,(k)}(X_1, \dots X_{2k+2}) | X_1 = s) \\ &= \frac{1}{2k+2}\Big(P\{X_{(k+1),X_2,\dots,X_{2k+1}} > -s\} - P\{X_{(k+1),X_2,\dots X_{2k+1}} < s\}\Big) \\ &+ \frac{2k+1}{2k+2}\Big(P\{X_{(k+1),s,\dots,X_{2k+1}} > -X_{2k+2}\} - P\{X_{(k+1),s,\dots X_{2k+1}} < X_{2k+2}\}\Big).\end{aligned} \tag{6.15}$$

Означимо са Y_j j-ту статистику поретка у узорку $X_2, ..., X_{2k+1}$. Тада је

$$\begin{aligned}\phi^{2,(k)}(s) = \frac{2k+1}{2k+2}\Big(&P\{s < Y_k, Y_k > -X_{2k+2}\} + P\{Y_k < s < Y_{k+1}, s > -X_{2k+2}\}\\ &+ P\{Y_{k+1} < s, Y_{k+1} > -X_{2k+2}\} - P\{s < Y_k, Y_k < X_{2k+2}\}\\ &- P\{Y_k < s < Y_{k+1}, s < X_{2k+2}\} - P\{Y_{k+1} < s, Y_{k+1} < X_{2k+2}\}\Big).\end{aligned} \tag{6.16}$$

Након рачунања добија се да је пројекција

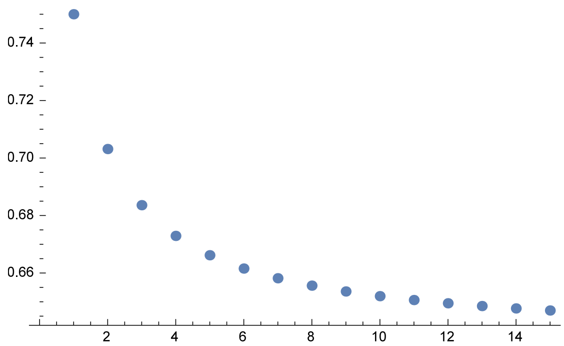
$$\phi^{2,(k)}(s) = \binom{2k+1}{k}\frac{1}{2^{2k+1}}\sum_{j=0}^{k}\frac{(-1)^j}{2j+1}\binom{k}{j}s^{2j+1}.$$

За $k = 1$ и $k = 2$ се добија

$$\begin{aligned}\phi^{2,(1)} &= \frac{s}{8}(3 - s^2)\\ \phi^{2,(2)} &= \frac{s}{48}(15 - 10s^2 + 3s^4).\end{aligned}$$

Показује да је $E\phi^{2,(k)}(X_1) = 0$ за свако $k \geq 1$, као и да је дисперзија пројекције позитивна.

Закључујемо да је језгро $\Phi^{2,(k)}$ недегенерисано па се применом Хефдингове теореме добија да је за свако k гранична расподела тест-статистике нормална са очекивањем нула и дисперзијом која је приказана на графику 6.2.



Слика 6.2: Асимптотска дисперзија статистике $I_n^{2,(k)}$

Нека је

$$K_n^{2,(k)} = \sup_{t\in R}\Big|H_n^{2,(k)}(t) - G_n^{2,(k)}(t)\Big|.$$

Тада је за свако фиксирано $t \in [-1,1]$ $K_n^{2,(k)*}(t) = H_n^{2,(k)}(t) - G_n^{2,(k)}(t)$ U-статистика са језгром

$$\Xi^{2,(k)}(X_1,...,X_{2k+1};t) = \mathrm{I}\{X_{(k+1),X_1,...,X_{2k+1}} > -t\} - \mathrm{I}\{X_{(k+1),X_1,...X_{2k+1}} < t\}.$$

Пројекција језгра на X_1 под претпоставком да важи H_0 и $s \in [-1,1]$ је

$$\xi^{2,(k)}(s,t) = P\{X_{(k+1),s,X_2...X_{2k+1}} > -t\} - P\{X_{(k+1),s,X_2...X_{2k+1}} < t\}.$$

Рачунањем одговарајућих вероватноћа добија се

$$\xi^{2,(k)}(s,t) = \binom{2k}{k}(1-t^2)^k 2^{-2k}\Big(\mathrm{I}\{s > t\} - \mathrm{I}\{s < -t\}\Big).$$

Дисперзија пројекције у опшетм случају је

$$\sigma^2_{K^{2,(k)}}(t) = \binom{2k}{k}^2 2^{-4k}(1-t^2)^{2k}((1-t)\mathrm{I}\{t \geq 0\} + (1+t)\mathrm{I}\{t < 0\}).$$

Максимум дисперзије се достиже за $t = 0$ и једнак је

$$\sigma^2_{K^{2,(k)}} = \binom{2k}{k}^2 2^{-4k}.$$

Користећи исте аргументе као и за тест-статистику $K_n^{1,(k)}$ закључујемо да $\sqrt{n}K_n^{2,(k)*}(t)$ конвергира центрираном Гаусовом процесу а да статистика $\sqrt{n}K_n^{2,(k)}$ корвергира случајној величини чију је расподелу тешко одредити, па таблице критичних вредности се могу добити Монте Карло методом.

За одређивање локалне Бахадурове ефикасности користимо следеће леме.

Лема 6.0.3. *За статистике $I_n^{2,(k)}$ и $K_n^{2,(k)}$ под претпоставком да важи H_0 функција великих одступања из теореме 3.2.2 постоји, аналитичка је за довољно мало $\varepsilon > 0$ и важи*

$$f_{I^{2,(k)}}(\varepsilon) = \frac{1}{2(2k+2)^2\sigma^2_{I^{2,(k)}}}\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0,$$
$$f_{K^{2,(k)}}(\varepsilon) = \frac{1}{2(2k+1)^2\sigma^2_{K^{2,(k)}}}\varepsilon^2 + o(\varepsilon^2),\ \varepsilon \to 0.$$

Доказ следи из теорема 3.2.6 и 3.2.9.

Лема 6.0.4. *За дату алтернативну расподелу $g(x;\theta)$ из $\mathcal{G}$ граничне вредности у вероватноћи P_θ тест-статистика $I^{2,(k)}$ и $K^{2,(k)}$ су*

$$b_{I^{2,(k)}}(\theta) = (2k+2)\int\limits_{-\infty}^{\infty} \psi^{2,(k)}(2G(x,0)-1)h(x)dx\cdot\theta + O(\theta^2), \quad \theta\to 0. \tag{6.17}$$

$$b_{K^{2,(k)}}(\theta) = (2k+1)\sup_t |\int\limits_{-\infty}^{\infty} \xi^{2,(k)}(2G(x,0)-1;t)h(x)dx|\cdot\theta + O(\theta^2), \quad \theta\to 0. \tag{6.18}$$

Доказ је аналоган доказу леме 6.0.2.

У табели 6.2 се налазе локалне Бахадурове ефикасности тестова $I_n^{2,(k)}$ и $K_n^{2,(k)}$. Примећујемо да су тестови упоредиви са претходним, као и да су ефикаснији у случају Кошијеве расподеле.

На исти начин као у случају теореме 6.0.2 може се доказати да су неке класе најповољнијих алтернатива за тест-статистике $I_n^{2,(k)}$ и $K_n^{2,(k)}$ дате следећом теоремом.

Теорема 6.0.3. *Нека је $g(x;\theta)$ функција густине из $\mathcal{G}_l$ за коју важи*

$$\int\limits_{-\infty}^{\infty} \frac{h^2(x)}{g(x,0)}dx < \infty.$$

Тада су, за мало $\theta>0$, густине алтернативе

$$g(x;\theta) = g(x,0) + C\psi^{2,(k)}(x), C\in\mathbb{R},\ g(x,0)\in\mathcal{H} \tag{6.19}$$

$$g(x;\theta) = g(x,0) + C\xi^{2,(k)}(x,0), C\in\mathbb{R},\ g(x,0)\in\mathcal{H} \tag{6.20}$$

локално оптималне за тест-статистике $I_n^{2,(k)}$ и $K_n^{2,(k)}$, редом.

Поглавље завршавамо још једним тестом симетрије који је заснован на карактеризацији из теореме 4.1.8. Предлажемо тест-статистику

$$S_n = \int_{-\infty}^{\infty} (R_n^{(1)}(t) - R_n^{(2)}(t))dF_n^*(t),$$

где су

$$R_n^{(1)}(t) = \frac{1}{n^2}\sum_{i,j} \mathrm{I}\{|X_i - X_j| < t\},$$

$$R_n^{(2)}(t) = \frac{1}{n^2}\sum_{i,j} \mathrm{I}\{|X_i + X_j| < t\},$$

Табела 6.2: Локална Бахадурова ефикасност тестова $I_n^{2,(k)}$ и $K_n^{2,(k)}$

Алтернатива	$I_n^{2,(1)}$	$I_n^{2,(2)}$	$I_n^{2,(3)}$	$I_n^{2,(4)}$	$K_n^{2,(k)}$
$\mathcal{N}_L$	0.906	0.873	0.849	0.838	0.637
$\mathcal{N}_A$	0.906	0.873	0.949	0.831	0.405
$\mathcal{L}_L$	0.988	0.972	0.957	0.945	0.75
$\mathcal{L}_A$	0.853	0.815	0.790	0.772	0.584
$\mathcal{C}_L$	0.709	0.764	0.800	0.818	0.811

V-емпиријске функције расподеле, а $F^*(t)$ функција расподеле збира две једнако расподељене случајне величине са функцијом расподеле F и $F_n^*(t)$ одговарајућа емпиријска функција. Након интеграције добијамо

$$S_n = \frac{1}{n^4} \sum_{i,j,a,b} \mathrm{I}\{|X_i - X_j| < X_a + X_b\} - \mathrm{I}\{|X_i + X_j| < X_a + X_b\}.$$

Може се лако показати да уколико важи нулта хипотеза симетрије око нуле расподела статистике S_n не зависи од расподеле. Зато можемо претпоставити да уколико важи H_0 узорак је из униформне $\mathcal{U}[-1,1]$ расподеле. Приметимо да је S_n V_n-статистика са симетричним језгром

$$\begin{aligned}\Phi(X_1, X_2, X_3, X_4) = \frac{1}{4!} \sum_{\pi(4)} & \mathrm{I}\{|X_{\pi_1} - X_{\pi_2}| < X_{\pi_3} + X_{\pi_4}\} \\ & - \mathrm{I}\{|X_{\pi_1} + X_{\pi_2}| < X_{\pi_3} + X_{\pi_4}\}.\end{aligned}$$

Прва пројекција језгра на X_1 ако важи H_0 је идентички једнака нула, док је друга на (X_1, X_2) једнака

$$\phi(s,t) = \begin{cases} -\frac{1}{12}(s+2)t, & -1 < s < t < 1, s+t < 0; \\ \frac{1}{12}s(2-t), & -1 < s < t < 1, s+t > 0; \\ -\frac{1}{12}(t+2)s, & 1 > s > t > -1, s+t < 0; \\ \frac{1}{12}(2-s)t, & 1 > s > t > -1, s+t > 0; \end{cases}$$

Јасно је да друга пројекција различита од нуле па је језгро слабо дегенерисано.

Да бисмо нашли граничну расподелу искористићемо репрезентацију V-статистика са симетричним језгром преко U-статистика (в. [53]). Наиме, важи

$$V = \sum_{k=1}^{m} n^{-m} \binom{n}{k} U_{nk},$$

где су

$$U_{nk} = \frac{1}{\binom{n}{k}} \sum_{1 \le i_1 < \cdots < i_k \le n} \Phi_{m,k}(X_{i_1}, ..., X_{i_k}),$$

$$\Phi_{m,k}(x_1, ..., x_k) = \sum_{\substack{\nu_1 + \cdots \nu_k = m \\ \nu_j \ge 1}} \frac{m!}{\nu_1! \cdots \nu_k!} \Phi(x_1^{\nu_1}, ..., x_k^{\nu_k}).$$

У нашем случају је

$$\begin{aligned} S_n = & \frac{1}{n^4} \sum_{1 \le i_1 \le n} \Phi(X_{i_1}, X_{i_1}, X_{i_1}, X_{i_1}) \\ & + \frac{1}{n^4} \sum_{1 \le i_1 < i_2 \le n} (4\Phi(X_{i_1}, X_{i_1}, X_{i_1}, X_{i_2}) + 4\Phi(X_{i_2}, X_{i_2}, X_{i_2}, X_{i_1}) \\ & + 6\Phi(X_{i_1}, X_{i_1}, X_{i_2}, X_{i_2})) \\ & + \frac{1}{n^4} \sum_{1 \le i_1 < i_2 < i_3 \le n} 12(\Phi(X_{i_1}, X_{i_1}, X_{i_2}, X_{i_3}) \\ & + \Phi(X_{i_2}, X_{i_2}, X_{i_1}, X_{i_3}) + \Phi(X_{i_3}, X_{i_3}, X_{i_1}, X_{i_2})) \\ & + \frac{1}{n^4} \sum_{1 \le i_1 < i_2 < i_3 < i_4 \le n} 24\Phi(X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}). \end{aligned}$$

Ако помножимо статистику са n и посматрамо граничну вредност кад n тежи бесконачности добијамо да прва два сабирка теже нули. Зато је гранична расподела статистике nS_n

$$\binom{4}{2} \sum_{i=1}^{\infty} \nu_i(\tau_i^2 - 1) + \frac{1}{3!} E(\Phi_{4,3}(X_1, X_2, X_3)) \tag{6.21}$$

где су $\{\nu_i\}$ сопствене вредности интегралног оператора $\hat{J} : L^2[-1,1] \mapsto L^2[-1,1]$ дефинисаног са

$$\hat{J}(x(t)) = \int_{-1}^{1} \frac{1}{2} x(t) \phi(s,t) dt.$$

Уколико важи H_0

$$\begin{aligned} E(\Phi_{4,3}(X_1, X_2, X_3)) &= 12(E(\Phi(X_1, X_1, X_2, X_3) + E(\Phi(X_1, X_2, X_2, X_3) \\ &+ E(\Phi(X_1, X_2, X_3, X_3)) \\ &= 12 \cdot 3 \cdot \frac{1}{6}\Big(P\{X_2 + X_3 > 0\} - P\{|2X_1| < X_2 + X_3\}\Big) \\ &= 2. \end{aligned}$$

Сада ћемо наћи сопствене вредности оператора, тј. вредности ν за које важи

$$\nu x(s) = \hat{J}(x(t)) = \int_{-1}^{1} \frac{1}{2} x(t)\phi(s,t)dt. \tag{6.22}$$

Користећи особине језгра може се показати да су све сопствене функције непарне. Развој сопствене функције у Фуријеов ред је

$$x(t) = \sum_{k=1}^{\infty} e_k(t) a_k, \tag{6.23}$$

где су $e_k(s) = \sqrt{2}\sin k\pi s$ ортонормиране функције.

Када оператор $\hat{J}$ дејствује на функцију x дату изразом (6.23) имамо да је једначина (6.22) еквивалентна са

$$\hat{J}(\sum_{k=1}^{\infty} a_k e_k(t)) = \mu \sum_{l=1}^{\infty} e_l(s) a_l. \tag{6.24}$$

Лева страна је непарна функција из $L^2[-1,1]$ па је можемо развити у Фуријеов ред, односно

$$\hat{J}(\sum_{k=1}^{\infty} a_k e_k(t)) = \sum_{l=1}^{\infty} b_l e_l(s).$$

Скаларно ћемо помножити претходну једнакост са e_l и добијамо

$$\begin{aligned} \int_{-1}^{1} \hat{J}(\sum_{k=1}^{\infty} a_k e_k(t)) e_l(s) \frac{1}{2} ds &= \int_{-1}^{1} \big(\sum_{k=1}^{\infty} a_k \hat{J}(e_k(t))\big) e_l(s) \frac{1}{2} ds \\ = \sum_{k=1}^{\infty} a_k \int_{-1}^{1}\int_{-1}^{1} \phi(t,s) e_k(t) e_l(s) \frac{1}{4} dt ds &= \sum_{k=1}^{\infty} J_{l,k} a_k = b_l. \end{aligned}$$

где је коефицијент

$$J_{l,k} = \int_{-1}^{1}\int_{-1}^{1} \phi(t,s) e_k(t) e_l(s) \frac{1}{4} dt ds.$$

Након израчунавања добијамо

$$J_{l,k} = \begin{cases} \frac{(-1)^{k+l}}{6kl\pi^2}, & k \neq l; \\ \frac{1}{3k^2\pi^2}, & k = l. \end{cases}$$

Приметимо да уколико заменимо $k = l$ у израз за $J_{l,k}$ за $k \neq l$ добијамо дупло мању вредност него што је $J_{l,l}$. Из израза (6.24) произилази

$$\sum_{l=1}^{\infty}\sum_{k=1}^{\infty} a_k J_{l,k} e_l = \nu \sum_{l=1}^{\infty} a_l e_l(s).$$

Изједначавајући коефицијенте уз функције e_l добијамо систем

$$\sum_{k=1}^{\infty} a_k J_{l,k} = \nu a_l.$$

Даље је

$$\nu a_l = \sum_{k=1}^{\infty} a_k \frac{(-1)^{k+l}}{6\pi^2 kl} + \frac{1}{6\pi^2 l^2} a_l = \frac{(-1)^l}{6\pi^2 l} C + \frac{1}{6\pi^2 l^2} a_l,$$

где је

$$C = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} a_k.$$

Из претходног добијамо

$$(\nu - \frac{1}{6\pi^2 l^2}) a_l = \frac{(-1)^l}{6\pi^2 l} C,$$

или еквивалентно

$$\frac{(-1)^l}{l} a_l = \frac{1}{6\pi^2 l^2} C \big(\nu - \frac{1}{6\pi^2 l^2}\big)^{-1} = \frac{1}{6\pi^2 l^2 \nu - 1} C.$$

Сумирајући леву и десну страну претходне једнакости по l добијамо

$$C = C \sum_{l=1}^{\infty} \frac{1}{6\pi^2 l^2 \nu - 1} = C\Big(\frac{1}{2} - \frac{\cot \frac{1}{\sqrt{6\nu}}}{2\sqrt{6\nu}}\Big).$$

Ова једначина се може решити нумерички. Највеће решење једначине је $\nu_1 = 0.0404938$. Зато је главна карактеристична вредност неопходна за рачунање функције великих одступања $\lambda_1 = 24.6952$. Приметимо још да је траг оператора $\hat{J}$

$$tr\hat{J} = \sum_{k=1}^{\infty} J_{k,k} = \sum_{k=1}^{\infty} \frac{1}{3k^2\pi^2} = \frac{1}{3\pi^2} \cdot \frac{\pi^2}{6} = \frac{1}{18}.$$

Сада можемо граничну расподелу статистике nS_n приказати у другачијем облику, односно важи

$$nS_n \xrightarrow{d} \binom{4}{2}\Big(\sum_{i=1}^{\infty}\nu_i(\tau_i^2-1)+tr\hat{J}\Big)=\binom{4}{2}\sum_{i=1}^{\infty}\nu_i\tau_i^2.$$

Користећи теорему 3.2.7 добијамо да важи следећа лема.

Лема 6.0.5. *За статистику S_n под претпоставком да важи H_0 функција великих одступања из теореме 3.2.2 постоји, аналитичка је за довољно мало $\varepsilon > 0$ и важи*

$$f_S(\varepsilon)=2.058\varepsilon+o(\varepsilon^{\frac{3}{2}}),\ \varepsilon\to 0.$$

Користећи теорему 3.2.11 добијамо да је гранична вредност у вероватноћи P_θ статистике S_n.

Лема 6.0.6. *За дату алтернативну расподелу $g(x;\theta)$ из фамилије $\mathcal{G}$ дефинисане у одељку 3.2.2 гранична вредност у вероватноћи P_θ тест статистике S_n је*

$$b_S(\theta)=6\int\limits_{R^2}\phi(x_1,x_2)g_\theta(x_1,0)g_\theta(x_1,0)dx_1dx_2\cdot\theta^2+o(\theta^2),\ \ \theta\to 0. \qquad (6.25)$$

У табели 6.3 су приказане локалне Бахадурове ефикасности статистике S_n уколико је алтернативна расподела нормална са параметром локације, нормална Азалинијевог типа и Кошијева расподела са параметром транслације.

Табела 6.3: Локална Бахадурова ефикасност тест-статистике S_n

Алт.	$\mathcal{N}_L$	$\mathcal{N}_A$	$\mathcal{C}_l$
	0.858	0.858	0.834

Поглавље 7

Тестови са оцењеним параметрима

За тестирање сагласности са фамилијом расподела која зависи од непознатих параметара, јако је важно предложити тест-статистику која уколико важи нулта хипотеза, не зависи од параметара расподеле. Зато у тест-статистикама веома често фигуришу оцене непознатих параметара. Неки од примера таквих тестова могу се видети између осталог у радовима [92], [64], [65], [52]. Ми ћемо предложити неколико нових тестова сагласности са експонециајлном расподелом овог типа. Тест-статистике ће бити конструисане коришћењем U-емпиријских Лапласових трансфорамција као и U-емпиријских момената расподеле.

7.1 Тестови засновани на Лапласовим трансформацијама расподела

До сад смо већ видели неколико приступа за конструкцију тест-статистика. Уз наведене постоји још много приступа међу којима је и онај који користи Лапласове трансформације. Могу се наћи многи радови на ову тему. Навешћемо неке од њих.

Барингхаус и Хензе [10] су констрисали тест-статистике на основу диференцијалне једначине коју Лапласова трансформација експоненцијалне расподеле задовољава. Сличан тест за Рејлијеву расподелу предожен је у раду Меинтаниса и Илиопулуса [64], а за гама расподелу у раду Хензеа, Меинтаниса и Ебнера [40]

Тест заснован на поређењу емпиријске и теоријске Лапласове трансформације конструисан је за експоненцијалну и инверзну Гаусову расподелу у радовима [38] и [39], редом.

За тестирање сагласности са расподелама чији носач је читав скуп реалних бројева, уместо емпиријске Лапласове трансформације, за конструкцију тестова коришћене су емпиријске карактеристичне функције (в. нпр. [41], [35]).

У овом одељку за тестирање сложене нулте хипотезе да је узорак из неке фамилије расподела приказаћемо тестове засноване на карактеризацијама једнако расподељених статистика облика

$$\omega_1(X_1, ..., X_m) \stackrel{d}{=} \omega_2(X_1, ..., X_m),$$

где су ω_1 и ω_2 линеарне комбинације елемената узорка и њихових статистика поретка. Резултати су део коауторског рада са М. Обрадовићем и приказани су на конференцији $PROBASTAT$ 2015 у Смоленицама. За тестирање сложене нулте хипотезе H_0 да је узорак из експоненцијалне расподеле предлажемо следећу фамилију тест-статистика:

$$J_{n,a} = \int_0^\infty (L_n^{(1)}(t) - L_n^{(2)}(t))\bar{X}e^{-a\bar{X}t}dt, \tag{7.1}$$

где је $\bar{X}$ узорачка средина, $a > 0$ позитивна константа а

$$L_n^{(k)}(t) = \frac{1}{n^{[m]}} \sum_{i_1<...<i_m} \sum_{\pi(m)} e^{-t\omega_k(X_{\pi(i_1)},...,X_{\pi(i_m)})}, \quad k = 1, 2.$$

U-емпиријске Лапласове трансформације статистика ω_1 и ω_2.

Напомињемо да се до сада у литератури не могу наћи примери оваквих тест-статистика.

Како бисмо одредили граничну расподелу тест-статистике $J_{n,a}$, под претпоставком да важи нулта хипотеза посматраћемо помоћну функцију

$$J^*_{n,a}(\mu) = \frac{\mu}{n^{[m]}} \sum_{i_1<...<i_m} \sum_{\pi(m)} \Big(\frac{1}{a\mu + \omega_1(X_{i_{\pi_1}}, .., X_{i_{\pi_m}})} - \frac{1}{a\mu + \omega_2(X_{i_{\pi_1}}, .., X_{i_{\pi_m}})}\Big).$$

где је $\mu = \lambda^{-1}$ параметар средње вредности. За свако фиксирано $\mu > 0$ $J^*_{n,a}(\mu)$ је U-статистика чија расподела не зависи од μ. Зато можемо претпоставити да је $\mu = 1$.. Симетрично језгро U-статистике $J^*_{n,a}(1)$ је

$$\Phi(X_1, ..., X_m; a) = \frac{1}{m!} \sum_{\pi(m)} \Big(\frac{1}{a\bar{X} + \omega_1(X_{\pi_1}, .., X_{\pi_m})} - \frac{1}{a\bar{X} + \omega_2(X_{\pi_1}, .., X_{\pi_m})}\Big).$$

У случају да се ради о недегенерисаном језгру можемо применити теорему 1.1.2 за U-статистике и добити да је гранична расподелу статистике

$\sqrt{n}J^*_{n,a}(1)$ нормална $\mathcal{N}(0, m^2\sigma^2_\Phi)$, где је σ^2_Φ дисперзија пројекције језгра на X_1 под претпоставком да важи нулта хипотеза.

Познато је да за граничну расподелу узорачке средине важи

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \mu^2).$$

Одавде закључујемо да су услови (2.4) из теореме 2.0.6 испуњени и закључујемо да се граничне расподеле статистика $\sqrt{n}J^*_{n,a}(\mu)$ и $\sqrt{n}J_{n,a}$ поклапају. Како расподела $\sqrt{n}J_{n,a}$ не зависи од параметра $\lambda = \mu^{-1}$ добијамо да је гранична расподела статистике $\sqrt{n}J_{n,a}$ нормална $\mathcal{N}(0, m^2\sigma^2_\Phi)$.

7.1.1 Приближна локална Бахадурова и Питманова асимптотска ефикасност тестова

Лема 7.1.1 (Милошевић, Обрадовић 2015)**.** *За дату алтернативну расподелу са функцијом густине $g(x;\theta)$ која припада фамилији $\mathcal{G}$ важи да је гранична вредност у вероватноћи тест-статистике $J_{n,a}$ кад $n \to \infty$ једнака*

$$b(\theta) = m\int_0^\infty \psi(x)h(x)dx \cdot \theta + o(\theta),\ \theta \to 0.$$

Доказ. Искористићемо познати резултат да узорачка средина скоро сигурно конвергира ка њеној очекиваној вредности $\mu(\theta)$ и применити закон великих бројева за U-статистике са оцењеним параметрима (в. теорему 2.0.7).

Закључујемо да се гранична вредност у вероватноћи тест-статистике $J_{n,a}$ кад $n \to \infty$ поклапа са граничном вредности у вероватноћи помоћне функције $J^*_{n,a}(\mu(\theta))$. С обзиром на то да је под претпоставком да важи нулта хипотеза ($\theta = 0$) тест статистика слободна од вредности параметра средње вредности, можемо претпоставити да је $\mu(0) = 1$. Добијамо да важи

$$\begin{aligned} b(\theta) &= E_\theta(\Phi(X_1, ..., X_m)) \\ &= \mu(\theta)\int\limits_{R^m}\Big(\frac{1}{a + \omega_1(x_1, ..., x_m)} - \frac{1}{a + \omega_2(x_1, ..., x_m)}\Big)dG_1(x_1, \theta)...dG_m(x_2, \theta). \end{aligned}$$

Први извод функције $b(\theta)$ по θ у нули је

$$\begin{aligned}
&b'_\theta(0)\\
&=\mu'(0)\int\limits_{R^m}\Big(\frac{1}{a+\omega_1(x_1,...,x_m)}-\frac{1}{a+\omega_2(x_1,...,x_m)}\Big)dG_1(x_1,0)...dG_m(x_m,0)\\
&+\frac{\partial}{\partial\theta}\int\limits_{R^m}\Big(\frac{1}{a+\omega_1(x_1,...,x_m)}-\frac{1}{a+\omega_2(x_1,...,x_m)}\Big)dG_1(x_1,\theta)...dG_m(x_m,\theta)\Big|_{\theta=0}.
\end{aligned}$$

Први сабирак је једнак нули због карактеризације. На други сабирак примењујемо теорему 3.2.10 и добијамо

$$b'_\theta(0)=m\int\limits_0^\infty h(x)\psi(x)dx.$$

Развијањем $b(\theta)$ у Маклоренов ред доказ је завршен. □

Претпоставимо да је алтернативна хипотеза $H_1:\theta_1=\frac{\delta}{\sqrt{n}},\ \ \delta>0$.

Како су функције $\sigma^2_\Psi(\theta)$ и $b(\theta)$ непрекидне по θ и $\sigma^2_\Psi(\theta)>0$ закључујемо да је језгро Ψ недегенерисано и под претпоставком да важи H_1, односно гранична расподеле статистике $J_{n,a}$ је нормална. При томе користимо теорему 2.0.6. Претпоставимо још да је $b(\theta)>0$. Закључујемо да су услови теореме 3.1.1 испуњени. На основу тога је Питманов нагиб

$$c_P=\frac{b'(0)^2}{m^2\sigma^2_\Psi}. \tag{7.2}$$

Слично, на основу граничне расподеле под нултом хипотезом налазимо да је коефицијент a из дефиниције приближног Бахадуровог нагиба 3.34 једнак $(\sigma^2_\Psi)^{-1}$ па је

$$c_B(\theta)=\frac{b'(0)^2\theta^2}{m^2\sigma^2_\Psi}. \tag{7.3}$$

На основу израза за локалне нагибе (7.2) и (7.3) закључујемо да се локалне релативне Питманове и Бахадурове ефикасности поклапају.

У тексту који следи представићемо три фамилије тест-статистика засноване редом на карактеризацијама експоненцијалне расподеле 4.1.2 и 4.1.1.

$$J_n^{\mathcal{D}}=\frac{\bar{X}}{n(n-1)}\sum_{i_1<i_2}\sum_{\pi(2)}\Big(\frac{1}{a\bar{X}+2\min(X_{i_{\pi_1}},X_{i_{\pi_2}})}-\frac{1}{a\bar{X}+X_{i_{\pi_1}}}\Big),$$

$$J_n^{\mathcal{P}}=\frac{\bar{X}}{n(n-1)}\sum_{i_1<i_2}\sum_{\pi(2)}\Big(\frac{1}{a\bar{X}+|X_{i_{\pi_1}}-X_{i_{\pi_2}}|}-\frac{1}{a\bar{X}+X_{i_{\pi_1}}}\Big)$$

Пројекције језгра U-статистика $J_{n,a}^{\mathcal{D}*}$, и $J_{n,a}^{\mathcal{P}*}$ на X_1 под претпоставком да важи H_0 су

$$\begin{aligned}\psi^{\mathcal{D}}(s) =&E(\Phi^{\mathcal{D}}(X_1,X_2)|X_1=s) = -\frac{1}{2(a+s)}+\frac{1}{2}e^aEi(-a)\\&+\frac{1}{2(a+2s)}e^{-s}\Big(2+ae^{a/2+s}\Gamma(0,\frac{a}{2})-2e^{\frac{a}{2}+s}s\Gamma(0,\frac{a}{2})\\&-ae^{a/2+s}\Gamma(0,\frac{a}{2}+s)+2e^{\frac{a}{2}+s}s\Gamma(0,\frac{a}{2}+s)\Big),\end{aligned}$$

$$\begin{aligned}\psi^{\mathcal{P}}(s) =&E(\Phi^{\mathcal{P}}(X_1,X_2)|X_1=s) = -\frac{1}{2(a+s)}+\frac{1}{2}e^aEi(-a)\\&-e^{-a-s}(e^{2a}Ei(-a)+Ei(a)-Ei(a+s)),\end{aligned}$$

где су $Ei(z)=\int_{-z}^{\infty}u^{-1}e^{-u}du$ Експоненцијални интеграл и $\Gamma(a,z)=\int_z^{\infty}t^{a-1}e^{-t}dt$ некомплетна Гама функција.

Може се показати да су језгра центрирана и недегенерисана за свако $a>0$.

Узимајући ово у обзир и претходна разматрања, можемо наћи граничне расподеле статистика $J_n^{\mathcal{D}}$, и $J_n^{\mathcal{P}}$ за свако $a>0$.

Упоредићемо ове тестове са тестовима интегралног типа заснованим на истим карактеризацијама. Конкурентске тест-статистике уместо разлике U-емпиријских Лапласових трансформација расподеле садрже разлике U-емпиријских функција расподеле,

$$\begin{aligned}I_n^{\mathcal{D}} &= \int_0^{\infty}(G_n(t)-F_n(t))dF_n(t),\\I_n^{\mathcal{P}} &= \int_0^{\infty}(H_n(t)-F_n(t))dF_n(t),\end{aligned}$$

где су

$$F_n(t) = \frac{1}{n}\sum_i \mathrm{I}\{X_i < t\}$$
$$G_n(t) = \frac{1}{\binom{n}{2}}\sum_{i<j} \mathrm{I}\{2\min(X_i, X_j) < t\},$$
$$H_n(t) = \frac{1}{\binom{n}{2}}\sum_{i<j} \mathrm{I}\{|X_i - X_j| < t\},$$

U-емпиријске функције расподела. На исти начин као у претходним поглављима може се показати да су ове тест-статистике еквивалентне U-статистикама са недегенерисаним језгрима, па ћемо тај поступак на овом месту изоставити.

Тестове ћемо упоредити за алтернативе (5.5), (5.6), (5.7) и (5.9).

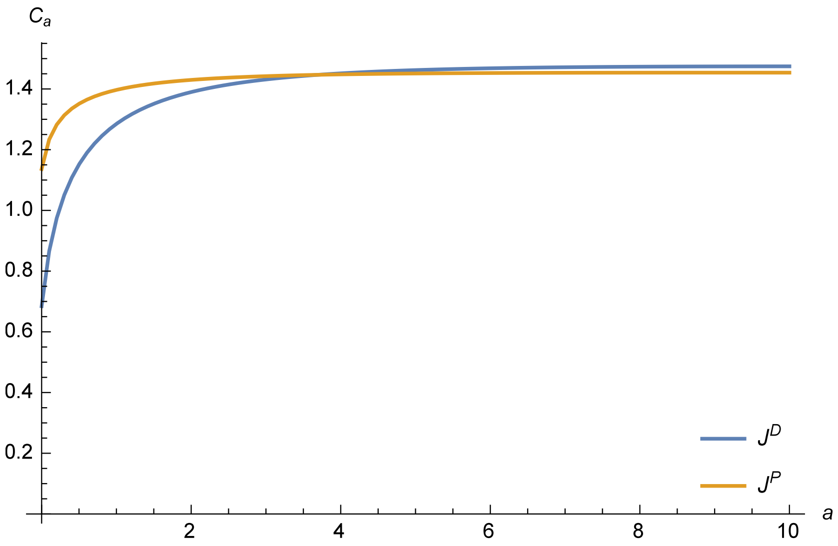
Табела 7.1: Приближна АРЕ $(J_{n,a}^{\mathcal{D}}, I_n^{\mathcal{D}})$

а	1	2	5
Вејбул	1.11	1.21	1.27
гама	1.09	1.08	1.03
Макехам	1.11	1.44	1.76
ЛСО	1.33	2.05	3.12

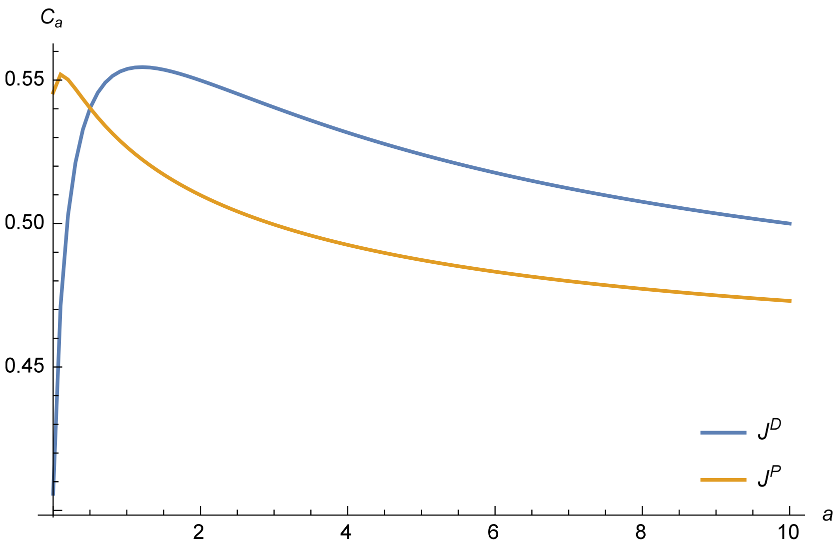
Табела 7.2: Приближна АРЕ $(J_{n,a}^{\mathcal{P}}, I_n^{\mathcal{P}})$

a	1	2	5
Вејбул	1.03	1.06	1.07
гама	1.04	1	0.96
Макехам	1	1.11	1.2
ЛСО	1.32	1.68	2.06

На графицима 7.1-7.4 је приказана зависност нагиба C_a од параметра a. Ради поређења, за сваку алтернативу, приказани су нагиби за обе статистике. Уколико је нагиб статистике већи тада је тај тест ефикаснији. Можемо приметити да у случају Мекехам и ЛСО алтернативе тест

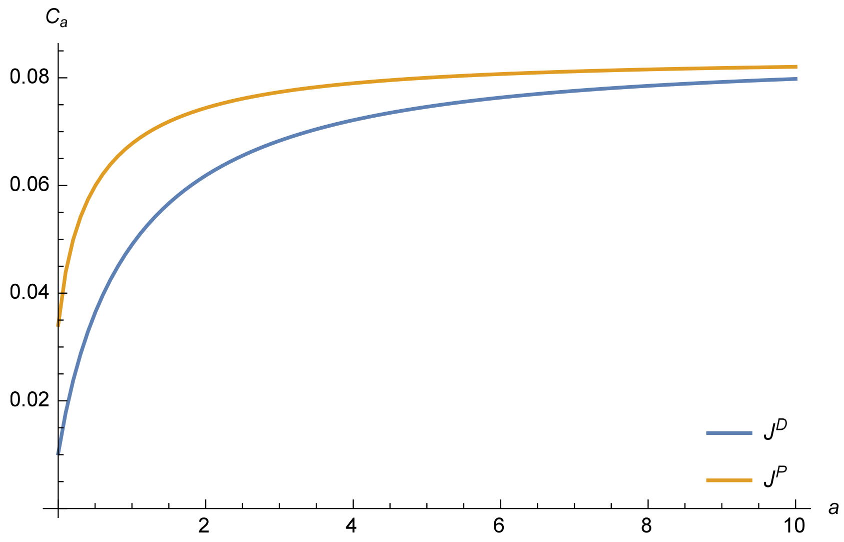


Слика 7.1: Приближан нагиб за Вејбул алтернативе

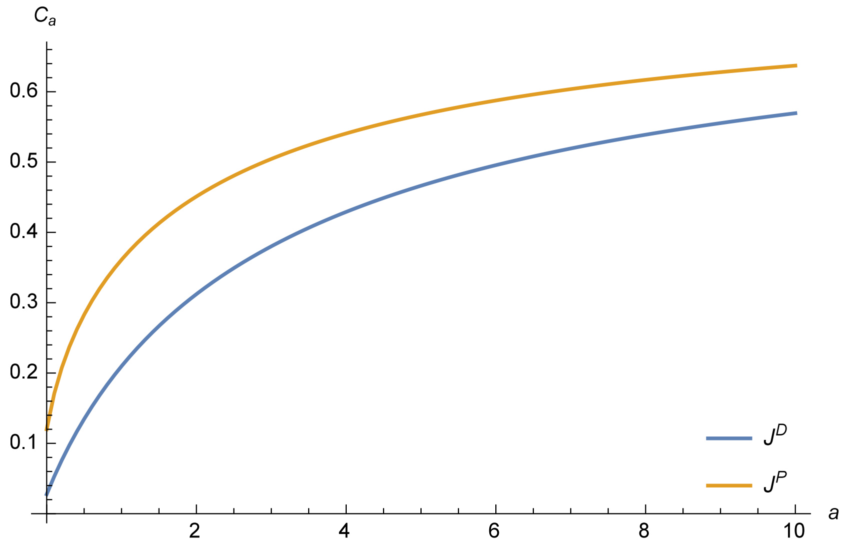


Слика 7.2: Приближан нагиб за гама алтернативе

заснован на Пури-Рубин је увек ефикаснији, док у случају гама алтернативе је тест заснован на Десуовој карактеризацији ефикаснији, осим за мање вредности параметра a. У случају Вејбулове алтернативе тест заснован на Пури-Рубин карактеризацији је ефикаснији за вредности параметра a мање од 3.5.



Слика 7.3: Приближан нагиб за Макехам алтернативе



Слика 7.4: Приближан нагиб за ЛСО алтернативе

7.2 Тестови засновани на U-емпиријским моментима

У овом одељку, као један од оригиналних резултата дисертације, предлажемо два теста експоненцијалности на основу наше карактеризације

4.1.3 за $n = 2, k = 2$ и $n = 3, k = 3$.

$$T_{m,(2)} = \frac{1}{\binom{n}{2}\bar{X}_n} \sum_{i<j} \Big(\min(X_i, X_j) + \frac{1}{2}X_i + \frac{1}{2}X_j - \max(X_i, X_j)\Big)$$

$$T_{m,(3)} = \frac{1}{\binom{n}{2}\bar{X}_n} \sum_{i<j<k} \Big(\mathrm{med}(X_i, X_j, X_k) + \frac{1}{3}X_i + \frac{1}{3}X_j + \frac{1}{3}X_k - \max(X_i, X_j, X_k)\Big).$$

Приметимо прво да уколико важи нулта хипотеза експоненцијалности предложене тест-статистике не зависе од параметра скалирања λ. Зато уколико тражимо граничну расподелу тест-статистика $\sqrt{n}T_{m,(2)}$ и $\sqrt{n}T_{m,(3)}$ можемо без умањења општости претпоставити да је $\mu = \frac{1}{\lambda} = 1$. Даље, приметимо да статистике $T_{m,2}$ и $T_{m,3}$ можемо представити у облику

$$\begin{aligned} T_{m,(2)} &= \frac{1}{\bar{X}_n} \cdot \frac{1}{\binom{n}{2}} \sum_{i<j} \Big(\min(X_i, X_j) + \frac{1}{2}X_i + \frac{1}{2}X_j - \max(X_i, X_j)\Big) \\ &= \frac{1}{\bar{X}_n} T^*_{m,(2)} \end{aligned} \tag{7.4}$$

$$\begin{aligned} T_{m,(3)} &= \frac{1}{\bar{X}_n} \cdot \frac{1}{\binom{n}{2}} \sum_{i<j<k} \Big(\mathrm{med}(X_i, X_j, X_k) + \frac{1}{3}X_i \\ &\quad + \frac{1}{3}X_j + \frac{1}{3}X_k - \max(X_i, X_j, X_k)\Big) = \frac{1}{\bar{X}_n} T^*_{m,(3)}. \end{aligned} \tag{7.5}$$

Статистике $T^*_{m,(2)}$ и $T^*_{m,(3)}$ су U-статистике. Њихова језгра су редом

$$\Upsilon(X_1, X_2) = \min(X_1, X_2) + \frac{1}{2}X_1 + \frac{1}{2}X_2 - \max(X_1, X_2)$$

$$\begin{aligned} \Upsilon(X_1, X_2, X_3) &= \mathrm{med}(X_1, X_2, X_3) + \frac{1}{3}X_1 \\ &\quad + \frac{1}{3}X_2 + \frac{1}{3}X_3 - \max(X_1, X_2, X_3). \end{aligned}$$

Прве пројекције језгра ових статистика, под претпоставком да важи нулта хипотеза су

$$\upsilon_2(s) = E\min(X_2, s) + \frac{s}{2} + \frac{1}{2} - E\max(X_2, s) = \frac{1}{2}(4e^{-s} + s - 3)$$

$$\begin{aligned} \upsilon_3(s) &= E\mathrm{med}(X_2, X_3, s) + \frac{s}{3} + \frac{2}{3} - E\max(X_2, X_3, s) \\ &= \frac{1}{6}(9e^{-2s} - 24e^{-s} - 4s + 13). \end{aligned}$$

Очекивања пројекција су нула а њихове дисперзије су редом

$$\sigma_2^2 = E(\upsilon_2^2(X_1)) = \frac{1}{12}, \tag{7.6}$$

$$\sigma_3^2 = E(\upsilon_2^2(X_1)) = \frac{4}{45}. \tag{7.7}$$

Применом теореме 1.1.2 закључујемо да су граничне расподеле помоћних статистика $\sqrt{n}T^*_{m,(2)}$ и $\sqrt{n}T^*_{m,(3)}$ нормалне $\mathcal{N}(0, \frac{1}{3})$ и $\mathcal{N}(0, \frac{4}{9})$. Како $\bar{X}_n \xrightarrow{p} \mu$, применом теореме Слуцког закључујемо да се граничне расподеле статистика $\sqrt{n}T_{m,(2)}$ и $\sqrt{n}T_{m,(3)}$ поклапају са расподелама одговарајућих помоћних статистика.

Тестове ћемо упоредити коршћењем релативне Питманове ефикасности, односно приближне локалне Бахадурове ефикасности, које су за ове тест-статистике поклапају.

Напомена: *На овом месту користили смо теорему Слуцког која се односи на производе низова случајних величина. Уколико за низове случајних величина* $\{A_n\}$ *и* $\{B_n\}$ *важи*

$$A_n \xrightarrow{d} A, \quad n \to \infty,$$
$$B_n \xrightarrow{p} b, \quad n \to \infty,$$

где је b *константа, тада*

$$A_n B_n \xrightarrow{d} Ab, \quad n \to \infty.$$

Претпоставимо да тестирање експонецијалности приказујемо у облику

$$H_0 : \theta = 0 \quad H_1 : \theta_1 = \frac{\delta}{\sqrt{n}}, \quad \delta > 0.$$

Уведимо следеће ознаке.

$$\begin{aligned}
\upsilon_2(s,\theta) &= E_\theta(\Upsilon_2(X_1, X_2)|X_1 = s),\\
\upsilon_3(s,\theta) &= E_\theta(\Upsilon_3(X_1, X_2, X_3)|X_1 = s),\\
\sigma_2^2(\theta) &= D(\upsilon_2(X_1,\theta)),\\
\sigma_3^2(\theta) &= D(\upsilon_3(X_1,\theta))
\end{aligned}$$

Како су функције $\sigma_2^2(\theta)$ и $\sigma_3^2(\theta)$ непрекидне по θ и $\sigma_2^2(0) > 0$ и $\sigma_3^2(0) > 0$, (на основу (7.6)) закључујемо да су језгра Υ_2 и Υ_3 недегенерисана и под претпоставком да важи H_1, односно граничне расподеле статистика $\sqrt{n}T_{m,(2)}$ и $\sqrt{n}T_{m,(3)}$ су нормалне. Тестове ћемо упоређивати против алтернатива за које је $\mu'_{\Upsilon_k}(0) > 0$. Због непрекидности функције $\mu_{\Upsilon_k}(\theta)$,за $k = 1, 2$ закључујемо да су и остали услови теореме 3.1.1 испуњени. На основу тога је Питманов нагиб

$$c_P = \frac{\mu'_{\Upsilon_k}(0)^2}{k^2\sigma_k^2}.$$

Истим резоновањем као за тест-статистике засноване на U-емпиријским Лапласовим трансформацијама закључујемо да је Бахадуров нагиб

$$c_B = \frac{\mu'_{\Upsilon_k}(0)^2\theta^2}{k^2\sigma_k^2}.$$

Зато се локалне релативне асимптотске Бахадурове ефикасности поклапају са Питмановим локалним асимптотским ефикасностима.

Упоредићемо тест-статистике (7.4) са тест-статистикама интегралног типа засноване на емпиријским функцијама расподеле односно са

$$\begin{aligned}
I_n^M &= \frac{1}{n^4}\sum_{i_1,..,i_4}\Big(\mathrm{I}\{\max(X_{i_2}, X_{i_3}) < X_{i_4}\}\\
&\quad - \mathrm{I}\{X_{i_1} + \min(X_{i_2}, X_{i_3}) < X_{i_4}\}\Big);\\
I_n &= \frac{1}{n^5}\sum_{i_1,..,i_5}\mathrm{I}\Big(\{\max(X_{i_2}, X_{i_3}, X_{i_4}) < X_{i_5}\}\\
&\quad - \mathrm{I}\{X_{i_1} + \mathrm{med}(X_{i_2}, X_{i_3}, X_{i_4}) < X_{i_5}\}\Big).
\end{aligned}$$

Статистика I_n је детаљно проучавана у одељку 5.1, а асимптотска својства статистике I_n^M се на сличан начин изводе па ћемо извођења изоставити и навести само резултате неопходне за одређивање релативне ефикасности.

Прва пројекција језгра Φ^M на X_1 под претпоставком да важи H_0 је

$$\phi^M(s) = \frac{5}{24} - \frac{e^{-2s}}{8} - \frac{e^{-s}}{6},$$

на основу кога добијамо да је гранична расподела тест-статистике I_n^M нормална $\mathcal{N}(0, \frac{2}{135})$. За граничну вредност тест-статистике у вероватноћи

P_θ кад $n \to \infty$ за алтернативе $g(x, \theta)$ које задовољавају услове регуларности важи

$$b'(0) = 4 \int_0^\infty \phi^M(s)h(s).$$

Сад имамо све што нам је потребно да одредимо Питманове нагибе и релативну ефикасност статистика против алтернатива (5.5),(5.7), (5.6) и (5.9). Резултати су приказани у табели 7.3. Интресантно је да се тест $T_{n,2}$ показао као ефикаснији од одговарајућег заснованог на U-емпиријским функцијама расподеле за све посматране алтернативе.

Табела 7.3: Приближна АРЕ ($T_{n,2}, I_n^M$ и $T_{n,3}, I_n$)

Стат.	$(T_{n,2}, I_n^M)$	$(T_{n,3}, I_n)$
Вејбул	1.169	0.980
гама	1.147	0.722
Макехам	1.600	1.214
ЛСО	3.600	2.818

7.3 Моћи тестова за мале обиме узорка

У табелама 7.4 и 7.5 приказане су моћи тестова предложених на основу U-емпиријских Лапласових трансформација и на основу U-емпиријских момената и упоређене са тестом интегралног типа (5.1). Можемо приметити да су за мале обиме узорака тестови засновани на Лапласовим трансформацијама доста бољи од одговарајућих заснованих на моментима, што се не дешава у асимптотском случају.

Табела 7.4: Проценат одбачених нултих хипотеза за различите тестове сагласности $n = 20$, $\alpha = 0.05$

Алт.	$J^{\mathcal{D},1}$	$J^{\mathcal{D},2}$	$J^{\mathcal{D},5}$	$J^{\mathcal{P},1}$	$J^{\mathcal{P},2}$	$J^{\mathcal{P},5}$	$T_{2,n}$	$T_{3,n}$	I_n	$I_n^{\mathcal{M}}$
$W(1.4)$	47	48	50	49	50	49	34	35	45	46
$\Gamma(2)$	70	67	67	65	64	62	54	52	61	59
HN	24	26	30	29	31	31	17	18	24	30
U	51	61	71	73	77	79	30	30	66	79
$CH(0.5)$	17	20	22	21	23	22	13	13	17	23
$CH(1.0)$	17	19	23	21	23	22	13	13	28	22
$CH(1.5)$	17	19	22	21	23	23	13	13	17	22
$LF(2.0)$	31	35	40	38	40	41	22	22	32	39
$LF(4.0)$	44	49	50	52	56	56	31	30	47	53
$EW(0.5)$	18	20	22	21	23	23	13	13	18	22
$EW(1.5)$	40	46	55	53	57	58	27	26	48	57

Табела 7.5: Проценат одбачених нултих хипотеза за различите тестове сагласности $n = 50$, $\alpha = 0.05$

Алт.	$J^{\mathcal{D},1}$	$J^{\mathcal{D},2}$	$J^{\mathcal{D},5}$	$J^{\mathcal{P},1}$	$J^{\mathcal{P},2}$	$J^{\mathcal{P},5}$	$T_{2,n}$	$T_{3,n}$	I_n	$I_n^{\mathcal{M}}$
$W(1.4)$	82	87	88	86	89	88	54	55	83	82
$\Gamma(2)$	97	98	97	97	97	96	83	83	96	94
HN	41	53	60	53	62	66	19	18	49	58
U	80	92	98	98	99	99	31	31	96	100
$CH(0.5)$	28	37	45	40	46	49	14	14	33	37
$CH(1.0)$	27	38	44	41	46	48	14	15	33	38
$CH(1.5)$	28	38	45	40	47	50	14	13	32	38
$LF(2.0)$	56	70	76	72	77	80	25	26	66	62
$LF(4.0)$	75	86	91	89	92	93	35	35	84	79
$EW(0.5)$	28	38	44	41	47	50	14	13	33	37
$EW(1.5)$	70	82	90	88	92	94	30	29	82	88

Поглавље 8

Примена у анализи временских серија

У овом поглављу приказаћемо примену неких од приказаних тестова за проверу коректности модела у анализи временских серија.

8.1 Провера коректности модела трајања

Напредак човечанства у протекле две деценије оставио је траг и на начин трговања свуда у свету. Највећи утицај на промене трговања имао је развој рачунарства. Значајно повећање моћи рачунара омогућило је да се трансакције сакупљају не само на дневном нивоу. Наиме, уместо бележења цене вредносних папира или робе (и осталих његових својстава) највише два пута дневно (на отварању и затварању берзе) омогућено је синхроно чување података, односно трансакције су се бележиле у реалном времену. Било је нужно променити приступ и у анализи таквих података јер модели временских серија којима се описивало неко обележје, претпостављали су да су временски тренуци у којима се дешавају промене истих, еквидистантни. Један од проблема био је и тај што чак ни просечна учесталост по јединици времена није била константа. Уколико би се одабрао мали интервал постојала је бојазан да се појаве интервали у којима нема трансакција. С друге стране ако је тај интервал велики онда би дошло до великог губитка информација што би се одразило на квалитет модела. Због свих ових недостатака претходног модела појавили су се стохастички модели за моделирање времена између трансакција. Ауторегресивни условни модели трајања (ACD) су први пут били предложени од стране Ингла и Расела (в. [24]) баш за моделирање неправилно расподељених финансијских трансакција. Од тада привлаче пажњу многих истраживача и њихова примена је све већа и већа.

У даљем тексту описаћемо неке ACD моделе.

Основни појам у ACD моделирању је *трајање*. Под трајањем подразумевамо интервал између две трансакције. Природно је да се трајање моделира ненегативном случајном величином.

Означимо са $X_i = T_i - T_{i-1}$, $i = 1, 2...$, дужину i-тог интервала. $ACD(p,q)$ моделом се претпоставља да је

$$X_i = \varepsilon_i \psi_i,$$

где је $\{\varepsilon_i\}$ низ независних и једнако расподељених случајних величина са позитивним носачем које испуњавању услов нормираности $E(\varepsilon_i) = 1$. Условна трајања ψ_i су описана једначином

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j X_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j},$$

где су p и q ненегативни природни бројеви који означавају ред модела, док су α_j и β_k позитивне константе такве да модел задовољава услове стационарности. Неопходан услов за то је да је сума коефицијената модела мања од 1.

Први предложени ACD модел подразумевао је да су случајне величине ε_i са експоненцијалном $\mathcal{E}(1)$ расподелом и назива се *експоненцијални ауторегресивни модел трајања* $EACD(p,q)$. Поред њега, у истом раду (в. [24]) је био предложен и модел у којима случајне величине ε_i имају Вејбулову расподелу $WACD(p,q)$. Касније су били предложени модели у којима случајне величине ε_i имају генерализовану гама расподелу (в. [62]) или Бурову расподелу (в. [33]).

Важна етапа у моделовању је провера коректности модела. Поред тестова корелисаности резидуала који се могу наћи у литератури нови приступ подразумева и проверу сагласности резидуала са претпостављеном расподелом, нпр. у случају $EACD(p,q)$ модела са експоненцијалном расподелом.

Посебну пажњу посветићемо $EACD(1,1)$ моделу. Претпоставимо дакле да је наш модел задат једначинама:

$$\begin{aligned} X_i &= \varepsilon \psi_i, \\ \psi_i &= \omega + \alpha X_{i-1} + \beta \psi_{i-1}, \quad i = 1, 2, ... \end{aligned} \tag{8.1}$$

Уколико је временска серија $\{X_t\}$ слабо стационарна онда њена дисперзија мора бити коначна и очекивање константно. Узимајући у обзир претпоставку модела добијамо да је

$$
\begin{aligned}
\mu_x = E(X_t) &= \frac{\omega}{1-\alpha_1-\beta_1} \\
\sigma_x^2 = D(X_t) &= \frac{\mu_x^2(1-(\alpha_1+\beta_1)^2)}{1-2\alpha_1^2-\beta_1^2-2\alpha_1\beta_1},
\end{aligned}
$$

на основу чега закључујемо да мора да важи

$$
\begin{aligned}
&0 \leq \alpha_1+\beta_1 \\
&2\alpha_1^2+\beta_1^2+2\alpha_1\beta_1 < 1.
\end{aligned}
$$

На сличан начин могу се добити потребни услови слабе стационарности за $EACD$ модел већег реда.

Уколико су познати параметри модела прогнозу можемо вршити на сличан начин као код $ARMA$ модела. Уколико то није случај потребно је параметре оценити. То радимо користећи метод условне максималне веродостојности.

Означимо са θ вектор свих непознатих параметара у моделу, $x_t = (x_1, ..., x_t)'$ вектор реализованих вредности временске серије до тренутка t и нека је $m = \max(p,q)$. Тада функцију веродостојности можемо приказати у облику

$$f(x_n|\theta) = f(x_m|\theta)\cdot f(x_{m+1}|\mathbf{x_m},\theta)\cdot f(x_n|\mathbf{x_{n-1}},\theta).$$

С обзиром на то да је m обично занемарљиво у односу на величину серије на основу које вршимо оцену непознатих параметара можемо занемарити утицај функције $f(x_m|\theta)$ на вредност функције веродостојности $f(x_n|\theta)$ а оцене параметара θ су оне које максимизирају функцију

$$L(\theta) = \prod_{i=m+1}^{n} f(x_i|\mathbf{x_{i-1}}\theta),$$

или, еквивалентно, али доста погодније за рад, функцију

$$l(\theta) = \log L(\theta). \tag{8.2}$$

Из поставке модела је

$$f(x_i|\mathbf{x_{i-1}},\theta) = \frac{1}{\psi_i}e^{-\frac{x_i}{\psi_i}}.$$

У том случају је израз 8.2 еквивалентан са

$$-\sum_{i=m+1}^{n} \log\psi_i + \frac{x_i}{\psi_i}.$$

Оцене добијене на овај начин називају се $QMLE$ *оцене квази методом максималне веродостојности.*

Оцењена дурација је тада

$$\hat{\psi}_i = \hat{\omega} + \hat{\alpha} x_{i-1} + \hat{\beta}\hat{\psi}_{i-1}$$

а резидуали модела су

$$\hat{\varepsilon}_i = \frac{x_i}{\hat{\psi}_i}.$$

Уколико је наша претпоставка о моделу коректна онда ће се оцењени резидуали $\{\varepsilon_i\}$ понашати као низ независних случајних величина са експоненцијалном $\mathcal{E}(1)$ расподелом. Један од начина да се испита некорелисаност резидуала је примена Љунг-Боксовог теста. За проверу сагласности резидуала са експоненцијалном расподелом предлажемо, поред стандардних тестова и тестове описане у претходним поглављима. У даљем тексту упоредићемо моћи тестова уколико је алтернативна расподела гама или Вејбулова расподела. Пре тога наглашавамо да модел 8.1 треба да задовољава услове регуларности који су гломазни и које нећемо наводити (в. [58]). За одређивање моћи тестова користићемо следећи алгоритам (в. [96]), који се ослања на параметарски бутстреп.

1. На основу полазне серије $X_1, ..., X_n$ $QMLE$ методом добију се оцене $\hat{\theta}^*$ непознатих параметара ACD модела, оцене резидуали модела и на основу њих добије вредност тест статистике $\hat{T}$.

2. Генеришу се $\varepsilon_1^*, \varepsilon_2^*, ..., \varepsilon_n^*$ са расподелом F_0 и на основу њих конструишу оцењене вредности елемената серије $X_1^*, X_2^*, ..., X_n^*$ користећи оцене параметара добијене у претходном кораку. За почетне вредности серије за генерисање се одаберу вредности $X_0^* = X_{-1}^* = \cdots = X_{-p+1}^* = \psi_0^* = \psi_{-1}^* = \cdots = \psi_{-q+1}^* = \bar{X}_n$

3. Оцене се параметри модела $\hat{\theta}$ и оцене резидуали на основу којих се рачуна вредност тест статистике $\hat{T}_n$.

4. Понове се кораци 2-4 B пута и на основу тога одреди $(1-\alpha)$-ти квантил расподеле тест статистике $\hat{q}$. Уколико је вредност $\hat{T}$ мања од $\hat{q}$ прихвата се нулта хипотеза.

5. Понове се кораци 1-5 N пута и моћ теста оцени уделом одбачених хипотеза.

Напомињемо да је ово алгоритам који се најчешће користи у овим ситуацијама.

Испитаћемо квалитет једног од наших тестова ($J^{\mathcal{P}}_{n,a}$), одређујући *емпиријску величину теста* (удео одбачених нултих хипотеза под претпоставком да важи она) и моћ теста (са $N = 1000$ понављања).

Генерисаћемо $EACD(1,1)$ серију обима $n = 300$ са параметрима $\theta = (0.3, 0.2, 0.4)$, где је $\omega = 0.3$, $\alpha = 0.2$ и $\beta = 0.4$. Поступак ћемо поновити са параметрима модела $\theta = (0.3, 0.2, 0.2)$.

Тестираћемо нулту хипотезу да су резидуали $\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n$ из експоненцијалне $\mathcal{E}(1)$ расподеле против алтернативе да се ради о гама $\Gamma(a,b)$ при чему због услова модела мора бити $a = b$, и против алтернативе да се ради о Вејбуловој $\mathcal{V}(\lambda, 1)$.

Конкурентски тестови су Колмогоров-Смирнов тест D_n, Крамер-фон Мизесов тест ω_n, тест заснован на Ђинијевом индексу са тест-статистиком

$$G'_n = 1 - G_n = \frac{1}{n-1}(2n - \frac{2}{n\bar{X}_n} jX_{(j)}),$$

тест експоненцијалности заснован на емпиријској карактеристичној функцији, односно њеном реалном ($s_n(t)$) и имагинарном делу ($c_n(t)$) са тест-статистиком

$$W_{n,a} = \int_{-\infty}^{\infty} (s_n(t) - ic_n(t))^2 e^{-at} dt.$$

Табела 8.1: Емпиријска величина и моћ тестова за модел са $\theta = (0.3, 0.2, 0.4)$

Тест	$\mathcal{E}(1)$	$\Gamma(1.2, 1.2)$	$\Gamma(1.5, 1.5)$	$\mathcal{V}(1.1)$	$\mathcal{V}(1.2)$
D_n	0.056	0.496	0.995	0.376	0.917
ω_n	0.041	0.581	0.999	0.488	0.965
G'_n	0.065	0.663	0.998	0.620	0.985
$W_{n,a}$	0.055	0.662	1	0.542	0.981
$J^{\mathcal{P}}_{n,a}$	0.056	0.735	1	0.580	0.976

Генерално говорећи видимо да се наш тест показао као најмоћнији за тестирање сагласности резидуала против гама алтернативе, док је у случају Вејбулове алтернативе један од водећих. Исто тако можемо приметити да се поредак тестова није променио након што смо мало изменили полазне параметре модела.

Табела 8.2: Емпиријска величина и моћ тестова за модел са $\theta = (0.3, 0.2, 0.2)$

Тест	$\mathcal{E}(1)$	$\Gamma(1.2, 1.2)$	$\Gamma(1.5, 1.5)$	$\mathcal{V}(1.1)$	$\mathcal{V}(1.2)$
D_n	0.043	0.509	0.990	0.407	0.906
ω_n	0.050	0.575	0.997	0.455	0.961
G'_n	0.048	0.655	0.997	0.642	0.988
$W_{n,a}$	0.058	0.651	1	0.550	0.982
$J^{\mathcal{P}}_{n,a}$	0.045	0.759	1	0.593	0.978

8.2 Детекција цикличне компоненте временске серије

Откривање цикличне компоненте важан је корак у анализи временских серија. У овом одељку приказаћемо како се неки од приказаних тестова могу користити у те сврхе.

Нека је $\{X_t,\ t \in N\}$ стационарна временска серија са очекивањем μ и коваријационом функцијом K за коју важи

$$\sum_{h=-\infty}^{\infty} |K(h)| < \infty.$$

Тада постоји спектрална густина временске серије $\{X_t,\ t \in N\}$ дефинисана са

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} K(k) e^{-ik\omega}, \ \ \omega \in [-\pi, \pi].$$

Једна од оцена спектралне густине на основу првих n чланова временске серије је *периодограм* дефинисан за Фуријеове фреквенције $\omega_j = \frac{2\pi j}{n}$, $j = -[(n-1)/2], ..., [n/2]$ са

$$I(\omega_j) = \begin{cases} \frac{1}{2\pi} \sum_{|k|<n} \hat{K}(k) e^{-ik\omega_j}, & \text{за } j \neq 0; \\ \frac{n}{2\pi} \bar{X}_n, & \text{иначе}, \end{cases} \tag{8.3}$$

где је

$$\hat{K}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_{t+|k|} - \bar{X}_n)(\bar{X}_{t+|k|} - \bar{X}_n).$$

Показује се да је израз (8.3) еквивалентан са

$$I(\omega_j) = \frac{1}{2\pi n} \Big| \sum_{t=1}^{n} X_t e^{-it\omega_t} \Big|^2.$$

Претпоставимо да је временска серија $\{X_t\}$ дефинисана са

$$X_t = \mu + \sum_{k=1}^{(p-1)/2} A_k \cos\frac{2\pi kt}{p} + B_k \sin\frac{2\pi kt}{p} + A_{\frac{p}{2}}(-1)^t + Z_t, \tag{8.4}$$

при чему је $A_{\frac{p}{2}} = 0$ уколико је p непаран број а $\{Z_t\}$ је Гаусов бели шум са дисперзијом σ^2. Желимо да тестирамо нулту хипотезу да посматрана серија нема сезонску компоненту, односно

$$H_0 : A_j = B_j = 0,\ j = 1, 2..., (p-1)/2.$$

У књизи Дејвиса и Броквела [15] показан је следећи резултат.

Теорема 8.2.1. *Нека је X_t стационарна временска серија дефинисана са (8.4). Тада су случајне величине*

$$Y_i = \frac{\sum_{k=1}^{i} I(\omega_k)}{\sum_{k=1}^{q} I(\omega_k)},\ \ i = 1, 2, ..., q-1,$$

где је $q = \left[\frac{n-1}{2}\right]$, расподељене као статистике поретка униформне $\mathcal{U}[0,1]$ расподеле ако и само ако је $\{X_t\}$ Гаусов бели шум.

Пример 8.2.1. *Посматрајмо временску серију*

$$X_t = \sin\frac{t}{3}\pi + Z_t,\ t = 1...100.$$

Желимо да тестирамо да је X_t Гаусов бели шум. Уколико је то тачно онда

$$Y_k = \frac{\sum_{i=1}^{k} I(\omega_i)}{\sum_{i=1}^{48} I(\omega_i)},\ \ k = 1.., 48,$$

су статистике поретка униформне $U[0,1]$ расподеле. Зато ћемо применити тестове T_n^1 и T_n^2 (5.39). Добијамо p-вредности теста 0.032 и 0.029 и закључујемо да циклична компонента постоји.

На основу овог примера видимо да је наш тест сагласности са униформном расподелом успешно детектовао цикличну компоненту генерисане серије.

Закључак

У дисертацији је предложено тринаест тестова сагласности који се могу поделити у две групе. Прву чине тестови сагласности засновани на U-статистикама и V-статистикама. Тестови су конструисани коришћењем карактеризација на основу једнако расподељених статистика, на основу независности статистика, момената статистика и каратеризација расподела на основу функционалних једначина које задовољава функција расподеле. У предложеним статистикама се јаваљају U-емпиријске функције расподеле и U-емпиријски моменти расподеле. Другу групу чине тестови сагласности засновани на U-статистикама и V-статистикама са оцењеним параметрима на основу карактеризација једнако расподељених статистика које су линеарне функције елемената узорака и његових статистика поретка. У предложеним тестовима се јављају U-емпиријски моменти као и U-емпиријске Лапласове трансформације. Поред тестова сагласности у раду је предложено и пет тестова симетрије.

За све предложене тест-статистике су нађене граничне расподеле. За њено одређивање коришћена су својства U-статистика и V-статистика која су наведена у првом делу дисертације.

Прва група предложених тестова је упоређивана са стандардним тестовима сагласности, као и са новијим тестовима који су били предложени у последњих неколико година. Главни критеријум за поређење је била локална Бахадурова асимптотска ефикасност. За одређивање исте коришћене су теореме о великим одступањима за неке класе U-статистика и одговарајућих фамилија U-статистика. Формулисана је и доказана нова теорема о функцији великих одступања чији значај произилази из чињенице да се њоме одређује функција великих одступања за широку класу тест-статистика. Неки од предложених тестова су упоређени и за узорка малих обима одређивањем емприсјке моћи тестова коришћењем Монте Карло метода.

На крају дисертације је приказана и примена предложених тестова у анализи временских серија.

Правци даљег истраживања су бројни а међу њима се издваја проналажење функције великих одступања за U-статистике са оцењеним параметрима са недегенерисаним или слабо дегенерисаним језгром. Уколико се то оствари решила би се велика класа проблема у непараметарској статистици. Још један од интресантних даљих правaца истраживања је проналажење граничне расподеле дегенерисане статистике Колмогоровљевог типа и одговрајуће функције великих одступања.

Литература

[1] I. G. Abrahamson, Exact Bahadur efficiences for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics, Ann. Math. Stat. (1967) vol.38 1475–90.

[2] 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.

[3] M. Ahsanullah, On some characteristic property of symmetric distributions. Pakist. J. Statist. vol.8 (1992) 19–22.

[4] M. Ahsanullah, G. G. Hamedani, *Exponential distribution: Theory and Methods*, NOVA Science, New York, 2010.

[5] 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.

[6] B. C. Arnold, N. Balakrishnan, H. N. Nagaraja, *A First Course in Order Statistics*, SIAM, Philadelphia, 2008.

[7] M. A. Arcones, A Bernstein-type inequality for U-statistics and U-processes, Statist. Probab. Lett. vol.22(3) (1995) 239–247.

[8] A. Azzalini with the collaboration of A. Capitanio, *The skew-normal and related families*, Cambridge University Press, New York 2014.

[9] N. Balakrishnan, C. R. Rao, *Order Statistics, Theory & Methods*, Elsevier, Amsterdam, 1998.

[10] L. Baringhaus, N. Henze, Test of fit for exponentiality based on a characterization via the mean residual life function, Statist. Papers vol.41(2) (2000) 225–236.

[11] L. Baringhaus, D. Gaigall, On an independence test approach to the goodness-of-fit problem, J. Multivar. Anal. (2015) DOI:10.1016/j.jmva.2015.05.013

[12] L. Baringhaus, N. Henze, A characterization of and new consistent tests for symmetry, Comm. Statist. Theory Methods vol.21(6) (1992) 1555–1566.

[13] 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.

[14] L. Beghin, Ya. Yu. Nikitin, Approximate asymptotic Bahadur efficiency of independence tests with random sample size, J. Ital. Statist. Soc. vol.8(1) (1999) 1–23.

[15] P. Brockwell. R. Davis, *Time Series: Theory and Methods* New York: Springer-Verlag, 1990.

[16] G. Burgio, Ya. Yu. Nikitin, The Combination of the Sign and Wilcoxon Tests for Symmetry and Their Pitman Efficiency, Asymptotic Methods in Probability and Statistics with Applications Birkhuser Boston, 2001. 395–407.

[17] G. Burgio, Ya. Yu. Nikitin, On the combination of the sign and Maesono tests for simmetry and its efficiency, Statistica vol.63(2) (2007) 213–222.

[18] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on sum of observation. Ann. Math. Stat. vol.23 (1952) 493–507.

[19] D. R. Cox, D. Oakes, *Analysis of survival data*, Chapman and Hall, New York 1984.

[20] S. Csörgo, J. L. Teugels, Empirical Laplace transform and approximation of compound distributions, J. Appl. Probab. vol.27(1) (1990) 88–101.

[21] R. D'Agostino , M. Stephens, *Goodness-of-fit techniques*, Marcel Dekker, Inc., New York (1986).

[22] M. M. Desu, A characterization of the exponential distribution by order statistics, Ann. Math. Statist. vol.42(2) (1971) 837–838.

[23] C. Donati-Martin, S. Song, M. Yor, On symmetric stable random variables and matrix transposition, Ann. Inst. H. Poincaré Probab. Statist. vol.30(3) (1994).

[24] R. F. Engle, J. R. Russell, Autoregressive conditional duration: a new model for irregularly spaced transaction data, Econometrica vol.66(5) (1998) 1127–1162.

[25] T. W. Epps, L. B. Pulley, A test for exponentiality vs. monotone hazard alternatives derived from the empirical characteristic function, J.R. Stat. Soc. Series B Stat. Methodol., vol.48 (1986) 206–213.

[26] J. Fortiana, A. Grané, Goodness-of-fit tests based on maximum correlations and their orthogonal decompositions, J. R. Stat. Soc. Series B Stat. Methodol. vol.65(1) (2003) 115–126.

[27] A. Feuerverger, R. A. Mureika, The empirical characteristic function and its applications, Ann. Appl. Stat. (1977) vol.5(1) 88–97.

[28] M. Fisz, Characterization of some probability distributions, Skand. Aktuarietidskr vol.41(1-2) (1958) 65–67.

[29] J. Galambos, S. Kotz, *Characterizations of Probability Distributions*, Springer-Verlag, Berlin-Heidelberg-New York, 1978.

[30] Y. M. Gomez, H. Bolfarine, H. W. Gomez, A New Extension of the Exponential Distribution, Rev. Columbiana Estadist., vol.37(1) (2014) 25–34.

[31] A. Grané, A. V. Tchirina, Asymptotic properties of a goodness-of-fit test based on maximum correlations, Statistics vol.47(1) (2013) 202–215.

[32] S. Ghosh, Some Tests of Normality Based on Transforms. Department of Statistics [University of North Carolina], 1988.

[33] J. Grammig, K. O. Maurer, Non-monotonic hazard functions and the autoregressive conditional duration model, Econom. J. vol.3(1) (2000) 16–38.

[34] P. Groeneboom, G. Shorack, Large deviation of goodness-of-fit statistics and liner combination of order statistics, Ann. Prob. vol.9, (1981) 971–87.

[35] N. Gürtler, N. Henze, Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function, Ann. Inst. Statist. Math. vol.52(2) (2000) 267–286.

[36] 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.

[37] 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.

[38] N. Henze, A new flexible class of omnibus tests for exponentiality, Comm. Statist. Theory Methods vol.22(1) (1993) 115–133.

[39] N. Henze, S. G. Meintanis. Tests of fit for exponentiality based on the empirical Laplace transform, Statistics vol.36(2) (2002) 147–161.

[40] N. Henze, S. G. Meintanis, B. Ebner, Goodness-of-fit tests for the gamma distribution based on the empirical Laplace transform, Comm. Statist. Theory Methods vol.41(9) (2012) 1543–1556.

[41] 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.

[42] N. Henze, S .G. Meintanis, Recent and classical tests for exponentiality: a partial review with comparisons, Metrika vol.61(1) (2005) 29–45.

[43] W. Hoeffding, A Class of Statistics with Asymptotically Normal Distribution, Ann. Math. Statist. vol.19(3) (1948) 293–395.

[44] M. Hollander, F. Proschan, Tests for the mean residual life, Biometrika vol.62(3) (1975) 585–593.

[45] H. K. Iverson, R. H. Randles, The effects on convergence of substituting parameter estimates into U-statistics and other families of statistics, Probab. Theory Related Fields, vol.81(3) (1989) 453–471.

[46] 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.

[47] P. L. Janssen, *Generalized empirical distribution functions with statistical applications*, Diepenbeek, Limburgs Universitair Centrum, 1988.

[48] V. Jevremović, A note on mixed exponential distribution with negative weights, Statist. Probab. Lett. vol.11(3) (1991) 259–265.

[49] M. Jovanović, B. Milošević, Ya. Yu. Nikitin, M. Obradović, K. Yu. Volkova, Tests of exponentiality based on Arnold-Villasenor characterization, and their efficiencies, Comput. Stat. Data Anal. vol.90 (2015) 100–113.

[50] A. M. Kagan, Yu. V. Linnik, C. R. Rao, *Characterization Problems in Mathematical Statistics* (translated from Russian), John Wiley and Sons, New York, 1973.

[51] E. Kamke, *Differentialgleichungen. Losungsmethoden und Losungen I* 7th edition. Akademische Verlagsgesellschaft Geest and Porting K.-G., Leipzig, 1961.

[52] B. Klar, On a test for exponentiality against Laplace order dominance, Statistics vol.37(6) (2003) 505–515.

[53] V. S. Korolyuk, Yu. V. Borovskih, *Theory of U-statistics*, Kluwer, Dordrecht, 1994.

[54] I. Kotlarski, On characterizing the gamma and the normal distribution, Pacific J. Math. vol.20(1) (1967) 69–76.

[55] H. L. Koul, A test for new better than used, Comm. Statist. Theory Methods vol.6(6) (1977) 563–574.

[56] H. L. Koul, Testing for new is better than used in expectation, Comm. Statist. Theory Methods vol.7(7) (1978) 685–701.

[57] E. Kremer, Approximate and local Bahadur efficiency of linear rank tests in the two-sample problem, Ann. Stat. vol.7(6) (1979) 1246–1255.

[58] S. Lee, H. Oh, Entropy test and residual empirical process for autoregressive conditional duration models, Comput. Stat. Data Anal. vol.86 (2015) 1–12.

[59] G. D. Lin, C. Y. Hu, On characterizations of the logistic distribution, J. Statist. Plann. Inference vol.138(4) (2008) 1147–1156.

[60] V. V. Litvinova, New nonparametric test for symmetry and its asymptotic efficiency, Vestnik St. Petersburg Univ. Math. vol.34(4) (2001) 12–14.

[61] E. Lukacs, A characterization of the gamma distribution, Ann. Math. Statist. vol.26(2) (1955) 319–324.

[62] A. Lunde, A generalized gamma autoregressive conditional duration model, Department of Economics, Politics and Public Administration, Aalbourg University, Denmark (1999).

[63] A. W. Marshall, I. Olkin, *Life distributions: Structure of nonparametric, semiparametric, and parametric families*, Springer, New York, 2007.

[64] S. G. Meintanis, G. Iliopoulos, Tests of fit for the Rayleigh distribution based on the empirical Laplace transform, Ann. Inst. Statist. Math. vol.55(1) 137–151 (2003).

[65] S. G. Meintanis, Ya. Yu. Nikitin, A. V. Tchirina, Testing exponentiality against a class of alternatives which includes the RNBUE distributions based on the empirical Laplace transform, J. Math. Sci. (N.Y.) vol.145(2) (2007) 4871–4879.

[66] B. Milošević, Asymptotic Efficiency of New Exponentiality Tests Based on a Characterization, Metrika vol.79(2) (2016) 221–236

[67] B. Milošević, Asymptotic Efficiency of Godness-of-fit Tests Based on Too-Lin Characterization, arXiv preprint (2015) arXiv:1508.05314.

[68] B. Milošević, M. Obradović, Some characterizations of exponential distributions based on order statistics, arXiv preprint (2014) arXiv:1412.5019.

[69] B. Milošević, M. Obradović, Two-dimensional Kolmogorov-type Goodness-of-fit Tests Based on Characterizations and their Asymptotic Efficiencies, J. Nonparametr. Stat. accepted for publication 1–17 (2015) arXiv:1505.07415.

[70] K. Morris, D. Szynal, Goodness of Fit Tests Based on Characterizations of Continuous Distributions, Appl. Math. vol.27(4) (2000) 475–488.

[71] K. W. Morris, D. Szynal, Goodness-of-fit tests using dual versions of characterizations via moments of order statistics, J. Math. Sci., vol.122(4), (2004) 3365–3383.

[72] P. Muliere, Ya. Yu. Nikitin, Scale-invariant test of normality based on Polya's characterization, Metron vol.60(1-2) (2002) 21–33.

[73] S. Nadarajah, F. Haghighi, An extension of the exponential distribution, Statistics, vol.45(6) (2010) 543–558.

[74] Ya. Yu. Nikitin, Bahadur efficiency of Watson-Darling goodness-of-fit tests. Zapiski Nauch.Semin. LOMI 158: 138-45; (1987b)transl.: J. Soviet Mathem vol.43 (1988) 2833–8.

[75] Ya. Yu. Nikitin, *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, New York, 1995.

[76] 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.

[77] Ya. Yu. Nikitin, Large Deviations of U-empirical Kolmogorov-Smirnov Test, and Their Efficiency, J. Nonparametr. Stat. vol.22(5) (2010) 649–668.

[78] 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.

[79] 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.

[80] 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.

[81] Ya. Yu. Nikitin, M. Ahsanullah, New U-empirical tests of symmetry based on extremal order statistics, and their efficiencies, Mathematical Statistics and Limit Theorems: Festschrift in Honour of Paul Deheuvels, Ed. M. Hallin, D. M. Mason, D. Pfeifer, J. G. Steinebach, Springer International Publishing, 2015. 231–248.

[82] Ya. Yu. Nikitin, A. V. Tchirina, Bahadur efficiency and local optimality of a test for the exponential distribution based on the Gini statistic, Statist. Methodol. Appl. vol.5(1) (1996) 163–175.

[83] M. Obradović, Three Characterizations of Exponential Distribution Involving the Median of Sample of Size Three, J. Statist. Theory Appl. vol.14(3) (2015) 257–264

[84] M. H. Obradović, Karakterizacije nekih raspodela i Bahadurova asimptotska efikasnost testova saglasnosti, Diss. Univerzitet u Beogradu-Matematički fakultet, (2015).

[85] M. Obradović, M. Jovanović, B. Milošević, Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics, Statistics vol.49(5) (2015) 1026–1041.

[86] N. Papadatos, Upper bound for the covariance of extreme order statistics from a sample of size three, Sankhya Ser. A vol.61 (1999) 229–240.

[87] V. Papathanasiou, Some characterizations of distributions based on order statistics, Statist. Probab. Lett. vol.9(2) (1990) 145–147.

[88] O. A. Podkorytova, Large deviation and Bahadur efficiency of Khmaladze-Aki statistic (ruski). Zapiski. Nauch, Semin. LOMIN (1990) vol.184 227–33.

[89] 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.

[90] M. Raghavachari, On theorem of Bahadur on the rate of convergence of test statistics, Ann. Math. Statist. vol.41(5) (1970) 1695–1699.

[91] C. R. Rao, Criteria of estimation in large samples, Sankhya Ser. A vol.25 (1963) 189–206.

[92] R. H. Randles, On the asymptotic normality of statistics with estimated parameters, Ann. Statist. vol.10(2) (1982) 462–474.

[93] F. H. Ruymgaart, M. C. A. van Zuijlen, Empirical U-statistics processes, J. Statist. Plann. Inference vol.32(2) (1992) 259–269.

[94] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York, 2002.

[95] B. W. Silverman, Convergence of a class of empirical distribution functions of dependent random variables, Ann. Probab. vol.11(3) (1983) 745–751.

[96] W. Stute, W. G. Manteiga, M. P. Quindimil, Bootstrap based goodness-of-fit-tests, Metrika vol.40(1) (1993) 243–256.

[97] A. V. Tchirina, Large deviations for a class of scale-free statistics under the gamma distribution, J. Math. Sci. vol.128(1) (2005) 2640–2655.

[98] A. V. Tchirina, Bahadur efficiency and local optimality of a test for exponentiality based on the Moran statistics, J. Math. Sci. vol.127(1) (2005) 1812–1819.

[99] Y. H. Too, G. D. Lin, Characterizations of uniform and exponential distributions, Statist. Probab. Lett. vol.7(5) (1989) 357–359.

[100] H. Volkmer, G. G. Hamedani, On distributions of order statistics and their applications to uniform distribution, Metrika vol.74(2) (2011) 287–295.

[101] K. Yu. Volkova, On asymptotic efficiency of exponentiality tests based on Rossbergs characterization, J. Math. Sci. (N.Y.) vol.167(4) (2010) 486–494.

[102] K. Yu. Volkova, Tests of exponentiality based on Yanev-Chakraborty characterization, and their efficiency, Ed. Nagy, S., Proc. 19th European Young Statistician Meeting, Prague, (2015) 156–159

[103] K. Yu. Volkova, On asymptotic efficiency of goodness-of-fit tests for the Pareto distribution based on its characterization, Statist. Methods Appl. (2015) DOI:10.1007/s10260-015-0330-y

[104] 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.

[105] H. S. Wieand, A condition under which the Pitman and Bahadur approaches to efficiency coincide, Ann. Statist. vol.4(5) (1976) 1003–1011.

Биографија аутора

Бојана Милошевић рођења је 15.2.1989. у Панчеву где је завршила основну школу "Исидора Секулић" и основну музичку школу "Јован Бандур". Школовање је наставила у Математичкој гимназији у Београду коју је завршила као вуковац. Током школовања је освојила више награда на републичким и савезним такмичењима из математике и физике. Основне студије на Математичком факултету, смер статистика, актуарска, финансијска и математика уписала је 2007. године и завршила са просечном оценом 10.0 као студент генерације. Мастер студије је завршила 2012. са истом просечном оценом. Исте године је уписала докторске студије, студијски правац математика.

У регионалном центру за младе таленте "Михајло Пупин" држала је додатну наставу од 2008. године до 2011. године.

У Математичкој гимназији ради од 2011. године до данас где је држала часове Геометрије и Вероватноће и математичке статистике.

На Математичком факултету ради од 2011. године до данас и то прве две године у звању сарадника у настави, а затим као асистент. Држала је вежбе из седам предмета: Вероватноћа и статистика Б, Вероватноћа и статистика (за информатичаре), Временске серије и примене у финансијама, Случајни процеси, Теорију вероватноће, Елементи финансијске математике и Одабрана поглавља случајних процеса.

Области њеног интресовања су теорија вероватноће и математичка статистика. Објавила је шест научних радова.

У часописима са SCI листе:

- M. Obradović, M. Jovanović, B. Milošević, V. Jevremović, Estimation of $P\{X \leq Y\}$ for Geometric-Poisson Model, Hacet. J. Math. Stat., vol.44(4) (2015) 949–964;

- M. Obradović, M. Jovanović, B. Milošević, Goodness-of-fit tests for Pareto distribution based on a characterization and their asymptotics, Statistics vol.49(5) (2015) 1026–1041;

- M. Jovanović, B. Milošević, Ya. Yu. Nikitin, M. Obradović, K. Yu. Volkova, Tests of exponentiality based on Arnold-Villasenor characterization, and their efficiencies, Comput. Stat. Data Anal. vol.90 (2015) 100–113;

- B. Milošević, Asymptotic Efficiency of New Exponentiality Tests Based on a Characterization, Metrika vol.79(2) (2016) 221–236;

- B. Milošević, M. Obradović, Two-dimensional Kolmogorov-type Goodness-of-fit Tests Based on Characterizations and their Asymptotic Efficiencies, J. Nonparametr. Stat. accepted for publication 1–17 (2015); arXiv:1505.07415.

У часопису ван SCI листе:

- 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, vol.11(1) (2014) 21–29.

На рецензији се налази још четири рада. Имала је следећа саопштења на научним конференцијама:

- M. Obradović, M. Jovanović, B. Milošević, Optimal unbiased estimates of $P\{X < Y\}$, Applied statistics, Ribno(Bled), Slovenija (2013);

- B. Milošević, The Ruin Time for The Sum of Two Compound Poisson Processes Perturbed by a Diffusion, The 13th Serbian Mathematical Congress, Vrnjačka banja (2014);

- Z. Vidović, B. Milošević, M. Obradović, K. Ilijević, Tests of normality and their sensitivity against particular alternatives Applied statistics, Ribno(Bled), Slovenija (2014);

- 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, (2014), 648–652;

- V. Jevremović, B. Milošević, M. Obradović, Karakterizacije raspodela verovatnoća s posebnim osvrtom na eksponencijalnu raspodelu, Peti simpozijum "Matematika i primene", Matematički fakultet, Beograd, Srbija (2014);

- B. Milošević, M. Obradović, Some characterization based exponentiality tests and their Bahadur efficiencies, PROBASTAT 2015-The 7th International Conference on Probability and Statistics, Smolenice, Slovačka (2015);
- B. Milošević, M. Obradović, Some goodness of fit tests based on U-empirical Laplace transforms, PROBASTAT 2015-The 7th International Conference on Probability and Statistics, Smolenice, Slovačka (2015)
- B. Milošević, M. Obradović, Testovi eksponencijalnosti zasnovani na empirijskim Laplasovim transformacijama, Šesti simpozijum "Matematika i primene", Matematički fakultet, Beograd, Srbija (2015);
- B. Milošević, M. Obradović, Characterization based symmetry tests and their asymptotic efficiencies, AMISTAT 2015, Prag, Češka (2015).

Учествовала је у следећим летњим школама:

- Summer School of Financial Mathematics, Ljubljana, Slovenija (2011);
- International Summer Academy 2012 on Advanced Stochastic Methods to Model Risk, Ulm, Nemačka (2012);
- Mathematical Models in Economics and Computer Implementation, Perm, Rusija (2013).

Учесник је пројекта 174012 Министарства просвете, науке и технолошког развоја Републике Србије од 2013. године.