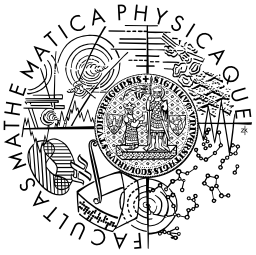

MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

**BAKALÁRSKA PRÁCA**

Romana Dvoranová

# Testování exponenciality

Katedra pravděpodobnosti a matematické statistiky

| | |
|---|---|
| Vedúci bakalárskej práce: | prof. RNDr. Anděl Jiří, DrSc. |
| Študijný program: | Matematika |
| Študijný obor: | Finanční matematika |

Praha 2016

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....  Podpis autora

Názov práce: Testování exponenciality

Autor: Romana Dvoranová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: prof. RNDr. Anděl Jiří, DrSc., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Táto bakalárska práca sa venuje podrobnému popisu a porovnaniu rady testov exponenciality. Zahŕňa klasické metódy testovania dobrej zhody pre exponenciálne rozdelenie, ako aj najnovšie testy exponenciality publikované v posledných rokoch. Podľa spôsobu charakterizácie exponenciálneho rozdelenia pokrýva $\chi^2$ testy dobrej zhody, testy založené na empirickej distribučnej funkcii, používajúce Kolmogorovovu-Smirnovovu a Cramérovu-von Misésovu testovú štatisku, ďalej testy založené na integrálnych transformáciach, entropii, strednej reziduálnej funkcii života, Giniho indexe a iné. Špeciálne sa táto bakalárska práca venuje testom založeným na charakterizácii pomocou rôznych typov entropie, ako napríklad Shannonovej, Rényiho alebo kumulatívnej reziduálnej entropii. V závere bakalárskej práce je zahrnutá simulačná štúdia porovnávajúca silu niekoľkých novších testov exponenciality, ktoré boli teoreticky popísané.

Kľúčové slová: exponenciálne rozdelenie, test dobrej zhody, empirická distribučná funkcia, entropia, kumulatívna reziduálna entropia

Title: Testing exponentiality

Author: Romana Dvoranová

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Anděl Jiří, DrSc., Department of Probability and Mathematical Statistics

Abstract: This bachelor thesis focuses on detailed review of a selection of tests for exponentiality and their comparison. This text presents classical methods for goodness-of-fit testing for exponentiality, as well as the most recent tests for exponentiality published in the last decades. Based on the characterisation of exponential distribution that is being used, the review includes $\chi^2$ goodness-of-fit tests, tests based on empirical distribution function using Kolmogorov-Smirnov and Cramér-von Misés test statistics, as well as tests based on integral transforms, entropy, mean residual life function, Gini index and others. In particular, this bachelor thesis focuses on tests for exponentiality based on entropy characterisation, e.g. using Shannon, Rényi or cumulative residual entropy. Finally, this thesis includes simulation study comparing power of several more recent tests for exponentiality that have been theoretically described.

Keywords: exponential distribution, goodness-of-fit test, empirical distribution function, entropy, cumulative residual entropy

Chcela by som poďakovať prof. RNDr. Jiřímu Anděloviovi, DrSc. ako vedúcemu mojej bakalárskej práce za jeho cenné rady a vedenie pri písaní tejto bakalárskej práce.

# Obsah

# Úvod

V tejto bakalárskej práci sa budem zaoberať rôznymi prístupmi k testovaniu exponenciality dát. Exponenciálne rozdelenie je veľmi dôležitým rozdelením v modelovaní v rôznych sférach nielen vedy ale aj inžinierstva, poisťovníctva a iných oborov. Preto je schopnosť spoľahlivo testovať, či reálne dáta pochádzajú z triedy exponenciálnych rozdelení, už niekoľko desaťročí predmetom výskumu matematikov.

Vychádzajúc z prác autorov Henze a Meintanis (2005) a Baratpour a Rad (2012), prevediem rozsiahlu rešerš literatúry s cieľom popísať klasické aj moderné testy exponenciality. Niektoré popísané testy na základe simulačnej štúdie porovnám z hľadiska ich empirickej sily proti vybraným alternatívam.

Ďalej bude táto bakalárska práca členená nasledovne. V kapitole 1 zavediem definície dôležitých pojmov a popíšem niektoré dôležité vlastnosti exponenciálneho rozdelenia. V kapitole 2 popíšem niekoľko testov exponenciality členených podľa charakterizácie exponenciálneho rozdelenia, ktorú využívajú. Nakoniec v kapitole 3 previedem simulačnú štúdiu s 1 000 simuláciami s cieľom porovnať empirickú silu vybraných modernejších testov exponenciality proti 6 vybraným alternatívnym rozdeleniam.

# 1. Základné pojmy

## 1.1 Exponenciálne rozdelenie

Na úvod definujeme exponenciálne rozdelenie.

**Definícia 1** (Exponenciálne rozdelenie)**.** *Spojitá náhodná veličina $X$ má exponenciálne rozdelenie s parametrom $\frac{1}{\lambda}, \lambda > 0$, ak jej hustota má tvar*

$$f(x) = \begin{cases} 0 & pre\ x \leq 0, \\ \lambda e^{-\lambda x} & pre\ x > 0. \end{cases}$$

*Označujeme $X \sim Exp(1/\lambda)$.*

Najčastejšie sa exponenciálne rozdelenie používa na modelovanie doby čakania na určitú udalosť alebo určitý jav. Ďalšou interpretáciou exponenciálneho rozdelenia je modelovanie doby medzi 2 udalosťami v postupnosti udalostí rovnakého typu (dopravné nehody, telefonáty do call centra, atd.). Parameter $\lambda$ nazývame intenzita. Distribučná funkcia exponenciálneho rozdelenia má tvar

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

Exponenciálne rozdelenie je spojitým rozdelením s nezáporným nosičom. Je špeciálnym prípadom gama rozdelenia s parametrami 1 a $\lambda$, t.j. $Exp(1/\lambda) = \Gamma(1,\lambda)$. Toto rozdelenie môžme taktiež chápať ako spojitú analógiu geometrického rozdelenia. Pre momenty exponenciálneho rozdelenia platí

$$\mathsf{E}\ X = \frac{1}{\lambda},$$
$$\mathsf{var}\, X = \frac{1}{\lambda^2}.$$

Ako jediné spomedzi spojitých rozdelení je exponenciálne rozdelenie označované ako rozdelenie bez pamäti. Hovoríme, že rozdelenie náhodnej veličiny $X$ je bez pamäti, ak splňuje

$$\mathsf{P}(X > s + t | X > t) = \mathsf{P}(X > s),$$

alebo ekvivalentne

$$\mathsf{P}(X > s + t) = \mathsf{P}(X > s)\, \mathsf{P}(X > t). \tag{1.1}$$

**Veta 1.** *Exponenciálne rozdelenie je jediné spojité rozdelenie bez pamäti.*

*Dôkaz.* Nech $X \sim Exp(1/\lambda)$, potom platí

$$\mathsf{P}(X > t) = 1 - \mathsf{P}(X \leq t) = 1 - F(t) = e^{-\lambda t}.$$

Analogicky $\mathsf{P}(X > s + t) = e^{-\lambda(s+t)}$. Z toho plynie

$$\mathsf{P}(X > s + t) = e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t} = \mathsf{P}(X > s)\ \mathsf{P}(X > t).$$

Teda platí (1.1) a tým je dokázané, že exponenciálne rozdelenie je rozdelenie bez pamäti. Teraz dokážeme, že exponenciálne rozdelenie je jediné spojité rozdelenie s touto vlastnosťou. Nech pre spojitú náhodnú veličinu $X$ s distribučnou funkciou $F$ platí (1.1), potom chceme dokázať, že $X$ má exponenciálne rozdelenie. Hľadáme funkciu $\bar{F}(t)$, ktorá spĺňa $\bar{F}(t+s) = \bar{F}(t)\bar{F}(s)$. Táto rovnica má dve triviálne riešenia $\bar{F}(t) = 0$ a $\bar{F}(t) = 1$. Funkcia $\bar{F}(t)$ má byť funkciou prežitia náhodnej veličiny, preto ani jedno z týchto triviálnych riešení nie je nami hľadaným riešením. Pre netriviálne riešenie má platiť $\bar{F}(t) = 1 - F(t)$, kde $F(t)$ je distribučná funkcia náhodnej veličiny $X$. Deriváciou distribučnej funkcie je hustota, preto funkcia $\bar{F}(t)$ musí byť diferencovateľná. Pozrime sa na deriváciu $\bar{F}(t)$

$$\begin{aligned}\bar{F}'(t) &= \lim_{h\to 0}\frac{\bar{F}(t+h)-\bar{F}(t)}{h} = \lim_{h\to 0}\frac{\bar{F}(t)\bar{F}(h)-\bar{F}(t)}{h}\\ &= \bar{F}(t)\lim_{h\to 0}\frac{\bar{F}(h)-1}{h} = a\bar{F}(t).\end{aligned}$$

Limita $a = \lim_{h\to 0}[\bar{F}(h)-1]/h$ musí existovať a byť nenulová, aby existovala nenulová derivácia $\bar{F}'(t) = -F'(t) = -f(t)$ pre nejaké $t$. Funkcia $\bar{F}(t)$ musí spĺňať nasledujúcu diferenciálnu rovnicu, pre $a \neq 0$

$$\bar{F}'(t) = a\bar{F}(t).$$

Ďalej musí byť funkcia $\bar{F}(t) \geq 0$, pretože je to funkcia prežitia. Triviálne riešenie $\bar{F}(t) = 0$ opäť nie je funkcia, ktorú hľadáme, pretože by neexistovala distribučná funkcia $F(t)$ splňujúca $\bar{F}(t) = 1 - F(t)$. Preto $\bar{F}(t)$ musí byť nenulová na nejakom intervale a na ňom pre $a \neq 0$ platí

$$\begin{aligned}\bar{F}'(t) &= a\bar{F}(t)\\ \frac{\bar{F}'(t)}{\bar{F}(t)} &= a\\ \int\frac{\bar{F}'(t)}{\bar{F}(t)}\,\mathrm{d}t &= \int a\,\mathrm{d}t\\ \log\bar{F}(t) &= at + C_1\\ \bar{F}(t) &= C_2 e^{at},\end{aligned}$$

kde $C_2 = e^{C_1} > 0$. Z definície konštanty $a$ dostávame

$$a = \lim_{h\to 0}\frac{\bar{F}(h)-1}{h} = \lim_{h\to 0}\frac{C_2 e^{ah}-1}{h} \neq 0.$$

Aby táto limita existovala a bola konečná, musí platiť

$$\lim_{h\to 0}(C_2 e^{ah} - 1) = 0,$$

v tom prípade limitu dopočítame pomocou l'Hospitalovho pravidla pre prípad „0/0". Odtiaľ dostávame hodnotu konštanty $C_2$ ako

$$1 = C_2\lim_{h\to 0} e^{ah} = C_2.$$

Dostávame riešenie $\bar{F}(t) = e^{at}$, $a \neq 0$. Pre $t = 0$ platí $\bar{F}(0) = \mathsf{P}(X > 0) = 1$. Z monotónnosti distribučnej funkcie plynie monotónnosť funkcie $\bar{F}(t)$, ktorá je nerastúca. Preto pre $t \leq 0$ platí

$$1 \geq \mathsf{P}(X > t) = \bar{F}(t) \geq \bar{F}(0) = 1.$$

Preto hľadaná funkcia má tvar

$$\bar{F}(t) = \begin{cases} 1 & \text{pre } t < 0, \\ e^{at} & \text{pre } t \geq 0,\ a \neq 0. \end{cases}$$

Prípustné hodnoty parametra $a$ dostaneme z

$$\lim_{t\to\infty} e^{at} = \lim_{t\to\infty} \bar{F}(t) = 1 - \lim_{t\to\infty} F(t) = 0.$$

Z tejto limity plynie, že $a < 0$, pretože exponenciála sa limitne blíži k 0 pre hodnoty argumentu blížiace sa $-\infty$. Označme $\lambda = -a$, $\lambda > 0$. Distribučná funkcia $F(t) = 1 - \bar{F}(t)$ náhodnej veličiny $X$ má tvar

$$F(t) = \begin{cases} 0 & \text{pre } t < 0, \\ 1 - e^{-\lambda t} & \text{pre } t \geq 0,\ \lambda > 0. \end{cases}$$

Z toho plynie, že náhodná veličina $X$ má exponenciálne rozdelenie $Exp(1/\lambda)$ a tým je tvrdenie vety dokázané.

□

## 1.2 Definície dôležitých pojmov

V tejto časti uvediem definície pojmov, na ktorých su založené štatistické testy exponenciality uvedené v kapitole 2 tejto bakalárskej práce.

**Definícia 2** (Empirická distribučná funkcia)**.** *Nech $X_1$,...,$X_n$ je náhodný výber. Empirická distribučná funkcia náhodného výberu $X_1$,...,$X_n$ je definovaná ako*

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}.$$

**Definícia 3** (Funkcia prežitia)**.** *Nech $X$ je náhodná veličina s distribučnou funkciou $F$. Potom jej funkciu prežitia definujeme ako*

$$\bar{F} := 1 - F.$$

**Definícia 4** (Integrovaná distribučná funkcia)**.** *Nech $X$ je kladná náhodná veličina s distribučnou funkciou $F$ a konečnou strednou hodnotou. Integrovaná distribučná funkcia náhodnej veličiny $X$ je definovaná pre $t > 0$ ako*

$$\Psi_X(t) := \int_t^\infty \bar{F}(x)\,\mathrm{d}x,$$

*kde $\bar{F}$ je funkcia prežitia z definície 3.*

Integrovaná distribučná funkcia charakterizuje rozdelenie náhodnej veličiny, čo využil Klar (2001) na konštrukciu testových štatistík, ako uvidíme neskôr.

**Definícia 5** (Empirická integrovaná distribučná funkcia)**.** *Nech $X_1$,...,$X_n$ je náhodný výber. Empirická integrovaná distribučná funkcia náhodného výberu je definovaná ako*

$$\Psi_n(t) := \int_t^\infty [1 - F_n(x)]\,\mathrm{d}x = \frac{1}{n}\sum_{i=1}^{n}(X_i - t)\mathbf{1}_{\{X_i > t\}}.$$

Druhá rovnosť v definícii 5 platí, pretože platí

$$\begin{aligned}\Psi_n(t) &= \int_t^\infty (1 - F_n(x))\,\mathrm{d}x = \frac{1}{n}\int_t^\infty \left(n - \sum_{i=1}^{n}\mathbf{1}_{\{X_i \le x\}}\right)\mathrm{d}x \\ &= \frac{1}{n}\int_0^\infty \sum_{i=1}^{n}\mathbf{1}_{\{X_i > x\}}\,\mathrm{d}x = \frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \mathbf{1}_{\{X_i > x\}}\,\mathrm{d}x \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i > t\}}\int_t^{X_i}\mathrm{d}x = \frac{1}{n}\sum_{i=1}^{n}(X_i - t)\mathbf{1}_{\{X_i - t\}}.\end{aligned}$$

**Definícia 6** (Stredná reziduálna funkcia života)**.** *Nech $X$ je kladná náhodná veličina s distribučnou funkciou $F$. Stredná reziduálna funkcia života náhodnej veličiny $X$ je definovaná pre $t > 0$ ako*

$$m(t) := \mathsf{E}(X - t|X > t) = \frac{1}{1 - F(t)}\int_t^\infty \bar{F}(x)\,\mathrm{d}x,$$

*kde $\bar{F}(x)$ je funkcia prežitia z definície 3.*

Ďalším pojmom, ktorý hrá kľúčovú rolu v rozvoji teórie testovania, či náhodný výber pochádza z určitej parametrickej triedy rozdelení, je pojem entropie. Ako prvý tento pojem ako mieru neurčitosti pokusu zaviedol Shannon v roku 1948. Ešte pred ním sa kvantifikáciou neurčitosti výsledkov pokusu zaoberal Hartley (1928), ktorý navrhol prirodzene modelovať túto mieru logaritmickou funkciou.

**Definícia 7** (Entropia pokusu)**.** *Majme pokus s $n$ možnými výsledkami $A_1$,...,$A_n$ s pravdepodobnosťami $p_1$,...,$p_n$. Entropia pokusu je definovaná ako*

$$H := -\sum_{i=1}^{n} p_i \log p_i.$$

Z definície 7 je zrejmé, že sa jedná o strednú hodnotu diskrétnej náhodnej veličiny $Z$ s hodnotami $-\log p_1$,..., $-\log p_n$ a ich pravdepodobnosťami $p_1$,...,$p_n$, t.j. $H = \mathsf{E}\ Z$ . Z tejto interpretácie môžme usúdiť, že Shannon vzal za mieru neurčitosti pokusu strednú hodnotu neurčitostí jednotlivých výsledkov $(-\log p_i)$. V tejto práci ďalej využijeme Shannonovu definíciu entropie spojitej náhodnej veličiny.

**Definícia 8** (Entropia spojitej náhodnej veličiny)**.** *Nech $X$ je spojitá náhodná veličina s hustotou $f$. (Shannonova) entropia náhodnej veličiny $X$ je definovaná ako*

$$H(f) := -\int_{-\infty}^{\infty} f(x) \log f(x) \, \mathrm{d}x.$$

Shannon (1948) ďalej uviedol nasledujúce tvrdenie charakterizujúce exponenciálne rozdelenie, ktoré samostatne dokážeme.

**Veta 2.** *Nech $X$ je spojitá náhodná veličina s hustotou $f$ a kladným nosičom ($f(x) = 0, x \leq 0$) a strednou hodnotou $\mathsf{E}\, X = a$. Potom $X$ má maximálnu entropiu $H(f)$ práve vtedy, keď $X \sim Exp(a)$.*

*Dôkaz.* Maximalizujeme entropiu $H(f) = -\int_0^{\infty} f(x) \log f(x) \, \mathrm{d}x$, za podmienok

$$a = \int_0^{\infty} x f(x) \, \mathrm{d}x \text{ a } 1 = \int_0^{\infty} f(x) \, \mathrm{d}x.$$

Pužijeme metódu Lagrangeových multiplikátorov a dostaneme Lagrangeovu funkciu v tvare

$$L(f) = \int_0^{\infty} -f(x) \log f(x) + \lambda_0 f(x) + \lambda_1 [x f(x)] \, \mathrm{d}x.$$

Deriváciou Lagrangeovej funkcie podľa $f$ dostaneme podmienku

$$-1 - \log f(x) + \lambda_0 + \lambda_1 x = 0.$$

Riešením tejto rovnice je funkcia

$$f(x) = e^{\lambda_0 + \lambda_1 x - 1}.$$

Toto riešenie dosadíme do druhej podmienky a riešime

$$1 = \int_0^{\infty} e^{\lambda_0 + \lambda_1 x - 1} = e^{\lambda_0 - 1} \left[ \frac{e^{\lambda_1 x}}{\lambda_1} \right]_0^{\infty} = -\frac{e^{\lambda_0 - 1}}{\lambda_1},$$

pričom $\lambda_1$ musí byť záporné, aby integrál bol konečný. Pre $\lambda_1$ dostávame vzťah

$$\lambda_1 = -e^{\lambda_0 - 1} < 0.$$

Dosadíme do prvej podmienky a pomocou integrácie per partes dostávame

$$\begin{aligned} a &= \int_0^{\infty} x e^{\lambda_0 + \lambda_1 x - 1} \, \mathrm{d}x = e^{\lambda_0 - 1} \left[ \frac{x e^{\lambda_1}}{\lambda_1} \Big|_0^{\infty} - \frac{1}{\lambda_1} \int_0^{\infty} e^{\lambda_1 x} \, \mathrm{d}x \right] \\ &= -\lambda_1 \frac{e^{\lambda_1 x}}{\lambda_1^2} \Big|_0^{\infty} \\ &= -\frac{1}{\lambda_1}. \end{aligned}$$

Z tejto rovnosti dostávame $\lambda_1 = -1/a$ a zo vzťahu z druhej podmienky dostávame $\lambda_0 = \log(\lambda_1) + 1 = \log(1/a) + 1$. Dosadením do $f(x)$ konečne dostávame

$$f(x) = e^{\log(1/a) + 1 - x/a - 1} = \frac{1}{a} e^{-x/a} = \lambda e^{-\lambda x}.$$

Tým sme dokázali, že hustota exponenciálneho rozdelenia maximalizuje entropiu na kladnej poloosi.

□

Veta 2 nám hovorí, že exponenciálne rozdelenie maximalizuje entropiu spojitej náhodnej veličiny na kladnej poloose.
Po tom ako Shannon zaviedol pojem entropie, niekoľko autorov nadviazalo na jeho prácu a zaviedli ďalšie miery tohto typu. Medzi najznámejšie patrí Rényiho entropia, ktorú Rényi (1961) zaviedol nasledovne.

**Definícia 9** (Rényiho entropia pokusu)**.** *Majme pokus s n možnými výsledkami* $A_1,...,A_n$ *s pravdepodobnosťami* $p_1,...,p_n$*. Rényiho entropia pokusu rádu* $\alpha, \alpha > 0, \alpha \neq 1$*, je definovaná ako*

$$H_\alpha := \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i^\alpha.$$

Shannonova entropia $H$ je špeciálnym prípadom Rényiho entropie $H_\alpha$ pre $\alpha \to 1$.

**Definícia 10.** *Nech X je spojitá náhodná veličina s hustotou f. Rényiho entropia rádu* $\alpha, \alpha > 0, \alpha \neq 1$*, náhodnej veličiny X je definovaná ako*

$$H_\alpha(f) := \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} f^\alpha(x)\,\mathrm{d}x.$$

Nevýhodou oboch predchádzajúcich definícii entropie spojitej náhodnej veličiny (definície 8, 10) je okrem iného nutnosť znalosti hustoty náhodnej veličiny. Aby sme dokázali napríklad odhadnúť Shannonovu entropiu pre spojitú náhodnú veličinu, potrebujeme odhad hustoty, čo je obecne zložité. Iný postup rozšírenia Shannonovej entropie na spojité náhodné veličiny bez znalosti hustoty zvolili Rao a kol. (2004) a zaviedli kumulatívnu reziduálnu entropiu.

**Definícia 11** (Kumulatívna reziduálna entropia)**.** *Nech X je spojitá náhodná veličina s distribučnou funkciou F. Kumulatívna reziduálna entropia náhodnej veličiny X je definovaná ako*

$$CRE(X) := -\int_0^{\infty} \bar{F}(x) \log \bar{F}(x)\,\mathrm{d}x,$$

*kde* $\bar{F}(x)$ *je funkcia prežitia z definície 3.*

Ďalším prístupom, ktorý autori využívajú na charakterizáciu exponenciálneho rozdelenia a jeho následné testovanie je použitie integrálnych transformácii. Dve z nich, ktoré neskôr uvidíme aplikované, si definujeme.

**Definícia 12** (Laplaceova transformácia)**.** *Nech X je reálna náhodná veličina s hustotou f. Potom Laplaceova transformácia hustoty náhodnej veličiny X je definovaná ako stredná hodnota*

$$\mathcal{L}_X(t) := \mathsf{E}(e^{-tX}), \quad t \in \mathbb{R}.$$

**Definícia 13** (Hankelova transformácia)**.** *Nech $X$ je reálna nezáporná náhodná veličina. Hankelova transformácia $X$ je definovaná ako reálna funkcia definovaná na $\mathbb{R}^+ \cup \{0\}$*

$$\mathcal{H}_X(t) := \mathsf{E}(J_0(2\sqrt{tX})), t \geq 0,$$

*kde $J_0$ je Besselova funkcia prvého druhu rádu 0, t.j.*

$$J_0(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!^2} \left(\frac{x}{2}\right)^{2k}.$$

Besselova funkcia prvého druhu rádu 0 má niekoľko integrálnych vyjadrení, čo dokazuje nasledujúca lemma.

**Lemma 3.** *Pre Besselovu funkciu prvého druhu rádu 0 platí*

$$J_0(x) = \frac{1}{\pi} \int_0^{\pi} \cos(x \sin\theta)\, \mathrm{d}\theta = \frac{1}{\pi} \int_0^{\pi} \cos(x \cos\theta)\, \mathrm{d}\theta.$$

*Dôkaz.* Označme funkciu $A(x) = 1/\pi \int_0^{\pi} \cos(x \sin\theta)\, \mathrm{d}\theta$. Pomocou rozvoja funkcie kosínus v mocninnú radu môžeme prepísať funkciu $A(x)$ do tvaru

$$A(x) = \frac{1}{\pi} \int_0^{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k (x \sin\theta)^{2k}}{(2k)!}\, \mathrm{d}\theta.$$

Keďže tento integrál je pre každé $x$ konečný (integrál z obmedzenej funkcie na konečnom intervale), zameníme poradie integrálu a sumy

$$\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{\pi(2k)!} \int_0^{\pi} \sin^{2k}\theta\, \mathrm{d}\theta.$$

Osobitne vypočítame hodnotu integrálu vovnútri sumácie. Funkcia sínus je symetrická na intervale (0,$\pi$) okolo bodu $\pi/2$, preto

$$\int_0^{\pi} \sin^{2k}\theta\, \mathrm{d}\theta = 2 \int_0^{\pi/2} \sin^{2k}\theta\, \mathrm{d}\theta = 2I_{2k}.$$

Pre integrály typu $I_n, n$ prirodzené, pomocou integrácie per partes získame vzťah

$$\begin{aligned}
I_n &= \int_0^{\pi/2} \sin^n\theta\, \mathrm{d}\theta = \int_0^{\pi/2} \sin^{n-1}\theta \sin\theta\, \mathrm{d}\theta \\
&= \Big[ -\sin^{n-1}\theta \cos\theta \Big]_0^{\pi/2} + \int_0^{\pi/2} (n-1) \sin^{n-2}\theta \cos^2\theta\, \mathrm{d}\theta \\
&= (n-1) \int_0^{\pi/2} (\sin^{n-2}\theta - \sin^n\theta)\, \mathrm{d}\theta \\
&= (n-1)(I_{n-2} - I_n).
\end{aligned}$$

Z tohto vzťahu ďalej dostávame rekurentný vzorec pre $I_n$ ako

$$I_n = \frac{n-1}{n} I_{n-2}.$$

Integrály sa líšia pre párne a nepárne $n$, pre náš vypočet potrebujeme vyjadriť $I_{2k}$, pre ktoré platí

$$\begin{aligned} I_{2k} &= \frac{(2k-1)}{2k}.\frac{(2k-3)}{2k-2}\dots\frac{1}{2}I_0 \\ &= \frac{(2k-1)(2k-3)\dots 1}{2^k[k(k-1)\dots 1]}\int_0^{\pi/2} 1\,\mathrm{d}\theta \\ &= \frac{(2k-1)(2k-3)\dots 1}{2^k k!}.\frac{\pi}{2}.\frac{2^k k!}{2^k k!} \\ &= \frac{(2k)!\pi}{(k!)^2 2^{2k+1}}. \end{aligned}$$

Dosadením tohto výsledku do $A(x)$ dostaneme

$$A(x) = \sum_{k=0}^{\infty}\frac{(-1)^k x^{2k}}{\pi(2k)!}2I_{2k} = \sum_{k=0}^{\infty}\frac{(-1)^k}{(k!)^2}\left(\frac{x}{2}\right)^{2k} = J_0(x).$$

Druhú rovnosť dokážeme použitím vlastností funkcí sínus a kosínus. Analogickou úpravou akú sme použili pre $A(x)$ upravíme druhý integrál do tvaru

$$\frac{1}{\pi}\int_0^{\pi}\cos(x\cos\theta)\,\mathrm{d}\theta = \sum_{k=0}^{\infty}\frac{(-1)^k x^{2k}}{\pi(2k)!}\int_0^{\pi}\cos^{2k}\theta\,\mathrm{d}\theta.$$

Integrál v sume vyjadríme pomocou $I_{2k}$ a tým ukážeme dokazovanú rovnosť

$$\int_0^{\pi}\cos^{2k}\theta\,\mathrm{d}\theta = 2\int_0^{\pi/2}\cos^{2k}\theta\,\mathrm{d}\theta = 2\int_{\pi/2}^{\pi}\sin^{2k}\theta\,\mathrm{d}\theta = 2I_{2k}.$$

Odtiaľ plynie rovnosť

$$\frac{1}{\pi}\int_0^{\pi}\cos(x\cos\theta)\,\mathrm{d}\theta = J_0(x)$$

a tým je tvrdenie dokázané.

□

# 2. Testovanie exponenciality

V kapitole 2 sa budeme venovať rôznym testovým štatistikám, ktoré sa používajú na testovanie hypotéz, či daný náhodný výber pochádza z triedy exponenciálnych rozdelení s jednorozmerným parametrom s hustotou ako v Definícii 1.

**Definícia 14** (Štatistika a štatistický test)**.** *Nech* $\boldsymbol{X} = X_1,...,X_n$ *je náhodný výber,* $S(\boldsymbol{X})$ *je reálna funkcia dát a* $C \subset \mathbb{R}$*. Štatistický test testujúci nulovú hypotézu* $H_0$ *proti alternatíve* $H_1$ *je definovaný pomocou testovej štatistiky* $S(\boldsymbol{X})$ *a kritického oboru* $C$*, pre ktoré platí:*
•$S(\boldsymbol{X}) \in C \Rightarrow H_0$ *zamietame v prospech* $H_1$,
•$S(\boldsymbol{X}) \notin C \Rightarrow H_0$ *nemôžme zamietnuť v prospech* $H_1$.

Testy uvedené v tejto kapitole, sú testami dobrej zhody (goodness-of-fit tests). Sú to testy, ktoré skúmajú, ako dobre sedí hypotetické rozdelenie (v našom prípade exponenciálne) na dáta v danom náhodnom výbere. Prvými testami tohto typu boli $\chi^2$ testy dobrej zhody, ktoré boli publikované už v roku 1900. Ďalšie testy dobrej zhody využívajú rôzne charakterizácie exponenciálneho rozdelenia na zostavenie testových štatistík, napríklad charakterizácie pomocou empirickej distribučnej funkcie (Kolmogorovov-Smirnovov test alebo Cramér-von Misés test)alebo novšie charakterizácie pomocou Hankelovej transformácie alebo kumulatívnej reziduálnej entropie.

## 2.1 $\chi^2$ testy dobrej zhody

Svoj súhrnný názov dostali $\chi^2$ testy dobrej zhody podľa asymptotického rozdelenia ich testových štatistík. Úplným priekopníkom v tejto oblasti bol Karl Pearson, ktorý v roku 1900 publikoval svoju slávnu $\chi^2$ štatistiku a položil základ v oblasti testovania vhodnosti použitia hypotetického modelu na pozorované dáta na ďalšie storočie.

Od ostatných testov dobrej zhody sa $\chi^2$ testy dobrej zhody líšia tým, že netestujú priamo či rozdelenie, z ktorého náhodný výber pochádza, je exponenciálne. Sú to testy, ktoré testujú hodnotu parametra $\boldsymbol{p}$ multinomického rozdelenia $Mult_K(N,\boldsymbol{p})$.

Nech $\boldsymbol{n} = (n_1,...,n_K)^T \sim Mult_K(N,\boldsymbol{p})$. Chceli by sme testovať hypotézu, či sa vektor pravdepodobností jednotlivých kategórii $\boldsymbol{p} = (p_1,...,p_K)^T$ rovná nejakému konkrétnemu vektoru pravdepodobností $\boldsymbol{p_0} = (p_{1,0},...,p_{K,0})^T, \sum_{i=1}^K p_{i,0} = 1$, t.j.

$$H_0 : \boldsymbol{p} = \boldsymbol{p_0} \text{ proti } H_1 : \boldsymbol{p} \neq \boldsymbol{p_0}.$$

**Pearsonova $\chi^2$ štatistika** Pearson v prelomovom článku (Pearson, 1900) navrhol použiť nasledovnú štatistiku:

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i - e_i)^2}{e_i},$$

kde $e_i = Np_{i,0}$ označuje očakávané četnosti v kategórii $i$ za platnosti nulovej hypotézy. Príliš veľká hodnota testovej štatistiky znamená príliš veľký rozdiel

pozorovaných a očakaváných hodnôt, čo naznačuje neplatnosť nulovej hypotézy. Ako názov napovedá, $\chi^2$ má za $H_0$ asymptoticky $\chi^2_{K-1}$ rozdelenie, čo dokážem v nasledujúcej vete.

**Veta 4.** *Pre Pearsonovu $\chi^2$ štatistiku za platnosti $H_0$ platí pre $N \to \infty$*

$$\chi^2 \xrightarrow{D} \chi^2_{K-1}.$$

*Dôkaz.* Testovú štatistiku môžeme prepísať do nasledovného tvaru

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i - Np_{i,0})^2}{Np_{i,0}} = \boldsymbol{Y}_N^T \boldsymbol{Y}_N =$$
$$= \left( \frac{(n_1 - Np_{1,0})}{\sqrt{Np_{1,0}}}, ..., \frac{(n_K - Np_{K,0})}{\sqrt{Np_{K,0}}} \right)^T \cdot \left( \frac{(n_1 - Np_{1,0})}{\sqrt{Np_{1,0}}}, ..., \frac{(n_K - Np_{K,0})}{\sqrt{Np_{K,0}}} \right).$$

Pre náhodný vektor $\boldsymbol{n} = (n_1,...,n_K)^T \sim Mult_K(N,\boldsymbol{p})$ za platnosti $H_0$ platí $\boldsymbol{p} = \boldsymbol{p_0}$. Náhodný vektor $\boldsymbol{n}$, ktorý predstavuje $N$ nezávislých náhodných pokusov, môžeme zapísať ako súčet $N$ nezávislých náhodných vektorov $\boldsymbol{m_j}$, $j = 1,...,N$, predstavujúcich jednotlivé pokusy, pre ktoré platí $\boldsymbol{m_j} \sim Mult_K(1,\boldsymbol{p_0})$. Z vlastností multinomického rozdelenia vieme, že pre každé $j = 1,...,N$ platí $\mathsf{E}\ \boldsymbol{m_j} = \boldsymbol{p_0}$ a $\mathsf{var}\, \boldsymbol{m_j} = diag(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^T$, kde $diag(\boldsymbol{p_0})$ predstavuje diagonálnu maticu s prvkami vektoru $\boldsymbol{p_0}$ na diagonále. Z centrálnej limitnej vety dostávame pre $N \to \infty$

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{N} (\boldsymbol{m_j} - \mathsf{E}\ \boldsymbol{m_j}) \xrightarrow{D} N_K(0, \mathsf{var}\, \boldsymbol{m_j}),$$
$$\frac{1}{\sqrt{N}} (\boldsymbol{n} - N\boldsymbol{p_0}) \xrightarrow{D} N_K(0, diag(\boldsymbol{p_0}) - \boldsymbol{p_0}\boldsymbol{p_0}^T).$$

Označme $\sqrt{\boldsymbol{p_0}}$ vektor s prvkami $\sqrt{p_{i,0}}$, potom variančnú maticu $\mathsf{var}\, \boldsymbol{m_j}$ môžeme upraviť do tvaru

$$\mathsf{var}\, \boldsymbol{m_j} = diag(\sqrt{\boldsymbol{p_0}})[I - (\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T] diag(\sqrt{\boldsymbol{p_0}}).$$

Keďže diagonálna matica je rovná svojej transpozícii, môžeme ďalej písať

$$\frac{1}{\sqrt{N}} diag(\sqrt{\boldsymbol{p_0}})^{-1} (\boldsymbol{n} - N\boldsymbol{p_0}) \xrightarrow{D} N_K(0, I - (\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T).$$

Inverz diagonálnej matice je diagonálna matica s inverznými prvkami na diagonále, v našom prípade s prvkami $1/\sqrt{p_{i,0}}$, preto náhodný vektor, ktorého asymptotické rozdelenie sme odvodili vyššie, je presne vektor $\boldsymbol{Y}_N$ s prvkami

$$Y_{i,N} = \frac{1}{\sqrt{Np_{i,0}}} (n_i - Np_{i,0}).$$

Matica $\Sigma = [I - (\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T]$ je idempotentná, lebo

$$\begin{aligned} \Sigma\Sigma &= [I - (\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T][I - (\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T] \\ &= I - 2(\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T + (\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T(\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T \\ &= I - 2(\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T + \sum_{i=1}^{K} p_{i,0}(\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T \\ &= [I - (\sqrt{\boldsymbol{p_0}})(\sqrt{\boldsymbol{p_0}})^T] = \Sigma. \end{aligned}$$

Odvodili sme, že vektor $\boldsymbol{Y}_N$ má asymptoticky rozdelenie $N_K(0,\Sigma)$, kde $\Sigma$ je idempotentná matica. Preto pre $N \to \infty$ platí

$$\chi^2 = \boldsymbol{Y}_N^T \boldsymbol{Y}_N \xrightarrow{D} \chi^2_{\mathrm{tr}},$$

kde stopa matice $\mathrm{tr}\Sigma = \sum_{i=1}^{K}(1 - p_{i,0}) = K - \sum_{i=1}^{K} = K - 1$. Tým je tvrdenie vety dokázané.

□

Veta 4 nám dáva prostriedok na určenie kritického oboru testu založeného na $\chi^2$ štatistike. Nulovú hypotézu zamietame pre hodnoty testovej štatistiky $\chi^2$ väčšie ako $(1-\alpha)$- kvantil $\chi^2_{K-1}$ rozdelenia.

Z vety 4 vieme, aké asymptotické rozdelenie má $\chi^2$ štatistika v prípade, že je hypotetické rozdelenie, ktoré testujeme presne dané. V prípade, že chceme testovať či náhodný výber pochádza z určitej parametrickej triedy rozdelení s neznámym $q$-rozmerným reálnym parametrom, $0 \leq q < K - 1$, tvrdenie vety 4 použiť nemôžeme. Chceme testovať hypotézu

$$H_0' : \boldsymbol{p} = \boldsymbol{p_0}(\boldsymbol{\theta}),\ (\boldsymbol{\theta}) \in \Theta \subset \mathbb{R}^q \text{ proti } H_1' : \boldsymbol{p} \neq \boldsymbol{p_0}(\boldsymbol{\theta}).$$

V takom prípade asymptotické rozdelenie $\chi^2$ štatistiky závisí na počte neznámych zložiek parametru. Aby sme mohli $\chi^2$ test dobrej zhody previesť, musíme neznámy parameter $\boldsymbol{\theta}$ aproximovať nejakým jeho odhadom. Ak za splnených podmienok regularity odhadneme neznámy parameter jeho maximálne vierohodným odhadom, $\chi^2$ štatistika má asymptoticky $\chi^2_{K-q-1}$ rozdelenie (Birch, 1964). Na základe toho, budeme nulovú hypotézu $H_0'$ zamietať pre hodnoty $\chi^2$ štatistiky väčšie ako $(1-\alpha)$-kvantil $\chi^2_{K-q-1}$ rozdelenia.

**Vierohodnostný pomer/ G-test** Vierohodnostná funkcia určuje pravdepodobnosť vyskytnutia sa pozorovaných hodnôt pri danej hodnote neznámeho parametra. Maximálne vierohodný odhad parametra túto pravdepodobnosť maximalizuje. Maximálna vierohodnosť za platnosti nulovej hypotézy musí byť menšia alebo rovná maximálnej vierohodnosti v prípade, že vektor pravdepodobností $\boldsymbol{p}$ môže mať ľubovoľnú hodnotu (Simonoff, 2003). Maximálne vierohodným odhadom parametra $\boldsymbol{p}$ je $\hat{\boldsymbol{p}}$ s hodnotami $\hat{p}_i = n_i/N$). Preto *vierohodnostný pomer (likelihood ratio)* je definovaný ako

$$\Lambda = \frac{\text{maximálna vierohodnosť za } H_0}{\text{maximálna vierohodnosť pri ľubovoľnom } \boldsymbol{p}}.$$

V našom prípade má tvar

$$\Lambda = \frac{\frac{N!}{n_1!\ldots n_K!}\prod_{i=1}^{K} p_{i,0}^{n_i}}{\frac{N!}{n_1!\ldots n_K!}\prod_{i=1}^{K}\left(\frac{n_i}{N}\right)^{n_i}} = \prod_{i=1}^{K}\left(\frac{e_i}{n_i}\right)^{n_i}.$$

Za platnosti nulovej hypotézy by mal byť maximálne vierohodný odhad $\hat{\boldsymbol{p}}$ vektora $\boldsymbol{p}$ „blízko"$\boldsymbol{p_0}$ a vierohodnostný pomer $\Lambda$ by nemal byť „príliš menší"ako 1 (Simonoff, 2003).

*Štatistika $G^2$ (loglikelihood statistics)*, ktorá je definovaná pomocou $\Lambda$, má tvar

$$G^2 = -2\log\Lambda = 2\sum_{i=1}^{K} n_i \log\left(\frac{e_i}{n_i}\right).$$

Rovnako ako Pearsonova štatistika nadobúda táto štatistika svoje minimum rovné 0 práve vtedy, keď $n_i = e_i$ pre všetky $i = 1,...,K$ a veľké hodnoty testovej štatistiky naznačujú neplatnosť nulovej hypotézy. Jej asymptotické rozdelenie určíme v ďalšom odseku.

**Štatistiky mocninnej divergencie** Obe vyššie predstavené štatistiky, $\chi^2$ a $G^2$, patria do obecnej triedy štatistík nazývaných *štatistiky mocninnej divergencie (power divergence statistics)*, ktoré majú asymptoticky $\chi^2_{K-1}$ rozdelenie. Túto triedu štatistík Cressie a Read (1984) definovali predpisom

$$2nI^\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{K} n_i \left[ \left( \frac{n_i}{e_i} \right)^\lambda - 1 \right], \text{pre } \lambda \in \mathbb{R}.$$

Aj keď $2nI^\lambda$ nie je definovaná pre $\lambda = 0$ a $\lambda = -1$, v týchto bodoch je spojito dodefinovaná limitami pre $\lambda \to 0$ a $\lambda \to -1$. Testové štatistiky $\chi^2$ a $G^2$ sú špeciálnymi prípadmi $2nI^\lambda$ pre $\lambda = 1$ a $\lambda = 0$, pretože platí

$$\begin{aligned} 2nI^1 &= \sum_{i=1}^{K} n_i \left[ \frac{n_i}{e_i} - 1 \right] = \sum_{i=1}^{K} \frac{n_i^2 - n_i e_i}{e_i} = \sum_{i=1}^{K} \left[ \frac{(n_i - e_i)^2}{e_i} + \frac{n_i e_i - e_i^2}{e_i} \right] \\ &= \chi^2 + \sum_{i=1}^{K} n_i - N \sum_{i=1}^{K} p_{i,0} = \chi^2 + N - N = \chi^2 \\ 2nI^0 &= \lim_{\lambda \to 0} 2nI^\lambda = \lim_{\lambda \to 0} \frac{2}{\lambda+1} \sum_{i=1}^{K} n_i \left[ \frac{\left(\frac{n_i}{e_i}\right)^\lambda - 1}{\lambda} \right] = 2 \sum_{i=1}^{K} n_i \log \left( \frac{n_i}{e_i} \right) = G^2 \end{aligned}$$

Všetky štatistiky z triedy $2nI^\lambda$ majú za platnosti nulovej hypotézy asymptoticky pre $n \to \infty$ rozdelenie $\chi^2_{K-1}$ (Cressie a Read, 1984). Preto nulovú hypotézu zamietame pre hodnoty $2nI^\lambda$ väčšie ako $(1-\alpha)$- kvantil rozdelenia $\chi^2_{K-1}$.

V prípade, že vektor pravdepodobností $\boldsymbol{p_0}$, nie je presne známy, ale závisí na $q$-rozmernom vektore parametrov $\boldsymbol{\theta}$, nahradzujeme očakávané četnosti $e_i$ ich odhadmi $\hat{e}_i$, ktoré sú založené na odhade $\hat{\boldsymbol{\theta}}$ vektora parametrov $\boldsymbol{\theta}$. Toto ovplyvňuje asymptotické rozdelenie štatistík $2nI^\lambda$ za $H_0$. Za predpokladu splnenia podmienok regularity a použitím odhadu $\boldsymbol{\theta}$, ktorý je najlepší asymptoticky normálny má $\chi^2$ štatistika pre $N \to \infty$ rozdelenie $\chi^2_{K-q-1}$. Medzi najlepšie asymptoticky normálne odhady parametrov patria aj maximálne vierohodné odhady (Birch, 1964), preto ich použitie v $\chi^2$ testoch dobrej zhody, založených na $2nI^\lambda$ štatistikách zachová asymptoticky $\chi^2_{K-q-1}$ rozdelenie. V tomto prípade hypotézu zamietame pre $2nI^\lambda$ väčšie ako $(1-\alpha)$- kvantil $\chi^2_{K-q-1}$ rozdelenia.

Zaujímavou otázkou je, ktorý test z tejto triedy je vhodné si vybrať. Cressie a Read (1984) odporúčajú používať štatistiku s $\lambda = 2/3$ v prípade, že $N \geq 10$ a $\min_i e_i \geq 1$. Ďalšou otázkou je, kedy je vhodné $\chi^2$ testy používať. Na túto otázku neexistuje presná odpoveď, ale časté doporučenie je, používať $\chi^2$ testy dobrej zhody v prípadoch, keď väčšina očakávaných četností v kategóriach $e_i$ má hodnotu aspoň 3 a všetky majú hodnotu aspoň 1 (Simonoff, 2003).

**Testovanie exponenciality so známym parametrom** Majme náhodný výber $X_1, \ldots, X_n$ z rozdelenia $F_X$. Chceme testovať hypotézu

$$\begin{aligned} &H_0 : X_1, \ldots, X_n \text{ je z } Exp(1/\lambda_0), \lambda_0 > 0 \text{ známe, proti} \\ &H_1 : X_1, \ldots, X_n \text{ nie je z } Exp(1/\lambda_0). \end{aligned}$$

Na prevedenie niektorého z $\chi^2$ testov dobrej zhody popísaných vyššie musíme vytvoriť z nášho náhodného výberu náhodný vektor z $Mult_K(n,\boldsymbol{p})$. Najprv rozdelíme reálnu os na $K$ disjunktných intervalov $(a_{k-1},a_k], k = 0,...,K$, $-\infty = a_0 < ... < a_K = \infty$, a vytvoríme náhodný vektor $\boldsymbol{n} = (n_1,...,n_K)^T$ z K-rozmerného multinomického rozdelenia predpisom:

$$n_i = \sum_{j=1}^{n} \mathbf{1}_{\{X_j \in (a_{k-1},a_k]\}}. \tag{2.1}$$

Četnosť $n_i$ určuje počet pozorovaní z náhodného výberu $X_1,...,X_n$, ktoré padli do $i$-tého intervalu. Aby sme mohli použiť jeden z $\chi^2$ testov, zostáva určiť vektor pravdepodobností jednotlivých kategórii za nulovej hypotézy. Ten jednoducho odvodíme z distribučnej funkcie exponenciálneho rozdelenia, ktoré testujeme. Označíme $\boldsymbol{p_0} = (p_{1,0},...,p_{K,0})^T$, kde $p_{i,0} = F(a_i) - F(a_{i-1}) = \lambda_0(e^{-\lambda_0 a_i} - e^{-\lambda_0 a_{i-1}}), i = 1,...,K$.

Prevedieme niektorý z vyššie popísaných $\chi^2$ testov dobrej zhody, na testovanie hypotézy či vektor $\boldsymbol{n} \sim Mult_K(n,\boldsymbol{p})$ pochádza z $Mult_K(n,\boldsymbol{p_0})$. Ako sme diskutovali vyššie, testová štatistika bude mať za platnosti nulovej hypotézy asymptoticky $\chi^2_{K-1}$ rozdelenie.

**Testovanie exponenciality s neznámym parametrom** V tomto prípade máme náhodný výber $X_1,...,X_n$ z rozdelenia $F_X$ a chceme testovať hypotézu

$$H_0 : X_1,...,X_n \text{ je z } Exp(1/\lambda), \lambda > 0, \text{proti}$$
$$H_1 : X_1,...,X_n \text{ nie je z } Exp(1/\lambda).$$

Vektor $\boldsymbol{n}$, vytvoríme rovnako ako v prípade so známym parametrom (podľa vzorca 2.1). Nevieme ale presne hodnotu parametra, ktorý by malo mať exponenciálne rozdelenie za platnosti nulovej hypotézy, preto nevieme určiť presne vektor $\boldsymbol{p_0}$. Strednú hodnotu $1/\lambda$ odhadneme výberovým priemerom $\overline{X}_n$, ktorý je maximálne vierohodným odhadom (a teda aj najlepším asymptoticky normálnym odhadom (Birch, 1964)) a dostaneme odhad vektora $\boldsymbol{p_0}$

$$\boldsymbol{p_0}(\hat{\lambda}) = (p_{1,0}(\hat{\lambda}),...,p_{K,0}(\hat{\lambda}))^T,$$

kde $p_{i,0}(\hat{\lambda}) = \frac{1}{\overline{X}_n}\left(e^{-\frac{a_i}{\overline{X}_n}} - e^{-\frac{a_{i-1}}{\overline{X}_n}}\right), i = 1,...,K.$

Budeme testovať pomocou jedného z $\chi^2$ testov dobrej zhody hypotézu, či vektor $\boldsymbol{n} \sim Mult_K(n,\boldsymbol{p})$ je z $Mult_K(n,\boldsymbol{p_0}(\hat{\lambda}))$. Testová štatistika bude mať za platnosti nulovej hypotézy asymptoticky $\chi^2_{K-2}$ rozdelenie, keďže sme odhadovali jeden neznámy parameter maximálne vierohodným odhadom (Birch, 1964).

## 2.2 Testy založené na empirickej distribučnej funkcii

Testy, ktoré porovnávajú emppririckú distribučnú funkciu náhodného výberu s hypotetickou distribučnou funkciou za platnosti nulovej hypotézy patria medzi klasické metódy testovania exponenciality. Medzi najznámejšie patria Kolmogorovov - Smirnovov test (KS test) a Cramérov-von Misesov test (CVM test) a ich modifikácie.

**Klasický Kolmogorovov-Smirnovov test** Kolmogorov vo svojom slávnom a prelomovom článku z roku 1933 formálne definoval empirickú distribučnú funkciu a skúmal ako dobre aproximuje $F_n(x)$ skutočnú distribučnú funkciu $F(x)$ pre veľké hodnoty $n$ (Stephens, 1992). Za účelom porovnania $F_n(x)$ a $F(x)$ definoval Kolmogorov (1933) štatistiku

$$D_n = \sup_x |F_n(x) - F(x)|.$$

Kolmogorov (1933) taktiež odvodil nasledovné asymptotické rozdelenie štatistiky $D_n$, ktoré nazývame *Kolmogorovovo rozdelenie.*

**Veta 5.** *Nech $F(x)$ je spojitá distribučná funkcia. Potom pre každé pevné $z \geq 0$ platí*

$$P(\sqrt{n}D_n \leq z) \overset{n\to\infty}{\longrightarrow} 1 - 2\sum_{i=1}^{\infty}(-1)^{(i-1)}e^{-2i^2z^2} = \frac{\sqrt{2\pi}}{z}\sum_{i=1}^{\infty}e^{-(2i-1)\pi^2/(8z^2)}.$$

*Dôkaz.* Táto veta bola dokázaná v (Kolmogorov, 1933). Alternatívny dôkaz bol publikovaný v (Feller, 1948).

□

Okrem toho, že Kolmogorov zaviedol veličinu $D_n$ a skúmal jej asymptotické vlastnosti, sám sa v článku z roku 1933 nevenoval aplikácii v testoch dobrej zhody. Kolmogorovovu prácu o niekoľko rokov rozšíril a doplnil napríklad o dvojvýberové testy Smirnov (1939a,b).

Kolmogorovov–Smirnovov test bol navrhnutý na testovanie hypotézy, či distribučná funkcia $F_X$ spojitého rozdelenia, z ktorého pochádza náhodný výber $X_1, ..., X_n$ je rovná nejakej konkrétnej vopred danej distribučnej funkcii $F$. Z vety 5 plynie, že testová štatistika $D_n$ má za platnosti nulovej hypotézy asymptoticky Kolmogorovovo rozdelenie. V našom prípade testovania exponenciality teda testujeme

$$H_0 : F_X(x) = 1 - e^{-\lambda_0 x}, \text{ pre } x \in \mathbb{R}, \lambda_0 > 0, \text{známe, proti}$$
$$H_1 : \exists\, x \in \mathbb{R} : F_X(x) \neq 1 - e^{-\lambda_0 x}.$$

Nulovú hypotézu zamietneme pre tie hodnoty testovej štatistiky $D_n$, ktoré sú väčšie ako $(1-\alpha)$-kvantil Kolmogorovovho rozdelenia. Tabuľky približných kritických hodnôt pre rôzne rozsahy náhodného výberu a rôzne hladiny testu publikoval Smirnov (1948).

**Lillieforsov test KS typu pri neznámom parametri** Kolmogorovov–Smirnovov test, testuje hypotézu o presnom exponenciálnom rozdelení a nedá sa použiť na situácie, kedy presná hodnota parametra $\lambda$ nie je známa. Lilliefors (1969) modifikoval tento test práve na prípady, kedy chceme testovať, či náhodný výber pochádza z parametrickej triedy exponeciálnych rozdelení s jednorozmerným parametrom. Lillieforsov test testuje hypotézu

$$H_0 : F_X(x) = 1 - e^{-\lambda x}, \text{ pre } x \in \mathbb{R}, \lambda > 0 \text{ neznáme, proti}$$
$$H_1 : \exists\, x \in \mathbb{R} : F_X(x) \neq 1 - e^{-\lambda x}.$$

Lilliefors navrhol upraviť štatistiku $D_n$ odhadnutím strednej hodnoty $1/\lambda$ výberovým priemerom $\overline{X}_n$ do tvaru

$$D_n^* = \max_{x \geq 0} \left| F_n(x) - \left(1 - e^{-\frac{x}{\overline{X}_n}}\right) \right|.$$

Nulovú hypotézu zamietame pre príliš veľké hodnoty testovej štatistiky $D*_n$. Približné kritického hodnoty pre rôzne hodnoty $n$ a $\alpha$, získané Monte Carlo simuláciami, publikoval Lilliefors (1969).

**Finkelsteinov-Schaferov test KS typu** Finkelstein a Schafer v roku 1971 prestavili ďalšiu modifikáciu klasického KS testu, ktorá sa v ich simulačnej štúdii ukázala byť silnejšia ako klasický KS test aj Lillieforsov test (Finkelstein a Schafer, 1971). Ich štatistika má dve verzie – na testovanie hypotézy o presnom exponenciálnom rozdelení a hypotézy o exponenciálnom rozdelení s neznámou strednou hodnotou. V prípade známeho parametra testujeme

$$H_0 : F_X(x) = 1 - e^{-\lambda_0 x}, prex \in \mathbb{R}, \lambda_0 > 0 \text{ známe, proti}$$
$$H_1 : \exists\, x \in \mathbb{R} : F_X(x) \neq 1 - e^{-\lambda_0 x}.$$

Testová štatistika má v tomto prípade tvar

$$S_n = \sum_{i=1}^{n} \max\left( \left| F(X_{(i)}) - \frac{i-1}{n} \right|, \left| F(X_{(i)}) - \frac{i}{n} \right| \right).$$

V prípade neznámeho parametra testujeme

$$H_0 : F_X(x) = 1 - e^{-\lambda x}, prex \in \mathbb{R}, \lambda > 0, \text{ proti}$$
$$H_1 : \exists\, x \in \mathbb{R} : F_X(x) \neq 1 - e^{-\lambda x}.$$

Testová štatistika má potom tvar

$$S_n^* = \sum_{i=1}^{n} \max\left( \left| F^*(X_{(i)}) - \frac{i-1}{n} \right|, \left| F^*(X_{(i)}) - \frac{i}{n} \right| \right),$$

kde $F^*(x) = 1 - e^{-\frac{x}{\overline{X}_n}}$.

Finkelstein a Schafer (1971) tiež publikovali tabuľky približných kritických hodnôt $S_n$ a $S_n^*$, ktoré získali pomocou Monte Carlo simulácii.

**Cramérov-von Misesov test** Cramer (1928) navrhol testovať hypotézu, že náhodný výber pochádza z konkrétneho rozdelenia s distribučnou funkciou $F(x)$ pomocou testovej štatistiky

$$\int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, \mathrm{d}K(x),$$

kde $K(x)$ je nejaká vhodná neklesajúca váhová funkcia. Von Mises (1931) nezávisle prišiel s ekvivalentným návrhom a odvodil niekoľko vlastností testu (Darling, 1957). Von Mises navrhol použiť nasledovnú štatistiku s váhovou funkciou $\psi(x)$.

$$\omega_n^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi(x) \, \mathrm{d}x$$

Ďalšiu modifikáciu tejto testovej štatistiky o niekoľko rokov neskôr zaviedol Smirnov (1936), ktorý navrhol použiť

$$\omega_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi[F(x)] \, \mathrm{d}F(x).$$

Smirnovova modifikácia $\omega_n^2$ s voľbou váhovej funkcie $\psi(x) = 1$, sa stala štatistikou $W^2$, ktorú dnes najčastejšie poznáme pod pojmom Cramérova-von Misesova štatistika a má tvar

$$W^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \, \mathrm{d}F(x).$$

**Lemma 6.** *Pre štatistiku $W^2$ platí*

$$W^2 = \sum_{i=1}^{n} \left[ F(X_{(i)}) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}.$$

*Dôkaz.* Anderson a Darling (1952) odvodili pomocou substitúcie a vlastností empirickej distribučnej funkcie, že

$$W^2 = 2 \sum_{i=1}^{n} \left[ \int_0^{F(X_{(i)})} t \, \mathrm{d}t - \frac{2i-1}{2n} \int_0^{F(X_{(i)})} 1 \, \mathrm{d}t \right] + n \int_0^1 (1-t)^2 \, \mathrm{d}t.$$

Ďalšími úpravami dostávame

$$\begin{aligned}
W^2 &= \sum_{i=1}^{n} \left[ \frac{F^2(X_{(i)})}{2} - 2\frac{2i-1}{2n} F(X_{(i)}) \pm \left( \frac{2i-1}{2n} \right)^2 \right] + n \left[ t - \frac{t^2}{2} \right]_0^1 \\
&= \sum_{i=1}^{n} \left[ F(X_{(i)}) - \frac{2i-1}{2n} \right]^2 - \frac{1}{4n^2} \sum_{i=1}^{n} (2i-1)^2 + \frac{n}{3} \\
&= \sum_{i=1}^{n} \left[ F(X_{(i)}) - \frac{2i-1}{2n} \right]^2 - \frac{1}{4n^2} (4n^2 - 1) + \frac{n}{3} \\
&= \sum_{i=1}^{n} \left[ F(X_{(i)}) - \frac{2i-1}{2n} \right]^2 + \frac{-4n^2 + 1 + 4n}{12} \\
&= \sum_{i=1}^{n} \left[ F(X_{(i)}) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12}.
\end{aligned}$$

□

Nulovú hypotézu zamietame pre veľké hodnoty testovej štatistiky, ktoré naznačujú, že empirická a hypotetická distribučná funkcia sa príliš líšia. Asymptotické rozdelenie testovej štatistiky diskutujú napríklad Anderson a Darling (1952). Dosadením distribučnej funkcie a hustoty podľa nami testovanej hypotézy o presnom exponenciálnom rozdelení dostávame

$$W^2 = \sum_{i=1}^{n} \left[ 1 - e^{-\lambda_0 X_{(i)}} - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}.$$

**Van Soestov test typu CVM** Van Soest (1969) analogickým spôsobom ako Finkelstein a Schafer (1971) rozšíril testovanie pomocou štatistiky $W^2$ na hypotézy o exponenciálnom rozdelení s neznámym parametrom odhadnutím strednej hodnoty $1/\lambda$ výberovým priemerom $\overline{X}_n$. Van Soestova $W^{2*}$ štatistika má tvar

$$W^{2*} = \sum_{i=1}^{n} \left[ F^*(X_{(i)}) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n},$$

kde $F^*(x) = 1 - e^{-\frac{x}{\overline{X}_n}}$.

## 2.3 Testy založené na vlastnosti bez pamäti

Ako sme dokázali vo vete 1, exponenciálne rozdelenie má unikátnu vlastnosť medzi spojitými rozdeleniami, že je bez pamäti. Ako prvý publikoval test exponenciality založený na vlastnosti bez pamäti (Angus, 1982). Predstavil dve testové štatistiky - jednu typu KS, druhú typu CVM. (Ahmad a Alwasel, 1999) publikovali lepší test exponenciality využívajúci túto charakterizáciu exponenciálneho rozdelenia. Zaviedli štatistiku, ktorá má za platnosti nulovej hypotézy normálne rozdelenie (na rozdiel od štatistík typu KS alebo CVM) a relatívne veľkú silu (v porovnaní s vtedy známymi asymptoticky normálnymi štatistikami založenými na Giniho indexe). Pomocou vlastnosti bez pamäti možno exponenciálne rozdelenie charakterizovať nasledovne.

**Veta 7.** *Nech $X$ je kladná spojitá náhodná veličina s distribučnou funkciou $F$. Potom $X \sim Exp(1/\lambda)$ práve vtedy keď $\bar{F}(2x) = \bar{F}^2(x)$ pre všetky $x \geq 0$.*

*Dôkaz.* Z vety 1 plynie

$$\bar{F}(2x) = \mathsf{P}(X > 2x) = \mathsf{P}(X > x)\,\mathsf{P}(X > x) = \bar{F}^2(x).$$

□

Použitím tejto charakterizácie môžeme definovať funkciu

$$\Delta_2(F) = \int_0^\infty [\bar{F}(2x) - \bar{F}^2(x)]^2 \,\mathrm{d}F(x). \tag{2.2}$$

Túto funkciu môžme ďalej upraviť do tvaru

$$\begin{aligned}\Delta_2(F) &= \int_0^\infty \bar{F}^2(2x)\,\mathrm{d}F(x) - 2\int_0^\infty \bar{F}(2x)\bar{F}^2(x)\,\mathrm{d}F(x) + \int_0^\infty \bar{F}(x)\,\mathrm{d}F(x) \\ &= \int_0^\infty \bar{F}^2(2x)\,\mathrm{d}F(x) - 2\int_0^\infty \bar{F}(2x)\bar{F}^2(x)\,\mathrm{d}F(x) + \frac{1}{5}.\end{aligned}$$

Z vety 7 vidíme, že pre exponenciálne rozdelenie je hodnota $\Delta_2(F)$ nulová, čo Ahmad a Alwasel využili k testovaniu. Miesto použitia odhadu $\Delta_2(F)$ spôsobom Cramér-von Mises, t.j. $\tilde{\Delta}_2(F) = \Delta_2(F_n)$, ktorý nemá za nulovej hypotézy asymptoticky normálne rozdelenie (Shorack a Wellner, 1986), zvolili autori iný postup. Navrhli odhad $\Delta_2(F)$, ktorý je asymptoticky normálny ako za platnosti

nulovej hypotézy, tak aj za platnosti alternatívy a odvodili jeho rozptyl. Navrhovaný odhad je založený na nasledujúcom odhade $F(x)$.

$$F_{n,\gamma}(x) = \frac{1}{n}\sum_{i=1}^{n} C_{i,n}(\gamma)\mathbf{1}_{\{x\leq X_i\}}, \tag{2.3}$$

kde $\{C_{i,n}(\gamma)\}_{i=1}^{n}, n \geq 1$, je trojrozmerné pole reálnych čísel závisejúce na parametri $\gamma, \gamma \in (0,1]$, splňujúce pre $n \to \infty$ a pre $\gamma \in (0,1]$

$$\frac{1}{n}\sum_{i=1}^{n} C_{i,n}(\gamma) \to 1,$$
$$\frac{1}{n}\sum_{i=1}^{n} C_{i,n}^2(\gamma) \to C^2(\gamma) > 1.$$

Aj keď $C_{i,n}(\gamma)$ môžme voliť ľubovoľne za splnenia daných podmienok, autori navrhujú jednoduchú voľbu pre $\gamma \in (0,1]$

$$C_{i,n}(\gamma) = \begin{cases} 1-\gamma & \text{ak je } i \text{ párne,} \\ 1+\gamma & \text{ak je } i \text{ nepárne.} \end{cases} \tag{2.4}$$

**Lemma 8.** *Pre* $\{C_{i,n}(\gamma)\}_{i=1}^{n}, n \geq 1$ *definované ako* (2.4) *platí*

$$C^2(\gamma) = 1 + \gamma^2 > 1.$$

*Dôkaz.* Pre $n$ párne platí

$$\begin{aligned} C^2(\gamma) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} C_{i,n}^2(\gamma) = \lim_{n\to\infty}\frac{1}{n}\left[\frac{n}{2}(1+\gamma)^2 + \frac{n}{2}(1-\gamma)^2\right] \\ &= \frac{1}{2}(1+2\gamma+\gamma^2+1-2\gamma+\gamma^2) = (1+\gamma^2). \end{aligned}$$

Pre $n$ nepárne platí

$$\begin{aligned} C^2(\gamma) &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} C_{i,n}^2(\gamma) = \lim_{n\to\infty}\frac{1}{n}\left[\frac{n+1}{2}(1+\gamma)^2 + \frac{n-11}{2}(1-\gamma)^2\right] \\ &= \lim_{n\to\infty}\frac{1}{2n}[(n+1)(1+2\gamma+\gamma^2) + (n-1)(1-2\gamma+\gamma^2)] \\ &= \lim_{n\to\infty}\frac{1}{n}(n+n\gamma^2+2\gamma) = (1+\gamma^2). \end{aligned}$$

□

Za odhad funkcie $\Delta_2(F)$ Ahmad a Alwasel navrhli vziať

$$\begin{aligned} \hat{\Delta}_2(F_{n,\gamma}) = \frac{1}{n}\sum_{i=1}^{n}\bar{F}_n(2X_{(i)}) - \frac{2}{n}\sum_{i=1}^{n}\bar{F}_n^2(2X_{(i)})\bar{F}_n^2(X_{(i)}) \\ +\frac{2\gamma}{n}\sum_{i=1}^{n}(-1)^i\bar{F}_n(2X_{(i)})\bar{F}_n^2(X_{(i)}) + o_p(\frac{1}{n}). \end{aligned} \tag{2.5}$$

Ahmad a Alwasel taktiež dokázali nasledujúcu vetu, ktorá nám umožňuje zostaviť testovú štatistiku, ktorá má asymptoticky štandardné normálne rozdelenie.

**Veta 9.** *Pre každé $i = 1,...,n$ označme*

$$\phi_2(X_i) = 2\int_0^{X_i/2} \bar{F}(2x)\,\mathrm{d}F(x) + \bar{F}^2(X_i) + \int_0^{\infty} \bar{F}(2x)\,\mathrm{d}F(x) + \frac{4}{5}$$
$$- 2C_{i,n}(\gamma)\left[2\int_0^{X_i} \bar{F}(x)\bar{F}(2x)\,\mathrm{d}F(x) + \frac{1}{3} - \frac{1}{3}\bar{F}^3\left(\frac{X_i}{2}\right) + \bar{F}^2(X_i)\bar{F}(2X_i)\right].$$

*Potom pre $n \to \infty$ platí*

$$\sqrt{n}(\hat{\Delta}_2(F_{n,\gamma}) - \Delta_2(F)) \xrightarrow{D} N(0,\sigma^2),$$

*kde $\sigma^2 = \mathsf{var}\left[\sum_{i=1}^n \phi_2(X_i)\right]$.*
*Špeciálne za platnosti nulovej hypotézy platí*

$$\sqrt{n}(\hat{\Delta}_2(F_{n,\gamma}) - \Delta_2(F)) \xrightarrow{D} N(0,\sigma_0^2(\gamma)),$$

*kde $\sigma_0^2(\gamma) = \frac{37}{2925}(C^2(\gamma) - 1)$.*

*Dôkaz.* Celý dôkaz aj s postupnými krokmi možno nájsť v (Ahmad a Alwasel, 1999).

□

Z Lemmy 8 plynie, že rozptyl za $H_0$ sa rovná $\sigma_0^2(\gamma) = \frac{37}{2925}\gamma^2$.

Chceme testovať hypotézu $H_0 : X_1, \ldots, X_n$ pochádza z $Exp(1/\lambda), \lambda > 0$, proti alternatíve $H_1 : X_1,...,X_n$ nepochádza z $Exp(1/\lambda)$. Na základe vety 9 zostavíme nasledujúcu testovú štatistiku pre $\gamma \in (0,1]$ pevné

$$Z_\gamma = \frac{\sqrt{n}}{\sigma_0(\gamma)}\hat{\Delta}_2(F_{n,\gamma}).$$

Nulovú hypotézu $H_0$ zamietame podľa tvrdenia vety 9 pre hodnoty testovej štatistiky $Z_\gamma > u_{1-\alpha}$, kde $u_{1-\alpha}$ je $(1-\alpha)$-kvantil rozdelenia $N(0,1)$.

Z testovej procedúry je jasné, že závisí na voľbe $\gamma \in (0,1]$. V (Ahmad a Alwasel, 1999) autori na základe experimentu v záujme dosiahnutia čo najväčšej sily testu odporúčajú voliť $\gamma \in [0.3,0.8]$. Jasnou výhodou tohto testu je znalosť asymptotického rozdelenia testovej štatistiky, ktoré je dokonca normálne so známym rozptylom.

## 2.4 Testy založené na integrálnych transformáciach

### 2.4.1 Laplaceova transformácia

Pomocou Laplaceovej transformácie náhodnej veličiny možno charakterizovať exponenciálne rozdelenie. Laplaceova transformácia veličiny s exponenciálnym rozdelením $Exp(1/\lambda)$, $\lambda > 0$, má pre $t \geq 0$ tvar

$$\mathcal{L}(t) = \mathsf{E}\left(e^{-tX}\right) = \int_0^{\infty} e^{-tx}\lambda e^{-\lambda x}\,\mathrm{d}x = \lambda\int_0^{\infty} e^{-(\lambda+t)x}\,\mathrm{d}x = \frac{\lambda}{\lambda+t}.$$

Baringhaus a Henze (1991) využili charakterizáciu pomocou diferenciálnej rovnice, ktorú spĺňuje Laplaceova transformácia náhodnej veličiny s exponenciálnym rozdelením. Uvádzajú, že diferenciálnu rovnicu

$$(\lambda + t)\mathcal{L}'(t) + \mathcal{L}(t) = 0, \quad t \geq 0,$$

s počiatočnou podmienkou $\mathcal{L}(0) = 1$, spĺňa jediná spojitá nezáporná náhodná veličina $X \sim Exp(1/\lambda)$, pretože Laplaceova transformácia charakterizuje rozdelenie nezápornej náhodnej veličiny. Toto tvrdenie samostatne dokážeme obsiahlejším spôsobom.

**Veta 10.** *Nech $X$ je nezáporná náhodná veličina s hustotou $f$. Potom Laplaceova transformácia náhodnej veličiny $X$ spĺňa diferenciálnu rovnicu*

$$(\lambda + t)\mathcal{L}'(t) + \mathcal{L}(t) = 0, \quad t \geq 0, \lambda > 0$$

*s počiatočnou podmienkou $\mathcal{L}(0) = 1$, práve vtedy, keď $X \sim Exp(1/\lambda)$.*

*Dôkaz.* Nech platí $X \sim Exp(1/\lambda), \lambda > 0$, potom Laplaceova transformácia $X$ je $\mathcal{L}(t) = \lambda/(\lambda + t)$, ako sme ukázali vyššie. Pre $\mathcal{L}(t)$ platí

$$(L)(0) = \frac{\lambda}{\lambda} = 1,$$
$$(\lambda + t)\left[ -\frac{\lambda}{(\lambda + t)^2} \right] + \frac{\lambda}{\lambda + t} = 0.$$

Tým sme ukázali prvú implikáciu. Nech naopak Laplaceova transformácia náhodnej veličiny $X$ spĺňa diferenciálnu rovnicu

$$(\lambda + t)\mathcal{L}'(t) + \mathcal{L}(t) = 0, \quad t \geq 0, \lambda > 0$$

s počiatočnou podmienkou $\mathcal{L}(0) = 1$. Jedná sa o separabilnú diferenciálnu rovnicu, ktorú môžeme upraviť ako

$$\begin{aligned}
(\lambda + t)\mathcal{L}'(t) + \mathcal{L}(t) &= 0 \\
\frac{1}{\mathcal{L}(t)}\mathcal{L}'(t) &= -\frac{1}{\lambda + t} \\
\frac{1}{\mathcal{L}(t)}\frac{\mathrm{d}\mathcal{L}(t)}{\mathrm{d}t} &= -\frac{1}{\lambda + t} \\
\int \frac{1}{\mathcal{L}(t)}\frac{\mathrm{d}\mathcal{L}(t)}{\mathrm{d}t}\,\mathrm{d}t &= -\int \frac{1}{\lambda + t}\,\mathrm{d}t \\
\log \mathcal{L}(t) &= -\log(\lambda + t) + C_1 \\
\log \mathcal{L}(t) &= \log\left(\frac{1}{\lambda + t}\right) + C_1 \\
\mathcal{L}(t) &= \frac{1}{\lambda + t}C_2,
\end{aligned}$$

kde $C_2 = e^{C_1}$. Integračnú konštantu dopočítame z počiatočnej podmienky

$$1 = \mathcal{L}(0) = \frac{1}{\lambda}C_2,$$

odkiaľ plynie, že $C_2 = \lambda$. Jednoznačným riešením tejto diferenciálnej rovnice je $\mathcal{L}(t) = \lambda/(\lambda + t).\, t \geq 0$, čo je rovné Laplaceovej transformácii veličiny s exponenciálnym rozdelením $Exp(1/\lambda)$. Keďže Laplaceova transformácia charakterizuje rozdelenie spojitej nezápornej náhodnej veličiny, náhodná veličina $X$ má exponenciálne rozdelenie $Exp(1/\lambda)$.

□

Majme náhodný výber $X_1, \ldots, X_n$ z neznámeho rozdelenia s nezáporných nosičom, potom jeho empirická Laplaceova transformácia je definovaná pre $t \geq 0$ ako

$$\mathcal{L}_n(t) = \frac{1}{n} \sum_{i=1}^{n} e^{-tX_i}.$$

Baringhaus a Henze (1991) navrhli na testovanie hypotézy

$$H_0 : X_1,...,X_n \text{ je z } Exp(1/\lambda), \lambda > 0, \text{proti}$$
$$H_1 : X_1,...,X_n \text{ nie je z } Exp(1/\lambda),$$

testovú štatistiku, ktorá má tvar váženého integrálu. Parameter exponenciálneho rozdelenia nie je známy, preto ho autori odhadli maximálne vierohodným odhadom $\hat{\lambda}_n = 1/\bar{X}_n$. Pre $a > 0$ má navrhovaná trieda testových štatistík formu

$$BH_{n,a} = n \int_0^\infty [(\hat{\lambda}_n + t)\mathcal{L}_n'(t) + \mathcal{L}_n(t)]^2 \bar{X}_n e^{-at\bar{X}_n} \, \mathrm{d}x.$$

Autori určili aj výpočetne výhodný tvar tejto testovej štatistiky, ktorý samostatne overíme v nasledujúcej vete.

**Veta 11.** *Nech $X_1, \ldots, X_n$ je náhodný výber z neznámeho rozdelenia s nezáporných nosičom a pre $i = 1, \ldots, n$ označme $Y_i = X_i/\bar{X}_n$. Pre testovú štatistiku $BH_{n,a}, a > 0$, platí*

$$BH_{n,a} = \frac{1}{n} \sum_{i,j=1}^{n} \left[ \frac{(1 - Y_i)(1 - Y_j)}{Y_i + Y_j + a} - \frac{Y_i + Y_j - 2Y_iY_j}{(Y_i + Y_j + a)^2} + \frac{2Y_iY_j}{(Y_i + Y_j + a)^3} \right].$$

*Dôkaz.* Testovú štatistiku môžeme jednoduchými upravami integrantu previesť na tvar

$$\begin{aligned} BH_{n,a} &= n \int_0^\infty [(\hat{\lambda}_n + t)\mathcal{L}_n'(t) + \mathcal{L}_n(t)]^2 \bar{X}_n e^{-at\bar{X}_n} \, \mathrm{d}x \\ &= \frac{1}{n} \int_0^\infty \left\{ \sum_{i=1}^{n} \left[ -(1 + \bar{X}_n t) \frac{X_i}{\bar{X}_n} e^{-X_i t} + e^{-X_i t} \right] \right\}^2 \bar{X}_n e^{-at\bar{X}_n} \, \mathrm{d}t \\ &= \frac{1}{n} \int_0^\infty \left\{ \sum_{i=1}^{n} e^{-X_i t} \left[ 1 - (1 + \bar{X}_n t) \frac{X_i}{\bar{X}_n}) \right] \right\}^2 \bar{X}_n e^{-at\bar{X}_n} \, \mathrm{d}t. \end{aligned}$$

Pomocou substitúcie $u = \bar{X}_n t$ a označením $Y_i = X_i/\bar{X}_n, i = 1, \ldots, n$, dostávame

$$BH_{n,a} = \frac{1}{n} \int_0^\infty \left[ \sum_{i=1}^{n} e^{-Y_i u}(1 - Y_i - uY_i) \right]^2 e^{-au} \, \mathrm{d}u.$$

Umocnením zátvorky v integrante a zámenou sumy a integrálu dostaneme

$$\begin{aligned}
BH_{n,a} &= \frac{1}{n}\int_0^\infty \sum_{i,j=1}^{n} e^{-(Y_i+Y_j)u}(1-Y_i-uY_i)(1-Y_j-uY_j)e^{-au}\,\mathrm{d}u \\
&= \frac{1}{n}\int_0^\infty \sum_{i,j=1}^{n} e^{-(Y_i+Y_j+a)u}[(1-Y_i-Y_j+Y_iY_j)+u(Y_i+Y_j-2Y_iY_j) \\
&\qquad + u^2Y_iY_j]\,\mathrm{d}u \\
&= \frac{1}{n}\sum_{i,j=1}^{n}\Bigg\{(1-Y_i-Y_j+Y_iY_j)\int_0^\infty e^{-(Y_i+Y_j+a)u}\,\mathrm{d}u \\
&\quad +(Y_i+Y_j-2Y_iY_j)\int_0^\infty ue^{-(Y_i+Y_j+a)u}\,\mathrm{d}u + Y_iY_j\int_0^\infty u^2e^{-(Y_i+Y_j+a)u}\,\mathrm{d}u\Bigg\}.
\end{aligned}$$

Integrály $\int_0^\infty ue^{-(Y_i+Y_j+a)u}\,\mathrm{d}u$ a $\int_0^\infty u^2e^{-(Y_i+Y_j+a)u}\,\mathrm{d}u$ vypočítame pomocou metódy integrácie per partes a dostaneme dokazovaný tvar štatistiky

$$BH_{n,a} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{(1-Y_i)(1-Y_j)}{(Y_i+Y_j+a)} - \frac{(Y_i+Y_j-2Y_iY_j)}{(Y_i+Y_j+a)^2} + \frac{2Y_iY_j}{(Y_i+Y_j+a)^3}\right].$$

□

Baringhaus a Henze (1991) odvodili skúmaním asymptotického správania štatistiky $BH_{n,a}$ asymptotické rozdelenie $BH_{n,a}$ za platnosti nulovej hypotézy. Na praktické použitie testu založeného na tejto štatistike autori určili približné kritické hodnoty na základe simulácii. Simulačná štúdia, ktorá porovnávala silu viacerých testov exponenciality, ukázala, že prijateľné výsledky pre úplne neznáme alternatívy majú štatistiky $BH_{n,1.5}$ a $BH_{n,2.5}$.

Henze a Meintanis (2002b) navrhli celú triedu testových štatistík, založenú na porovnaní Laplaceovej transformácie exponenciálneho rozdelenia a empirickej Laplaceovej transformácie náhodného výberu. Jednotlivé prvky triedy sa líšia voľbou váhovej funkcie. Návrh Henzeho a Meintanisa bol motivovaný pozorovaním, že v prípade, že náhodný výber $X_1,\ldots,X_n$ pochádza z exponenciálneho rozdelenia, by sa empirická Laplaceova transformácia pozorovaní $Y_1,\ldots,Y_n$, kde $Y_i = X_i/\bar{X}_n$, nemala príliš líšiť od Laplaceovej transformácie rozdelenia $Exp(1)$, ktorá je rovná $\mathcal{L}(t)1/(1+t), t \geq 0$.

Na testovanie hypotézy

$$\begin{aligned}
&H_0 : X_1,...,X_n \text{ je z } Exp(1/\lambda), \lambda > 0, \text{proti} \\
&H_1 : X_1,...,X_n \text{ nie je z } Exp(1/\lambda),
\end{aligned}$$

navrhujú Henze a Meintanis (2002b) použiť testovú štatistiku

$$W_{n,a} = n\int_0^\infty [(1+t)\mathcal{L}_n(t)-1]^2 w(t)e^{-at}\,\mathrm{d}t,$$

kde $a > 0$ a $w(t)$ je nejaká reálna spojitá váhová funkcia splňujúca rovnosť $w(t) = O(t^r), r > 0$ pre $t \to \infty$. Špeciálnym prípadom tejto testovej štatistiky je štatistika, ktorú navrhol Henze (1993), s voľbou váhovej funkcie $w(t) = (1+t)^{-2}$,

ktorá má tvar

$$HE_{n,a} = n \int_0^\infty \left[\mathcal{L}_n(t) - \frac{1}{1+t}\right]^2 e^{-at}\,\mathrm{d}t$$
$$= n \int_0^\infty [(1+t)\mathcal{L}_n(t) - 1]^2 (1+t)^{-2} e^{-at}\,\mathrm{d}t.$$

Ďalšou špeciálnou voľbou, ktorú navrhli Henze a Meintanis (2002b) použiť, je štatistika s váhovou funkciou $w(t) = (1+t)^2$, ktorá má tvar

$$HM_{n,a} = n \int_0^\infty [(1+t)\mathcal{L}_n(t) - 1]^2 e^{-at}\,\mathrm{d}t$$
$$= n \int_0^\infty \left[\mathcal{L}_n(t) - \frac{1}{1+t}\right]^2 (1+t)^2 e^{-at}\,\mathrm{d}t.$$

Výhodou štatistiky $HM_{n,a}$ oproti $HE_{n,a}$ je jednoduchosť výpočtu zjednodušeného tvaru testovej štatistiky. Výpočetne výhodný tvar štatistiky $HE_{n,a}$, ktorý odvodil Henze (1993), je rovný

$$HE_{n,a} = \frac{1}{n}\sum_{i,j=1}^{n} \frac{1}{Y_i + Y_j + a} - 2\sum_{i=1}^{n} e^{Y_i+a} E_1(Y_i + a) + n(1 - ae^a E_1(a)),$$

kde $E_1(z) = \int_z^\infty \frac{e^{-at}}{t}\,\mathrm{d}t, z > 0$, je exponenciálny integrál. $E_1(z)$ sa musí numericky približne vypočítať, čo môže mať negatívny vplyv na presnosť testu. Naproti tomu $HM_{n,a}$ možno zapísať ako

$$HM_{n,a} = \frac{1}{n}\sum_{i,j=1}^{n} \frac{1 + (Y_i + Y_j + a + 1)^2}{(Y_i + Y_j + a)^3} - 2\sum_{i=1}^{n} \frac{1 + Y_i + a}{(Y_i + a)^2} + \frac{n}{a}.$$

Oba tieto zjednodušené tvary možno overiť z definície štatistík analogickým spôsobom ako sme odvodili zjednodušený tvar testovej štatistiky $BH_{n,a}$. Preto jednoducho aplikovateľnou testovou štatistikou z tejto triedy štatistík je $HM_{n,a}$. Henze a Meintanis (2002b) na základe simulačnej štúdie určili približné kritické hodnoty tejto testovej štatistiky. Nulovú hypotézu zamietame pre veľké hodnoty testovej štatistiky $HM_{n,a}$, ktoré naznačujú signifikantný rozdiel medzi empirickou Laplaceovou transformáciou náhodného výberu a Laplaceovou transformáciou exponenciálneho rozdelenia. Autori doporučujú použitie štatistík $HM_{n,0.75}$ a $HM_{n,1.0}$, ktoré sa ukázali byť empiricky najsilnejšie v ich simulačnej štúdii.

### 2.4.2 Fourierova transformácia

Fourierovou transformáciou náhodnej veličiny $X$ s hustotou $f$ je jej charakteristická funkcia, ktorá je definovaná pre $t \in \mathbb{R}$ ako

$$\phi(t) = \mathsf{E}\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx} f(x)\,\mathrm{d}x = C(t) + iS(t),$$

kde $C(t) = \mathsf{E}\left[\cos(tX)\right]$ je reálna časť charakteristickej funkcie a $S(t) = \mathsf{E}\left[\sin(tX)\right]$ je jej imaginárna časť. Charakteristickú funkciu exponenciálneho rozdelenia s hustotou $f(x) = \lambda e^{-\lambda x}, \lambda > 0, x \geq 0$, odvodíme pomocou integrácie per partes. Pre reálnu časť máme

$$C(t) = \mathsf{E}\left[\cos(tX)\right] = \lambda \int_0^\infty \cos(tx) e^{-\lambda x}\,\mathrm{d}x = \lambda I,$$

pričom integrál $I$ je rovný

$$\begin{aligned} I &= \int_0^\infty \cos(tx)e^{-\lambda x}\,\mathrm{d}x = \Big[-\frac{\cos(tx)e^{-\lambda x}}{\lambda}\Big]_0^\infty - \frac{t}{\lambda}\int_0^\infty \sin(tx)e^{-\lambda x}\,\mathrm{d}x \\ &= \frac{1}{\lambda} - \frac{t}{\lambda}\Big\{\Big[-\frac{\sin(tx)e^{-\lambda x}}{\lambda}\Big]_0^\infty - \frac{t}{\lambda}\int_0^\infty \cos(tx)e^{-\lambda x}\,\mathrm{d}x\Big\} \\ &= \frac{1}{\lambda} - \frac{t^2}{\lambda^2}I. \end{aligned}$$

Odtiaľ dostávame, že $I = \frac{1}{\lambda + t^2/\lambda}$ a preto pre $t \in \mathbb{R}$

$$C(t) = \lambda I = \frac{1}{1 + t^2/\lambda^2}.$$

Analogickým použitím integrácie per partes pre imaginárnu časť odvodíme, že pre $t \in \mathbb{R}$

$$S(t) = \frac{t/\lambda}{1 + t^2/\lambda^2}.$$

Charakteristická funkcia exponenciálneho rozdelenia je rovná pre $t \in \mathbb{R}$

$$\phi(t) = \frac{1}{1 + t^2/\lambda^2} + i\frac{t/\lambda}{1 + t^2/\lambda^2}.$$

Z tohto vzťahu dostávame prvú charakterizáciu exponenciálneho rozdelenia, keďže pre nejaké $\lambda > 0$ a všetky $t \in \mathbb{R}$ platí

$$S(t) = \frac{t}{\lambda}C(t).$$

Túto charakterizáciu dokázali Meintanis a Iliopoulos (2003) a dôkaz ich tvrdenia ďalej detailne rozpracujeme.

**Veta 12.** *Zo všetkých spojitých nezáporných náhodných veličín s hladkými hustotami s konečnou limitou pre $x \to 0^+$ a absolútne integrovateľnými deriváciami, náhodná veličina s exponenciálnym rozdelením $Exp(1/\lambda), \lambda > 0$ je jediná, ktorá spĺňa*

$$S(t) = \frac{t}{\lambda}C(t), \quad \textit{pre } t \in R.$$

*Dôkaz.* Majme náhodnú veličinu s hustotou $f$ spĺňajúcu predpoklady vety. Vynásobením rovnice v tvrdení vety parametrom $\lambda$ dostávame

$$\lambda\int_0^\infty \sin(tx)f(x)\,\mathrm{d}x = t\int_0^\infty \cos(tx)f(x)\,\mathrm{d}x.$$

Integráciou per partes pravej strany dostávame

$$\begin{aligned} t\int_0^\infty \cos(tx)f(x)\,\mathrm{d}x &= \int_0^\infty [t\cos(tx)]f(x)\,\mathrm{d}x = \int_0^\infty [\sin(tx)]'f(x)\,\mathrm{d}x \\ &= \big[\sin(tx)f(x)\big]_0^\infty - \int_0^\infty \sin(tx)f'(x)\,\mathrm{d}x \\ &= -\int_0^\infty \sin(tx)f'(x)\,\mathrm{d}x, \end{aligned}$$

pričom sme použili konečnosť limity hustoty v 0 zprava a absolútnu integrovateľnosť derivácie hutoty, ktoré predpokladáme. Teda pre všetky $t \in \mathbb{R}$ platí

$$\int_0^\infty [\lambda f(x) + f'(x)] \sin(tx)\,\mathrm{d}x = 0.$$

Aby platila táto rovnica pre každé reálne $t$, musí byť funkcia $\lambda f(x) + f'(x)$ rovná 0. Teda hľadáme riešenie diferenciálnej rovnice

$$-\frac{1}{\lambda} = f(x),$$

s počiatočnou podmienkou $\int_0^\infty f(x)\,\mathrm{d}x = 1$, keďže riešením má byť hustota nezápornej náhodnej veličiny. Úpravou a riešením diferenciálnej rovnice dostávame

$$\begin{aligned}
\frac{\mathrm{d}f(x)}{\mathrm{d}x} &= -\lambda f(x) \\
\frac{1}{f(x)}\frac{\mathrm{d}f(x)}{\mathrm{d}x} &= -\lambda \\
\int \frac{1}{f(x)}\frac{\mathrm{d}f(x)}{\mathrm{d}x}\,\mathrm{d}x &= \int -\lambda\,\mathrm{d}x \\
\log[f(x)] &= -\lambda x + C_1 \\
f(x) &= C_2 e^{-\lambda x},
\end{aligned}$$

kde $C_2 = e^{C_1}$. Z počiatočnej podmienky dostávame

$$\int_0^\infty f(x)\,\mathrm{d}x = C_2\left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^\infty = \frac{C_2}{\lambda} = 1.$$

Odtiaľ plynie, že hodnota integračnej konštanty je $C_2 = \lambda$ a jednoznačným riešením tejto diferenciálnej rovnice je hustota exponenciálneho rozdelenia $f(x) = \lambda e^{-\lambda x}, x \geq 0$.

□

Druhú charakterizáciou exponenciálneho rozdelenia pomocou charakteristickej funkcie, ktorú dokázali taktiež Meintanis a Iliopoulos (2003), dostaneme z druhej mocniny modulu charakteristickej funkcie rozdelenia $Exp(1/\lambda)$

$$|\phi(t)|^2 = C^2(t) + S^2(t) = \frac{1}{(1 + t^2/\lambda^2)^2} + \frac{t^2/\lambda^2}{(1 + t^2/\lambda^2)^2} = \frac{1}{1 + t^2/\lambda^2} = C(t).$$

Meintanis a Iliopoulos (2003) dokázali, že pre nezápornú náhodnú veličinu $X$ platí $|\phi(t)|^2 = C(t)$ pre $t \in \mathbb{R}$ práve vtedy, keď $X$ pochádza z exponenciálneho rozdelenia.

Henze a Meintanis (2002a, 2005) zaviedli testy exponenciality založené na vyššie uvedených charakterizácii exponenciálneho rozdelenia pomocou charakteristickej funkcie a tieto testy ďalej predstavíme.

Najprv predstavme test, ktorý zostavili Henze a Meintanis (2002a) na základe prvej charakterizácie. Nech $X_1, \ldots, X_n$ je náhodný výber s neznámym rozdelením a označme $Y_i = X_i/\bar{X}_n, i = 1, \ldots, n$. Chceme testovať nulovú hypotézu, že výber $X_1, \ldots, X_n$ pochádza z exponenciálneho rozdelenia s neznámym parametrom. Empirická charakteristická funkcia je definovaná pre $t \in \mathbb{R}$ ako

$$\phi_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itY_j} = C_n(t) + iS_n(t),$$

kde $C_n(t) = \frac{1}{n}\sum_{j=1}^{n}\cos(tY_j)$ a $S_n(t) = \frac{1}{n}\sum_{j=1}^{n}\sin(tY_j)$ predstavujú reálnu a imaginárnu časť empirickej charakteristickej funkcie. Henze a Meintanis (2002a) navrhujú zamietnuť nulovú hypotézu v prospech alternatívy, že náhodný výber nemá exponenciálne rozdelenie, pre veľké hodnoty štatistiky

$$CF_n^1 = n\int_0^\infty |S_n(t) - tC_n(t)|^2 w(t)\,\mathrm{d}t,$$

kde $w(t)$ je nezáporná váhová funkcia splňujúca $\int_0^\infty t^2 w(t)\,\mathrm{d}t < \infty$. Autori skúmali dve triedy štatistík so špeciálnymi voľbami váhovej funkcie. V simulačnej štúdii, v ktorej Henze a Meintanis (2002a) porovnávali tieto dve triedy s inými testami exponenciality, preukázali obe triedy veľmi podobné empirické výsledky, preto uvedieme len jednu z nich. Pre triedu testových štatistík s váhovou funkciou $w(t) = e^{-at}, a > 0$, odvodili Henze a Meintanis (2002a) úpravou integrálu výpočetne výhodný tvar

$$CF_{n,a}^1 = \frac{a}{2n}\sum_{j=1}^{n}\sum_{k=1}^{n}\Bigg[\frac{1}{a^2+(Y_j-Y_k)^2} - \frac{1}{a^2+(Y_j+Y_k)^2} - \frac{4(Y_j+Y_k)}{(a^2+(Y_j+Y_k))^2} + \frac{2a^2-6(Y_j-Y_k)^2}{(a^2+(Y_j-Y_k))^3} + \frac{2a^2-6(Y_j+Y_k)^2}{(a^2+(Y_j+Y_k))^3}\Bigg].$$

Test založený na štatistike $CF_{n,a}^1$ zamieta na hladine $\alpha$ nulovú hypotézu pre hodnoty testovej štatistiky väčšie ako približná kritická hodnota, ktorú môžeme pre danú voľbu $n, a$ nájsť v tabuľke, ktorú zostavili a publikovali Henze a Meintanis (2002a). Približné kritické hodnoty autori získali na základe Monte Carlo simulácii.

Henze a Meintanis (2005) pomocou druhej charakterizácie exponenciálneho rozdelenia na základe charakteristickej funkcie navrhli triedu štatistík

$$CF_n^2 = n\int_0^\infty [|\phi_n(t)|^2 - C_n(t)]^2 w(t)\,\mathrm{d}t,$$

kde $w(t)$ je nezáporná váhová funkcia. Autori skúmali dve váhové funkcie a ako výhodnejšia sa ukázala byť váhová funkcia $w(t) = e^{-at}, a > 0$, pre ktorú odvodili zjednodušený predpis

$$\begin{aligned} CF_{n,a}^2 = {} & \frac{a}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left[\frac{1}{a^2+(Y_j-Y_k)^2} + \frac{1}{a^2+(Y_j+Y_k)^2}\right] \\ & - \frac{2a}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\left[\frac{1}{a^2+(Y_j-Y_k-Y_l)^2} + \frac{1}{a^2+(Y_j-Y_k+Y_l)^2}\right] \\ & + \frac{a}{n^3}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\left[\frac{1}{a^2+(Y_j-Y_k-Y_l+Y_m)^2} + \frac{1}{a^2+(Y_j-Y_k+Y_l-Y_m)^2}\right]. \end{aligned}$$

Testy založené na $CF_{n,a}^2$ zamietajú nulovú hypotézu pre hodnoty testovej štatistiky väčšie ako približné kritické hodnoty, získané simuláciami a uvedené pre rôzne hodnoty $n, a$, ktoré uviedli Henze a Meintanis (2005) v tabuľke.

### 2.4.3 Hankelova transformácia

Medzi najnovšie testy exponenciality patrí test založený na Hankelovej transformácii, ktorý publikovali Baringhaus a Taherizadeh (2013). Hankelova transformácia náhodnej veličiny $X \sim Exp(1/\lambda)$ je rovná $\mathcal{H}_X(t) = e^{\frac{-t}{\lambda}}$ (Baringhaus a Taherizadeh, 2013). Charakterizácia exponenciálneho rozdelenia pomocou Hankelovej transformácie, ktorú autori použili na odvodenie testovej štatistiky vychádza z jednoznačnosti Hankelových transformácii, ktoré sú rovné $e^{-t}$ pre $t \in [0,\epsilon], \epsilon > 0$.

**Veta 13.** *Nech $X$ je nezáporná náhodná veličina s Hankelovou transformáciou $\mathcal{H}_X(t) = e^{-t}$ pre všetky $t \in [0,\epsilon], \epsilon > 0$. Potom $X \sim Exp(1)$.*

*Dôkaz.* Uvažujme nezápornú náhodnú veličinu $Z$, ktorá má beta rozdelenie $B(1/2,1/2)$ s hustotou $1/\pi x^{-1/2}(1-x)^{-1/2}$, pre $x \in (0,1)$ a 0 inak, ktoré je nezávislá na $Y = 2\sqrt{X}$. Označme náhodnú veličinu $V = 2Z - 1$, ktorá je symetricky rozdelená okolo 0, pretože rozdelenie $B(1/2,1/2)$ je symetrické okolo bodu $1/2$ a lineárna transformácia nezmení symetriu rozdelenia. Ďalej hľadáme rozdelenie náhodnej veličiny $V$, teda jej distribučnú funkciu

$$\begin{aligned} F_V(t) &= \mathsf{P}(V \leq t) = \mathsf{P}(2Z - 1 \leq t) = \mathsf{P}(Z \leq (t+1)/2) \\ &= \int_0^{\frac{t+1}{2}} \frac{1}{\pi} x^{-1/2}(1-x)^{-1/2}\,\mathrm{d}x. \end{aligned}$$

Podľa Leibnitzovho pravidla o derivácii integrálu s parametrom dostávame hustotu pre $t \in (-1,1)$

$$\begin{aligned} f_V(t) &= \frac{d}{dt}F_V(t) = \frac{1}{\pi}\left(\frac{t+1}{2}\right)^{-1/2}\left(1 - \frac{t+1}{2}\right)^{-1/2}\frac{1}{2} \\ &= \frac{1}{2\pi}2(t+1)^{-1/2}(1-t)^{-1/2} = \frac{1}{\pi}(1-t^2)^{-1/2}. \end{aligned}$$

Fourierova transformácia náhodnej veličiny $V$ je

$$\begin{aligned} \phi_V(t) &= \mathsf{E}\left[e^{itV}\right] = \int_{-1}^{1} e^{itv}\frac{1}{\pi}(1-v^2)^{-1/2}\,\mathrm{d}v \\ &= \frac{1}{\pi}\int_{-1}^{1}[\cos(tv) + i\sin(tv)](1-v^2)^{-1/2}\,\mathrm{d}v \\ &= \frac{1}{\pi}\int_{0}^{\pi}[\cos(t\cos\theta) - i\sin(t\cos\theta)]\frac{1}{\sqrt{1-\cos^2\theta}}\sin\theta\,\mathrm{d}\theta \\ &= \frac{1}{\pi}\int_0^{\pi}\cos(t\cos\theta)\,\mathrm{d}\theta - \frac{i}{\pi}\int_0^{\pi}\sin(t\cos\theta)\,\mathrm{d}\theta = J_0(t), \end{aligned}$$

pričom sme použili substitúciu $v = -\cos\theta$. Druhý integrál v predposlednej rovnosti je nulový pretože integrál z funkcie kosínus na intervale $(0,\pi)$ je nulový. Prvý integrál je integrálnym vyjadrením Besselovej funkcie prvého druhu rádu 0, ktoré sme dokázali v lemme 3. Ďalej preto platí pre $|t| \leq \epsilon$

$$e^{-t^2} = \mathcal{H}_X(t^2) = \mathsf{E}\left[J_0(tY)\right] = \mathsf{E}\left[\phi_V(tY)\right] = \mathsf{E}\left(e^{iVY}\right).$$

Z charakteristickej funkcie normálneho rozdelenia vieme, že náhodná veličina $VY \sim N(0,2)$. Z momentov normálneho rozdelenia vieme, že pre všetky prirozdené $k$ platí

$$\begin{aligned}\mathsf{E}\,[(VY)^{2k-1}] &= 0,\\ \mathsf{E}\,[(VY)^{2k}] &= \frac{(2k)!}{2^k k!}2^k = \frac{(2k)!}{k!}.\end{aligned}$$

Z nezávislosti $V$ a $Y$ môžeme ďalej písať

$$\mathsf{E}\,[(2V\sqrt{X})^{2k}] = \mathsf{E}\,[(2V)^{2k}X^k] = \frac{(2k)!}{k!}.$$

Odtiaľ dostávame

$$\begin{aligned}\mathsf{E}\ X^k &= \frac{1}{\mathsf{E}\,[(2V)^2 k]}\frac{(2k)!}{k!}\\ &= \frac{1}{2^{2k}\,\mathsf{E}\,[(V^2)^k]}\frac{(2k)!}{k!}.\end{aligned}$$

Preto je nutné nájsť momenty a rozdelenie náhodnej veličiny $V^2$. Jej distribučná funkcia má tvar

$$\begin{aligned}F_{V^2}(t) &= \mathsf{P}(V^2 \le t) = \mathsf{P}(V \le \sqrt{t}|V \ge 0) + \mathsf{P}(V \le -\sqrt{t}|V \le 0)\\ &= \int_0^{\sqrt{t}} f_V(x)\,\mathrm{d}x + \int_{-\sqrt{t}}^0 f_V(x)\,\mathrm{d}x = 2\int_0^{\sqrt{t}} f_V(x)\,\mathrm{d}x\\ &= 2\int_0^{\sqrt{t}} \frac{1}{\pi}(1-x^2)^{-1/2}\,\mathrm{d}x.\end{aligned}$$

Pomocou distribučnej funkcie a derivácie integrálu s parametrom dostávame hustotu pre $t \in (0,1)$

$$\begin{aligned}f_{V^2}(t) &= \frac{d}{dt}F_{V^2}(t) = 2\frac{1}{\pi}(1-t)^{-1/2}\frac{1}{2\sqrt{t}}\\ &= \frac{1}{\pi}t^{-1/2}(1-t)^{-1/2} = f_Z(t).\end{aligned}$$

Teda náhodná veličina $V^2$ má taktiež beta rozdelenie $B(1/2,1/2)$. Momenty tohto beta rozdelenia pre prirodzené $k$ spĺňajú

$$\mathsf{E}\,[(V^2)^k] = \frac{1}{2^{2k}}\frac{(2k)!}{k!k!}.$$

Odtiaľ dostávame momenty náhodnej veličiny $X$ ako

$$\mathsf{E}\ X^k = \frac{1}{2^{2k}}\frac{2^{2k}k!k!}{(2k)!}\frac{(2k)!}{k!} = k!,$$

čo sú momenty rozdelenia $Exp(1)$. Rozdelenie $Exp(1)$ je jednoznačne určené svojimi momentami, preto $X \sim Exp(1)$.

□

Na základe tohto výsledku navrhli Baringhaus a Taherizadeh (2013) testovú štatistiku typu KS na testovanie hypotézy

$$H_0 : X_1,...,X_n \text{ je z } Exp(1/\lambda), \lambda > 0, \text{proti}$$
$$H_1 : X_1,...,X_n \text{ nie je z } Exp(1/\lambda).$$

Nimi navrhovaná štatistika porovnáva empirickú verziu Hankelovej transformácie s hodnotou za platnosti nulovej hypotézy. Má tvar

$$L_n = \sqrt{n} \sup_{0 \leq t \leq 1} |\mathcal{H}_n(t) - e^{-t}|,$$

kde $\mathcal{H}_n(t) = \frac{1}{n} \sum_{i=1}^{n} J_0(2\sqrt{tY_i}), 0 \leq t \leq 1$ je empirická Hankelova transformácia náhodného výberu $Y_1 = X_1/\overline{X}_n, ..., Y_n = X_n/\overline{X}_n$. Interval [0,1] na ktorom hľadáme suprémum môžeme nahradiť iným konečným intervalom $[0,\epsilon], 0 < \epsilon < \infty$, autori volili interval [0,1] pre jednoduchosť.

Test založený na $L_n$ zamieta $H_0$ pre veľké hodnoty testovej štatistiky, naznačujúce veľký rozdiel medzi Hankelovou transformáciou transformovaných dát a Hankelovou transformáciou náhodnej veličiny s exponenciálnym rozdelením. Konkrétne zamietame $H_0$ pre $L_n > c_{n,\alpha}$, kde $c_{n,\alpha}$ je $(1-\alpha)$-kvantil rozdelenia $L_n$ za $H_0$. Keďže pre náhodné veličiny $cX_i, c > 0, i = 1,...,n$, platí

$$Y_i' = \frac{cX_i}{\frac{1}{n}\sum_{i=1}^{n} cX_i} = Y_i,$$

hodnota testovej štatistiky $L_n$ je pre $X_1,...,X_n$ a $cX_i,...,cX_n$ taktiež rovná. Preto testová štatistika $L_n$ nezávisí na neznámom parametri $\lambda$ exponenciálneho rozdelenia.

Baringhaus a Taherizadeh (2013) odvodili asymptotické rozdelenie testovej štatistiky za platnosti nulovej hypotézy. Na základe simulácii zostavili Baringhaus a Taherizadeh (2013) tabuľku približných kritických hodnôt $c_{n,\alpha}$ pre rôzne hodnoty $n$ a $\alpha$.

## 2.5 Testy založené na entropii

Testy exponenciality založené na entropii boli hlavne v posledných rokoch záujmom výskumu niekoľkých autorov. Používajú rôzne typy entropie a prístupy, preto niekoľko takýchto testov bude predstavených v nasledujúcej časti práce.

### 2.5.1 Shannonova entropia

Charakterizáciu exponenciálneho rozdelenia pomocou Shannonovej entropie, sme získali v kapitole 1, dokázaním, že exponenciálne rozdelenie maximalizuje Shannonovu entropiu na kladnej poloosi. Otázkou ale je, akým spôsobom porovnať hustotu náhodného výberu s hustotou exponenciálneho rozdelenia za využitia tejto charakterizácie. Kullback a Leibler (1951) zaviedli veličinu, ktorá porovnáva vzdialenosť dvoch hustôt, založenú na Shannonovej entropii. Nech $f$ a $g$ sú hustoty rozdelenia s kladným nosičom, potom Kullbackova-Leiblerova vzdialenosť hustôt $f$ a $g$ je definovaná ako

$$D_{KL}(f : g) = \int_0^\infty f(x) \log \left[ \frac{f(x)}{g(x)} \right] \mathrm{d}x.$$

Platí $D_{KL}(f : g) \geq 0$, pričom rovnosť nastáva práve vtedy, keď $f = g$ skoro všade Kullback a Leibler (1951).

Ebrahimi a kol. (1992) navrhli spôsob ako použiť Kullbackovu-Leiblerovu vzdialenosť na testovanie exponenciality. Nech $X$ je náhodná veličina s hustotou $f$ a konečnou strednou hodnotou $\mathsf{E}\ X = 1/\lambda, \lambda > 0$, a označme $f_0(x) = \lambda e^{-\lambda x}, x > 0$, hustotu exponenciálneho rozdelenia. Kullbackova-Leiblerova vzdialenosť rozdelenia náhodnej veličiny $X$ a exponenciálneho rozdelenia je

$$\begin{aligned} D_{KL}(f : f_0) &= \int_0^\infty f(x) \log \left[ \frac{f(x)}{f_0(x)} \right] \mathrm{d}x \\ &= \int_0^\infty f(x) \log f(x)\, \mathrm{d}x - \int_0^\infty f(x) \log f_0(x)\, \mathrm{d}x \\ &= -H(f) - \log \lambda \int_0^\infty f(x)\, \mathrm{d}x + \lambda \int_0^\infty x f(x)\, \mathrm{d}x \\ &= -H(f) - \log \lambda + 1, \end{aligned}$$

kde $H(f)$ je Shannonova entropia rozdelenia s hustotou $f$.

Nech $X_1, \ldots, X_n$ je náhodný výber zo spojitého rozdelenia s hustotou $f$, kladným nosičom a konečnou strednou hodnotou $\mathsf{E}\ X = 1/\lambda, \lambda > 0$ neznáme. Chceli by sme testovať hypotézu

$$\begin{aligned} &H_0 : X_1,...,X_n \text{ je z } Exp(1/\lambda), \lambda > 0, \text{proti} \\ &H_1 : X_1,...,X_n \text{ nie je z } Exp(1/\lambda). \end{aligned}$$

Ako sme odvodili vyššie, Kullbackova-Leiblerova vzdialenosť hustoty $f$ a hustoty exponenciálneho rozdelenia je rovná $D_{KL}(f : f_0) = -H(f) - \log \lambda + 1$ a za platnosti nulovej hypotézy je rovná 0. K výpočtu $D_{KL}(f : f_0)$ je nutná presná znalosť oboch rozdelení, avšak hustotu $f$ ani parameter $\lambda$ nepoznáme. Preto je nutné použiť odhady parametra $\lambda$ a Shannonovej entropie $H(f)$. Konkrétne parameter $\lambda$ odhadneme pomocou $\hat{\lambda} = 1/\bar{X}_n$. Za odhad Shannonovej entropie vzali Ebrahimi a kol. (1992) známy Vasickov odhad entropie, ktorý ďalej odvodíme.

Majme náhodný výber $X_1,...,X_n$ s rozsahom $n \geq 3$ zo spojitého rozdelenia s distribučnou funkciou $F$ a hustotou $f(x)$ a nosičom $(a,b)$, $-\infty \leq a < b \leq \infty$. Vasicek (1976) navrhol upraviť vzťah pre Shannonovu entropiu rozdelenia $H(f)$ pomocou substitúcie $p = F(x)$ a jeho výsledok a vzťah pre odhad entropie teraz odvodíme a overíme. Pre danú substitúciu platí

$$\begin{aligned} &p = F(x) \\ &p \in (0,1), \quad x \in (a,b) \\ &\mathrm{d}p = f(x)\, \mathrm{d}x \\ &F^{-1}(p) = x \\ &\frac{\mathrm{d}}{\mathrm{d}p} F^{-1}(p) = \frac{1}{F'(x)} = \frac{1}{f(x)}, \end{aligned}$$

pričom posledná rovnosť plynie z vety o derivácii inverznej funkcie a z faktu, že distribučná funkcia spojitého rozdelenia je na nosiči ostro rastúca, teda tu existuje

jej inverz. Pomocou tejto substitúcie upravíme Shannonovu entropiu do tvaru

$$\begin{aligned} H(f) &= \int_a^b f(x) \log \left[ \frac{1}{f(x)} \right] \mathrm{d}x \\ &= \int_0^1 \log \left[ \frac{\mathrm{d}}{\mathrm{d}p} F^{-1}(p) \right] \mathrm{d}p. \end{aligned}$$

Keďže distribučnú funkciu $F$ nepoznáme presne, odhadneme ju pomocou empirickej distribučnej funkcie $F_n$. Zvoľme prirodzené číslo $m$ tak, že $m < n/2$. Pre každé $p \in (0,1)$ chceme získať odhad derivácie $F^{-1}(p)$. Empirická distribučná funkcia je po častiach konštantná skokovitá funkcia, so skokmi v bodoch $X_{(i)}$ veľkosti $1/n$. Označme $X_{(0)} = 0$, potom postupnosť $\{F_n(X_{(j)})\}_{j=0}^n = \{j/n\}_{j=0}^n$ je delením intervalu (0,1). Preto pre každé $p \in (0,1)$ existuje $i \in 1, \ldots, n$ také, že $(i-1)/n < p \leq i/n$. Odtiaľ a z montónnosti $F_n$ plynie, že platí aj $(i-m)/n < p \leq (i+m)/n$. Chceli by sme odhadnúť hodnotu derivácie inverzu distribučnej funkcie v $p$. V bodoch $\{j/n\}_{i=0}^n$ delenia intervalu (0,1) môžeme odhadnúť funkčnú hodnotu inverzu pomocou

$$F_n^{-1}\left(\frac{j}{n}\right) = F_n^{-1}[F_n(X_{(j)})] = X_{(j)}.$$

Odtiaľ dostávame odhad derivácie inverzu distribučnej funkcie tak, že nahradíme operátor derivácie operátorom rozdielu

$$\widehat{\frac{\mathrm{d}}{\mathrm{d}p} F^{-1}}(p) = \frac{F_n^{-1}(\frac{i+m}{n}) - F_n^{-1}(\frac{i-m}{n})}{\frac{2m}{n}} = \frac{n(X_{(i+m)} - X_{(i-m)})}{2m}.$$

Pre špeciálne prípady $i > n$, resp. $i < 1$ označíme $X_{(i)} = X_{(1)}$, resp. $X_{(n)}$. Aj tento odhad je po častiach konštantná funkcia v premennej $p$, s konštantnou hodnotou na každom intervale $((j-1)/n, j/n), j = 1, \ldots, n$. Dosadením do upraveného vzťahu pre Shannonovu entropiu dostávame Vasickov odhad v tvare

$$HV_{m,n} = \frac{1}{n} \sum_{i=1}^{n} \log \left[ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right],$$

kde $m < n/2$ je vopred dané prirodzené čislo, $X_{(i)} = X_{(1)}$, pre $i < 1$ a $X_{(i)} = X_{(n)}$ pre $i > n$.

Požitím odhadov $\hat{\lambda}$ a $HV_{m,n}$ dostávame odhad Kullbackovej-Leiblerovej vzdialenosti hustôt

$$D_{m,n} = -HV_{m,n} + \log \bar{X}_n + 1.$$

Ebrahimi a kol. (1992) navrhli vziať za testovú štatistiku monotónnu transformáciu $D_{m,n}$ v tvare

$$KLV_{m,n} = \exp\{-D_{m,n}\} = \frac{\exp\{HVm,n\}}{\exp\{\log \bar{X}_n + 1\}}.$$

Testová štatistika $KLV_{m,n}$ závisí na rozsahu náhodného výberu $n$ a tiež na parametri $m, m < n/2$, z Vasickovho odhadu entropie. Približné kritické hodnoty testu $c_{m,n}(\alpha)$ autori získali pomocou Monte Carlo simulácii. Nulovú hypotézu

zamietame pre hodnoty testovej štatistiky $KLV_{m,n} < c_{m,n}(\alpha)$, svedčiace v prospech alternatívy. Pre daný rozsah výberu zistili autori, že najväčšiu silu má test pri použití hodnoty $m$ takej, ktorá maximalizuje kritickú hodnotu $c_{m,n}(\alpha)$. Ebrahimi a kol. (1992) zostavili tabuľku doporučenej hodnoty $m$ pri danom rozsahu výberu ako aj tabuľku približných kritických hodnôt $c_{m,n}(\alpha)$ pre dané $n$ a $\alpha$ a doporučené $m$.

Choi a kol. (2004) použitím rovnakého postupu ako Ebrahimi a kol. (1992), ale odlišného odhadu Shannonovej entropie navrhli 2 ďalšie testové štatistiky na testovanie exponenciality. Použili Van Esov a Correov odhad entropie.Van Esov odhad entropie má tvar

$$E_{m,n} = \frac{1}{n-m}\sum_{i=1}^{n-m}\log\left[\frac{n+1}{m}(X_{(i+m)} - X_{(i-m)})\right] + \sum_{i=m}^{n}\frac{1}{k} + \log\left(\frac{m}{n+1}\right),$$

kde $m < n/2$ je vopred dané prirodzené číslo, $X_{(i)} = X_{(1)}$, pre $i < 1$ a $X_{(i)} = X_{(n)}$ pre $i > n$. Correov odhad entropie je modifikáciou Vasickovho odhadu a je definovaný ako

$$C_{m,n} = -\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{\sum_{j=i-m}^{i+m}(X_{(j)} - \bar{X}_{(i)})(j-i)}{n\sum_{j=i-m}^{i+m}(X_{(j)} - \bar{X}_{(i)})^2}\right],$$

kde $\bar{X}_{(i)} = \sum_{j=i-m}^{i+m} X_{(j)}/(2m+1)$, $m < n/2$ je vopred dané prirodzené číslo, $X_{(i)} = X_{(0)}$, pre $i < 1$ a $X_{(i)} = X_{(n)}$ pre $i > n$.

Odhady Kullbackovej-Leiblerovej vzdialenosti, ktoré navrhli Choi a kol. (2004) sú

$$DE_{m,n} = -E_{m,n} + \log\bar{X}_n - 1$$

a

$$DC_{m,n} = -C_{m,n} + \log\bar{X}_n - 1.$$

A navrhované testové štatistiky, získané rovnakou monotónnou transformáciou ako štatistika $KLV_{m,n}$ sú

$$KLE_{m,n} = \frac{\exp\{E_{m,n}\}}{\exp\{\log\bar{X}_n + 1\}}$$

a

$$KLC_{m,n} = \frac{\exp\{C_{m,n}\}}{\exp\{\log\bar{X}_n + 1\}}.$$

Malé hodnoty testových štatistík $KLE_{m,n}$ a $KLC_{m,n}$ naznačujú, že náhodný výber nepochádza z exponenciálneho rozdelenia. Preto testy založené na $KLE_{m,n}$ a $KLC_{m,n}$, rovnako ako test založený na $KL_{m,n}$, s približnou hladinou $\alpha$ zamietajú $H_0$ v prospech alternatívy pre hodnoty testových štatistík menšie ako $\alpha$-kvantil rozdelenia danej testovej štatistiky za platnosti nulovej hypotézy. Kvôli nutnosti voliť hodnotu parametra $m$, nie je možné určiť rozdelenie za platnosti nulovej hypotézy analyticky. Preto Choi a kol. (2004) určili približné kritické hodnoty na základe Monte Carlo simulácii empiricky. Pre každý rozsah výberu vybrali autori hodnotu $m$, ktorá maximalizuje približnú kritickú hodnotu a tieto hodnoty taktiež možno nájsť v tabuľke.

### 2.5.2 Rényiho entropia

Na základe Rényiho entropie môžeme podobne ako v prípade Shannonovej entropie a Kullbackovej-Leiblerovej vzdialenosti definovať asymetrickú Rényiho vzdialenosť rádu $r$ medzi dvomi hustotami $f(x)$ a $g(x)$ pre $r > 0, r \neq 1$, ako

$$D_R^r(f:g) = \frac{1}{r-1} \log \int_{-\infty}^{\infty} \left[\frac{f(x)}{g(x)}\right]^{r-1} f(x)\,\mathrm{d}x.$$

Abbasnejad (2012) zaviedol niekoľko testov exponenciality založených na Rényiho vzdialenosti. Previedol problém testovania exponenciality pomocou transformácie dát na problém testovania uniformity na intervale (0,1). Podobnú stratégiu transformácie dát a použitia Shannonovej entropie pred ním použil na zostavenie testov exponenciality Taufer (2002). Avšak v rozsiahlej simulačnej štúdii porovnávajúcej testy exponenciality založené na entropii, ktorú previedol Taufer (2002), sa ukázalo, že transformácia náhodného výberu má v tomto prípade značne negatívny vplyv na silu testu. Testy zavedené Tauferom neobstáli v porovnaní s testami založenými na entropii, ktoré zaviedli Ebrahimi a kol. (1992). Pre testy založené na Rényiho entropii, ktoré zostavil Abbasnejad (2012), sa takýto vplyv nepreukázal a preto tieto testy predstavíme.

Na charakterizáciu exponenciálneho rozdelenia používa Abbasnejad (2012) nasledujúce transformácie nezávislých náhodných veličín na veličiny s rovnomerným rozdelením.

**Veta 14.** *Nech $X_1$ a $X_2$ sú náhodné veličiny s distribučnou funkciou $F$. Potom platí, že*

$$W = \frac{X_1}{X_1 + X_2} \sim R(0,1) \text{ práve vtedy, keď } X_i \sim Exp(1/\lambda)\, \lambda > 0, \text{ a}$$
$$Z = \frac{X_1 - X_2}{X_1 + X_2} \sim R(-1,1) \text{ práve vtedy, keď } X_i \sim Exp(1/\lambda)\, \lambda > 0.$$

*Dôkaz.* Dôkaz možno nájsť v Abbasnejad (2012). □

Na zostavenie testových štatistík transformujeme náhodný výber $X_1, \ldots, X_n$ z rozdelenia s kladným nosičom s distribučnou funkciou $F$ na dva nové náhodné výbery

$$W_{ij} = \frac{X_{(i)}}{X_{(i)} - X_{(j)}} \quad i \neq j, \quad i,j = 1,\ldots,n$$

a

$$Z_{ij} = \frac{X_{(i)} - X_{(j)}}{X_{(i)} + X_{(j)}} \quad i > j, i,j = 1,\ldots,n.$$

Nosičom rozdelení oboch nových náhodných výberov je (0,1). V prípade výberu $Z_{ij}$ to vyplýva z vlastnosti poriadkových štatistík, ktoré sú usporiadané podľa veľkosti vzostupne. Chceme testovať hypotézu

$$H_0 : X_1, \ldots, X_n \text{ je z } Exp(1/\lambda), \lambda > 0, \text{proti}$$
$$H_1 : X_1, \ldots, X_n \text{ nie je z } Exp(1/\lambda).$$

Z vety 14 plynie, že za platnosti hypotézy $H_0$, majú $W_{ij}, i,j = 1,\ldots,n, i \neq j$, aj $Z_{ij}, i,j = 1,..,n, i > j$ rovnomerné rozdelenia na intervale (0,1). Preto môžeme formulovať ekvivalentnú hypotézu

$$H_0^* : W_{ij}(\text{resp. } Z_{ij}) \text{ je z rozdelenia } R(0,1), \text{ proti}$$
$$H_1^* : W_{ij}(\text{resp. } Z_{ij}) \text{ nie je z rozdelenia } R(0,1).$$

Na testovanie uniformity na intervale (0,1) môžeme použiť nasledovný vzťah pre Rényiho vzdialenosť rozdelení. Nech $Y$ je náhodná veličina s hustotou $g(x)$ pre $x \in (0,1)$ a nech $g_0(x) = 1, x \in (0,1)$, označuje hustotu rozdelenia $R(0,1)$. Potom pre Rényiho vzdialenosť rádu $r, r > 0, r \neq 1$, hustôt $g$ a $g_0$ platí

$$D_R^r(g : g_0) = \frac{1}{r-1} \log \int_0^1 \left[\frac{g(x)}{g_0(x)}\right]^{r-1} g(x)\,\mathrm{d}x = \frac{1}{r-1} \log \int_0^1 g^r(x)\,\mathrm{d}x$$
$$= -H_r(g),$$

kde $H_r(g)$ je Rényiho entropia hustoty $g$ rádu $r$. Keďže hustotu $g$ nepoznáme presne, musíme hodnotu jej Rényiho entropie odhadnúť. Abbasnejad (2012) uvažuje 4 rôzne odhady Rényiho entropie a na transformáciu dát používa transformácie $W$ a $Z$, ktoré boli predstavené vyššie. V závere svojho článku Abbasnejad porovnáva v simulačnej štúdii všetkých 8 takto vzniknutých štatistík, pričom 2 z nich, ktoré ďalej predstavíme, patrili medzi empiricky najsilnejšie.

Využívaný odhad Rényiho entropie pre náhodný výber $Y_1,...,Y_n$ zaviedli Wachowiak a kol. (2005) a jedná sa o Vasickov typ odhadu entropie, ktorý má tvar

$$HV_{r,m,n} = -\frac{1}{r-1} \log \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{n}{2m} (Y_{(i+m)} - Y_{(i-m)}) \right]^{1-r} \right\},$$

kde $m$ je prirodzené číslo splňujúce $m < n/2$, $Y_{(i)} = Y_{(1)}$ pre $i < 1$ a $Y_{(i)} = Y_{(n)}$ pre $i > n$. Tento odhad bol odvodený z Rényiho entropie analogickým spôsobom ako Vasickov odhad Shannonovej entropie. Dve nami uvažované testové štatistiky sú definované ako

$$TV^W = -HV_{r,m,n'}$$

a

$$TV^Z = -HV_{r,m,n''},$$

kde index $Z$ alebo $W$ udáva, ktorá transformácia dát bola použitá. Transformované náhodné výbery sa líšia svojimi rozsahmi, pretože rozsah výberu $W_{ij}$ je $n' = n(n-1)$ a rozsah výberu $Z_{ij}$ je $n'' = n(n-1)/2$. Pri výpočte testových štatistík vypočitame hodnoty Vasickovho typu odhadu Rényiho entropie $HV_{r,m,n}$ pre transformované dáta s poriadkovými štatistikami $W_{(1)},\ldots,W_{(n')}$, resp. $Z_{(1)},\ldots,Z_{(n'')}$. Nulovú hypotézu zamietame pre veľké hodnoty testových štatistík, ktoré odhadujú Rényiho vzdialenosť skutočnej hustoty transformovaných dát a hustoty rovnomerného rozdelenia, svedčiace v prospech alternatívy.

Testové štatistiky aj kritické hodnoty závisia na voľbe parametrov $r$ a $m$. Abbasnejad (2012) podobne ako autori ostatných testov založených na odhade entropie, tvrdí, že neexistuje kritérium na výber optimálneho $r^*$ a $m^*$, ktoré závisia na alternatíve a rozsahu výberu. Autor však doporučuje voľbu $r^* = 1.2$ a $m^* = \lfloor \sqrt{n} + 0.5 \rfloor$, ktorá sa ukázala ako vhodná na základe empirických výsledkov simulácii. Približné kritické hodnoty pre túto voľbu parametrov a rôzne voľby $n$ a $\alpha$ uviedol Abbasnejad (2012) v tabuľke.

### 2.5.3 Lin-Wongova vzdialenosť

Abbasnejad a kol. (2012) definovali novú vzdialenosť medzi dvomi hustotami $f(x)$ a $g(x)$ ako

$$D_{LW}(f:g) = \int_{-\infty}^{\infty} f(x) \log \left[ \frac{2f(x)}{f(x)+g(x)} \right] \mathrm{d}x.$$

Platí, že $D_{LW}(f:g) \geq 0$, pričom rovnosť platí práve vtedy, keď $f(x) = g(x)$ (Abbasnejad a kol., 2012). Preto je možné použiť Lin-Wongovu vzdialenosť na testovanie dobrej zhody a špeciálne na testovanie exponenciality. Chceme testovať hypotézu $H_0$, že hustota $f(x)$ náhodného výberu $X_1, \ldots, X_n$ je rovná hustote $f_0(x) = \lambda e^{-\lambda x}, x \geq 0$, pre nejaké $\lambda > 0$, proti alternatíve, že náhodný výber nepochádza z exponenciálneho rozdelenia. Lin-Wongova informácia má v tomto prípade tvar

$$D_{LW}(f:f_0) = \int_{0}^{\infty} f(x) \log \left[ \frac{2f(x)}{f(x)+\lambda e^{-\lambda x}} \right] \mathrm{d}x,$$

pričom za platnosti nulovej hypotézy je $D_{LW}(f:f_0)$ nulová. Veľké hodnoty Lin-Wongovej vzdialenosti svedčia v prospech alternatívy. Abbasnejad a kol. (2012) odhadli Lin-Wongovu vzdialenosť podobným spôsobom aký sme videli vo Vasickovom odhade entropie, t.j. použitím substitúcie $F(x) = p$, a zostavili testovú štatistiku

$$LW = -\frac{1}{n} \sum_{i=1}^{n} \log \left[ \frac{1}{2} + \frac{n}{4m\bar{X}_n} (X_{(i+m)} - X_{(i-m)}) e^{-X_{(i)}/\bar{X}_n} \right],$$

kde $X_{(i)} = X_{(0)}$ pre $i < 1$ a $X_{(i)} = X_{(n)}$ pre $i > n$ a $m$ je parameter. Abbasnejad a kol. (2012) taktiež uviedli tabuľku približných kritických hodnôt, ktorú získali pomocou Monte Carlo simulácii pre voľbu parametra $m = \lfloor \sqrt{n} + 1/2 \rfloor$.

### 2.5.4 Kumulatívna reziduálna entropia

Rao a kol. (2004) zaviedli novú mieru informácie, ktorá rozširuje pojem Shannonovej entropie a nazvali ju kumulatívna reziduálna entropia (CRE). Namiesto hustoty v definícii Shannonovej entropie používa funkciu prežitia. CRE má teda pre nejakú distribučnú funkciu $F$ tvar

$$CRE(F) = -\int_{0}^{\infty} \bar{F}(x) \log \bar{F}(x) \, \mathrm{d}x.$$

Baratpour a Rad (2012) definovali novú vzdialenosť medzi dvomi rozdeleniami založenú na kumulatívnej reziduálnej entropii, tak ako je definovaná Kullbackova-Leiblerova vzdialenosť na základe Shannonovej entropie. Táto nová vzdialenosť bola nazvaná kumulatívna Kullbackova-Leiblerova vzdialenosť.

**Definícia 15.** *Nech $X, Y$ sú nezáporné absolútne spojité náhodné veličiny s distribučnou funkciou $F$, resp. $G$. Potom kumulatívna Kullbackova-Leiblerova vzdialenosť medzi týmito rozdeleniami je definovaná ako*

$$CKL(F:G) = \int_{0}^{\infty} \bar{F}(x) \log \left[ \frac{\bar{F}(x)}{\bar{G}(x)} \right] \mathrm{d}x - (\mathsf{E}\, X - \mathsf{E}\, Y).$$

Platí, že $CKL(F:G) \geq 0$, pričom rovnosť platí práve vtedy, keď $F = G$ (Baratpour a Rad, 2012). Preto sa $CKL$ ponúka ako dobrý prostriedok na testovanie zhody rozdelení.

Nech $X_1, \ldots, X_n \geq 0$, je náhodný výber z rozdelenia s distribučnou funkciou $F$ a s konečným $1/\lambda = \mathsf{E}\,(X_1^2)/(2\,\mathsf{E}\,X_1)$. Nech je ďalej $F_0(x) = 1 - e^{-\lambda x}$, pre $x > 0, \lambda > 0, \lambda$ neznáme. Testujeme hypotézu

$$\begin{aligned} &H_0 : F(x) = F_0(x) \text{ pre všetky } x > 0, \text{ proti} \\ &H_1 : \exists x : F(x) \neq F_0(x). \end{aligned}$$

Za platnosti nulovej hypotézy je $CKL(F : F_0) = 0$ a veľké hodnoty $CKL(F : F_0)$ svedčia v prospech alternatívy. Keďže $F$ nepoznáme presne musíme hodnotu $CKL(F : F_0)$ nejakým spôsobom odhadnúť. V prípade tejto konkrétnej testovanej nulovej hypotézy má $CKL(F : F_0)$ tvar

$$\begin{aligned} CKL(F:F_0) &= -CRE(F) - \int_0^\infty \bar{F}(x) \log \bar{F}_0(x)\,\mathrm{d}x - \mathsf{E}\,X_1 + \frac{1}{\lambda} \\ &= -CRE(F) + \lambda \int_0^\infty x\bar{F}(x)\,\mathrm{d}x - \mathsf{E}\,X_1 + \frac{1}{\lambda} \\ &= -CRE(F) + \frac{\lambda}{2}\left[x^2(1-F(x))\big|_0^\infty + \int_0^\infty x^2 f(x)\,\mathrm{d}x\right] - \mathsf{E}\,X_1 + \frac{1}{\lambda} \\ &= -CRE(F) - \frac{\lambda}{2}\,\mathsf{E}\,X_1^2 - \mathsf{E}\,X_1 + \frac{1}{\lambda} \\ &= -CRE(F) + \frac{1}{\lambda}. \end{aligned}$$

V poslednej rovnosti sme využili predpoklad o rozdelení náhodného výberu, že $1/\lambda = \mathsf{E}\,(X_1^2)/(2\,\mathsf{E}\,X_1)$. Na odhadnutie $CKL(F : F_0)$ je nutné odhadnúť veličiny $CRE(F)$ a $1/\lambda$. Ako odhad $1/\lambda$ vezmeme konzistentný odhad veličiny, ktorý má tvar

$$\widehat{\frac{1}{\lambda}} = \frac{\sum_{i=1}^n X_i^2}{2\sum_{i=1}^n X_i}.$$

Hodnotu $CRE(F)$ odhadneme kumulatívnou reziduálnou entropiou empirickej distribučnej funkcie, ktorou je

$$\widehat{CRE}(F) = -\int_0^\infty \bar{F}_n(x) \log \bar{F}_n(x)\,\mathrm{d}x = -\sum_{i=1}^{n-1} \frac{n-i}{n} \log\left(\frac{n-1}{n}\right)(X_{(i+1)} - X_{(i)}).$$

Druhá rovnosť platí, pretože empirická distribučná funkcia aj $\bar{F}_n(x)$ sú konštantné na každom intervale $[X_{(i)}, X_{(i+1)}), i = 1, \ldots, n-1$, s funkčnými hodnotami $F_n(x) = i/n$ a $\bar{F}_n(x) = (n-i)/n$. Baratpour a Rad (2012) navrhli použiť testovú štatistiku

$$BR_n = \frac{\sum_{i=1}^{n-1} \frac{n-1}{n} \log\left(\frac{n-1}{n}\right)(X_{(i+1)} - X_{(i)}) + \frac{\sum_{i=1}^n X_i^2}{2\sum_{i=1}^n X_i}}{\frac{\sum_{i=1}^n X_i^2}{2\sum_{i=1}^n X_i}}.$$

Nulovú hypotézu zamietame pre veľké hodnoty testovej štatistiky $BR_n$. Na hladine $\alpha$ zamietame $H_0$ pre hodnoty $BR_n \geq C_{n,1-\alpha}$, kde $C_{n,1-\alpha}$ je $(1-\alpha)-$kvantil rozdelenia štatistiky $BR_n$ za platnosti $H_0$. Keďže rozdelenie za platnosti nulovej

hypotézy nebolo možné odvodiť, uviedli Baratpour a Rad (2012) tabuľku približných kritických hodnôt testovej štatistiky, získaných pomocou Monte Carlo simulácii.

Zardasht a kol. (2015) použili na porovnanie rozdelení dvoch náhodných veličín $X$ a $Y$ s distribučnými funkciami $F$ resp. $G$ porovnávaciu distribučnú funkciu $D(u) = F^{-1}[G(u)], 0 \leq u \leq 1$. V prípade rovnosti distribučných funkcii $F$ a $G$ je $D(u) = F^{-1}[F(u)] = u, 0 \leq u \leq 1$, rovná distribučnej funkcii rovnomerného rozdelenia na intervale [0,1]. Zardasht a kol. (2015) navrhli použiť na testovanie zhodnosti rozdelení náhodných veličín $X$ a $Y$ kumulatívnu reziduálnu entropiu ich porovnávacej distribučnej funkcie $D(u)$. Bez dôkazu uvádzajú nasledujúce tvrdenie, ktoré dokážem.

**Veta 15.** *Pre kumulatívnu reziduálnu entropiu porovnávacej distribučnej funkcie $D(u)$ distribučných funkcii $F$ a $G$ platí*

$$CRE(D) = -\int_0^\infty \bar{F}(x) \log \bar{F}(x) \, \mathrm{d}G(x).$$

*Dôkaz.* Z definície kumulatívnej reziduálnej entropie pomocou substitúcie dostaneme

$$\begin{aligned} CRE(D) &= -\int_0^\infty \bar{D}(u) \log \bar{D}(u) \, \mathrm{d}u \overset{S}{=} \left[x = G^{-1}(u), u \in (0,1)\right] \\ &\overset{S}{=} -\int_0^\infty \bar{F}[G^{-1}(G(u))] \log \bar{F}[G^{-1}(G(u))] \, \mathrm{d}G(u) \\ &= -\int_0^\infty \bar{F}(x) \log \bar{F}(x) \, \mathrm{d}G(x). \end{aligned}$$

□

Nech $Y$ je nezáporná náhodná veličina s distribučnou funkciou $G$ a náhodná veličina $X$ pochádza z exponenciálneho rozdelenia $Exp(1/\lambda), \lambda > 0$ neznáme. Chceme testovať hypotézu $H_0$, že $G$ je rovná distribučnej funkcii exponenciálneho rozdelenia $Exp(1/\lambda), \lambda > 0$ neznáme, proti alternatíve, že $Y$ nepochádza z exponenciálneho rozdelenia. V tomto prípade náhodných veličín $X$ a $Y$ je kumulatívna reziduálna entropia ich porovnávacej distribučnej funkcie rovná

$$\begin{aligned} C(exp,Y) := CRE(D) &= -\int_0^\infty e^{-\lambda x} \log e^{-\lambda x} \, \mathrm{d}G(x) \\ &= \int_0^\infty \lambda x e^{-\lambda x} dG(x). \end{aligned}$$

Ako sme ukázali vyššie, ak má $Y$ exponenciálne rozdelenie, je $D(u)$ rovná distribučnej funkcii rovnomerného rozdelenia $R(0,1)$. Preto za platnosti nulovej hypotézy $H_0$ je $C(exp,Y)$ rovná kumulatívnej reziduálnej entropii rovnomerného rozdelenia $R(0,1)$ a jej presný tvar odvodíme.

Nech je náhodná veličina $Z \sim R(0,1)$, potom jej funkcia prežitia je $\bar{F}_Z(x) = 1 - x$ pre $x \in (0,1)$. Jej kumulatívnu reziduálnu entropiu upravíme pomocou

integrácie per partes na tvar

$$\begin{aligned} CRE(F_Z) &= -\int_0^\infty \bar{F}_Z(x)\log\bar{F}_Z(x)\,\mathrm{d}x = -\int_0^1 (1-x)\log(1-x)\,\mathrm{d}x \\ &= -\left[-\frac{(1-x)^2\log(1-x)}{2}\right]_0^1 + \frac{1}{2}\int_0^1 (1-x)\,\mathrm{d}x \\ &= \frac{1}{2}\lim_{x\to 1}(1-x)^2\log(1-x) + \frac{1}{2}\left[-\frac{(1-x)^2}{2}\right]_0^1 \\ &= \frac{1}{4}. \end{aligned}$$

Limita v predposlednej rovnosti je nulová, pretože podľa l'Hospitalovho pravidla

$$\lim_{x\to 1}(1-x)^2\log(1-x) = \lim_{x\to 1}\frac{\log(1-x)}{\frac{1}{(1-x)^2}} = \lim_{x\to 1}\frac{(1-x^3)}{2(1-x)} = 0.$$

Preto Zardasht a kol. (2015) zostavili test exponenciality, ktorý používa ako mieru odlišnosti rozdelenia náhodnej veličiny $Y$ a exponenciálneho rozdelenia vzťah $C(exp,Y) - 1/4$.

Nech $X_1,\ldots,X_n$ je náhodný výber z rozdelenia s distribučnou funkciou $F$, nech sú všetky $X_i \geq 0$ a nech náhodná veličina $X$ pochádza z rovnakého rozdelenia ako náhodný výber. Zardasht a kol. (2015) navrhli nasledovný odhad $C(exp,X)$, v ktorom parameter $\lambda$ odhadli pomocou $1/\bar{X}_n$

$$C_n = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{\bar{X}_n}e^{-\frac{X_i}{\bar{X}_n}}.$$

Odhad $C_n$ je zároveň testovou štatistikou, pričom na testovanie použijeme fakt, že za platnosti $H_0$ je $C(exp,X) = 1/4$. Príliš veľké a príliš malé hodnoty $C_n - 1/4$ svedčia v prospech alternatívy.

Testová štatistika $C_n$ má za platnosti nulovej hypotézy asymptoticky normálne rozdelenie, pretože Zardasht a kol. (2015) dokázali za $H_0$

$$\sqrt{n}(C_n - 1/4) \xrightarrow{D} N\left(0,\frac{5}{382}\right).$$

Odtiaľ odvodíme približný konfidenčný interval testovej štatistiky ako

$$\begin{aligned} u_{\alpha/2} \leq \sqrt{\frac{382n}{5}}(C_n - \frac{1}{4}) \leq u_{1-\alpha/2} \\ -u_{1-\alpha/2} \leq \sqrt{\frac{382n}{5}}(C_n - \frac{1}{4}) \leq u_{1-\alpha/2}. \end{aligned}$$

Preto nulovú hypotézu zamietame pre hodnoty $\sqrt{\frac{382n}{5}}|C_n - 1/4| > u_{1-\alpha/2}$, kde $u_{1-\alpha/2}$ je $(1-\alpha/2)-$kvantil štandardného normálneho rozdelenia $N(0,1)$. Jasnou výhodou tohto testu je jednoduchosť testovej štatistiky a známosť asymptotického rozdelenia za platnosti nulovej hypotézy, ktoré je normálne.

## 2.6 Testy založené na strednej reziduálnej funkcii života

Pomocou strednej reziduálnej funkcie života možno jednoznačne charakterizovať exponenciálne rozdelenie, pretože je jediné s konštantnou funkciou $m(u)$, ako dokazuje nasledujúca. Nasledujúca veta bola stručne dokázaná v Shanbhag (1970), my jej dôkaz doplníme o chýbajúce kroky.

**Veta 16.** *Nech $X_1, \ldots, X_n$ je náhodný výber z rozdelenia s neznámou distribučnou funkciou $F$ s nezáporným nosičom a konečnou kladnou strednou hodnotou a nech ich $X$ je náhodná veličina s rovnakou distribučnou funkciou. Potom náhodný výber pochádza z exponenciálneho rozdelenia práve vtedy, keď je stredná reziduálna funkcia života náhodnej veličiny $X$ konštantná, t.j. keď pre všetky $z > 0$ platí*

$$\mathsf{E}(X - z | X > z) = \mathsf{E}\, X.$$

*Dôkaz.* Nech je náhodný výber (aj náhodná veličina $X$) z exponenciálneho rozdelenia s hustotou $f(x) = \lambda e^{-\lambda x}, x \geq 0, \lambda > 0$. Potom stredná reziduálna funkcia života má pre $z > 0$ tvar

$$\begin{aligned}\mathsf{E}\,(X - z | X > z) &= \frac{1}{\bar{F}(z)} \int_z^\infty \bar{F}(x)\,\mathrm{d}x = \frac{1}{e^{-\lambda z}} \int_z^\infty e^{-\lambda x}\,\mathrm{d}x \\ &= \frac{1}{e^{-\lambda z}} \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^\infty = \frac{1}{\lambda} = \mathsf{E}\, X.\end{aligned}$$

Tým sme dokázali prvú implikáciu. Teraz predpokladajme, že pre strednú reziduálnu funkciu života $m(z)$ náhodnej veličiny $X$ platí pre $z > 0$

$$m(z) = \mathsf{E}\,(X - z | X > z) = \mathsf{E}\, X.$$

Tento vzťah môžeme prepísať do tvaru

$$\int_z^\infty (x - z)\,\mathrm{d}F(x) = \bar{F}(z)\,\mathsf{E}\, X.$$

Zo spojitosti a diferencovateľnosti ľavej strany plynie spojitosť a diferencovateľnosť pravej strany. Derivovaním ľavej strany podľa $z$ dostaneme použitím derivácie integrálu podľa parametru

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z} \int_z^\infty (x - z)\,\mathrm{d}F(x) &= \frac{\mathrm{d}}{\mathrm{d}z} \int_z^\infty (x - z) f(x)\,\mathrm{d}x \\ &= \int_z^\infty \frac{\partial}{\partial z}(x - z) f(x)\,\mathrm{d}x - (z - z) f(z) \\ &= -\int_z^\infty \mathrm{d}F(x) = -(1 - F(z)) = -\bar{F}(z).\end{aligned}$$

Derivovaním pôvodnej rovnice podľa premennej $z$ a dosadením posledného výsledku dostávame diferenciálnu rovnicu

$$-\frac{1}{\mathsf{E}\, X} \bar{F}(z) = \frac{\mathrm{d}}{\mathrm{d}z} \bar{F}(z).$$

Za počiatočnú podmienku vezmeme vlastnosť funkcie prežitia nezápornej náhodnej veličiny, pre ktorú platí

$$\lim_{z\to 0} \bar{F}(z) = \lim_{z\to 0}[1 - F(z)] = 1 \text{ a } \lim_{z\to\infty} \bar{F}(z) = 0.$$

Diferenciálnu rovnicu vyriešime nasledujúcim spôsobom

$$\begin{aligned}
\frac{1}{\bar{F}(z)} \frac{\mathrm{d}}{\mathrm{d}z} \bar{F}(z) &= -\frac{1}{\mathsf{E}\ X} \\
\int \frac{1}{\bar{F}(z)} \frac{\mathrm{d}}{\mathrm{d}z} \bar{F}(z)\, \mathrm{d}z &= \int -\frac{1}{\mathsf{E}\ X}\, \mathrm{d}z \\
\log \bar{F}(z) &= -\frac{1}{\mathsf{E}\ X} z + C_1 \\
\bar{F}(z) &= C_2 e^{-\frac{1}{\mathsf{E}\ X} z},
\end{aligned}$$

kde $C_2 = e^{C_1}$. Hodnotu konštanty dopočítame z limity $\bar{F}(z)$ pre $z \to 0$ ako

$$1 = \lim_{z\to 0} \bar{F}(z) = C_2 \lim_{z\to 0} e^{-\frac{1}{\mathsf{E}\ X} z} = C_2,$$

pretože stredná hodnota náhodnej veličiny $X$ je z predpokladu vety kladná. Označme $\lambda = 1/\mathsf{E}\ X$, potom jednoznačným riešením tejto diferenciálnej rovnice je $\bar{F}(z) = e^{-\lambda z}, z \geq 0$. Odtiaľ plynie, že náhodná veličina $X$ aj náhodny výber pochádzajú z rozdelenia $Exp(1/\lambda)$.

□

Baringhaus a Henze (2000) tvrdia, že podmienka $\mathsf{E}\,(X - z|X > z) = \mathsf{E}\ X$ je ekvivalentná podmienke

$$[\min(X,z)] = F(z)\,\mathsf{E}\ X. \tag{2.6}$$

Toto tvrdenie dokážeme nasledovne

$$\begin{aligned}
\frac{1}{\bar{F}(z)} \int_z^\infty (x - z)\, \mathrm{d}F(x) &= \mathsf{E}\ X \\
\int_z^\infty (x - z)\, \mathrm{d}F(x) &= [1 - F(z)]\,\mathsf{E}\ X \\
F(z)\,\mathsf{E}\ X &= \int_0^\infty x\, \mathrm{d}F(x) - \int_z^\infty (x - z)\, \mathrm{d}F(x) \\
F(z)\,\mathsf{E}\ X &= \int_0^z x\, \mathrm{d}F(x) + \int_z^\infty z\, \mathrm{d}F(x) \\
F(z)\,\mathsf{E}\ X &= \mathsf{E}\,[\min(X,z)].
\end{aligned}$$

Charakterizáciu exponenciálneho rozdelenia pomocou (2.6) využili Baringhaus a Henze (2000) na zostavenie dvoch testových štatistík- jednu typu KS a druhú typu CVM.

Majme náhodný výber $X_1, \ldots, X_n$ s rozdelenia s nezáporným nosičom a kladnou strednou hodnotou $EX$. Označme $Y_i = X_i/\bar{X}_n, i = 1, \ldots, n,$. Chceme testovať hypotézu

$$\begin{aligned}
&H_0 : X_1,\ldots,X_n \text{ je z } Exp(1/\lambda), \lambda > 0, \text{proti} \\
&H_1 : X_1,\ldots,X_n \text{ nie je z } Exp(1/\lambda).
\end{aligned}$$

Za platnosti nulovej hypotézy sa pre veľké $n$ správajú $Y_1,\ldots,Y_n$ približne ako $n$ nezávislých náhodných veličín s rozdelením $Exp(1)$. Pre náhodnú veličinu $Y \sim Exp(1)$ má podmienka (2.6) tvar

$$\mathsf{E}\left[\min(Y,z)\right] = F(z).$$

Skúmaním rozdielu empirických verzii pravej a ľavej strany, získaných odhadnutím neznámej distribučnej funkcie náhodného výberu $Y_1,\ldots,Y_n$ empirickou distribučnou funkciou, možno zostaviť testové štatistiky na testovanie exponenciality pôvodného náhodného výberu $X_1,\ldots,X_n$. Preto Baringhaus a Henze (2000) navrhli skúmať testovú štatistiku typu KS

$$L_n = \sqrt{n}\sup_{z\geq 0}\left|\frac{1}{n}\sum_{i=1}^{n}\min(Y_i,z) - \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{Y_i\leq z\}}\right|,$$

a štatistiku typu CVM

$$G_n = n\int_0^\infty \left[\frac{1}{n}\sum_{i=1}^{n}\min(Y_i,z) - \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{Y_i\leq z\}}\right]^2 e^{-z}\,\mathrm{d}z.$$

Autori tiež odvodili asymptotické rozdelenie testových štatistík za platnosti nulovej hypotézy. Prišli k výsledku, že pre $n\to\infty$ majú testové štatistiky $L_n$ a $G_n$ rovnaké asymptotické rozdelenie ako klasické štatistiky KS a CVM na testovanie uniformity na intervale [0,1] pre náhodný výber s rozsahom $n-1$. Pre praktické použitie testov založených na $L_n$ a $G_n$ nasimulovali Baringhaus a Henze (2000) približné kritické hodnoty testov pre rôzne rozsahy výberov. Empirické kritické hodnoty získané simuláciou sa príliš nelíšia pre $n \geq 20$ od skutočných kritických hodnôt asymptotického rozdelenia.

Jammalamadaka a Taufer (2006) predstavili celú triedu testových štatistík $JT_\alpha, \alpha \in (0,1)$, založených na charakterizácii exponenciálneho rozdelenia pomocou strednej reziduálnej funkcie života. Využili vyššie dokázané tvrdenie, že konštantnosť $m(u)$ charakterizuje exponenciálne rozdelenie. Simulačná štúdia, ktorú previedli Jammalamadaka a Taufer (2006), porovnávala silu niekoľkých testov exponenciality so silou celej škály reprezentatov triedy $JT_\alpha$. Táto štúdia ukázala veľmi silnú volatilitu výsledkov novo zavedenej triedy $JT_\alpha$ v závislosti na voľbe parametra $\alpha$. Túto triedu by bolo možné doporučiť na testovanie proti špecifickým alternatívam, kedy je voľba parametra $\alpha$ jednoduchšia. V prípade testovania proti úplne neznámej alternatíve ponúkajú iné testy založené na charakterizácii pomocou $m(u)$ spoľahlivejšiu silu testy, ktoré zaviedli Baringhaus a Henze (2000) alebo Aboukhmaseen a Aly (2016).

Aboukhmaseen a Aly (2016) vyšli z výsledkov, ktoré publikovali Jammalamadaka a Taufer (2006), a zaviedli ďalšiu triedu testových štatistík, ktorá ponúka empiricky lepšie výsledky ako trieda štatistík $JT_\alpha$. Použili rovnakú charakterizáciu pomocou konštantnosti strednej reziduálnej funkcie života.

Majme náhodný výber $X_1,\ldots,X_n$ z neznámeho rozdelenia a označme $X_(0) = 0$. Označme normalizované vzdialenosti $Y_i = (n-i+2)(X_{(i)} - X_{(i-1)})$, pre $i = 1,\ldots,n-1$. Nové testové štatistiky sú založené na veličine

$$T_n(u;\gamma) = \sqrt{n+1}\left[\frac{\frac{1}{n+1}\sum_{i=n-\lfloor(n+1)u\rfloor+2}^{n+1} Y_i^\gamma}{\left(\frac{1}{n+1}\sum_{i=1}^{n+1} Y_i\right)^\gamma} - \frac{\lfloor(n+1)u\rfloor}{n+1}\Gamma(1+\gamma)\right],$$

kde $\gamma > 0$ je pevné. Aboukhmaseen a Aly (2016) doporučujú ako najsilnejšie z nimi zostavených štatistík štatistiky

$$A_{n,1} = \frac{1}{n+1} \sum_{k=1}^{n+1} T_n^2(k/(n+1); 1)$$

a

$$A_{n,2} = (n+1) \sum_{k=1}^{n+1} \frac{T_n^2(k/(n+1); 1)}{k(n-k+1)}.$$

Skúmaním asymptotického správania testových štatistík odvodili Aboukhmaseen a Aly (2016), že testová štatistika $A_{n,1}$ má rovnaké asymptotické rozdelenie, ako štatistika $G_n$, ktorú zaviedli Baringhaus a Henze (2000), ktoré je zároveň limitným rozdelením klasickej CVM štatistiky na testovanie uniformity na intervale (0,1).Pre praktické použitie testov uviedli Aboukhmaseen a Aly (2016) na základe Monte Carlo simulácii tabuľku približných kritických hodnôt pre konečné rozsahy náhodného výberu.

## 2.7 Test založený na integrovanej distribučnej funkcii

Inovatívny spôsob testovania exponenciality zvolil vo svojej práci Klar (2001), ktorý použil na charakterizáciu triedy jednoparametrických exponenciálnych rozdelení integrovanú distribučnú funkciu náhodnej veličiny $X$ tak, ako bola definovaná v definícii 4. Pre exponenciálne rozdelenie $Exp(1/\lambda), \lambda > 0$, a $t > 0$ má tvar

$$\Psi(t,\lambda) = \int_t^\infty [1 - F(x)]\,\mathrm{d}x = \int_0^\infty e^{-\lambda x}\,\mathrm{d}x = \frac{e^{-\lambda t}}{\lambda}.$$

Majme náhodný výber $X_1, \ldots, X_n$ z neznámeho rozdelenia s kladným nosičom a konečnou strednou hodnotou $\mathsf{E}\ X$. Testujeme hypotézu

$$H_0 : X_1,...,X_n \text{ je z } Exp(1/\lambda), \lambda > 0 \text{ neznáme, proti}$$
$$H_1 : X_1,...,X_n \text{ nie je z } Exp(1/\lambda).$$

Klar (2001) navrhol testovať exponencialitu náhodného výberu porovnaním empirickej integrovanej distribučnej funkcie $\Psi_n(t)$, ktorá bola definovaná v definícii 5, s integrovanou distribučnou funkciou za platnosti hypotézy. Keďže parameter testovaného exponenciálneho rozdelenia nie je známy, autor navrhol použiť jeho maximálne vierohodný odhad $\hat{\lambda}_n = 1/\bar{X}_n$ a porovnávať $\Psi_n(t)$ s $\Psi(t,\hat{\lambda}_n)$. Testová, štatistika, ktorú zostavil Klar (2001), má tvar

$$K_n = \hat{\lambda}_n^3 \int_0^\infty \{\sqrt{n}[\Psi_n(t) - \Psi(t,\hat{\lambda}_n)]\}^2\,\mathrm{d}t.$$

Veľké hodnoty testovej štatistiky naznačujú, že sa empirická integrovaná distribučná funkcia príliš líši od integrovanej distribučnej funkcie za platnosti hypotézy, a teda svedčia v prospech alternatívy. Autor uviedol výpočetne výhodný tvar testovej štatistiky $K_n$, ktorý v nasledujúcej vete odvodíme a overíme.

**Veta 17.** *Označme $Y_i = \hat{\lambda}_n X_i, i = 1, \ldots, n$. Potom pre testovú štatistiku $K_n$ platí*

$$K_n = \frac{n}{2} - 2\sum_{i=1}^{n} e^{-Y_i} - \frac{1}{3n}\sum_{i=1}^{n}(n-i-1)Y_{(i)}^3 + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i}^{n} Y_{(i)}^2 Y_{(j)}.$$

*Dôkaz.* Najprv upravme tvar testovej štatistiky $K_n$ dosadením funkcii a vložením parametra $\hat{\lambda}_n^3$ do vnútra integrálu na

$$\begin{aligned} K_n &= \hat{\lambda}_n^3 \int_0^\infty \left\{ \sqrt{n}\left[ \frac{1}{n}\sum_{i=1}^{n}(X_i - t)\mathbf{1}_{\{X_i > t\}} - \frac{e^{-\hat{\lambda}_n t}}{\hat{\lambda}_n} \right] \right\}^2 \mathrm{d}t \\ &= \int_0^\infty \left\{ \sqrt{n}\left[ \frac{1}{n}\sum_{i=1}^{n}(\hat{\lambda}_n X_i - \hat{\lambda}_n t)\mathbf{1}_{\{\hat{\lambda}_n X_i > \hat{\lambda}_n t\}} - e^{-\hat{\lambda}_n t} \right] \right\}^2 \hat{\lambda}_n \, \mathrm{d}t. \end{aligned}$$

Použitím substitúcie $u = \hat{\lambda}_n t$ a označením náhodných veličín $Y_i = \hat{\lambda}_n X_i, i = 1, \ldots, n$, prevedieme štatistiku $K_n$ do tvaru

$$\begin{aligned} K_n &= \int_0^\infty \left\{ \sqrt{n}\left[ \frac{1}{n}\sum_{i=1}^{n}(Y_i - u)\mathbf{1}_{\{Y_i > u\}} - e^{-u} \right] \right\}^2 \mathrm{d}u \\ &= \frac{1}{n}\int_0^\infty \left\{ \sum_{i=1}^{n}\left[ (Y_i - u)\mathbf{1}_{\{Y_i > u\}} - e^{-u} \right] \right\}^2 \mathrm{d}u. \end{aligned}$$

Výraz v integrante umocníme a podľa vzorca pre druhú mocninu. Pri násobení dvoch indikátorov $\mathbf{1}_{\{Y_i>u\}}\mathbf{1}_{\{Y_j>u\}}$ dostaneme nenulovú hodnotu práve vtedy keď $Y_i > u$ a $Y_j > u$, čo je rovné $\mathbf{1}_{\{\min(Y_i,Y_j)>u\}}$. Dostávame tvar

$$K_n = \frac{1}{n}\int_0^\infty \sum_{i=1}^{n}\sum_{j=1}^{n}\left[ (Y_i - u)(Y_j - u)\mathbf{1}_{\{\min(Y_i,Y_j)>u\}} - 2e^{-u}(Y_i - u)\mathbf{1}_{\{Y_i>u\}} + e^{-2u} \right] \mathrm{d}u.$$

Zámenou sumy a integrálu a následným výpočtom integrálov vo vnútri sumy upravíme štatistiku do tvaru

$$\begin{aligned} K_n &= \frac{1}{n}\sum_{i,j=1}^{n}\left\{ \int_0^{\min(Y_i,Y_j)} (Y_i - u)(Y_j - u)\,\mathrm{d}u - 2\int_0^{Y_i} (Y_i - u)e^{-u}\,\mathrm{d}u \right. \\ &\qquad \left. + \int_0^\infty e^{-2u}\,\mathrm{d}u \right\} \\ &= \frac{1}{n}\sum_{i,j=1}^{n}\left\{ \left[ Y_i Y_j u - \frac{(Y_i + Y_j)}{2}u^2 + \frac{u^3}{3} \right]_0^{\min(Y_i,Y_j)} - 2\left[ e^{-u}(1 + u - Y_i) \right]_0^{Y_i} + \frac{1}{2} \right\} \\ &= \frac{1}{n}\sum_{i,j=1}^{n}\left\{ Y_i Y_j \min(Y_i,Y_j) - \frac{(Y_i + Y_j)}{2}\min(Y_i,Y_j)^2 + \frac{\min(Y_i,Y_j)^3}{3} \right. \\ &\qquad \left. - 2(e^{-Y_i} + Y_i - 1) + \frac{1}{2}. \right\} \end{aligned}$$

Posledné dva členy sčítanca, môžeme jednoducho upraviť, použitím vzťahu, že pre $Y_i$, platí $1/n\sum_{i=1}^{n} Y_i = 1$, do tvaru

$$\frac{1}{n}\sum_{i,j=1}^{n}\left[ -2(e^{-Y_i} + Y_i - 1) + \frac{1}{2} \right] = -2\sum_{i=1}^{n} e^{-Y_i} - 2n + 2n + \frac{n}{2} = -2\sum_{i=1}^{n} e^{-Y_i} + \frac{n}{2}.$$

Zvyšné členy upravíme tým, že nahradíme $Y_i$ ich poriadkovými štatistikami, aby sme mohli určiť minimum, keďže poriadkové štatistiky sú usporiadané vzostupne. Dvojice indexov $(i,j)$ rozdelíme na diagonálne $(i,i)$ a mimodiagonálne $(i,j), i \neq j$. Očividne skúmaný výraz je symetrický v $(i,j)$, preto jeho hodnota pre dvojice indexov nad diagonálou $(i,j), i < j$, je rovná hodnote pre dvojice pod diagonálou $(i,j), i > j$. Preto platí

$$\begin{aligned} S :&= \frac{1}{n}\sum_{i,j=1}^{n}\left\{Y_iY_j\min(Y_i,Y_j) - \frac{(Y_i+Y_j)}{2}\min(Y_i,Y_j)^2\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(\frac{Y_{(i)}^2Y_{(j)}}{2} - \frac{Y_{(i)}^3}{6}\right).2 + \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i^3}{3} \\ &= -\frac{1}{3n}\sum_{i=1}^{n}(n-i-1)Y_{(i)}^3 + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n}Y_{(i)}^2Y_{(j)}. \end{aligned}$$

A nakoniec dostávame zjednodušený tvar testovej štatistiky $K_n$ v tvare

$$\begin{aligned} K_n &= S - 2\sum_{i=1}^{n}e^{-Y_i} + \frac{n}{2} \\ &= \frac{n}{2} - 2\sum_{i=1}^{n}e^{-Y_i} - \frac{1}{3n}\sum_{i=1}^{n}(n-i-1)Y_{(i)}^3 + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i}^{n}Y_{(i)}^2Y_{(j)}, \end{aligned}$$

a tým je tvrdenie vety dokázané.

□

Nech $\alpha \in (0,1)$ a $z_n(\alpha)$ značí $(1-\alpha)$-kvantil rozdelenia $K_n$ za platnosti nulovej hypotézy. Test exponenciality založený na testovej štatistike $K_n$ zamieta nulovú hypotézu na hladine $\alpha$ v prospech alternatívy v prípade, že $K_n > z_n(\alpha)$. Klar (2001) tvrdí, že asymptotické rozdelenie testovej štatistiky za platnosti nulovej hypotézy sa zdá byť nemožné získať analyticky. Preto autor spolu s testom publikoval tabuľku empirických kritických hodnôt, pre rôzne rozsahy výberu a hladiny testu. Tieto približné kritické hodnoty boli získané Monte Carlo simuláciami.

So zámerom vylepšiť silu testu pridal Klar (2001) do testovej štatistiky váhovú funkciu $e^{-at}, a > 0$, a modifikoval testovú štatistiku na tvar

$$K_{n,a} = (a\hat{\lambda}_n)^3\int_0^\infty\{\sqrt{n}[\Psi_n(t) - \Psi(t,\hat{\lambda}_n)]\}^2e^{-a\hat{\lambda}_nt}\,\mathrm{d}t.$$

Analogickým spôsobom ako sme použili v dôkaze vety 17 upravil autor testovú štatistiku na výpočetne výhodný tvar

$$\begin{aligned} K_{n,a} =& \frac{2(3a+2)n}{(2+a)(1+a)^2} - 2a^3\sum_{i=1}^{n}\frac{e^{-(1+a)Y_i}}{(1+a)^2} - \frac{2}{n}\sum_{i=1}^{n}e^{-aY_i} \\ &+ \frac{2}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n}[a(Y_{(j)} - Y_{(i)})]e^{-aY_{(i)}}. \end{aligned}$$

Rovnakým spôsobom ako v prípade štatistiky $K_n$ získal Klar (2001) pomocou Monte Carlo simulácii empirické kritické hodnoty pre testy založene na $K_{n,a}$ pre $a = 1, 5, 10, 20$.

## 2.8 Test založený na Giniho indexe

Gail a Gastwirth (1978) navrhli test exponenciality založený na známom Giniho indexe. Majme náhodný výber $X_1, \ldots, X_n$ z rozdelenia s konečným druhým momentom. Chceme testovať hypotézu, že tento náhodný výber pochádza z exponenciálneho rozdelenia s neznámym parametrom proti alternatíve, že výber nemá exponenciálne rozdelenie. Potom Giniho štatistika, ktorú navrhnujú Gail a Gastwirth (1978) použiť, má tvar

$$GI_n = \frac{\sum_{i=1}^{n-1}[i(n-1)(X_{(i+1)} - X_{(i)})]}{(n-1)\sum_{i=1}^{n} X_i}.$$

Ako dokázali Gail a Gastwirth (1978), testová štatistika $GI_n$ má za platnosti nulovej hypotézy asymptoticky normálne rozdelenie

$$\sqrt{12(n-1)}(GI_n - 1/2) \xrightarrow{D} N(0,1).$$

Tento výsledok o asymptoticky normálnom rozdelení štatistiky využijeme na výpočet obojstranného približného konfidenčného intervalu. Nulovú hypotézu zamietame na hladine približne $\alpha$ pre hodnoty testovej štatistiky

$$\left|\sqrt{12(n-1)}(GI_n - 1/2)\right| > u_{1-\alpha/2},$$

kde $u_{1-\alpha/2}$ označuje $(1-\alpha/2)$-kvantil štandardného normálneho rozdelenia. Autori odvodili aj presné rozdelenie testovej štatistiky, avšak pre jednoduchosť aplikácie testu je výhodnejšie použiť symptotickú normalitu testovej štatistiky. Ako ukázali Gail a Gastwirth (1978) v simulačnej štúdii, aproximácia rozdelenia štatistiky normálnym rozdelením je veľmi presné už pre malé $n$, napr. $n = 10$.

## 2.9 Test založený na normalizovaných vzdialenostiach dát

Jammalamadaka a Taufer (2003) zaviedli dve testové štatistiky na testovanie exponenciality založené na charakterizácii exponenciálneho rozdelenia pomocou normalizovaných vzdialeností pozorovaní. Obe testové štatistiky preukázali v simulačnej štúdii porovnávajúcej silu rôznych testov exponenciality, ktorú previedli Jammalamadaka a Taufer (2003) takmer identickú silu, preto bude ďalej predstavená len jedna z nich.

Seshari a kol. (1969) dokázali nasledujúcu vlastnosť, ktorá charakterizuje exponenciálne rozdelenie.

**Veta 18.** *Nech $X_1, \ldots, X_n$ je náhodný výber z rozdelenia $Exp(1/\lambda), \lambda > 0$, a označme $X_{(0)} = 0$. Potom normalizované vzdialenosti*

$$Y_i = (n - i + 1)(X_{(i)} - X_{(i-1)}), \quad i = 1, \ldots, n,$$

*sú nezávislé a rovnako rozdelené a platí $Y_i \sim Exp(1/\lambda)$, pre $i = 1, \ldots, n$. Táto vlastnosť charakterizuje exponenciálne rozdelenie.*

Na základe tejto charakterizácie autori zostavili testovú štatistiku typu KS. Keďže za platnosti nulovej hypotézy pochádzajú $X_1, \ldots, X_n$ a $Y_1, \ldots, Y_n$ z rovnakého exponenciálneho rozdelenia, prirodzene sa ponúka porovnať ich empirické distribučné funkcie. Porovnáme vzdialenosť medzi empirickou distribučnou funkciou $F_n$ výberu $X_1, \ldots, X_n$ a empirickou distribučnou funkciou $G_n$ transformovaného výberu $Y_1, \ldots, Y_n$ pomocou testovej štatistiky

$$T_{1,n} = \sqrt{\frac{n}{2}} \sup_{0 \leq t < \infty} |F_n(t) - G_n(t)|.$$

Za platnosti nulovej hypotézy, že pôvodný náhodný výber pochádza z exponenciálneho rozdelenia s neznámym parametrom, by mali byť empirické distribučné funkcie $F_n$ a $G_n$ blízko seba a hodnota testovej štatistiky by mala byť malá. Hodnota testovej štatistiky sa vypočítava rovnako ako v prípade dvojvýberového KS testu, kde testované dva porovnávané výbery sú $X_1, \ldots, X_n$ a $Y_1, \ldots, Y_n$, ktoré však nie sú nezávislé.

Jammalamadaka a Taufer (2003) dokázali, že $T_1$,$n$ má za platnosti nulovej hypotézy asymptoticky rovnaké rozdelenie ako klasická KS štatistika, ktorá má asymptoticky Kolmogorovovo rozdelenie. Dokázali, že pre testovú štatistiku $T_1$,$n$ a $t \geq 0$ teda platí

$$\lim_{n \to \infty} \mathsf{P}(T_{1,n} > t) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 t^2}.$$

Nulovú hypotézu zamietame na hladine približne $\alpha$ pre hodnoty testovej štatistiky $T_1$,$n$ väčšie ako $(1 - \alpha)$-kvantil Kolmogorovovho rozdelenia. Hodnoty týchto kvantilov je nutné dopočítať numericky, alebo nájsť v tabuľkách.

# 3. Porovnanie testov

V tejto kapitole som sa rozhodla pomocou simulácii porovnať silu niektorých testov, ktoré boli predstavené v kapitole 2. Henze a Meintanis (2005) publikovali rozsiahlu Monte Carlo simulačnú štúdiu porovnávajúcu celú škálu testov exponenciality a väčšina z nich bola predstavená v predchádzajúcej časti tejto práce. Autori uvažovali dva rôzne rozsahy náhodného výberu $n = 20$ a $n = 50$ pre hladinu testu $\alpha = 0{,}05$. Vychádzajúc z ich výsledkov som vybrala pre $n = 20$ tri, ktoré sa ukázali byť najsilnejšie proti uvažovaným alternatívam. Konkrétne sú to testy založené na Laplaceovej transformácii, $BH_{n,a}$, charakteristickej funkcii, $CF^2_{n,a}$, a strednej reziduálnej funkcii života, $G_n$. K týmto trom štatistikám som sa rozhodla pridať ešte štatistiku $KLV_{m,n}$, založenú na charakterizácii pomocou Kullbackovej-Leiblerovej vzdialenosti využívajúc Vasickov odhad entropie. Silu tejto štatistiky som sa rozhodla skúmať z dôvodu, že testové štatististiky $KLE_{m,n}$ a $KLC_{m,n}$, ktoré publikovali Choi a kol. (2004), založené na rovnakej charakterizácii exponenciálneho rozdelenia pomocou Shannonovej entropie, neboli v simulačnej štúdii porovnané s testovou štatistikou $KLV_{m,n}$. Ako ukázali Choi a kol. (2004) na základe simulácii, test založený na štatistike $KLC_{m,n}$, využívajúcej Correov odhad entropie, vykazuje empiricky väčšiu silu ako test založený na $KLE_{m,n}$, využívajúcej Van Esov odhad, preto v nasledujúcej simulačnej štúdii bude zahrnutá štatistika $KLC_{m,n}$. Henze a Meintanis (2005) uvažovali v simulačnej štúdii celkovo až 9 alternatívnych rozdelení, ja som sa rozhodla vybrať 6 z nich, pre ktoré som previedla menšiu simulačnú štúdiu s 1 000 simuláciami.

Cieľom mojej simulačnej štúdie bolo určiť empirickú silu niekoľkých testov exponenciality proti daným alternatívam a porovnať nové metódy testovania exponenciality so staršími testami na základe ich empirickej sily. Zároveň bolo cieľom porovnať skúmané testy medzi sebou podľa charakterizácie exponenciálneho rozdelenia, ktorú používajú (napr. jednotlivé testy založené na entropii).

Na porovnanie výsledkov uvádzam výsledky časti simulačnej štúdie, ktorú previedli Henze a Meintanis (2005), pre 4 vyššie spomínané testové štatistiky(viz Tabulka 3.1). V tabulke nájdeme percentuálny podiel, zaokrúhlený na najbližšie celé číslo, nasimulovaných náhodných výberov o rozsahu $n = 20$ z jednotlivých alternatívnych rozdelení označených ako signifikantný vrámci Monte Carlo simulačnej štúdie s 10 000 simuláciami pre $\alpha = 0{,}05$. Je to podiel simulovaných náhodných výberov z celkového počtu 10 000, pre ktoré daný test zamietol nulovú hypotézu $H_0$ v prospech alternatívy $H_1$ na hladine približne $\alpha$, vyjadrený ako percento zaokrúhlené na najbližší percentuálny bod.

V simulačnej štúdii testujem pre každý náhodný výber hypotézu

$$H_0 : X_1,...,X_{20} \text{ je z } Exp(1/\lambda), \lambda > 0 \text{ neznáme, proti}$$
$$H_1 : X_1,...,X_{20} \text{ nie je z } Exp(1/\lambda).$$

Simulácie aj jednotlivé testy som implementovala v jazyku R, čo popíšem v ďalšej časti. Uvažované alternatívy sú

- Weibullovo rozdelenie $W(\theta)$ s hustotou $\theta x^{\theta-1}e^{-x^\theta}, x \geq 0$, pre hodnoty parametra $\theta$ rovné 0,8 a 1,4,
- Gamma rozdelenie $\Gamma(\theta)$ s hustotou $\Gamma(\theta)^{-1}x^{\theta-1}e^{-x}, x \geq 0$, pre hodnoty parametra $\theta$ rovné 1, 2 a 0,4,

- rovnomerné rozdelenie $R(0,1)$ na intervale $[0,1]$.

Medzi porovnávané testové štatistiky nebola zaradená jediná nová testová štatistika $L_n$ založená na Hankelovej transformácii, ktorá nemá výpočetne výhodný tvar. Rovnako medzi porovnávané testy neboli zahrnuté $\chi^2$ testy dobrej zhody, ktoré sú známe tým, že svojou silou moderným metódam nekonkurujú. Porovnávanými testami sú testy používajúce štatistiky

- $BH_{n,\,1,5}$, založená na Laplaceovej transformácii (Baringhaus a Henze, 1991),
- $G_n$ typu CVM, založená na strednej reziduálnej funkcii života (Baringhaus a Henze, 2000),
- $CF^2_{n,\,2,5}$, založená na charakteristickej funkcii (Henze a Meintanis, 2005),
- $KLV_{4,n}$, založená na Vasickovom odhade entropie (Ebrahimi a kol., 1992),
- $KLC_{4,n}$, založená na Correovom odhade entropie (Choi a kol., 2004),
- $LW_{n,m^*}$, kde $m^* = \lfloor \sqrt{n} + 1/2 \rfloor$, založená na Lin-Wongovej vzdialenosti (Abbasnejad a kol., 2012),
- $TV^Z$, založená na Rényiho entropii (Abbasnejad, 2012),
- $BR_n$, založená na kumulatívnej reziduálnej entropii (Baratpour a Rad, 2012),
- $C_n$, založená na kumulatívnej reziduálnej entropii (Zardasht a kol., 2015),
- $A_{n,1}$ a $A_{n,2}$, založené na strednej reziduálnej funkcii života (Aboukhmaseen a Aly, 2016),
- $GI_n$, založená na Giniho indexe (Gail a Gastwirth, 1978),
- $T_{1,n}$, založená na normalizovaných vzdialenostiach (Jammalamadaka a Taufer, 2003).

V prípade, že testová štatistika závisí na voľbe parametra, zvolila som v každom prípade buď hodnotu doporučovanú autormi, v prípade $KLV_{4,n}, KLC_{4,n}$ a $LW_{n,m^*}$, alebo hodnotu, ktorá bola simulačnou štúdiou overená ako najvhodnejšia pre neznámu alternatívu, v prípade $BH_{n,\,1,5}$ a $CF^2_{n,\,2,5}$.

## 3.1 Implementácia testov v jazyku R

Simulačnú štúdiu som implementovala v jazyku R. Niektoré z testov, ktoré uvažujem, sú implementované v balíčku EWGoF a tieto funkcie som využila, konkrétne testy založené na štatistikách $GI_n, BH_{n,\,1,5}, CF^2_{n,\,2,5}$, a $G_n$, je možné previesť pomocou nasledujúcich príkazov.

```
>x<-rexp(20,1)
>EDF_NS.test(x,type="G")
>LRI.test(x,type="BH",a=1.5)
>CF.test(x,type="T1",a=2.5)
>LRI.test(x,type="BHC")
```

| **Alternatíva** | $BH_{n,1,5}$ | $G_n$ | $CF^2_{n,2,5}$ | $KLV_{4,n}$ |
|---|---|---|---|---|
| $W(0{,}8)$ | 24 | 22 | 4 | 11 |
| $W(1{,}4)$ | 37 | 35 | 45 | 20 |
| $\Gamma(0{,}4)$ | 80 | 75 | 33 | 51 |
| $\Gamma(1)$ | 5 | 5 | 5 | 5 |
| $\Gamma(2)$ | 51 | 47 | 55 | 33 |
| $R(0{,}1)$ | 61 | 70 | 86 | 31 |

Tabulka 3.1: Percentuálny podiel simulovaných náhodných výberov označených ako signifikantný v Monte Carlo štúdii s 10 000 simuláciami ($n = 20$, $\alpha = 0.05$) (Henze a Meintanis, 2005).

Ostatné štatistiky som implementovala samostatne a testy založené na nich som previedla pomocou testovacieho postupu a približných kritických hodnôt, ktoré pre každý test určili jeho autori a boli popísané v kapitole 2. Výnimkou sú testy založené na testových štatistikách $TV^Z$ a $LW_n$. Testy založené na empirický kritických hodnotách publikovaných pre dané testy nedosahovali ani približnú hladinu $\alpha = 0{,}05$. Preto som pre tieto testové štatistiky pomocu simulácie s 10 000 simuláciami určila približné kritické hodnoty (viz Tabulka 3.2). Tieto kritické hodnoty sa líšia od tých, ktoré určili autori. Testy prevedené pomocou nasimulovaných približných kritických hodnôt už dosahujú hladinu približne $\alpha = 0{,}05$.

Teraz uvediem príklad prevedenia simulačnej štúdie s 1 000 simuláciami prevedenú pre testovú štatistiku $C_n$ a alternatívu $\Gamma(0{,}4)$. Ukážky simulácii pre zvyšné štatistiky možno nájsť v Prílohe A tejto bakalárskej práce.

```
> n<-20
> asim<-1000
> p<-0
> for(i in 1:asim)
+ {set.seed(i)
+  x<-rgamma(n,0.4)
+  xn<-mean(x)
+  y<-x/xn
+  z<-c()
+  for(j in 1:n){
+    zi<-y[j]*exp(-y[j])
+    z<-c(z,zi)
+  }
+  Cn<-sum(z)/n
+  test<-abs(sqrt(382*n/5)*(Cn-0.25))
+  r<-ifelse(test>1.96,1,0)
+  p<-p+r}
> print(p/asim*100)
[1] 79.5
```

| $n$ | $LW_{n,\alpha}$ | $T^Z_\alpha$ |
|---|---|---|
| 5 | 0,3968 | 1,1670 |
| 10 | 0,2218 | 0,5990 |
| 20 | 0,0917 | 0,3298 |
| 30 | 0,0330 | 0,2295 |

Tabulka 3.2: Približné kritické hodnoty pre štatistiky $LW_n$ a $TV^Z$ získané simulačnou štúdiou s 10 000 simuláciami, $\alpha = 0{,}05$.

## 3.2 Výsledky simulačnej štúdie

Výsledné hodnoty pre jednotlivé štatistiky možno nájsť v Tabulke 3.3. Výsledky sú konzistentné s výsledkami v Tabulke 3.1, avšak líšia sa pre testovú štatistiku $CF^2_{n,\,2,5}$. Výsledky, ktoré prezentovali Henze a Meintanis (2005), naznačujú menšiu silu testovej štatistiky $CF^2_{n,\,2,5}$, ako výsledky v mojej simulačnej štúdii. Z výsledkov prezentovaných v Tabulke 3.3 pre jednotlivé štatistiky možno utvoriť nasledujúce závery:

- Spomedzi všetkých uvažovaných testov ukázali najväčšiu empirickú silu testy založené na testových štatistikách $BH_{n,\,1,5}$, $KLV_{4,n}$, $LW_{n,m^*}$ a $A_{n,1}$. Je však nutné dodať, že testová štatistika $LW_{n,m^*}$ preukázala takmer nulovú silu v prípade alternatív $W(0{,}8)$ a $\Gamma(0{,}4)$.

- Proti všetkým uvažovaným alternatívam za ostatnými testami zaostáva test založený na $T_{1,n}$, ktorý nedosahuje ani porovnateľné výsledky s ostatnými testami.

- Všetky skúmané testy majú značne menšiu silu proti alternatíve $W(0{,}8)$, v porovnaní s ostatnými alternatívami.

- Porovnaním testov založených na charakterizácii pomocou strednej reziduálnej funkcii života má test založený na $A_{n,2}$ menšiu empirickú silu ako zvyšné dva. Najlepším testom v tejto skupine testov je test založený na $A_{n,1}$, ktorý patrí medzi najsilnejšie aj v porovnaní s ostatnými testami.

- Porovnaním testov založených na entropii má test založený na $KLV_{4,n}$ najväčšiu silu, pretože $LW_n$ zlyháva v prípade dvoch alternatív. Testy používajúce štatistiky $TV^Z$ a $C_n$ majú výrazne lepšie výsledky v prípade alternatív $W(0{,}8)$ a $\Gamma(0{,}4)$, ale zaostávajú v prípade ostatných alternatív.

| **Testová štatistika** | $W(0,8)$ | $W(1,4)$ | $\Gamma(0,4)$ | $\Gamma(1)$ | $\Gamma(2)$ | $R(0,1)$ |
|---|---|---|---|---|---|---|
| $BH_{n,\,1,5}$ | 25 | 39 | 82 | 5 | 52 | 63 |
| $G_n$ | 23 | 38 | 77 | 5 | 48 | 72 |
| $CF^2_{n,\,2,5}$ | 19 | 34 | 67 | 5 | 39 | 81 |
| $KLV_{4,\,n}$ | 4 | 43 | 34 | 5 | 54 | 90 |
| $KLC_{4,\,n}$ | 2 | 41 | 24 | 5 | 51 | 89 |
| $LW_{n,m^*}$ | 1 | 46 | 0 | 5 | 53 | 91 |
| $TV^Z$ | 20 | 14 | 83 | 5 | 32 | 42 |
| $BR_n$ | 7 | 36 | 26 | 5 | 39 | 92 |
| $C_n$ | 19 | 35 | 80 | 5 | 50 | 50 |
| $A_{n,\,1}$ | 25 | 37 | 82 | 5 | 47 | 75 |
| $A_{n,\,2}$ | 21 | 30 | 75 | 5 | 39 | 67 |
| $GI_n$ | 24 | 38 | 79 | 5 | 48 | 70 |
| $T_{1,\,n}$ | 4 | 23 | 2 | 5 | 30 | 47 |

Tabulka 3.3: Percentuálny podiel simulovaných náhodných výberov označených ako signifikantný v simulačnej štúdii ($n = 20$, $\alpha = 0.05$, 1 000 simulácii).

# Záver

Cieľom tejto bakalárskej práce bolo prezentovať prehľad najrôznejších prístupov k testovaniu hypotéz o exponencialite náhodného výberu a prezentované testy porovnať. V kapitole 2 boli popísané ako klasické metódy testovania dobrej zhody, tak aj moderné metódy používajúce špecifické charakterizácie exponenciálneho rozdelenia. Medzi klasické metódy, ktoré boli popísané patria $\chi^2$-testy dobrej zhody. Najmä Pearsonova $\chi^2$ štatistika, patrila k bodom zvratu v testovaní dobrej zhody. Avšak silou sa dnešným modernejším metódam $\chi^2$-testy dobrej zhody nevyrovnávajú. Ďalším spôsobom, ktorý možno zaradiť medzi klasický, je testovanie pomocou Kolmogorovovej-Smirnovovej alebo Cramérovej-von Misésovej štatistiky. Napriek tomu, že vo svojej klasickej podobe testujú vopred presne známe exponenciálne rozdelenie, autori, ako napríklad Lilliefors (1969), Finkelstein a Schafer (1971) alebo Van Soest (1969), boli schopní ich klasické verzie rozšíriť na testovanie exponenciálneho rozdelenia s neznámym parametrom. Použitie štatistík typu Kolmogorov-Smirnov alebo Cramér-von Misés je dodnes jednou zo základných metód testovania dobrej zhody.

Ďalej boli predstavené viaceré moderné prístupy testovania exponenciality. Medzi metódy používajúce integrálne transformácie boli zaradené testy založené na Laplaceovej a Hankelovej transformácii náhodnej veličiny, alebo charakteristickej funkcii. Ďalej boli spomedzi testov využívajúcich entropiu predstavené testy založené na Shannonovej, Rényiho a kumulatívnej reziduálnej entropii a Lin-Wongovej vzdialenosti. Novšie prístupy testovania charakterizujú exponenciálne rozdelenie pomocou strednej reziduálnej funkcie života. V závere kapitoly 2 boli predstavené aj testy využívajúce Giniho index, empirickú integrovanú distribučnú funkciu a normalizované vzdialenosti.

Novšie testy exponenciality, ktoré boli teoreticky popísané v kapitole 2, boli v kapitole 3 porovnané na základe simulačnej štúdie. V nadväznosti na výsledky rozsiahlej Monte Carlo simulačnej štúdie, ktorú publikovali Henze a Meintanis (2005), bola prevedená simulačná štúdia pre rozsah náhodného výberu $n = 20$ a hladinu testu $\alpha = 0{,}05$ porovnávajúca silu 14 testov exponenciality proti 6 alternatívam. Ako najsilnejšie proti uvažovaným alternatívam sa ukázali byť testy založené na štatistikách $BH_{n,1,5}$, $KLV_{4,n}$, $LW_{n,m^*}$ a $A_{n,1}$. Štatistiku $BH_{n,1,5}$ zaviedli Baringhaus a Henze (1991) a je založená na charakterizácii exponenciálneho rozdelenia pomocou Laplaceovej transformácie. Testová štatistika $KLV_{4,n}$, ktorú navrhli Ebrahimi a kol. (1992), je založená na charakterizácii pomocou Vasickovho odhadu Shannonovej entropie. Obe testové štatistiky $LW_{n,m^*}$ a $A_{n,1}$ patria medzi najnovšie publikované testy exponenciality. Prvý z nich, založený na Lin-Wongovej vzdialenosti, publikoval Abbasnejad a kol. (2012). Druhý publikoval Aboukhmaseen a Aly (2016) a je založený na charakterizácii exponenciálneho rozdelenia pomocou strednej reziduálnej funkcie života.

Testovanie exponenciality dát je využívané v rôznych sférach vedy a techniky a preto sa dá očakávať, že zostane aj naďalej predmetom záujmu odbornej verejnosti. Preto rešerš, ktorá bola prevedená v tejto bakalárskej práci, bude pravdepodobne možné rozšíriť o nové spôsoby testovania exponenciality už v najbližších rokoch.

# Zoznam použitej literatúry


Abbasnejad, M. (2012). Testing exponentiality based on Renyi entropy of transformed data. *Journal of Statistical Research of Iran*, **8**(2), 149–162.

Abbasnejad, M., Arghami, N. a Tavakoli, M. (2012). A goodness-of-fit test for exponentiality based on Lin-Wong information. *Journal of the Iranian Statistical Society*, **11**(2), 191–202.

Aboukhmaseen, S. a Aly, E. (2016). On some tests for exponentiality based on the mean residual life function. *Kuwait Journal of Science*, **43**(2).

Ahmad, I. a Alwasel, I. (1999). A goodness-of-fit test for exponentiality based on the memoryless property. *Journal of the Royal Statistical Society, Series B*, **61**(3), 681–689.

Anderson, T. a Darling, D. (1952). Asymptotic theory of certain "goodness of fit"criteria based on stochastic processes. *The Annals of Mathematical Statistics*, **23**(2), 193–212.

Angus, J. (1982). Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation. *Journal of Statistical Planning and Inference*, **6**, 241–251.

Baratpour, S. a Rad, A. (2012). Testing goodness-of-fit for exponential distribution based on cumulative residual entropy. *Communications in Statistics, Theory and Methods*, **41**(8), 1387–1396.

Baringhaus, L. a Henze, N. (1991). A class of consistent tests for exponentiality based on the empirical Laplace transform. *Annals of the Institute of Statistical Mathematics*, **43**, 551–564.

Baringhaus, L. a Henze, N. (2000). Tests of fit for exponentiality based on a characterization via the mean residual life function. *Statistical Papers*, **41**, 225–236.

Baringhaus, L. a Taherizadeh, F. (2013). A K-S type test for exponentiality based on Hankel transform. *Communications in Statistics, Theory and Methods*, **42**, 3781–3792.

Birch, M. (1964). A new proof of the Pearson-Fisher theorem. *The Annals of Mathematical Statistics*, **35**(2), 817–824.

Choi, B., Keeyoung, K. a Song, S. (2004). Goodness-of-fit test for exponentiality based on Kullback-Leibler information. *Communications in Statistics, Simulation and Computation*, **33**(2), 525–536.

Cramer, H. (1928). On the composition of elementary errors. *Scandinavian Actuarial Journal*, (1), 13–74.

Cressie, N. a Read, T. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, **46**, 440–464.

Darling, D. (1957). The Kolmogorov-Smirnov, Cramer-von Mises tests. *The Annals of Mathematical Statistics*, **28**(4), 823–838.

Ebrahimi, N., Habibullah, M. a Soofi, E. (1992). On information and sufficiency. *Journal of the Royal Statistical Society, Series B*, **54**(3), 739–748.

Feller, W. (1948). On the Kolmogorov-Smirnov limit theorems for empirical distributions. *The Annals of Mathematical Statistics*, **19**(2), 177–189.

Finkelstein, J. a Schafer, R. (1971). Improved goodness-of-fit tests. *Biometrika*, **58**(3), 641–645.

Gail, M. a Gastwirth, J. (1978). A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. *Journal of the Royal Statistical Society, Series B (Methodological)*, **40**(3), 350–357.

Hartley, R. (1928). Transmission of information. *The Bell System Technical Journal*, **7**(3), 535–563.

Henze, N. (1993). A new flexible class of omnibus tests for exponentiality. *Communications in Statistics, Theory and Methods*, **22**, 115–133.

Henze, N. a Meintanis, S. (2002a). Goodness-of-fit tests based on a new characterization of the exponential distribution. *Communications in Statistics, Theory and Methods*, **31**(9), 1479–1497.

Henze, N. a Meintanis, S. (2002b). Tests of fit for exponentiality based on the empirical Laplace transform. *Statistics*, **36**(2), 147–161.

Henze, N. a Meintanis, S. (2005). Recent and classical tests for exponentiality: a partial review with comparisons. *Metrika*, **61**, 29–45.

Jammalamadaka, S. a Taufer, E. (2003). Testing exponentiality by comparing the empirical distribution function of the normalized spacings with that of the original data. *Journal of Nonparametric Statistics*, **15**, 719–729.

Jammalamadaka, S. a Taufer, E. (2006). Use of mean residual life in testing departure from exponentiality. *Journal of Nonparametric Statistics*, **18**, 277–292.

Klar, B. (2001). Goodness-of-fit tests for the exponential and the normal distribution based on the integrated distribution function. *Annals of the Institute of Statistical Mathematics*, **53**(2), 338–353.

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 83–91.

Kullback, S. a Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, **22**(1), 79–86.

Lilliefors, H. (1969). On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*, **64**(325), 387–389.

Meintanis, S. a Iliopoulos, G. (2003). Characterizations of the exponential distribution based on certain properties of its characteristic function. *Kybernetika*, **39**(3), 295–298.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosphical Magazine Series 5*, **50**, 157–175.

Rao, M., Chen, Y., Vemuri, B. a Wang, F. (2004). Cummulative residual entropy: A new measure of information. *IEEE Transactions on Information Theory*, **50**(6), 1220–1228.

Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkley Symposium on Mathematical Statistics and Probability*, pages 547–561, Berkley, California, 1961. University of California Press.

Seshari, V., Csörgő, M. a Stephens, M. (1969). Tests for the exponential distribution using Kolmogorov-type statistics. *Journal of the Royal Statistical Society, Series B (Methodological)*, **31**, 449–509.

Shanbhag, D. N. (1970). The characterizations for exponential and geometric distributions. *Journal of the American Statistical Association*, **65**, 1256–1259.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 379–423, 623–656.

Shorack, G. a Wellner, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York. ISBN 978-0-898716-84-9.

Simonoff, J. (2003). *Analyzing Categorical Data*. Springer texts in statistics. Springer-Verlag, New York. ISBN 0-387-00749-0.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, **19**(2), 279–281.

Smirnov, N. (1936). Sur la distribution de w2. *Comptes Rendus de l'Académie des Science Paris*, **202**, 449–452.

Smirnov, N. (1939a). Ob uklonenijah empiriceskoi krivoi raspredelenija. *Sbornik: Mathematics (Matematicheskii Sbornik)*, **48**(6), 13–26.

Smirnov, N. (1939b). On the estimation of the discrepancy between empirical curves of distributions for two independent samples. *Bulletin mathématiques de l'Université de Moscou*, **2**(2).

Stephens, M. (1992). An appreciation of Kolmogorov's 1933 paper. Technical Report 453, Department of Statistics, Stanford University, USA.

Taufer, E. (2002). On entropy based tests for exponentiality. *Communications in Statistics, Simulation and Computation*, **31**(2), 189–200.

Van Soest, J. (1969). Some goodness of fit tests for exponential distributions. *Statistica Neerlandica*, **23**(1), 41–51.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society, Series B*, **38**(1), 54–59.

Von Mises, R. (1931). *Vorlesungen aus dem Gebiete der Angewandten Mathematik*. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik. Dueticke, Leipzig and Wien.

Wachowiak, M., Smolíková, R., Tourassi, G. a Elmaghraby, A. (2005). Estimation of generalized entropies with sample spacing. *Pattern Analysis and Application*, **8**, 95–101.

Zardasht, V., Parsi, S. a Mousazadeh, M. (2015). On empirical cumulative residual entropy and a goodness-of-fit test for exponentiality. *Statistical Papers*, **56**(3), 677–688.

# Zoznam tabuliek

# Prílohy

## A. Ukážky simulačnej štúdie

V tejto prílohe uvediem ukážky výstupov simulácii, ktoré boli popísané v kapitole 3. Simulácia kritických hodnôt testovej štatistiky $C_n$ pre $n = 10$:

```
> n<-10
> sim<-10000
> d<-c()
> for(i in 1:sim)
+ {set.seed(i)
+  x<-rgamma(n,1)
+  r<-1.2
+  m<-trunc(sqrt(n)+0.5)
+  n2<-n*(n-1)/2
+  xi<-sort(x)
+  z<-c()
+  for(j in 1:n){
+    for(k in (j+1):n){
+      zj<-(xi[j]-xi[k])/(xi[j]+xi[k])
+      z<-c(z,zj)
+    }
+  }
+  zi<-sort(z)
+  y<-c()
+  for(j in 1:n2){
+    a<-ifelse(j+m>n2,zi[n2],zi[j+m])
+    b<-ifelse(j-m<1,zi[1],zi[j-m])
+    yi<-a-b
+    yi<-(n2*yi/(2*m) ) ^ (1-r)
+    y<-c(y,yi)
+  }
+  test<-sum(y)/n2
+  test<-log(test)/(r-1)
+  d<-c(d,test)
+ }
> d<-sort(d)
> crit<-quantile(d,0.95)
> print(crit)
      95%
0.5990363
```

Simulácia kritických hodnôt testovej štatistiky $LW_n$ pre $n = 10$:

```
> n<-10
> sim<-10000
> d<-c()
> for(i in 1:sim)
+ {set.seed(i)
+  x<-rgamma(n,1)
+  xi<-sort(x)
+  xn<-mean(x)
+  m<-trunc(sqrt(n)+0.5)
+  z<-c()
+  for(j in 1:n){
```

```
+    a<-ifelse(j+m>n,xi[n],xi[j+m])
+    b<-ifelse(j-m<1,xi[1],xi[j-m])
+    yi<-a-b
+    koef<-4*m*xn
+    f<--x[j]/xn
+    zi<-n*yi*exp(f)/koef
+    zi<-log(zi+0.5)
+    z<-c(z,zi)
+  }
+  test<--sum(z)/n
+  d<-c(d,test)
+ }
> d<-sort(d)
> crit<-quantile(d,0.95)
> print(crit)
      95%
0.2217874
```

Teraz uvediem ukážky výstupov pre mnou implementované štatistiky pre alternatívu $W(1.4)$. Pre porovnanie ide o údaje v druhom stĺpci Tabulky 3.3. Simulácia pre testovú štatistiku $KLV_{4,n}$ pre alternatívu $W(1{,}4)$

```
> n<-20
> asim<-1000
> p<-0
> for(k in 1:asim)
+ {set.seed(k)
+  x<-rweibull(n,1.4)
+  m=4
+  xi<-sort(x)
+  pom1<-rep(xi[1],m)
+  pom2<-rep(xi[n],m)
+  xi<-c(pom1,xi,pom2)
+  E<-c()
+  for(i in 1:n){
+    Ei<-(xi[i+2*m]-xi[i])*n/(2*m)
+    Ei<-log(Ei)
+    E<-c(E,Ei)
+  }
+  F<-sum(E)/n
+  xn<-mean(x)
+  test<-exp(F)/exp(log(xn)+1)
+  r<-ifelse(test<0.6799,1,0)
+  p<-p+r}
> print(p/asim*100)
[1] 42.6
```

Simulácia pre testovú štatistiku $KLC_{4,n}$ pre alternatívu $W(1{,}4)$

```
> n<-20
> asim<-1000
> p<-0
> for(k in 1:asim)
+ {set.seed(k)
+  x<-rweibull(n,1.4)
+  m=4
+  xi<-sort(x)
```

```
+  pom1<-rep(xi[1],m)
+  pom2<-rep(xi[n],m)
+  xi<-c(pom1,xi,pom2)
+  E<-c()
+  for(i in 1:n){
+    A<-c()
+    for(j in (i-m):(i+m)){
+      Ai<-xi[j+m]
+      A<-c(A,Ai)
+    }
+    B<-sum(A)/(2*m+1)
+    C<-c()
+    D<-c()
+    for(j in (i-m):(i+m)){
+      Ci<-A[j-i+m+1]
+      Ci<-(Ci-B)*(j-i)
+      C<-c(C,Ci)
+      Di<-(A[j-i+m+1]-B)^2
+      D<-c(D,Di)
+    }
+    Ei<-sum(C)/sum(D)
+    Ei<-log(Ei/n)
+    E<-c(E,Ei)
+  }
+  F<--sum(E)/n
+  xn<-mean(x)
+  test<-exp(F)/exp(log(xn)+1)
+  r<-ifelse(test<0.77276,1,0)
+  p<-p+r}
> print(p/asim*100)
[1] 40.9
```

Simulácia pre testovú štatistiku $LW_n$ pre alternatívu $W(1{,}4)$

```
> n<-20
> asim<-1000
> p<-0
> for(i in 1:asim)
+ {set.seed(i)
+  x<-rweibull(n,1.4)
+  xi<-sort(x)
+  xn<-mean(x)
+  m<-trunc(sqrt(n)+0.5)
+  z<-c()
+  for(j in 1:n){
+    a<-ifelse(j+m>n,xi[n],xi[j+m])
+    b<-ifelse(j-m<1,xi[1],xi[j-m])
+    yi<-a-b
+    koef<-4*m*xn
+    f<--x[j]/xn
+    zi<-n*yi*exp(f)/koef
+    zi<-log(zi+0.5)
+    z<-c(z,zi)
+  }
+  test<--sum(z)/n
+  r<-ifelse(test>0.0917,1,0)
+  p<-p+r}
```

```
> print(p/asim*100)
[1] 45.9
```

Simulácia pre testovú štatistiku $TV^Z$ pre alternatívu $W(1,4)$

```
> n<-20
> asim<-100
> r<-1.2
> m<-trunc(sqrt(n)+0.5)
> p<-0
> for(i in 1:asim)
+ {set.seed(i)
+  x<-rweibull(n,1.4)
+  n2<-n*(n-1)/2
+  xi<-sort(x)
+  z<-c()
+  for(j in 1:n){
+    for(k in (j+1):n){
+      zj<-(xi[j]-xi[k])/(xi[j]+xi[k])
+      z<-c(z,zj)
+    }
+  }
+  zi<-sort(z)
+  y<-c()
+  for(j in 1:n2){
+    a<-ifelse(j+m>n2,zi[n2],zi[j+m])
+    b<-ifelse(j-m<1,zi[1],zi[j-m])
+    yi<-a-b
+    yi<-(n2*yi/(2*m) ) ^ (1-r)
+    y<-c(y,yi)
+  }
+  test<-sum(y)/n2
+  test<-log(test)/(r-1)
+  res<-ifelse(test>0.3298,1,0)
+  p<-p+res}
> print(p/asim*100)
[1] 14
```

Simulácia pre testovú štatistiku $BR_n$ pre alternatívu $W(1,4)$

```
> n<-20
> asim<-1000
> p<-0
> for(i in 1:asim)
+ {set.seed(i)
+  x<-rweibull(n,1.4)
+  xi<-sort(x)
+  s<-x*x
+  A<-sum(s)/(2*sum(x))
+  B<-c()
+  for(i in 1:n-1){
+    Bi<-xi[i+1]-xi[i]
+    a<-(n-i)/n
+    Bi<-a*log(a)*Bi
+    B<-c(B,Bi)
+  }
+  test<-(sum(B)+A)/A
```

```
+  r<-ifelse(test>=0.162147,1,0)
+  p<-p+r}
> print(p/asim*100)
[1] 36.3
```

Simulácia pre testovú štatistiku $A_{n,1}$ pre alternatívu $W(1,4)$

```
> n<-20
> asim<-1000
> p<-0
> for(i in 1:asim)
+ {set.seed(i)
+  n2<-n+1
+  x<-rweibull(n2,1.4)
+  xi<-sort(x)
+  y<-c(n2*xi[1])
+  for(j in 2:n2){
+    yi<-(n-j+2)*(xi[j]-xi[j-1])
+    y<-c(y,yi)
+  }
+  yn<-sum(y)
+  z<-c()
+  for(j in 1:n2){
+    pom<-0
+    ni<-n-j+2
+    for(k in ni:n2){
+      pom<-pom+y[k]
+    }
+    zi<-pom/yn-j/n2
+    z<-c(z,zi)
+  }
+  z<-z*z
+  test<-sum(z)
+   r<-ifelse(test>0.44,1,0)
+  p<-p+r}
> print(p/asim*100)
[1] 37.3
```

Simulácia pre testovú štatistiku $A_{n,2}$ pre alternatívu $W(1,4)$

```
> n<-20
> asim<-1000
> p<-0
> for(i in 1:asim)
+ {set.seed(i)
+  n2<-n+1
+  x<-rweibull(n2,1.4)
+  xi<-sort(x)
+  y<-c(n2*xi[1])
+  for(j in 2:n2){
+    yi<-(n-j+2)*(xi[j]-xi[j-1])
+    y<-c(y,yi)
+  }
+  yn<-sum(y)
+  z<-c()
+  for(j in 1:n2){
+    pom<-0
```

```
+    ni<-n-j+2
+    for(k in ni:n2){
+      pom<-pom+y[k]
+    }
+    zi<-pom/yn-j/n2
+    h<-n2-j+1
+    pom2<-j*h
+    zi<-zi*zi/pom2
+    z<-c(z,zi)
+  }
+  l<-n2^2
+  test<-l*sum(z)
+  r<-ifelse(test>2.41,1,0)
+  p<-p+r}
> print(p/asim*100)
[1] 30.4
```

Simulácia pre testovú štatistiku $T_{1,n}$ pre alternatívu $W(1,4)$

```
> n<-20
> asim<-1000
> p<-0
> for(i in 1:asim)
+ {set.seed(i)
+  x<-rweibull(n,1.4)
+  xi<-sort(x)
+  y1<-n*xi[1]
+  y<-c(y1)
+  for (i in 1:n){
+    yi<-(n-i+1)*(xi[i]-xi[i-1])
+    y<-c(y,yi)
+  }
+  test<-ks.test(x,y)$statistic
+  r<-ifelse(test>0.3041,1,0)
+  p<-p+r}
> print(p/asim*100)
   D
22.8
```