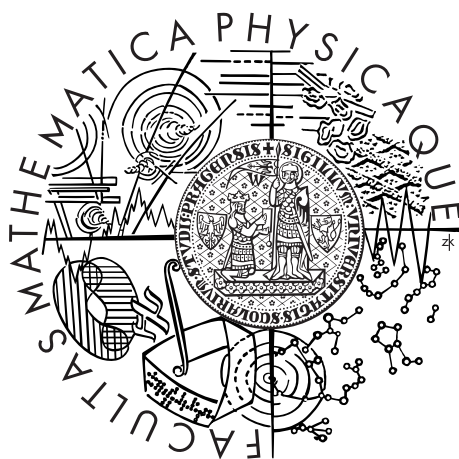


Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## DIPLOMOVÁ PRÁCE



Zuzana Jirmanová

### Testy založené na empirických charakteristických funkcích

Katedra pravděpodobnosti a matematické statistiky  
Vedoucí diplomové práce: Prof. RNDr. Marie Hušková, DrSc.  
Studijní obor: Matematika, Matematická statistika

## Poděkování

Ráda bych zde poděkovala především své vedoucí diplomové práce Prof. RNDr. Marii Huškové, DrSc. za cenné rady, návrhy a připomínky, které příznivě ovlivnily kvalitu této práce.

Děkuji také své rodině za podporu v celém období studia.

Prohlašuji, že jsem svoji diplomovou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 13. srpna 2007

Zuzana Jirmanová

# Obsah

<b>1</b>	<b>Úvod</b>	<b>4</b>
1.1	Motivace . . . . .	4
1.2	Hlavní myšlenka . . . . .	6
1.2.1	Asymptotické vlastnosti dané statistiky . . . . .	8
<b>2</b>	<b>Dvouvýběrový problém</b>	<b>13</b>
2.1	Testová statistika . . . . .	13
2.2	Asymptotické vlastnosti . . . . .	17
2.2.1	Chování při platnosti $H_0$ . . . . .	17
2.2.2	Chování při platnosti pevné alternativy . . . . .	19
<b>3</b>	<b>K-výběrový problém</b>	<b>22</b>
3.1	Testová statistika . . . . .	22
3.2	Mnohonásobná porovnávání . . . . .	24
3.3	Asymptotické vlastnosti . . . . .	26
3.3.1	Chování při platnosti $H_0$ . . . . .	26
3.3.2	Chování při platnosti pevné alternativy . . . . .	27
3.3.3	Chování při platnosti lokální alternativy . . . . .	27
<b>4</b>	<b>Simulace</b>	<b>29</b>
4.1	Citlivost testu na změnu parametrů . . . . .	30
4.1.1	Pravděpodobnosti zamítnutí . . . . .	30
4.1.2	Aproximace kritických hodnot . . . . .	38
4.2	Aplikace testu na diskrétní rozdělení . . . . .	46
4.3	Aplikace testu na spojitá rozdělení . . . . .	51
<b>5</b>	<b>Závěr</b>	<b>56</b>
<b>A</b>	<b>Příloha</b>	<b>57</b>
A.1	Citace continuous mapping theorem . . . . .	57
A.2	Zdrojový kód programu . . . . .	57

**Název práce:** Testy založené na empirických charakteristických funkcích

**Autor:** Zuzana Jirmanová

**Katedra (ústav):** Katedra pravděpodobnosti a matematické statistiky

**Vedoucí diplomové práce:** Prof. RNDr. Marie Hušková, DrSc.

**e-mail vedoucího:** huskova@karlin.mff.cuni.cz

**Abstrakt:** Důležitým statistickým problémem je testovat, zda dva výběry (či obecněji  $k$ -výběrů,  $k \geq 2$ ) pocházejí z rozdělení se stejnou distribuční funkcí. V této práci si představíme dvouvýběrový a  $k$ -výběrový test založený na váženém funkcionálu rozdílu empirických charakteristických funkcí. Tyto testy jsou analogií testů typu Kolmogorov-Smirnov, ale díky tomu, že charakteristická funkce je spojitá pro všechna  $t \in \mathbb{R}$ , se nemusíme při testování omezovat pouze na výběry se spojitou distribuční funkcí. U obou testů vyšetřujeme asymptotické vlastnosti daných statistik a konzistenci testů. Pro dvouvýběrový problém navíc provedeme užitím metody bootstrap s vrácením několik druhů simulací, na nichž si ověříme sílu testu pro různá rozdělení.

**Klíčová slova:** Neparametrické testy, empirické charakteristické funkce, dvouvýběrový problém,  $k$ -výběrový problém

**Title:** Tests based on empirical characteristic functions

**Author:** Zuzana Jirmanová

**Department:** Department of Probability and Mathematical Statistics

**Supervisor:** Prof. RNDr. Marie Hušková, DrSc.

**Supervisor's e-mail address:** huskova@karlin.mff.cuni.cz

**Abstract:** An important problem in statistics is to test whether two (or generally  $k$ ,  $k \geq 2$ ) sets of observations come from the same parent population. In this paper we review two-sample and  $k$ -sample tests based on weighted integrals involving the empirical characteristic functions. These tests are analogical cases of tests of Kolmogorov-Smirnov types, but due to the fact, that characteristic function is continuous for  $\forall t \in \mathbb{R}$ , they are extensions of these tests in the case of discrete populations. The consistency of the tests as well as their asymptotic distribution are investigated. We perform a simulation experiment conducted to assess the power of the two-sample test, using the bootstrap with replacement method.

**Keywords:** Nonparametric tests, empirical characteristic function, two-sample problem,  $k$ -sample problem

# Kapitola 1

## Úvod

### 1.1 Motivace

Začneme s obecným problémem testovat typ a tvar rozdělení nějakého výběru. Mějme výběr  $X = (X_1, \dots, X_n)$ , kde  $X_i$ ,  $i = 1, \dots, n$  jsou nezávislé náhodné veličiny z rozdělení, které má distribuční funkci  $F$ . Testujeme hypotézu

$$(1.1) \quad H_0 : F = F_0 \quad \text{proti alternativě} \quad H_1 : H_0 \quad \text{neplatí,}$$

pro danou distribuční funkci  $F_0$ .

Je několik typů testů zabývajících se tímto problémem, my si zde však zmiňme pouze tu třídu testů, která je založena na následující větě.

**Věta 1** (Glivenko-Cantelli). *Nechť  $F(x)$  značí distribuční funkci výběru, tj. platí pro ni  $F(x) = P(X_i \leq x)$ ,  $F_n(x)$  buď empirická distribuční funkce výběru definovaná vztahem*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I\{X_i < x\}. \quad \text{Označme} \quad H_n = \sup_x |F_n(x) - F(x)|.$$

*Pak platí*

$$P \left( \lim_{n \rightarrow \infty} H_n = 0 \right) = 1.$$

Při těchto testech je často konstruována testová statistika založená právě na rozdílu empirické a hypotetické distribuční funkce, případně na funkcionálu tohoto rozdílu. Mezi nejznámější z těchto testů patří například Kolmogorovův-Smirnovův test, zamítající  $H_0$ , pokud

$$D_n = \sup_x |F_n(x) - F_0(x)| > C_n,$$

pro danou kritickou hodnotu  $C_n$ .

Důležitou vlastností této statistiky je, že její rozdělení zůstává při platnosti  $H_0$  stejné pro všechna spojitá rozdělení  $F_0$ , tj. je nezávislé na hypotetické distribuční funkci a je funkcí pouze velikosti výběru  $n$ . Hodnoty pro malá  $n$ , řekněme  $n \leq 100$ , jsou natabelovány, pro větší  $n$  se využívá skutečnosti, že

$$P(\sqrt{n}D_n \leq z) \rightarrow L(z) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j+1} e^{-2j^2 z^2}.$$

Dalšími z řady těchto testů jsou např. Cramerův von Mises test s oborem kritických hodnot

$$W_n = n \int_{\mathbb{R}} [F_n(x) - F_0(x)]^2 f_0(x) dx \geq C_n,$$

kde  $f_0$  značí hustotu rozdělení  $F_0$ , nebo Andersonův-Darlingův test zamítající  $H_0$  při

$$A_n = n \int_{\mathbb{R}} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} f_0(x) dx \geq C_n.$$

Pro oba tyto testy platí stejně jako pro Kolmogorovův-Smirnovův test nezávislost testové statistiky na hypotetické distribuční funkci  $F_0$ , ale rovněž pouze v případě, že je tato funkce spojitá. Pokusme se proto na daný problém podívat z jiného úhlu a zkonstruovat třídu testů, které by byly rovněž konzistentní, ale které by nebyly omezeny pouze na případ testování spojitých distribučních funkcí.

Proveďme transformaci distribuční funkce  $F$  dle vzorce

$$K(z) = \int_{-\infty}^{+\infty} k(z, x) dF(x), \quad z = s + it, \quad s, t \in \mathbb{R}, \quad i = \sqrt{-1}$$

a transformaci k ní příslušné empirické distribuční funkce

$$K_n(z) = \int_{-\infty}^{+\infty} k(z, x) dF_n(x).$$

Předpokládejme, že  $K(z)$  jednoznačně určuje rozdělení  $X$ , které testujeme v  $H_0$ . Testy založené na těchto transformacích rovněž častou vedou ke konzistentním testům, některé z nich si ukážeme v dalších kapitolách. Nyní si ještě ukažme nejvýznamnější z těchto transformací:

- položíme-li  $k(t, x) = e^{zx}$ , dostaneme pro  $t = 0$  a  $s \in \mathbb{R}$  momentovou vytvořující funkci danou vztahem

$$\Psi(x) = \int_{\mathbb{R}} e^{sx} dF(x)$$

- pro totéž jádro transformace  $k(t, x) = e^{zx}$ , při  $t = 0$  a  $s < 0$ , získáme Laplaceovu transformaci
- transformace daná jádrem  $k(z, x) = z^x, 0 < z < 1$  dává pravděpodobnostní vytvořující funkci
- Mellinova transformace vznikne dosazením  $k(z, x) = x^z, z > 0$
- položíme-li  $k(t, x) = e^{zx}$ , pak pro imaginární část  $z$  dostáváme přímo charakteristickou funkci veličiny  $X$

Zaměříme se na poslední jmenovanou transformaci-charakteristickou funkci náhodné veličiny  $X$ . Proč že je výhodné pracovat právě s charakteristickými funkcemi náhodných veličin? Co činí testy na nich založené tak významnými? Jsou to především některé vlastnosti těchto funkcí:

charakteristická funkce dané náhodné veličiny jednoznačně určuje její distribuční funkci a naopak. Tuto vzájemnou a jednoznačnou souvislost popisuje následující vztah:

$$\varphi(t) = Ee^{itX} = E \cos(tX) + iE \sin(tX) = \int_{\mathbb{R}} e^{itx} dF(x).$$

Stejně důležitým a potěšujícím faktem je i to, že charakteristická funkce existuje pro všechna  $t \in \mathbb{R}$ . Tyto funkce jsou též spojité (dokonce stejnoměrně) a v absolutní hodnotě nabývají hodnoty maximálně jedna na celé reálné ose, proto jsou testy založené na těchto funkcích výborným pomocníkem například v případě, že testovaná distribuční funkce v (1.1) je diskrétní. Další významné vlastnosti může najít čtenář např. v [10], kde se Nikolai G. Ushakov věnoval jejich studiu mnohem podrobněji.

Závěrem je nutno zdůraznit, že testujeme hypotézu  $H_0 : F = F_0$ , tedy hypotéza nám přesně určuje hypotetické rozdělení  $F_0$  a nevystupují zde žádné rušivé parametry. Pokud by tomu tak bylo, tj. pokud by hypotéza pouze určovala nějakou parametrickou třídu  $F_{\Theta} = \{F_{\gamma}; \gamma \in \Theta\}$ , byla by situace značně komplikovanější a uvedené postupy by nešly použít. Toto ale není tématem této diplomové práce, proto se zde tímto případem nezabývám.

## 1.2 Hlavní myšlenka

Zkusme si nyní popsat hlavní myšlenku konstrukce obecné testové statistiky založené na empirických charakteristických funkcích. Mějme náhodný výběr  $X_1, \dots, X_n$  nezávislých náhodných veličin z rozdělení s distribuční funkcí  $F$

a charakteristickou funkcí  $\varphi(t) = Ee^{itX_1}$ ,  $t \in \mathbb{R}$ . Připomeňme, že empirická charakteristická funkce pro tento výběr je definována vztahem

$$\varphi_n(t) = \frac{1}{n} \sum_{j=1}^n \exp(itX_j), \quad t \in \mathbb{R}, \quad i = \sqrt{-1}.$$

Testujeme hypotézu

$$(1.2) \quad H_0 : F = F_0 \quad \text{proti alternativě} \quad H_1 : H_0 \quad \text{neplatí,}$$

pro danou distribuční funkci  $F_0$ .

Díky jednoznačnému vztahu mezi funkcemi distribučními a charakteristickými, lze očekávat i jistou analogii mezi testy založenými na empirických distribučních a empirických charakteristických funkcích.

Protože již víme, že  $\sup_x |F_n(x) - F(x)| \rightarrow 0$  skoro jistě, nabízí se postavit naši testovou statistiku rovněž na vzdálenosti mezi hypotetickou charakteristickou funkcí, označme ji  $\varphi_0$  a empirickou charakteristickou funkcí spočtenou na základě výběru  $X_1, \dots, X_n$ , tj. na veličině

$$D_n(t) = |\varphi_n(t) - \varphi_0(t)|.$$

Při této volbě získáváme přímo i konzistenci statistiky  $D_n(t)$ . Podívejme se však blíže na její možné chování vzhledem k reálné ose. Přestože pro charakteristickou funkci náhodné veličiny, jejíž distribuční funkce je absolutně spojitá, platí  $\lim_{|t| \rightarrow \infty} |\varphi(t)| = 0$ , obecně to pro charakteristickou funkci platit nemusí: může totiž obsahovat některé periodické komponenty, navíc empirická charakteristická funkce tyto komponenty obsahuje zcela určitě. Je tedy nutné naši testovou statistiku doplnit o nějaký druh integrovatelné váhové funkce, minimálně pro velké hodnoty  $t$ .

Jednou z možných testových statistik může tedy být

$$T_{n,w} = n \int_{-\infty}^{+\infty} D_n^2(t) w(t) dt,$$

kde

$$D_n(t) = |\varphi_n(t) - \varphi_0(t)|,$$

$\varphi_0(t)$  je charakteristická funkce příslušející distribuční funkci  $F_0$  z testované hypotézy (1.2) a  $w(t)$ ,  $t \in \mathbb{R}$ , je vhodná váhová funkce. Vidíme, že statistika  $T_{n,w}$  (resp. její odmocnina) je vlastně vzdálenost mezi empirickou charakteristickou a hypotetickou charakteristickou funkcí. Je proto důležité



volit váhovou funkci  $w(t)$  nezápornou; další doporučené požadavky na tuto funkci uvedeme později.

Odvoďme si ještě jiné vyjádření této testové statistiky, které je nejen vhodnější pro další počítání, ale zároveň poslouží i k lepšímu pochopení tohoto tématu.

Buď  $\varphi_0(t) = C_0(t) + iS_0(t)$  výše zmíněná hypotetická charakteristická funkce. Pak

$$\begin{aligned} T_{n,w} &= n \int_{\mathbb{R}} D_n^2(t) w(t) dt \\ &= n \int_{\mathbb{R}} \left| \frac{1}{n} \sum_{j=1}^n (\cos(tX_j) - C_0(t)) + i \frac{1}{n} \sum_{j=1}^n (\sin(tX_j) - S_0(t)) \right|^2 w(t) dt \\ &= \frac{1}{n} \int_{\mathbb{R}} \left[ \left( \sum_{j=1}^n (\cos(tX_j) - C_0(t)) \right)^2 + \left( \sum_{j=1}^n (\sin(tX_j) - S_0(t)) \right)^2 \right] w(t) dt. \end{aligned}$$

Předpokládejme dále, že platí  $w(t) = w(-t)$  pro  $\forall t \in \mathbb{R}$ . Můžeme tedy psát

$$\begin{aligned} T_{n,w} &= \frac{1}{n} \int_{\mathbb{R}} \left[ \left( \sum_{j=1}^n (\cos(tX_j) - C_0(t)) \right)^2 + \left( \sum_{j=1}^n (\sin(tX_j) - S_0(t)) \right)^2 \right] w(t) dt = \\ (1.3) \quad &= \int_{\mathbb{R}} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n (\cos(tX_j) + \sin(tX_j) - S_0(t) - C_0(t)) \right)^2 w(t) dt, \end{aligned}$$

neboť pro sudou  $w(t)$  je

$$2 \int_{\mathbb{R}} \sum_{j=1}^n (\cos(tX_j) - C_0(t)) \sum_{k=1}^n (\sin(tX_k) - S_0(t)) w(t) dt = 0.$$

### 1.2.1 Asymptotické vlastnosti dané statistiky

Vhodným prostředím pro zkoumání asymptotických vlastností nám buď separabilní Hilbertův prostor  $\mathcal{H} = L^2(\mathbb{R}, \mathcal{B}, w(t)dt)$  měřitelných funkcí  $f : \mathbb{R} \rightarrow \mathbb{R}$  takových, že platí  $\int_{\mathbb{R}} f^2(t)w(t)dt < \infty$ . Skalární součin a normu na  $\mathcal{H}$  zde značíme pořadě

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t)g(t)w(t)dt \quad \text{a} \quad \|f\| = \left( \int_{\mathbb{R}} f^2(t)w(t)dt \right)^{1/2},$$

konvergencí v  $\mathcal{H}$  myslíme konvergenci v této normě. Konvergenci v distribuci budeme v následujícím textu značit výrazem  $\xrightarrow{D}$  a výrazem  $\xrightarrow{P}$  pak konvergenci v pravděpodobnosti.

Prozkoumejme nyní chování obecné testové statistiky pro dvě základní situace.

### Za platnosti $H_0$

Předpoklad sudé váhové funkce  $w(t)$ ,  $t \in \mathbb{R}$  je pro vyšetřování asymptotického chování obecné testové statistiky při platnosti  $H_0$  velice důležitý. Pokud totiž ve vyjádření testové statistiky v (1.3) položíme

$$(1.4) \quad Z_n(t) := \frac{1}{\sqrt{n}} \sum_{j=1}^n (\cos(tX_j) + \sin(tX_j) - S_0(t) - C_0(t)), \quad t \in \mathbb{R}$$

máme

$$T_{n,w} = \int_{\mathbb{R}} (Z_n(t))^2 w(t) dt,$$

tedy ve zmiňovaném Hilbertově prostoru je pro sudou  $w(t)$

$$T_{n,w} = \|Z_n(t)\|^2.$$

Nyní už můžeme využít matematický aparát a ověřené limitní věty, které nám Hilbertův prostor nabízí.

**Věta 2.** *Bud'  $X = (X_1, \dots, X_n)$  náhodný výběr nezávislých stejně rozdělených náhodných veličin s charakteristickou funkcí  $\varphi(t) = C(t) + iS(t)$ . Nechť pro váhovou funkci z vyjádření testové statistiky platí  $w(t) = w(-t), \forall t \in \mathbb{R}$ . Pak*

$$(1.5) \quad Z_n \xrightarrow{D} \mathcal{Z},$$

kde  $Z_n$  je definováno v (1.4) a  $\mathcal{Z} \in \mathcal{H}$  je gaussovský proces s kovarianční funkcí

$$K(s, t) = C(s - t) + S(s + t) - C(s)C(t) - C(s)S(t) - S(s)C(t) - S(s)S(t).$$

Dále

$$(1.6) \quad T_{n,w} \xrightarrow{D} \|\mathcal{Z}\|^2.$$

*Důkaz.*  $Z_n(t)$  definované v (1.4) můžeme rovněž zapsat následujícím způsobem:

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n W_j(t),$$

kde  $W_j(t) = \cos(tX_j) + \sin(tX_j) - S_0(t) - C_0(t)$ ,  $j = 1, \dots, n$  jsou nezávislé stejně rozdělené náhodné veličiny a prvky Hilbertova prostoru takové, že  $E \|W_1\| = 0$  a  $E \|W_1\|^2 < \infty$  (poznamenejme, že  $|W_1(t)| \leq 4 \forall t \in \mathbb{R}$ ).

Pak ale z centrální limitní věty pro Hilbertovy prostory (přesné znění této věty najdeme v [3] na str.450 pod názvem 'Functional central limit theorem') dostáváme, že  $Z_n \xrightarrow{D} \mathcal{Z}$  pro  $n \rightarrow \infty$ , kde  $\mathcal{Z}$  je gaussovský proces s kovarianční funkcí

$$K(s, t) = E \mathcal{Z}(s) \mathcal{Z}(t) = C(s-t) + S(s+t) - C(s)C(t) - C(s)S(t) - S(s)C(t) - S(s)S(t),$$

tím jsme dokázali tvrzení (1.5).

Dále protože  $Z_n$  je prvkem Hilbertova prostoru  $\mathcal{H}$ ,  $Z_n \xrightarrow{D} \mathcal{Z}$ ,  $T_{n,w} = \|Z_n\|^2$  a zobrazení  $\|\cdot\|^2 : \mathcal{H} \rightarrow \mathbb{R}$  je spojité pro  $\forall t \in \mathbb{R}$ ,  $H \in \mathcal{H}$  plyne tvrzení (1.6) z tzv. *Continuous mapping theorem* (toto tvrzení je citované v příloze na konci práce).  $\square$

### Při pevné alternativě

**Věta 3.** *Mějme  $X = (X_1, \dots, X_n)$  výběr nezávislých stejně rozdělených náhodných veličin, označme  $\varphi(t) = C(t) + iS(t)$  charakteristickou funkci tohoto výběru a  $\varphi_0 = C_0(t) + iS_0(t)$  charakteristickou funkci příslušející distribuční funkci  $F_0$  z hypotézy (1.2). Pak*

$$n^{-1}T_{n,w} \xrightarrow{P} \int_{\mathbb{R}} D^2(t)w(t)dt,$$

kde

$$D^2(t) = |\varphi(t) - \varphi_0(t)|^2.$$

*Důkaz.* Víme, že

$$n^{-1}T_{n,w} = \int_{\mathbb{R}} D_n^2(t)w(t)dt,$$

pro

$$D_n^2(t) = |\varphi_n(t) - \varphi_0(t)|^2.$$

Označme si dále pro jednoduchost

$$C_n(t) := \frac{1}{n} \sum_{j=1}^n (\cos(tX_j)) \quad a \quad S_n(t) := \frac{1}{n} \sum_{j=1}^n (\sin(tX_j)).$$

Máme tedy

$$\begin{aligned} D_n^2(t) &= |\varphi_n(t) - \varphi_0(t)|^2 = |\varphi_n(t) - \varphi(t) + \varphi(t) - \varphi_0(t)|^2 = \\ &= |C_n(t) + iS_n(t) - C(t) - iS(t) + C(t) + iS(t) - C_0(t) - iS_0(t)|^2. \end{aligned}$$

Označíme-li

$$\begin{aligned} A &:= A(t) := \varphi_n(t) - \varphi(t) = (C_n(t) - C(t)) + i(S_n(t) - S(t)), \\ B &:= B(t) := \varphi(t) - \varphi_0(t) = (C(t) - C_0(t)) + i(S(t) - S_0(t)), \end{aligned}$$

a  $\bar{A}$ , resp.  $\bar{B}$  komplexně sdružené jednotky k  $A$ , resp.  $B$ , dostáváme

$$\begin{aligned} D_n^2(t) &= (A + B)(\bar{A} + \bar{B}) = |A|^2 + A\bar{B} + \bar{A}B + |B|^2 \\ &= |A|^2 + |B|^2 + 2\operatorname{Re}[A]\operatorname{Re}[B] + 2\operatorname{Im}[A]\operatorname{Im}[B]. \end{aligned}$$

Dále ze Schwarzovy nerovnosti plyne

$$|(\operatorname{Re}[A])(\operatorname{Re}[B])| \leq \sqrt{|\operatorname{Re}[A]|^2 |\operatorname{Re}[B]|^2}$$

a

$$\int_{\mathbb{R}} \sqrt{|\operatorname{Re}[A]|^2 |\operatorname{Re}[B]|^2} w(t) dt \leq \sqrt{\int_{\mathbb{R}} |\operatorname{Re}[A]|^2 w(t) dt \int_{\mathbb{R}} |\operatorname{Re}[B]|^2 w(t) dt},$$

totéž platí analogicky pro  $|\operatorname{Im}[A]\operatorname{Im}[B]|$ . Protože však v námi zvoleném Hilbertově prostoru platí

$$\|\varphi_n - \varphi\| \xrightarrow{P} 0, \quad \text{tj.} \quad \int_{\mathbb{R}} |\varphi_n(t) - \varphi(t)|^2 w(t) dt \xrightarrow{P} 0,$$

máme

$$\int_{\mathbb{R}} \left[ \underbrace{(C_n(t) - C(t))^2}_{\geq 0} + \underbrace{(S_n(t) - S(t))^2}_{\geq 0} \right] w(t) dt \xrightarrow{P} 0,$$

a tedy i  $\|C_n - C\| \xrightarrow{P} 0$  a  $\|S_n - S\| \xrightarrow{P} 0$ .

Odtud získáváme

$$\int_{\mathbb{R}} (C_n(t) - C(t))(C(t) - C_0(t))w(t)dt = \int_{\mathbb{R}} (\operatorname{Re}[A])(\operatorname{Re}[B])w(t)dt \xrightarrow{P} 0,$$

neboť

$$(1.7) \quad \int_{\mathbb{R}} (\operatorname{Re}[A])(\operatorname{Re}[B])w(t)dt \leq \sqrt{\underbrace{\int_{\mathbb{R}} |\operatorname{Re}[A]|^2 w(t)dt}_{\xrightarrow{P} 0} \underbrace{\int_{\mathbb{R}} |\operatorname{Re}[B]|^2 w(t)dt}_{\text{konstanta}}},$$

analogicky

$$\int_{\mathbb{R}} (S_n(t) - S(t))(S(t) - S_0(t))w(t)dt = \int_{\mathbb{R}} (\operatorname{Im}[A])(\operatorname{Im}[B])w(t)dt \xrightarrow{P} 0.$$

Celkem proto

$$\begin{aligned} \int_{\mathbb{R}} D_n^2(t)w(t)dt &= \underbrace{\int_{\mathbb{R}} |\varphi_n(t) - \varphi(t)|^2 (t)w(t)dt}_{\xrightarrow{P} 0} + \int_{\mathbb{R}} |\varphi(t) - \varphi_0(t)|^2 (t)w(t)dt \\ &\quad + 2 \underbrace{\int_{\mathbb{R}} (C_n(t) - C(t))(C(t) - C_0(t))w(t)dt}_{\xrightarrow{P} 0} \\ &\quad + 2 \underbrace{\int_{\mathbb{R}} (S_n(t) - S(t))(S(t) - S_0(t))w(t)dt}_{\xrightarrow{P} 0}, \end{aligned}$$

tedy

$$n^{-1}T_{n,w} = \int_{\mathbb{R}} D_n^2(t)w(t)dt \xrightarrow{P} \int_{\mathbb{R}} |\varphi(t) - \varphi_0(t)|^2 w(t)dt = \int_{\mathbb{R}} D^2 w(t)dt.$$

□

V dalších kapitolách se zaměříme na testy založené na empirických charakteristických funkcích pro dvouvýběrový a k-výběrový problém. Popíšeme testové statistiky a vyšetříme jejich základní limitní vlastnosti pro různé alternativy.

# Kapitola 2

## Dvouvýběrový problém

### 2.1 Testová statistika

Častým statistickým problémem je testovat, zda dva náhodné výběry pocházejí z rozdělení se stejnou distribuční funkcí.

Uvažujme tedy dva náhodné navzájem nezávislé výběry nezávislých stejně rozdělených náhodných veličin  $X_1, \dots, X_{n_1}$  a  $Y_1, \dots, Y_{n_2}$ .

Označme po řadě  $F$  a  $G$  distribuční funkce výběru veličin  $X_j$ ,  $j = 1, \dots, n_1$  a  $Y_k$ ,  $k = 1, \dots, n_2$ , tedy funkce, pro které platí:  $F(x) = P(X_i \leq x)$  a  $F(y) = P(Y_i \leq y)$ . Testujeme hypotézu

$$H_0 : F \equiv G \quad \text{proti alternativě} \quad H_1 : H_0 \text{ neplatí.}$$

Vzhledem k jednoznačnému vztahu mezi distribuční a charakteristickou funkcí dané náhodné veličiny však můžeme tuto hypotézu ekvivalentně zapsat ve tvaru

$$(2.1) \quad H_0 : \varphi_1 \equiv \varphi_2,$$

kde  $\varphi_1(t) = E \exp(itX_1)$  a  $\varphi_2(t) = E \exp(itY_1)$ ,  $t \in \mathbb{R}$ ,  $i = \sqrt{-1}$  značí po řadě charakteristické funkce prvního a druhého výběru. V celé této kapitole dále předpokládejme, že  $n_j/(n_1+n_2) \rightarrow \nu_j \in (0, 1)$ ,  $n_j \rightarrow \infty$ ,  $j = 1, 2$  a označme  $N = n_1 + n_2$ .

Naše testová statistika má pro tento případ tvar

$$(2.2) \quad T_{n_1, n_2} = \frac{n_1 n_2}{N} \int_{-\infty}^{+\infty} D_{n_1, n_2}^2(t) w(t) dt \quad \text{pro } \forall t \in \mathbb{R},$$

kde

$$(2.3) \quad D_{n_1, n_2}^2(t) = |\varphi_{n_1}(t) - \varphi_{n_2}(t)|^2$$

a  $w(t)$  je vhodná nezáporná integrovatelná váhová funkce.  $\varphi_{n_1}$  a  $\varphi_{n_2}$  ve (2.3) značí empirické charakteristické funkce výběrů definovaných vztahy

$$\varphi_{n_1} = \frac{1}{n_1} \sum_{l=1}^{n_1} \exp(itX_l) \quad \text{a} \quad \varphi_{n_2} = \frac{1}{n_2} \sum_{j=1}^{n_2} \exp(itY_j).$$

Předpokládáme-li sudou váhovou funkci  $w(t)$ , podobně jako v (1.3) můžeme i tuto statistiku dále upravit na tvar

$$\begin{aligned} T_{n_1, n_2} &= \frac{n_1 n_2}{N} \int_{-\infty}^{+\infty} D_{n_1, n_2}^2(t) w(t) dt = \\ &= \int_{\mathbb{R}} \left( \frac{\sqrt{n_2}}{\sqrt{n_1 N}} \sum_{j=1}^{n_1} (\cos(tX_j) + \sin(tX_j)) - \frac{\sqrt{n_1}}{\sqrt{n_2 N}} \sum_{k=1}^{n_2} (\cos(tY_k) + \sin(tY_k)) \right)^2 w(t) dt, \end{aligned}$$

neboť

$$\begin{aligned} D_{n_1, n_2}^2 &= \left| \frac{1}{n_1} \sum_{j=1}^{n_1} (\cos(tX_j) + i \sin(tX_j)) - \frac{1}{n_2} \sum_{k=1}^{n_2} (\cos(tY_k) + i \sin(tY_k)) \right|^2 = \\ &= \left( \frac{1}{n_1} \sum_{j=1}^{n_1} \cos(tX_j) - \frac{1}{n_2} \sum_{k=1}^{n_2} \cos(tY_k) \right)^2 + \left( \frac{1}{n_1} \sum_{j=1}^{n_1} \sin(tX_j) - \frac{1}{n_2} \sum_{k=1}^{n_2} \sin(tY_k) \right)^2 = \\ (2.4) \quad &= \left( \frac{1}{n_1} \sum_{j=1}^{n_1} (\cos(tX_j) + \sin(tX_j)) - \frac{1}{n_2} \sum_{k=1}^{n_2} (\cos(tY_k) + \sin(tY_k)) \right)^2. \end{aligned}$$

Tento tvar testové statistiky lze ještě v případě potřeby přepsat do ekvivalentního tvaru, který je vhodný zvláště pro teoretické úvahy:

$$\begin{aligned} T_{n_1 n_2} &= \frac{n_1 n_2}{N} \int_{-\infty}^{+\infty} \left( \frac{1}{n_1} \sum_{j=1}^{n_1} [\cos(tX_j) + \sin(tX_j) - E_{H_0}(\cos(tX_j) + \sin(tX_j))] - \right. \\ (2.5) \quad &\left. \frac{1}{n_2} \sum_{j=1}^{n_2} [\cos(tY_k) + \sin(tY_k) - E_{H_0}(\cos(tY_k) + \sin(tY_k))] \right)^2 w(t) dt. \end{aligned}$$

Podívejme se nyní na volbu váhové funkce  $w(t)$ . Přestože většina vlastností testové statistiky  $T_{n_1, n_2}$  (zvláště a hlavně pak tvar asymptotického rozdělení a konzistence při pevných či lokálních alternativách) se dá prokázat pro poměrně obecnou funkci  $w(t)$ , Meintanis [9] doporučuje pro snadnější výpočty a odvozování především tyto dvě konkrétní:

$$(2.6) \quad w_a^1(t) = \exp(-a|t|), \quad a > 0, \quad t \in \mathbb{R}$$

a

$$(2.7) \quad w_a^2(t) = \exp(-at^2), \quad a > 0, \quad t \in \mathbb{R}.$$

Tyto funkce mají oproti obecným funkcím několik výhod. Předně se jedná o charakteristické funkce-tedy z vlastností charakteristických funkcí dostáváme, že jsou definované a spojité na celé reálné ose. Dále jsou nezáporné a v neposlední řadě sudé, což umožňuje přepsat testovou statistiku do početně příjemnějšího tvaru tak, jako jsme to již učinili v (2.4), a na základě kterého můžeme v Hilbertově prostoru vyšetřovat asymptotické vlastnosti této statistiky při platnosti  $H_0$ .

Ukažme si, jaké získáme vyjádření testové statistiky pro tyto dvě funkce, které je vhodné zvláště pro praktické počítání. Testovou statistiku  $T_{n_1, n_2}$  si nejdříve vyjádříme ještě jiným způsobem: umocněním členu  $D_{n_1, n_2}$  a použitím součtových vzorců pro goniometrické funkce dostaneme pro další úpravy její výhodnější tvar:

$$\begin{aligned} T_{n_1, n_2} = \int_{\mathbb{R}} \left[ \frac{n_2}{Nn_1} \sum_{k=1}^{n_1} \sum_{l=1}^{n_1} \cos(t(X_k - X_l)) - \right. \\ \left. - \frac{2}{N} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} \cos(t(X_k - Y_l)) + \right. \\ \left. + \frac{n_1}{Nn_2} \sum_{k=1}^{n_2} \sum_{l=1}^{n_2} \cos(t(Y_k - Y_l)) \right] w(t) dt. \end{aligned}$$

Protože pro obecné  $Z \in \mathbb{C}$  platí

$$\begin{aligned} \int_{\mathbb{R}} \cos(tZ) \cdot \exp(-a|t|) dt &= 2 \int_0^{\infty} \cos(tZ) \cdot \exp(-at) dt = \\ &= 2 \operatorname{Re} \left[ \int_{-\infty}^0 \exp(t(a - iZ)) dt \right] = 2 \operatorname{Re} \left[ \frac{1}{a - iZ} \right] = \frac{2a}{a^2 + Z^2} \end{aligned}$$



a

$$\begin{aligned} \int_{\mathbb{R}} \cos(tZ) \cdot \exp(-at^2) dt &= \operatorname{Re} \int_{\mathbb{R}} \exp(iZt - at^2) dt = \\ &= \exp\left(\frac{Z^2}{4a}\right) \operatorname{Re} \left[ \int_{\mathbb{R}} \exp\left(-\left(\sqrt{at} - \frac{iZ}{2\sqrt{a}}\right)^2\right) dt \right] = \\ &= \operatorname{Re} \int_{\mathbb{R}} \exp(-x^2) \frac{1}{\sqrt{a}} dx = \sqrt{\frac{\pi}{a}}, \end{aligned}$$

obdržíme po dosazení váhových funkcí (2.6) a (2.7) do původního vyjádření  $T_{n_1, n_2}$  (2.8) následující testové statistiky:

$$\begin{aligned} (2.8) \quad T_{n_1, n_2}^{(1)} &= \frac{n_1 n_2}{N} \int_{\mathbb{R}} D_{n_1, n_2}^2(t) w_a^1(t) dt \\ &= \frac{2an_2}{n_1 N} \sum_{k=1}^{n_1} \sum_{l=1}^{n_1} \left[ \frac{1}{a^2 + (X_k - X_l)^2} \right] - \\ &\quad - \frac{4a}{N} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} \left[ \frac{1}{a^2 + (X_k - Y_l)^2} \right] + \\ &\quad + \frac{2an_1}{n_2 N} \sum_{k=1}^{n_2} \sum_{l=1}^{n_2} \left[ \frac{1}{a^2 + (Y_k - Y_l)^2} \right], \end{aligned}$$

a

$$\begin{aligned} (2.9) \quad T_{n_1, n_2}^{(2)} &= \frac{n_1 n_2}{N} \int_{\mathbb{R}} D_{n_1, n_2}^2(t) w_a^2(t) dt = \\ &= \frac{n_2}{n_1 N} \sqrt{\frac{\pi}{a}} \sum_{k=1}^{n_1} \sum_{l=1}^{n_1} \exp\left(\frac{-(X_k - X_l)^2}{4a}\right) - \\ &\quad - \frac{2}{N} \sqrt{\frac{\pi}{a}} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} \exp\left(\frac{-(X_k - Y_l)^2}{4a}\right) + \\ &\quad + \frac{n_1}{n_2 N} \sqrt{\frac{\pi}{a}} \sum_{k=1}^{n_2} \sum_{l=1}^{n_2} \exp\left(\frac{-(Y_k - Y_l)^2}{4a}\right). \end{aligned}$$

Zmiňme si ještě význam parametru  $a$ . Tento parametr kontroluje rychlost klesání hodnot váhových funkcí k ose reálných čísel. Zřejmě pro vyšší hodnoty  $a$  klesají obě váhové funkce rychleji, pro hodnoty blízké 0 je klesání těchto funkcí pomalejší. V kapitole 4 jsou provedeny simulace obou testových statistik pro několik konkrétních voleb parametru  $a$ , čtenář se tedy sám může blíže ujistit o závislosti velikosti testových statistik  $T_{n_1, n_2}^{(1)}$  a  $T_{n_1, n_2}^{(2)}$  na volbě tohoto parametru.

## 2.2 Asymptotické vlastnosti

Asymptotické vlastnosti dané testové statistiky zkoumáme opět v Hilbertově prostoru  $\mathcal{H} = L^2(\mathbb{R}, \mathcal{B}, w(t)dt)$  měřitelných funkcí  $f : \mathbb{R} \rightarrow \mathbb{R}$  takových, že  $\int_{\mathbb{R}} f^2(t)w(t)dt < \infty$ . Značení zachováváme stejně jako v první kapitole. Pokusme se opět nejprve zapsat testovou statistiku v podobě normy tohoto prostoru. Ve vyjádření statistiky (2.5) označme

$$Z_{n_1} = \frac{1}{\sqrt{n_1}} \sum_{j=1}^{n_1} [\cos(tX_j) + \sin(tX_j) - E_{H_0}(\cos(tX_j) + \sin(tX_j))]$$

a

$$Z_{n_2} = \frac{1}{\sqrt{n_2}} \sum_{k=1}^{n_2} [\cos(tY_k) + \sin(tY_k) - E_{H_0}(\cos(tY_k) + \sin(tY_k))].$$

Pak pro

$$(2.10) \quad Z_{n_1 n_2} = \sqrt{\frac{n_2}{N}} Z_{n_1} - \sqrt{\frac{n_1}{N}} Z_{n_2}$$

získáváme

$$T_{n_1 n_2} = \|Z_{n_1 n_2}\|^2.$$

### 2.2.1 Chování při platnosti $H_0$

Následující věta popisuje limitní chování testové statistiky za platnosti nulové hypotézy (2.1).

**Věta 4.** *Nechť  $X_j$ ,  $j = 1, \dots, n_1$  a  $Y_j$ ,  $j = 1, \dots, n_2$  jsou dva nezávislé výběry nezávislých stejně rozdělených náhodných veličin se společnou charakteristickou funkcí  $\varphi(t) = C(t) + iS(t)$ . Předpokládejme, že funkce  $w$  z vyjádření testové statistiky (2.2) je sudá, tj.  $w(t) = w(-t)$ ,  $\forall t \in \mathbb{R}$ . Pak existuje gaussovský proces  $\mathcal{Z}$  takový, že při  $n_1, n_2 \rightarrow \infty$  platí, že*

$$(2.11) \quad Z_{n_1, n_2} \xrightarrow{D} \mathcal{Z},$$

kde  $Z_{n_1, n_2}$  je definováno v (2.10) a  $\mathcal{Z} \in \mathcal{H}$  má kovarianční funkci tvaru  $K(s, t) = C(s-t) + S(s+t) - C(s)C(t) - C(s)S(t) - S(s)C(t) - S(s)S(t)$  pro všechna  $s, t \in \mathbb{R}$  a platí

$$(2.12) \quad T_{n_1 n_2} \xrightarrow{D} \|\mathcal{Z}\|^2.$$

*Důkaz.* Rozepišme si  $Z_{n_1, n_2}$  do následujícího tvaru:

$$\begin{aligned} Z_{n_1 n_2}(t) &= \sqrt{\frac{n_2}{N}} Z_{n_1} - \sqrt{\frac{n_1}{N}} Z_{n_2} \\ &= \sqrt{\frac{n_2}{N}} \frac{1}{\sqrt{n_1}} \sum_{j=1}^{n_1} W_{1j}(t) \\ &\quad - \sqrt{\frac{n_1}{N}} \frac{1}{\sqrt{n_2}} \sum_{j=1}^{n_2} W_{2j}(t), \end{aligned}$$

kde

$$\begin{aligned} W_{1j}(t) &= (\cos(tX_j) - C(t)) + (\sin(tX_j) - S(t)) \\ \text{a } W_{2j}(t) &= (\cos(tY_j) - C(t)) + (\sin(tY_j) - S(t)) \end{aligned}$$

jsou nezávislé stejně rozdělené náhodné veličiny a prvky Hilbertova prostoru takové, že  $E \|W_{11}\| = 0$  a  $E \|W_{11}\|^2 < \infty$  (poznamenejme, že  $|W_{11}(t)| \leq 4$ ). Pak z centrální limitní věty pro Hilbertovy prostory dostáváme

$$Z_{n_m} \xrightarrow{D} Z_m, \quad \text{pro } n \rightarrow \infty, \quad m = 1, 2,$$

kde  $Z_m$  je gaussovský proces s kovarianční funkcí  $K_m(s, t)$ . Protože ale  $Z_{n_1}$  a  $Z_{n_2}$  jsou nezávislé, jsou nezávislé rovněž  $Z_1$  a  $Z_2$  a za platnosti  $H_0$  mají  $Z_1$  a  $Z_2$  dokonce stejné rozdělení. Proto pro gaussovský proces  $\mathcal{Z}$  z (2.11) platí  $\mathcal{Z}(t) = \sqrt{\frac{n_2}{N}} Z_1(t) - \sqrt{\frac{n_1}{N}} Z_2(t)$ . Kovarianční funkce tohoto procesu je dána vztahem  $K(s, t) = E \mathcal{Z}(s) \mathcal{Z}(t) = \frac{n_2}{N} K_1(s, t) + \frac{n_1}{N} K_1(s, t) = K_1(s, t)$ . *Continuous mapping theorem* nám obdobně jako v důkazu věty 1 dává vztah (2.12), tedy zbytek požadovaného tvrzení. □

Rozdělení  $\|\mathcal{Z}\|^2$  je stejné (více o této aproximaci se lze dočíst např. v [2]) jako rozdělení  $\sum_{j \geq 1} \lambda_j(w) N_j^2$ , kde  $N_1, N_2, \dots$  jsou nezávislé veličiny s normálním rozdělením a  $(\lambda_j(w))_{j \geq 1}$  značí nenulová vlastní čísla integrálního operátoru definovaného předpisem

$$(2.13) \quad Ah(s) = \int_{-\infty}^{+\infty} K(s, t) h(t) w(t) dt.$$

Přestože nedokážeme tato vlastní čísla přesně spočítat, můžeme představu o limitním rozdělení veličiny  $\|\mathcal{Z}\|^2$  získat alespoň výpočtem

$$E \|\mathcal{Z}\|^2 = \int_{-\infty}^{+\infty} K(t, t) w(t) dt,$$

kde  $K(t, t)$  je kovarianční funkce limitního rozdělení testové statistiky z předešlé věty.

Při pohledu na kovarianční strukturu vidíme, že při platnosti  $H_0$  závisí asymptotické rozdělení testovací statistiky  $T_n(w, k)$  na hypotetické charakteristické funkci. Tento důležitý fakt značně omezuje aplikaci uvedené věty a je výraznou nevýhodou testů založených na empirických charakteristických funkcích, nicméně ve čtvrté kapitole se přesvědčíme, že pro dané rozdělení  $F_0$  lze velice uspokojivě aproximovat kritické hodnoty pro zamítnutí  $H_0$ . Tím se pokusíme tento nedostatek alespoň částečně odstranit.

## 2.2.2 Chování při platnosti pevné alternativy

Nyní prověříme asymptotické chování  $T_{n_1, n_2}$  v obecné situaci, tzn. nejen za platnosti  $H_0$ .

**Věta 5.** *Nechť  $X_j$ ,  $j = 1, \dots, n_1$  a  $Y_j$ ,  $j = 1, \dots, n_2$  jsou dva nezávislé výběry nezávislých stejně rozdělených náhodných veličin,  $\varphi_1(t) = C_1(t) + iS_1$  a  $\varphi_2(t) = C_2(t) + iS_2$  buďte po řadě charakteristické funkce těchto výběrů. Pak pro  $T_{n_1, n_2}$  platí:*

$$N(n_1 n_2)^{-1} T_{n_1, n_2} \xrightarrow{P} \int_{-\infty}^{+\infty} D^2(t) w(t) dt,$$

kde  $D(t) = |\varphi_1(t) - \varphi_2(t)|$ .

*Důkaz.* Z vyjádření testové statistiky v (2.2) okamžitě dostáváme, že

$$(2.14) \quad \frac{N T_{n_1, n_2}}{n_1 n_2} = \int_{\mathbb{R}} D_{n_1, n_2}^2(t) w(t) dt.$$

Dále

$$\begin{aligned} D_{n_1, n_2}^2 &= |\varphi_{n_1}(t) - \varphi_{n_2}(t)|^2 = |\varphi_{n_1}(t) - \varphi_1(t) - \varphi_{n_2}(t) + \varphi_2(t) + \varphi_1(t) - \varphi_2(t)|^2 \\ &= |A - B + C|^2 \end{aligned}$$

pro

$$\begin{aligned} A &:= A(t) = \varphi_{n_1}(t) - \varphi_1(t) = (C_{n_1} - C_1) + (S_{n_1} - S_1), \\ B &:= B(t) = \varphi_{n_2}(t) - \varphi_2(t) = (C_{n_2} - C_2) + (S_{n_2} - S_2), \\ C &:= C(t) = \varphi_1(t) - \varphi_2(t) = (C_1 - C_2) + (S_1 - S_2). \end{aligned}$$

Máme tedy

$$\begin{aligned} D_{n_1, n_2}^2 &= (A - B + C)(\overline{A - B + C}) = |A|^2 - A\overline{B} + A\overline{C} - \overline{A}B + |B|^2 \\ &\quad - B\overline{C} + \overline{A}C - \overline{B}C + |C|^2 = |A|^2 + |B|^2 + |C|^2 + E, \end{aligned}$$

položíme-li

$$E := 2\operatorname{Re}[A]\operatorname{Re}[C] - 2\operatorname{Re}[B]\operatorname{Re}[C] - 2\operatorname{Re}[A]\operatorname{Re}[B] \\ 2\operatorname{Im}[A]\operatorname{Im}[C] - 2\operatorname{Im}[B]\operatorname{Im}[C] - 2\operatorname{Im}[A]\operatorname{Im}[B].$$

Podobně jako v důkazu věty 2 odhadněme shora součiny reálných a imaginárních složek  $A, B$  a  $C$  dle vztahů

$$|(\beta)(\delta)| \leq \sqrt{|\beta|^2 |\delta|^2}$$

a

$$\int_{\mathbb{R}} \sqrt{|\beta|^2 |\delta|^2} w(t) dt \leq \sqrt{\int_{\mathbb{R}} |\beta|^2 w(t) dt} \sqrt{\int_{\mathbb{R}} |\delta|^2 w(t) dt}, \quad \beta, \delta \in \mathcal{H}.$$

Pak z

$$\|\varphi_{n_m} - \varphi_m\| \xrightarrow{P} 0, \quad \|C_{n_m} - C_m\| \xrightarrow{P} 0 \quad \text{a} \quad \|S_{n_m} - S_m\| \xrightarrow{P} 0, \quad m = 1, 2,$$

získáváme

$$\int_{\mathbb{R}} Ew(t) dt \xrightarrow{P} 0,$$

neboť

$$\int_{\mathbb{R}} Ew(t) dt = \int_{\mathbb{R}} 2\operatorname{Re}[A]\operatorname{Re}[C]w(t) dt - \int_{\mathbb{R}} 2\operatorname{Re}[B]\operatorname{Re}[C]w(t) dt \\ - \int_{\mathbb{R}} 2\operatorname{Re}[A]\operatorname{Re}[B]w(t) dt + \int_{\mathbb{R}} 2\operatorname{Im}[A]\operatorname{Im}[C]w(t) dt \\ - \int_{\mathbb{R}} 2\operatorname{Im}[B]\operatorname{Im}[C]w(t) dt - \int_{\mathbb{R}} 2\operatorname{Im}[A]\operatorname{Im}[B]w(t) dt$$

a jednotlivé sčítance tohoto výrazu konvergují v pravděpodobnosti k nule analogicky dle (1.7).

Celkem tedy

$$\int_{\mathbb{R}} D_{n_1, n_2}^2(t) w(t) dt = \underbrace{\int_{\mathbb{R}} |\varphi_{n_1}(t) - \varphi_1(t)|^2 w(t) dt}_{\xrightarrow{P} 0} + \underbrace{\int_{\mathbb{R}} |\varphi_{n_2}(t) - \varphi_2(t)|^2 w(t) dt}_{\xrightarrow{P} 0} \\ + \int_{\mathbb{R}} |\varphi_2(t) - \varphi_1(t)|^2 w(t) dt + \underbrace{\int_{\mathbb{R}} Ew(t) dt}_{\xrightarrow{P} 0}.$$

Tím dostáváme požadované

$$\frac{NT_{n_1, n_2}}{n_1 n_2} = \int_{\mathbb{R}} D_{n_1, n_2}^2(t) w(t) dt \xrightarrow{P} \int_{\mathbb{R}} D^2 w(t) dt = \int_{\mathbb{R}} |\varphi_2(t) - \varphi_1(t)|^2 w(t) dt.$$

□

Z uvedeného vztahu je zřejmé, že pro  $F \equiv G$  je  $\Delta := \int_{-\infty}^{+\infty} D^2(t)w(t)dt$  rovno nule, v opačném případě je tato hodnota ostře větší od nuly. Proto test, zamítající  $H_0$  pro velké hodnoty  $T_{n_1, n_2}$  na hladině  $\alpha$ , je konzistentní vůči každé dvojici pevných alternativních rozdělení.

Velikost integrálu  $\Delta$  vlastně určuje vzdálenost mezi distribučními funkcemi  $F_1$  a  $F_2$  (pro  $F_i$  značí distribuční funkci  $i$ -tého výběru) a jeho hodnota ukazuje na sílu testu. Pro konkrétní volbu váhové funkce je pak síla testu funkcí pouze parametru  $a$ , tedy  $\Delta = \Delta_a$ . Hodnota tohoto integrálu pro charakteristické funkce obou rozdělení tedy ukazuje na citlivost testu pro konkrétní dvojici rozdělení.

# Kapitola 3

## K-výběrový problém

### 3.1 Testová statistika

Přejdeme nyní od dvouvýběrového problému k obecnějšímu zadání-necht' je naším úkolem porovnávat rozdělení více výběrů, řešme úlohu pro obecné  $k$ .

Necht'  $\mathbf{Y}_j = (Y_{j,1}, \dots, Y_{j,n_j})^T$ ,  $j = 1, \dots, k$  je  $k$ -nezávislých náhodných výběrů, kde  $j$ -tý výběr obsahuje  $n_j$  pozorování a přísluší mu distribuční funkce  $F_j$ . Testujeme hypotézu

$$(3.1) \quad H_0 : F_1 = \dots = F_k \quad \text{proti alternativě} \quad H_1 : H_0 \quad \text{neplatí,}$$

kteřou můžeme opět ekvivalentně zapsat ve tvaru

$$(3.2) \quad H_0 : \varphi_1 = \dots = \varphi_k \quad \text{proti alternativě} \quad H_1 : H_0 \quad \text{neplatí,}$$

Označme

$$\hat{\phi}_j(t) = \frac{1}{n_j} \sum_{s=1}^{n_j} \exp\{itY_{js}\}, \quad t \in \mathbb{R}, \quad j = 1, \dots, k$$

empirickou charakteristickou funkci  $j$ -tého výběru a

$$\hat{\phi}(t) = \sum_{j=1}^k \frac{n_j}{n} \hat{\phi}_j(t), \quad t \in \mathbb{R}$$

empirickou charakteristickou funkci sdruženého výběru  $\mathbf{Y}_1, \dots, \mathbf{Y}_k$ .

Předpokládejme, že velikost výběru splňuje podmínku

$$(3.3) \quad \lim_{n \rightarrow \infty} \frac{n_j}{n} = p_j \in (0, \infty), \quad \text{kde } n = \sum_{j=1}^k n_j$$

a že pro nenulovou váhovou funkci  $w(\cdot)$  platí:

$$(3.4) \quad w(t) = w(-t) \quad \text{a} \quad 0 < \int_{\mathbb{R}} w(t)dt < \infty, \quad t \in \mathbb{R}.$$

Testová statistika má pak tvar

$$(3.5) \quad T_n(w, k) = \int_{\mathbb{R}} \sum_{j=1}^k n_j \left| \widehat{\phi}_j(t) - \widehat{\phi}(t) \right|^2 w(t)dt.$$

Hypotézu  $H_0$  zamítáme pro velké hodnoty testové statistiky  $T_n(w, k)$ .

Tuto statistiku můžeme opět vyjádřit pomocí goniometrických funkcí, ukažme si jak.

Vzdálenost  $\left| \widehat{\phi}_j(t) - \widehat{\phi}(t) \right|^2 =: \beta$  můžeme rozepsat podobně jako v předchozí kapitole následujícím způsobem:

$$\begin{aligned} \beta = & \left( \frac{1}{n_j} \sum_{s=1}^{n_j} [\cos(tY_{js}) + \sin(tY_{js}) - E_{H_0}(\cos(tY_{js}) + \sin(tY_{js}))] - \right. \\ & \left. - \sum_{j=1}^k \frac{1}{n} \sum_{s=1}^{n_j} [\cos(tY_{js}) + \sin(tY_{js}) - E_{H_0}(\cos(tY_{js}) + \sin(tY_{js}))] \right)^2. \end{aligned}$$

Položíme-li

$$(3.6) \quad Z_{jn}(t) := \frac{1}{\sqrt{n}} \sum_{s=1}^{n_j} \left( (\cos(tY_{js}) + \sin(tY_{js})) - E_{H_0}(\cos(tY_{js}) + \sin(tY_{js})) \right),$$

dostaneme

$$\beta = \left| \frac{\sqrt{n}}{n_j} Z_{jn}(t) - \sum_{j=1}^k \frac{1}{\sqrt{n}} Z_{jn}(t) \right|^2.$$

Máme tedy

$$\begin{aligned} T_n(w, k) &= \int_{\mathbb{R}} \sum_{j=1}^k \left| \frac{\sqrt{n}}{\sqrt{n_j}} Z_{jn}(t) - \sum_{j=1}^k \frac{\sqrt{n_j}}{\sqrt{n}} Z_{jn}(t) \right|^2 w(t)dt \\ &= \int_{\mathbb{R}} \sum_{j=1}^k \frac{n}{n_j} \left| Z_{jn}(t) - \sum_{j=1}^k \frac{n_j}{n} Z_{jn}(t) \right|^2 w(t)dt, \end{aligned}$$

kde pro

$$\bar{Z}_n(t) := \frac{1}{n} \sum_{j=1}^k Z_{jn}(t)$$



získáváme výsledný tvar testové statistiky pro  $k$ -výběrový problém

$$T_n(w, k) = \int_{\mathbb{R}} \sum_{j=1}^k \frac{n}{n_j} (Z_{jn}(t) - n_j \bar{Z}_n(t))^2 w(t) dt.$$

Vektor  $Z_{jn}$ ,  $j = 1, \dots, k$  si v dalším textu označme jako

$$(3.7) \quad \mathbf{Z}_n(t) := (Z_{1n}(t), \dots, Z_{kn}(t))^T,$$

pro zjednodušení zápisu můžeme dále položit

$$(3.8) \quad \mathbf{Z}_n := \{\mathbf{Z}_n(t), t \in \mathbb{R}\}.$$

## 3.2 Mnohonásobná porovnávání

Porovnáváme-li více než dva výběry, je namístě diskutovat hladinu  $\alpha$ , na které je test prováděn. Pokud bychom se totiž snažili tuto situaci převést na případ dvouvýběrového testu, tj. z daných  $k$ -výběrů vytvořili  $k(k-1)/2$  dvojic a na každou z nich aplikovali předchozí dvouvýběrový test, nesplňoval by tento postup testování podmínku udržení hladiny  $\alpha$ , tj. velikost pravděpodobnosti chyby prvního druhu.

Za platnosti nulové hypotézy by nám zřejmě dvouvýběrový test pro každou z dvojic dal signifikantní výsledek s pravděpodobností  $\alpha$ . My však chceme  $H_0$  zamítnout, pokud dá alespoň jeden z  $k(k-1)/2$  testů signifikantní výsledek (vidíme tedy, že pravděpodobnost zamítnutí  $H_0$ , je-li správná (tedy pravděpodobnost chyby 1. druhu), by byla při  $k > 2$  vyšší než požadovaná  $\alpha$ ).

Pokud  $H_0$  zamítneme, je třeba také rozhodnout, které výběry se od sebe významně liší. To se provádí pomocí metod mnohonásobného porovnávání.

Prověřme tedy, zda námi zvolená testová statistika pro  $k$ -výběrový problém odpovídá dané hladině  $\alpha$ . Nejprve si vyjádříme jádro naší testové statistiky (které ještě podělíme  $n$ ) jako kvadratickou formu:

$$(3.9) \quad \sum_{j=1}^k \frac{1}{n_j} (Z_{jn}(t) - n_j \bar{Z}_n(t))^2 = (Z_{1n}^0(t), \dots, Z_{kn}^0(t))^T \mathbf{A} (Z_{1n}^0(t), \dots, Z_{kn}^0(t)),$$

kde

$$Z_{jn}^0(t) = Z_{jn}(t) / \sqrt{n_j}, \quad j = 1, \dots, k,$$

a  $\mathbf{A}$  je idempotentní matice s prvky  $a_{ij} = I_{\{j=v\}} - \frac{\sqrt{n_j n_v}}{n}$ ,  $j, v = 1, \dots, k$ , pro  $I_{\{K\}}$  značící indikátor množiny  $K$ .

Nyní na daný problém aplikujeme Scheffého metodu (její přesný popis čtenář najde např. v [1]). Ta je postavena na speciálním tvaru Schwarzovy nerovnosti která říká, že pro matici  $\mathbf{M}_{t \times t} > 0$  a pro libovolný vektor  $\mathbf{x} \in \mathbb{R}_t$  platí ekvivalence

$$(3.10) \quad [\mathbf{x}^T \mathbf{M} \mathbf{x} \leq 1] \iff [(\mathbf{h}^T \mathbf{x})^2 \leq \mathbf{h}^T \mathbf{M}^{-1} \mathbf{h} \quad \forall \mathbf{h} \in \mathbb{R}_t].$$

Připomeňme, že matice  $\mathbf{B}$  se nazývá idempotentní, platí-li  $\mathbf{B}^2 = \mathbf{B}$ . Naše matice  $\mathbf{A}$  tuto podmínku splňuje, neboť  $\mathbf{A} = \mathbf{I} - \frac{\sqrt{n_j n_s}}{n} \mathbf{a}$

$$\begin{aligned} \mathbf{A}^2 &= \mathbf{I} - 2 \frac{\sqrt{n_j n_s}}{n} + \sum_{i=1}^k \frac{\sqrt{n_j n_i}}{n} \frac{\sqrt{n_i n_s}}{n} = \mathbf{I} - 2 \frac{\sqrt{n_j n_s}}{n} + \frac{\sqrt{n_j n_s}}{n} \\ &= \mathbf{I} - \frac{\sqrt{n_j n_s}}{n}. \end{aligned}$$

Protože je  $\mathbf{A}$  idempotentní, lze za její pseudoinverzní matici vzít matici jednotkovou, tj.  $\mathbf{A}^- = \mathbf{I}$ .

Nyní tedy na daný rozklad použijeme zmíněnou Schwarzovu nerovnost (3.10) a dostáváme, že

$$(3.11) \quad \mathbf{Z}^{0T} \mathbf{A} \mathbf{Z}^0 \geq \max_{\mathbf{a}} \frac{(\mathbf{a}^T \mathbf{Z}^0)^2}{\mathbf{a}^T \mathbf{a}} \quad \text{pro všechna } \mathbf{a} \in \mathbb{R}^k.$$

Položme pro pevné  $j$  a  $v$ ,  $1 \leq j < v \leq k$ , vektor  $\mathbf{a}$  tak, aby

$$\mathbf{a} = (0, \dots, 0, \underbrace{1/\sqrt{n_j}}_{j\text{-tá pozice}}, 0, \dots, 0, \underbrace{-1/\sqrt{n_v}}_{v\text{-tá pozice}}, 0, \dots, 0).$$

Pak získáme po dosazení do (3.11)

$$\sum_{j=1}^k \frac{1}{n_j} (Z_{jn}(t) - n_j \bar{Z}_n(t))^2 \geq \max_{1 \leq j < v \leq k} \frac{\left| \frac{Z_{jn}(t)}{n_j} - \frac{Z_{vn}(t)}{n_v} \right|^2}{\frac{1}{n_j} + \frac{1}{n_v}}.$$

Pronásobíme-li nyní obě strany číslem  $n$ , dostaneme se k výsledné nerovnici

$$(3.12) \quad \sum_{j=1}^k \frac{n}{n_j} (Z_{jn}(t) - n_j \bar{Z}_n(t))^2 \geq \max_{1 \leq j < v \leq k} \frac{n_j n_v}{n_j + n_v} |\widehat{\varphi}_j(t) - \widehat{\varphi}_v(t)|^2.$$

Hypotézu  $H_0$  (3.2) zamítneme na hladině  $\alpha$ , pokud

$$T_n(w, k) \geq t(\alpha, n),$$

kde  $t(\alpha, n)$  je daná kritická hodnota (resp. její asymptotický odhad). Pokud tedy  $H_0$  zamítáme na hladině  $\alpha$ , můžeme na základě výše odvozené nerovnosti (3.12) na stejné hladině  $\alpha$  a při stejné kritické hodnotě  $t(\alpha, n)$  rovněž testovat hypotézu  $H_{jv} : \varphi_j \equiv \varphi_v$  s oborem kritických hodnot

$$\frac{n_j n_v}{n_j + n_v} \int_{\mathbb{R}} |\widehat{\varphi}_j(t) - \widehat{\varphi}_v(t)|^2 w(t) d(t) \geq t(\alpha, n), \quad 1 \leq j < v \leq k.$$

Takto upraveným dvouvýběrovým testem tedy můžeme zkoumat  $k$ -výběrový problém na dané hladině  $\alpha$  postupem uvedeným na začátku tohoto paragrafu. Zároveň odsud dostáváme, která dvojice z  $k$  výběrů se od sebe významně liší.

### 3.3 Asymptotické vlastnosti

Asymptotické vlastnosti statistiky  $T_n(w, k)$  zkoumejme v Hilbertově prostoru  $\mathcal{H}$  definovaném v kapitole 1.2.1.

#### 3.3.1 Chování při platnosti $H_0$

**Věta 6.** *Budte  $\mathbf{Y}_j = (Y_{j,1}, \dots, Y_{j,n_j})^T$ ,  $j = 1, \dots, k$ ,  $k$ - nezávislých výběrů nezávislých stejně rozdělených náhodných veličin se společnou distribuční funkcí  $F$ . Nechť jsou splněny podmínky (3.3) a (3.4). Pak pro  $n \rightarrow \infty$  platí:*

$$\mathbf{Z}_n \xrightarrow{D} \mathbf{Z}$$

a

$$T_n(w, k) \xrightarrow{D} \int_{\mathbb{R}} \sum_{j=1}^k \frac{1}{p_j} (Z_j(t) - p_j \bar{Z}(t))^2 w(t) dt,$$

kde  $\mathbf{Z}_n$  je definováno v (3.8),  $\bar{Z}(t) = \sum_{j=1}^k Z_j(t)$  a

$\mathbf{Z} = \{(Z_1(t), \dots, Z_k(t))^T\}$ ,  $t \in \mathbb{R}$  je  $k$ -rozměrný gaussovský proces s nulovou střední hodnotou, nezávislými přírůstky a kovarianční strukturou  $\text{cov}(Z_j(t_1), Z_j(t_2)) = p_j \text{cov}(\cos(t_1 Y) + \sin(t_1 Y), \cos(t_2 Y) + \sin(t_2 Y))$ ,  $t_1, t_2 \in \mathbb{R}$ , kde  $Y$  má distribuční funkci  $F$ .

*Důkaz.* Důkaz je analogií důkazu věty 2, proto jej zde už nebudu blíže rozepisovat. Pouze uveďme, že při bližším pohledu na strukturu veličiny  $Z_{jn}$  definovanou v (3.6) vidíme, že jde opět o součet nezávislých náhodných veličin, jejichž střední hodnota je při platnosti nulové hypotézy rovna nule a která má konečné druhé momenty. Dále je tato veličina spojitá pro  $\forall t \in \mathbb{R}$ . Tyto vlastnosti veličiny  $Z_{jn}(t)$  nám v Hilbertově prostoru umožní použít centrální

větu a *continuous mapping theorem* pro získání limitního rozdělení testové statistiky  $T_n(w, k)$ .  $\square$

Víme už tedy, jaké je asymptotické rozdělení  $T_n(w, k)$  za platnosti hypotézy  $H_0$ , známe-li testovanou distribuční funkci výběru. Ale co její chování při obecnějších alternativách?

### 3.3.2 Chování při platnosti pevné alternativy

Vyslovme nyní větu zaručující konzistenci uvedeného dvouvýběrového testu, zamítajícího hypotézu (3.1) pro velké hodnoty testové statistiky.

**Věta 7.** *Budte  $\{Y_{m,j}\}_{j=1}^{n_m}$ ,  $m = 1, \dots, k$ , na sobě nezávislé náhodné výběry nezávislých náhodných veličin, necht' pro velikost  $j$ -tého výběru platí podmínka (3.3) a  $\varphi_m$  značí charakteristickou funkci tohoto výběru,  $j = 1, \dots, k$ . Pak pro  $n \rightarrow \infty$  platí*

$$\frac{1}{n} T_n(w, k) \xrightarrow{P} \int_{\mathbb{R}} \sum_{j=1}^k p_j |\varphi_j(t) - \bar{\varphi}(t)|^2 w(t) dt, \quad \text{pro } \forall t \in \mathbb{R},$$

kde  $\bar{\varphi}(t) = \sum_{j=1}^k p_j \varphi_j(t)$ .

*Důkaz.* Důkaz věty je založen na stejném principu jako důkaz věty 2 a 4. Proto jej zde již nebudu uvádět.  $\square$

### 3.3.3 Chování při platnosti lokální alternativy

Omezme se nyní jen na případy, kdy je distribuční funkce  $F_0$  testovaného výběru při platnosti nulové hypotézy symetrická a absolutně spojitá s hustotou  $f_0$ . Testujme hypotézu  $H_0$  oproti obecné alternativě:

$$(3.13) \quad H_n : Y_{j,1}, \dots, Y_{j,n_j} \quad \text{jsou nezávislé stejně rozdělené veličiny s hustotou } g_{n,j},$$

kde

$$(3.14) \quad g_{n,j}(x) = \left(1 + \frac{\kappa}{\sqrt{n}} u_j(x)\right) f_0(x), \quad j = 1, \dots, k, \quad x \in \mathbb{R},$$

$\kappa \neq 0$  a  $u_j$ ,  $j = 1, \dots, k$  jsou měřitelné funkce takové, že

$$(3.15) \quad \int u_j(x) f_0(x) dx = 0 \quad \text{a} \quad 0 < \int u_j^2(x) f_0(x) dx < \infty, \quad j = 1, \dots, k, \quad x \in \mathbb{R}.$$

Při těchto omezeních již můžeme vyslovit následující větu:

**Věta 8.** *Nechť  $Y_{j,1}, \dots, Y_{j,n_j}$ ,  $j = 1, \dots, k$ , splňují model (3.1), kde  $Y_{j,s}$  jsou nezávislé stejně rozdělené veličiny s hustotou  $g_{n,j}$  definovanou v (3.14) a splňující (3.15). Nechť pro váhovou fci  $w(t), t \in \mathbb{R}$  platí (3.4). Pak existuje gaussovský proces  $\mathbf{Z} = \{\mathbf{Z}(t); t \in \mathbb{R}\}$  s nulovou střední hodnotou takový, že při  $n \rightarrow \infty$  platí*

$$\{\mathbf{Z}_n(t); t \in \mathbb{R}\} \xrightarrow{D} \{\mathbf{Z}(t) + \kappa \boldsymbol{\mu}(t); t \in \mathbb{R}\}$$

a

$$\|T_n(w, k)\|^2 \xrightarrow{D} \int_{\mathbb{R}} \sum_{j=1}^k \frac{1}{p_j} (Z_j(t) - p_j \bar{Z}(t) - \kappa p_j \mu_j^0(t))^2 w(t) dt.$$

Proces  $\{\mathbf{Z}(t); t \in \mathbb{R}\}$  je gaussovský proces z věty 6 a  $\boldsymbol{\mu}(t)$  a  $\boldsymbol{\mu}^0(t)$  jsou vektory se složkami

$$\mu_j(t) = \int (\cos(tx) + \sin(tx)) u_j(x) f_0(x) dx, \quad j = 1, \dots, k,$$

a

$$\mu_j^0(t) = \mu_j(t) - p_j \sum_{s=1}^k p_s \mu_s(t), \quad j = 1, \dots, k, \quad t \in \mathbb{R}.$$

*Důkaz.* Důkaz věty zde nebudeme uvádět, neb matematický aparát použitý při jeho konstrukci potřebuje větší teoretické zázemí, které však přesahuje rámec této diplomové práce. Čtenář si jej však může vyhledat ve článku [5].  $\square$

Všimněme si, že při rostoucím  $\kappa$  (tvrzení platí pro libovolné  $\kappa \neq 0$ ) a pro  $\|\boldsymbol{\mu}^0\|^2$  výrazně převažuje vliv nenáhodné složky  $T_n(w, k)$ , tj.  $\kappa \|\boldsymbol{\mu}^0\|^2$  nad náhodnou a konverguje k nekonečnu. Proto je tento test konzistentní při  $\kappa \rightarrow \infty$ .

# Kapitola 4

## Simulace

Uvedené testové statistiky závisí na hypotetickém rozdělení i při platnosti nulové hypotézy, což nás může od užívání těchto testů trochu odrazovat. Je proto vhodné si pomocí simulací ukázat i výhody těchto testů. Omezíme se zde pouze na dvouvýběrový problém, tj. na statistiku

$$T_{n_1, n_2} = \frac{n_1 n_2}{n_1 + n_2} \int_{-\infty}^{+\infty} D_{n_1, n_2}^2(t) w(t) dt \quad \text{pro } \forall t \in \mathbb{R},$$

kde

$$D_{n_1, n_2}^2(t) = |\varphi_{n_1}(t) - \varphi_{n_2}(t)|^2.$$

Pro určení kritických hodnot dané dvojice výběrů můžeme použít např. bootstrap s vrácením nebo bez vrácení, my se zaměříme na první jmenovanou metodu-tedy bootstrap s vrácením.

Popis metody je jednoduchý:

1. Z původních výběrů  $X_1, \dots, X_{n_1}$  a  $Y_1, \dots, Y_{n_2}$  vytvoříme sdružený výběr  $Z_1, \dots, Z_N$ ,  $N = n_1 + n_2$ . Z tohoto výběru postupně a náhodně vybereme s vrácením nový výběr-říkejme mu bootstrapový a označme jej  $X_1^*, \dots, X_{n_1}^*$  a  $Y_1^*, \dots, Y_{n_2}^*$ .
2. Takových bootstrapových výběrů provedeme celkem  $B$  - a pro každý z nich spočteme testovou statistiku  $T_{n_1, n_2}^*$ . Máme tedy celkem  $B$  hodnot testových statistik pro konkrétní bootstrapové verze, na základě kterých lze získat aproximaci pro podmíněné rozdělení (podmíněné sdruženým výběrem  $Z_1, \dots, Z_N$ ) zkoumaných statistik.
3. Kritické hodnoty pro danou dvojici výběrů se poté určí jako  $(1 - \alpha)$  kvantily tohoto rozdělení.

O vhodnosti použití této aproximace pojednává následující věta.

**Věta 9.** *Budte  $\{X_{m,j}\}_{j=1}^{n_m}$ ,  $m = 1, 2$ , dva nezávislé náhodné výběry, každému výběru nechť přísluší charakteristická funkce  $\varphi_m$ . Bud' dále  $w(t)$ ,  $t \in \mathbb{R}$  taková, že  $w(t) = w(-t) \forall t \in \mathbb{R}$  a  $0 < \int_{\mathbb{R}} w(t) dt < \infty$ ,  $t \in \mathbb{R}$ . Předpokládejme, že pro velikost výběrů platí  $\lim_{n \rightarrow \infty} n_j/n = p_j \in (0, 1)$ ,  $j = 1, \dots, k$ . Pak pro  $n \rightarrow \infty$  a pro  $\forall x \in \mathbb{R}$  platí*

$$P(T_{n_1, n_2}^* \leq x | X_{m,j}; j = 1, \dots, n_m, m = 1, 2) - P_{\bar{F}}(T_{n_1, n_2} \leq x) \longrightarrow 0 \quad \text{sk. j.},$$

kde  $P_{\bar{F}}$  je pravděpodobnost počítaná při rozdělení  $\bar{F}(x) = \sum_{j=1}^k p_j F_j(x)$ ,  $x \in \mathbb{R}$ , kde  $F_j$  značí distribuční funkci  $j$ -tého výběru.

*Důkaz.* Věta je převzata z článku [6], kde je rovněž dokázána. □

Vidíme tedy, že za platnosti  $H_0$  má testová statistika vypočtena na základě bootstrapových výběrů  $X_1^*, \dots, X_{n_1}^*$  a  $Y_1^*, \dots, Y_{n_2}^*$  stejné limitní rozdělení (které je ovšem podmíněné původní dvojicí výběrů) jako testová statistika vypočtená z původního výběru  $X_1, \dots, X_{n_1}$  a  $Y_1, \dots, Y_{n_2}$ .

Za platnosti alternativy je statistika  $T_{n_1, n_2}^*$  ohraničena v pravděpodobnosti,  $T_{n_1, n_2}^* = O_P(1)$ , zatímco statistika spočtená z původní dvojice výběrů v tomto případě nabývá velkých hodnot (konverguje v pravděpodobnosti k nekonečnu). Proto aproximace kritických hodnot pomocí metody bootstrap s vrácením vede ke konzistentním testům, tj. při použití takto získaných kritických hodnot zamítneme nulovou hypotézu při platnosti pevné alternativy s pravděpodobností jedna pro  $n \rightarrow \infty$ .

## 4.1 Citlivost testu na změnu parametrů

Prozkoumejme nyní citlivost dvouvýběrového testu na změnu parametrů některých konkrétních rozdělení. Zaměříme se na normální, Cauchyho a dvojité exponenciální (tzv. Laplaceovo) rozdělení. Pro každé z těchto rozdělení zkoumáme podobnou situaci - zatímco první výběr generujeme vždy se stejnými parametry  $\mu=0$  a  $\sigma^2=1$ , v druhém tyto parametry postupně měníme. Nejprve při pevném  $\mu = 0$  volíme různé parametry  $\sigma^2$  a poté situaci obrátíme, tj. pro pevné  $\sigma^2=1$  zvětšujeme parametr  $\mu$ .

### 4.1.1 Pravděpodobnosti zamítnutí

Podívejme se nejprve, jak se při konkrétních změnách daných parametrů mění tzv. *rejection rates*, tedy pravděpodobnosti zamítnutí hypotézy  $H_0$ .

Pro každou danou kombinaci parametrů postupně generujeme 2000 dvojic náhodných výběrů o velikostech  $n_1 = n_2 = 30$ . Pro každou z těchto dvojic dále simulujeme  $B = 2000$  bootstrapových výběrů, na jejichž základě spočteme kritické hodnoty pro danou dvojici na hladině  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ . Obdržené kritické hodnoty porovnáme s testovou statistikou spočítanou na základě původního výběru a rozhodneme, zda pro danou dvojici hypotézu  $H_0$  zamítneme či nikoliv. Z takto získaných hodnot spočteme pravděpodobnost zamítnutí pro daný testovací vzorek, uvádět ji budeme v procentech.

Tento postup aplikujeme při dané testové statistice pro obě doporučené váhové funkce  $w_a^1(t) = \exp(-a|t|)$ ,  $a > 0$ ,  $t \in \mathbb{R}$ , a  $w_a^2(t) = \exp(-at^2)$ ,  $a > 0$ ,  $t \in \mathbb{R}$ , tj. pro testové statistiky (2.8) a (2.9). Simulace dále provádíme pro různé hodnoty parametru  $a$  a to pro  $a = 0,25; 0,5; 1; 2$  a  $3$ .

Při pohledu na výsledné tabulky (4.1)-(4.6) vidíme, že při platnosti  $H_0$  (tedy v případě, kdy jsou oba výběry brány ze stejného rozdělení beze změny parametrů), nám ve všech šesti případech procentuální podíl zamítnutých hypotéz víceméně věrně kopíruje hladiny  $\alpha_1$  a  $\alpha_2$ , na kterých byl test prováděn, a to pro všechny parametry  $a$ . Dle očekávání nám rovněž toto procento stoupá, čím více se od  $H_0$  vzdalujeme-tedy čím je rozdíl mezi danými parametry výraznější. Z testovaných rozdělení test nejcitlivěji reaguje na změnu parametrů u normálního rozdělení. Nejpomalejší reakci na změnu parametrů vidíme u Cauchyho rozdělení, nicméně přesto lze říci, že je test vůči změnám velikosti parametrů velice citlivý-a to bez ohledu na to, zda měníme parametr střední hodnoty  $\mu$  či rozptylu  $\sigma^2$ , přestože na změnu  $\mu$  reaguje rychleji. Citlivost testu rovněž očekávaně stoupá při vzrůstajícím parametru  $a$ , neboť jak jsme již diskutovali v kapitole 1.2., je v tomto případě konvergence obou testových statistik k ose reálných čísel rychlejší.



N(0,1) N( $\mu,1$ )	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\mu = 0.0$										
$T_{n_1 n_2}^{(1)}$	4.2	9.1	5.0	9.6	6.3	11.1	4.6	9.8	4.8	9.0
$T_{n_1 n_2}^{(2)}$	4.1	9.1	5.3	10.9	5.1	11.8	4.5	9.7	4.6	9.4
$\mu = 0.1$										
$T_{n_1 n_2}^{(1)}$	5.3	10.9	5.4	11.1	5.2	10.5	6.5	12.3	5.7	10.5
$T_{n_1 n_2}^{(2)}$	5.6	11.7	6.9	13.0	5.7	11.2	6.6	13.0	6.0	11.0
$\mu = 0.2$										
$T_{n_1 n_2}^{(1)}$	6.3	12.1	8.3	15.1	9.1	16.1	10.7	17.7	11.2	19.2
$T_{n_1 n_2}^{(2)}$	9.7	14.9	10.2	17.2	10.4	18.0	12.0	19.8	11.6	19.2
$\mu = 0.3$										
$T_{n_1 n_2}^{(1)}$	8.6	16.7	11.3	18.3	16.0	25.7	17.9	26.8	18.3	28.5
$T_{n_1 n_2}^{(2)}$	13.8	23.3	15.1	23.3	19.6	30.1	19.2	28.4	19.0	30.1
$\mu = 0.4$										
$T_{n_1 n_2}^{(1)}$	13.5	22.8	18.8	29.6	24.5	36.6	29.7	42.1	32.3	44.2
$T_{n_1 n_2}^{(2)}$	21.0	32.0	26.9	38.6	31.4	43.2	32.8	44.4	33.3	45.8
$\mu = 0.5$										
$T_{n_1 n_2}^{(1)}$	18.3	29.8	27.0	38.1	35.3	48.1	43.4	55.7	45.3	58.5
$T_{n_1 n_2}^{(2)}$	31.7	44.0	37.8	50.4	43.4	55.5	45.9	59.7	46.6	60.0
$\mu = 0.6$										
$T_{n_1 n_2}^{(1)}$	28.2	39.8	36.1	48.7	45.6	58.9	55.1	68.4	59.1	71.6
$T_{n_1 n_2}^{(2)}$	43.5	56.9	49.9	62.4	55.3	68.6	58.7	71.0	60.5	73.6
$\mu = 0.8$										
$T_{n_1 n_2}^{(1)}$	46.3	60.0	59.7	73.0	72.8	82.3	82.7	89.7	84.1	90.5
$T_{n_1 n_2}^{(2)}$	67.3	78.9	74.9	83.6	80.5	88.5	85.3	91.5	84.9	91.2
$\mu = 1.0$										
$T_{n_1 n_2}^{(1)}$	67.7	80.0	82.2	89.6	90.3	94.9	95.3	97.6	95.3	97.7
$T_{n_1 n_2}^{(2)}$	87.1	93.6	91.2	95.2	94.7	97.3	96.7	98.0	95.6	98.1

Tabulka 4.1: Pravděpodobnosti zamítnutí (v %) pro N(0,1) při změně parametru  $\mu$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  byly počítány pro  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

C(0,1) C( $\mu,1$ )	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\mu = 0.0$										
$T_{n_1 n_2}^{(1)}$	4,4	9,9	4,8	9,8	4,8	9,9	4,8	9,2	4,2	8,7
$T_{n_1 n_2}^{(2)}$	5,7	12,0	5,0	11,3	4,8	9,1	4,9	9,3	4,6	8,8
$\mu = 0.1$										
$T_{n_1 n_2}^{(1)}$	4,1	10,3	4,8	9,8	5,7	10,3	5,1	10,5	3,5	8,9
$T_{n_1 n_2}^{(2)}$	5,1	10,3	5,9	10,0	5,8	10,2	5,2	10,1	3,5	9,4
$\mu = 0.2$										
$T_{n_1 n_2}^{(1)}$	5,2	10,3	6,7	12,8	6,4	11,5	7,3	13,9	6,1	11,1
$T_{n_1 n_2}^{(2)}$	6,3	12,4	7,3	13,1	6,1	11,4	7,2	13,4	5,8	11,1
$\mu = 0.3$										
$T_{n_1 n_2}^{(1)}$	6,7	12,5	8,2	14,9	8,8	15,8	8,9	15,4	8,0	15,8
$T_{n_1 n_2}^{(2)}$	8,8	15,7	9,0	15,6	9,2	15,9	8,9	14,6	7,7	15,9
$\mu = 0.4$										
$T_{n_1 n_2}^{(1)}$	10,6	18,3	11,1	18,8	14,1	23,7	13,6	21,3	11,8	19,9
$T_{n_1 n_2}^{(2)}$	15,4	23,8	13,1	21,2	13,3	23,1	11,9	19,8	11,6	19,8
$\mu = 0.5$										
$T_{n_1 n_2}^{(1)}$	11,8	20,6	17,3	27,3	18,2	28,8	18,8	28,5	16,7	27,9
$T_{n_1 n_2}^{(2)}$	16,7	25,2	20,3	31,0	18,5	28,7	17,4	27,2	16,0	26,7
$\mu = 0.6$										
$T_{n_1 n_2}^{(1)}$	16,1	26,1	21,9	33,0	23,1	34,4	24,4	36,9	21,9	34,1
$T_{n_1 n_2}^{(2)}$	22,9	34,8	25,0	37,4	23,0	34,4	22,9	35,4	20,8	32,8
$\mu = 0.8$										
$T_{n_1 n_2}^{(1)}$	30,4	43,1	35,9	49,1	42,0	55,5	40,0	52,7	37,0	50,8
$T_{n_1 n_2}^{(2)}$	41,9	56,3	41,5	54,4	42,2	55,0	37,3	51,7	35,2	49,0
$\mu = 1.0$										
$T_{n_1 n_2}^{(1)}$	43,2	57,2	52,4	65,8	59,6	71,7	58,2	70,0	55,5	67,9
$T_{n_1 n_2}^{(2)}$	58,0	70,0	60,4	71,9	59,6	70,8	53,9	66,8	53,8	65,3

Tabulka 4.2: Pravděpodobnosti zamítnutí (v %) pro C(0,1) při změně parametru  $\mu$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  byly počítány pro  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

DEx(0,1)	a=0.25		a=0.5		a=1		a=2		a=3	
DEx( $\mu,1$ )	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\mu = 0.0$										
$T_{n_1 n_2}^{(1)}$	4,3	9,6	4,7	10,3	5,6	10,3	5,0	10,1	4,2	9,8
$T_{n_1 n_2}^{(2)}$	5,4	10,5	5,1	10,3	5,8	10,0	4,7	10,2	4,3	9,9
$\mu = 0.1$										
$T_{n_1 n_2}^{(1)}$	5,6	10,2	5,3	10,7	5,5	11,1	6,0	11,7	4,9	11,0
$T_{n_1 n_2}^{(2)}$	5,7	12,3	4,8	10,7	5,5	11,1	5,6	11,4	4,9	11,0
$\mu = 0.2$										
$T_{n_1 n_2}^{(1)}$	6,6	12,4	8,6	15,3	10,5	17,2	9,0	16,4	10,7	18,1
$T_{n_1 n_2}^{(2)}$	7,5	13,9	8,8	15,7	10,1	17,4	9,1	16,2	10,4	17,3
$\mu = 0.3$										
$T_{n_1 n_2}^{(1)}$	10,1	18,0	12,2	21,2	14,3	23,2	15,3	23,2	15,0	22,8
$T_{n_1 n_2}^{(2)}$	13,8	23,0	14,3	23,3	15,2	22,8	14,8	22,8	14,7	23,2
$\mu = 0.4$										
$T_{n_1 n_2}^{(1)}$	15,7	24,9	20,9	31,7	22,0	32,1	24,4	35,4	21,9	33,4
$T_{n_1 n_2}^{(2)}$	20,2	30,5	23,9	34,3	22,9	31,7	24,4	34,5	21,6	33,4
$\mu = 0.5$										
$T_{n_1 n_2}^{(1)}$	22,9	34,0	29,5	41,7	31,5	44,4	33,9	46,5	34,9	46,6
$T_{n_1 n_2}^{(2)}$	30,2	42,7	33,5	45,0	32,4	44,8	33,0	46,1	34,5	45,5
$\mu = 0.6$										
$T_{n_1 n_2}^{(1)}$	31,7	44,0	37,4	50,0	43,1	56,0	47,4	59,5	45,2	58,1
$T_{n_1 n_2}^{(2)}$	41,0	54,6	41,5	53,6	44,3	57,1	45,7	58,6	45,0	56,9
$\mu = 0.8$										
$T_{n_1 n_2}^{(1)}$	51,6	64,3	59,0	70,0	67,5	77,9	69,1	79,3	69,7	79,1
$T_{n_1 n_2}^{(2)}$	64,7	74,3	64,5	75,4	69,0	78,8	67,9	78,7	68,0	78,1
$\mu = 1.0$										
$T_{n_1 n_2}^{(1)}$	66,8	77,8	77,4	85,7	85,4	91,1	86,8	92,2	85,8	92,0
$T_{n_1 n_2}^{(2)}$	79,8	87,9	84,0	90,4	86,6	92,3	86,1	91,9	84,4	91,2

Tabulka 4.3: Pravděpodobnosti zamítnutí (v %) pro DEx(0,1) při změně parametru  $\mu$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  byly počítány pro  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

N(0,1) N(0,σ <sup>2</sup> )	a=0.25		a=0.5		a=1		a=2		a=3	
	α <sub>1</sub>	α <sub>2</sub>	α <sub>1</sub>	α <sub>2</sub>	α <sub>1</sub>	α <sub>2</sub>	α <sub>1</sub>	α <sub>2</sub>	α <sub>1</sub>	α <sub>2</sub>
σ <sup>2</sup> = 1.0										
$T_{n_1 n_2}^{(1)}$	4.2	9.1	5.0	9.6	6.3	11.1	4.6	9.8	4.8	9.0
$T_{n_1 n_2}^{(2)}$	4.1	9.1	5.3	10.9	5.1	11.8	4.5	9.7	4.6	9.4
σ <sup>2</sup> = 1.1										
$T_{n_1 n_2}^{(1)}$	4,4	10,6	5,8	11,9	5,3	11,1	5,2	10,4	4,9	11,0
$T_{n_1 n_2}^{(2)}$	5,5	10,9	5,9	11,8	5,5	10,9	4,9	10,5	4,8	10,5
σ <sup>2</sup> = 1.2										
$T_{n_1 n_2}^{(1)}$	5,5	11,6	7,0	13,5	7,5	14,3	7,0	14,4	6,5	13,2
$T_{n_1 n_2}^{(2)}$	8,5	15,0	7,8	15,2	7,2	14,0	6,5	13,1	5,4	12,0
σ <sup>2</sup> = 1.3										
$T_{n_1 n_2}^{(1)}$	9,7	17,8	9,9	17,3	12,0	21,5	10,2	19,9	8,1	17,0
$T_{n_1 n_2}^{(2)}$	13,4	21,9	11,8	22,2	11,9	21,6	8,1	17,5	7,0	15,4
σ <sup>2</sup> = 1.4										
$T_{n_1 n_2}^{(1)}$	12,1	21,9	15,8	25,8	18,0	29,4	17,0	30,1	11,8	23,8
$T_{n_1 n_2}^{(2)}$	18,5	29,4	19,1	30,7	18,3	30,4	13,3	26,5	9,2	19,7
σ <sup>2</sup> = 1.5										
$T_{n_1 n_2}^{(1)}$	14,1	25,9	20,6	33,0	24,9	36,8	23,7	38,8	19,9	33,6
$T_{n_1 n_2}^{(2)}$	24,9	37,3	26,0	40,2	24,2	39,5	19,6	33,3	15,1	28,8
σ <sup>2</sup> = 1.6										
$T_{n_1 n_2}^{(1)}$	20,0	33,6	26,1	41,4	33,0	47,5	32,3	50,3	24,3	41,0
$T_{n_1 n_2}^{(2)}$	33,0	45,2	36,2	50,0	35,1	50,4	26,3	46,3	19,0	35,8
σ <sup>2</sup> = 1.8										
$T_{n_1 n_2}^{(1)}$	30,3	44,9	41,6	56,0	50,6	64,5	51,3	67,9	43,8	62,5
$T_{n_1 n_2}^{(2)}$	47,8	62,8	53,5	67,1	53,3	69,2	47,4	65,6	37,3	57,8
σ <sup>2</sup> = 2.0										
$T_{n_1 n_2}^{(1)}$	42,0	57,6	53,3	69,3	64,5	76,5	69,5	82,5	63,6	79,9
$T_{n_1 n_2}^{(2)}$	62,6	75,2	68,1	80,3	69,7	81,8	67,1	82,3	59,0	77,3

Tabulka 4.4: Pravděpodobnosti zamítnutí(v %) pro N(0,1) při změně parametru σ<sup>2</sup>. Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách α<sub>1</sub> = 5% a α<sub>2</sub> = 10%, při velikostech výběrů n<sub>1</sub> = n<sub>2</sub> = 30.

C(0,1) C(0,σ <sup>2</sup> )	a=0.25		a=0.5		a=1		a=2		a=3	
	α <sub>1</sub>	α <sub>2</sub>	α <sub>1</sub>	α <sub>2</sub>	α <sub>1</sub>	α <sub>2</sub>	α <sub>1</sub>	α <sub>2</sub>	α <sub>1</sub>	α <sub>2</sub>
σ <sup>2</sup> = 1.0										
$T_{n_1 n_2}^{(1)}$	4,4	9,9	4,8	9,8	4,8	9,9	4,8	9,2	4,2	8,7
$T_{n_1 n_2}^{(2)}$	5,7	12,0	5,0	11,3	4,8	9,1	4,9	9,3	4,6	8,8
σ <sup>2</sup> = 1.1										
$T_{n_1 n_2}^{(1)}$	4,5	9,3	4,7	10,0	4,4	9,1	4,8	9,9	4,4	9,2
$T_{n_1 n_2}^{(2)}$	4,8	9,1	4,6	10,3	4,7	9,3	5,2	9,9	4,4	9,9
σ <sup>2</sup> = 1.2										
$T_{n_1 n_2}^{(1)}$	4,2	9,6	5,3	11,5	6,7	12,1	5,7	12,3	5,5	11,5
$T_{n_1 n_2}^{(2)}$	6,2	11,7	6,2	12,7	6,6	12,9	5,5	12,5	5,7	12,6
σ <sup>2</sup> = 1.3										
$T_{n_1 n_2}^{(1)}$	5,4	10,8	6,8	13,9	8,1	14,3	7,8	14,1	6,3	12,6
$T_{n_1 n_2}^{(2)}$	8,2	15,0	7,8	15,0	8,1	15,1	7,7	13,7	6,3	12,3
σ <sup>2</sup> = 1.4										
$T_{n_1 n_2}^{(1)}$	7,2	13,0	9,3	16,3	11,1	18,8	10,5	18,3	10,8	18,1
$T_{n_1 n_2}^{(2)}$	9,3	16,4	10,3	17,7	11,8	19,5	10,3	17,8	10,7	17,8
σ <sup>2</sup> = 1.5										
$T_{n_1 n_2}^{(1)}$	8,2	15,4	11,0	20,3	12,3	21,0	13,8	22,5	11,3	20,5
$T_{n_1 n_2}^{(2)}$	13,3	20,8	13,6	23,0	12,5	21,1	13,2	21,6	11,5	20,8
σ <sup>2</sup> = 1.6										
$T_{n_1 n_2}^{(1)}$	8,2	15,4	11,0	20,3	12,3	21,0	13,8	22,5	11,3	20,5
$T_{n_1 n_2}^{(2)}$	13,3	20,8	13,6	23,0	12,5	21,1	13,2	21,6	11,5	20,8
σ <sup>2</sup> = 1.8										
$T_{n_1 n_2}^{(1)}$	13,9	24,3	19,1	29,8	24,0	35,3	23,8	36,2	23,7	35,1
$T_{n_1 n_2}^{(2)}$	22,9	33,3	24,2	35,5	24,9	36,2	23,4	34,9	23,4	34,3
σ <sup>2</sup> = 2.0										
$T_{n_1 n_2}^{(1)}$	19,2	32,1	26,4	38,4	30,9	43,8	34,5	47,3	31,6	44,4
$T_{n_1 n_2}^{(2)}$	31,6	45,1	32,5	44,8	31,8	45,2	33,4	47,2	31,1	43,6

Tabulka 4.5: Pravděpodobnosti zamítnutí (v %) pro C(0,1) při změně parametru  $\sigma^2$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

DEx(0,1)	a=0.25		a=0.5		a=1		a=2		a=3	
DEx(0, $\sigma^2$ )	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\sigma^2 = 1.0$										
$T_{n_1 n_2}^{(1)}$	4,3	9,6	4,7	10,3	5,6	10,3	5,0	10,1	4,2	9,8
$T_{n_1 n_2}^{(2)}$	5,4	10,5	5,1	10,3	5,8	10,0	4,7	10,2	4,3	9,9
$\sigma^2 = 1.1$										
$T_{n_1 n_2}^{(1)}$	4,9	9,7	5,7	10,7	5,6	11,4	5,7	10,7	6,1	12,1
$T_{n_1 n_2}^{(2)}$	5,1	11,1	5,6	11,2	5,5	10,4	5,4	11,2	5,6	12,4
$\sigma^2 = 1.2$										
$T_{n_1 n_2}^{(1)}$	5,3	10,4	6,7	13,2	7,1	13,1	6,4	13,6	7,0	13,6
$T_{n_1 n_2}^{(2)}$	7,1	13,2	6,5	13,6	6,8	13,5	6,1	13,0	6,8	13,5
$\sigma^2 = 1.3$										
$T_{n_1 n_2}^{(1)}$	7,1	13,9	8,3	14,7	9,7	18,1	8,7	16,5	8,3	16,1
$T_{n_1 n_2}^{(2)}$	10,2	17,5	9,4	17,1	10,4	19,2	8,6	17,4	7,7	15,4
$\sigma^2 = 1.4$										
$T_{n_1 n_2}^{(1)}$	8,9	16,1	12,0	21,0	13,6	23,0	13,7	23,6	11,9	22,0
$T_{n_1 n_2}^{(2)}$	12,8	20,7	14,1	24,7	14,5	24,6	13,3	23,5	10,8	21,9
$\sigma^2 = 1.5$										
$T_{n_1 n_2}^{(1)}$	10,9	19,5	13,9	24,8	17,6	28,4	17,0	28,1	15,5	26,9
$T_{n_1 n_2}^{(2)}$	18,3	27,9	18,4	29,0	18,3	29,1	16,0	28,4	14,5	25,9
$\sigma^2 = 1.6$										
$T_{n_1 n_2}^{(1)}$	13,0	22,0	18,5	30,2	22,6	33,6	22,2	35,3	22,1	34,9
$T_{n_1 n_2}^{(2)}$	17,9	29,4	24,4	35,3	24,7	37,2	22,4	36,6	21,3	34,4
$\sigma^2 = 1.8$										
$T_{n_1 n_2}^{(1)}$	18,9	31,1	27,3	38,3	32,4	46,0	36,0	51,4	33,5	49,2
$T_{n_1 n_2}^{(2)}$	31,0	42,4	34,3	46,2	35,8	49,8	35,6	52,1	33,7	49,3
$\sigma^2 = 2.0$										
$T_{n_1 n_2}^{(1)}$	25,6	37,4	34,1	47,1	43,2	56,5	47,7	61,3	46,6	62,0
$T_{n_1 n_2}^{(2)}$	40,2	54,9	44,5	57,6	47,4	62,8	48,4	62,6	45,5	61,3

Tabulka 4.6: Pravděpodobnosti zamítnutí (v %) pro DEx(0,1) při změně parametru  $\sigma^2$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

### 4.1.2 Aproximace kritických hodnot

Prostudujme nyní odhady kritických hodnot pro dané případy. Pro stejné kombinace parametrů, pro jaké jsme počítali pravděpodobnosti zamítnutí, opět generujeme 2000 dvojic náhodných výběrů daných rozdělení. Pro každou z nich simulujeme dalších  $B = 2000$  výběrů, na jejichž základě spočteme kritické hodnoty (opět pro hladiny  $\alpha_1 = 5\%$  a  $\alpha_2 = \%$ ) pro konkrétní dvojici výběrů. Z takto získaných 2000 kritických hodnot vezmeme  $(1-\alpha_1)$  a  $(1-\alpha_2)$  kvantil a dostáváme požadované odhady kritických hodnot pro  $\alpha_1$  a  $\alpha_2$ .

**Pozn.:** Aproximace kritických hodnot by samozřejmě šla udělat i jednodušeji - není nutně třeba simulovat 2000 dvojic, pro každou z nich spočítat aproximaci kritické hodnoty a z těchto hodnot ještě jednou odhadovat kritické hodnoty pro dané kvantily-teoreticky by stačilo vzít pouze jednu danou dvojici výběrů konkrétních rozdělení s danými parametry a aproximovat kritickou hodnotu jen na základě první aplikace bootstrapu (aplikací bootstrapu myslím odhadování kritické hodnoty na základě  $B$  bootstrapových výběrů). Protože však použitá metoda bootstrapu dobře aproximuje kritické hodnoty pouze pro konkrétní danou dvojici výběrů, je poněkud ošidné chtít tuto kritickou hodnotu zobecnit pro daný tvar rozdělení. Protože však smyslem tohoto odstavce je aproximované kritické hodnoty mezi sebou nějakým způsobem porovnávat (při platnosti  $H_0$  i při pevných alternativách), rozhodla jsem se pro výše uvedený způsob odhadu, na jehož základě už lze tyto hodnoty alespoň částečně zobecnit.

Aproximované kritické hodnoty pro daná rozdělení zachycují tabulky (4.7) - (4.12). Při jejich studiu zjišťujeme velice zajímavou věc - pro velikost aproximovaných kritických hodnot (v dalším textu je už budu pro jednoduchost nazývat pouze 'kritické hodnoty') není v případě malých hodnot parametru  $a$  pro testovou statistiku  $T_{n_1 n_2}^{(1)}$  příliš důležité, z jakého rozdělení a při jakých kombinacích parametrů byly výběry simulovány. Kritická hodnota této testové statistiky je pro všechna tři rozdělení a pro všechny testované kombinace parametrů zhruba stejná, mění se až na desetinném místě a to v rámci jednoho rozdělení (pro různé parametry  $\mu$  a  $\sigma^2$ ) ne víc než o dvě desetiny, v porovnání různých kombinací parametrů bez ohledu na rozdělení výběrů pak maximálně o šest desetin. To je vzhledem k tomu, že limitní rozdělení testové statistiky závisí na hypotetické distribuční funkci, celkem hezký výsledek. Znamená to totiž, že test můžeme bez větších obav použít i tam, kde si nejsme jisti, z jakého rozdělení výběry pocházejí - a přesto nám metoda bootstrap s vrácením dá poměrně přesnou aproximaci kritické hodnoty. Mluvím teď samozřejmě pouze o testovaných třech rozděleních, ostatní by bylo třeba

ještě prozkoumat.

Se vzrůstajícím  $a$  však tato vlastnost (nezávislost na rozdělení náhodných výběrů a velikosti změny parametru) mizí-je to dané tím, že pro větší  $a$  se rychleji blížíme limitnímu rozdělení dané statistiky a ta už se pro konkrétní případy přeci jen liší v závislosti na rozdělení, ze kterých výběry pocházejí. Přesto nám v rámci jednoho rozdělení pro změny obou parametrů opět vychází přibližně stejná hodnota, lišící se v případě změny parametru  $\mu$  pouze v řádu několika setin, v případě změny  $\sigma^2$  pak v řádu cca. dvou desetin. Při vzájemném porovnávání získaných kritických hodnot mezi kombinacemi všech testovaných rozdělení a parametrů vidíme, že změny ve velikosti obou testových statistik rovněž nejsou nijak výrazné, liší se maximálně o tři desetiny. Problém je však v tom, že už se zde při velikosti kritických hodnot objevuje jistá závislost na velikosti změny parametru-a to pro každé rozdělení různě velká, zatímco v případě malé hodnoty  $a$  pro  $T_{n_1 n_2}^{(1)}$  aproximované kritické hodnoty pouze oscilovaly kolem určité konstanty. Tato závislost je patrná hlavně u změny parametru  $\sigma^2$ , u parametru  $\mu$  se objevuje až pro  $\mu \geq 0.5$ .

Oba testy při větším parametru  $a$  tedy poměrně citlivě reagují na změnu obou parametrů, na změnu parametru  $\sigma^2$  rychleji. S významnějším rozdílem ve velikosti daného parametru se však tato citlivost stává závislou na rozdělení, ze kterého výběry pochází. Nevíme-li proto, z jakého rozdělení výběry pochází, je bootstrap s vracením užitečným nástrojem hlavně v případě, kdy se rozdíl ve změně parametru nepředpokládá příliš významný.



N(0,1) N( $\mu,1$ )	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\mu = 0.0$										
$T_{n_1 n_2}^{(1)}$	12.660	10.623	6.071	4.886	2.538	1.939	0.816	0.596	0.367	0.260
$T_{n_1 n_2}^{(2)}$	5.559	4.255	3.365	2.498	1.846	1.336	0.905	0.633	0.575	0.397
$\mu = 0.1$										
$T_{n_1 n_2}^{(1)}$	12.622	10.621	6.069	4.896	2.530	1.952	0.819	0.599	0.367	0.259
$T_{n_1 n_2}^{(2)}$	5.557	4.256	3.366	2.513	1.852	1.339	0.904	0.639	0.575	0.396
$\mu = 0.2$										
$T_{n_1 n_2}^{(1)}$	12.626	10.600	6.073	4.901	2.522	1.945	0.818	0.598	0.370	0.260
$T_{n_1 n_2}^{(2)}$	5.511	4.265	3.362	2.512	1.849	1.345	0.909	0.637	0.583	0.397
$\mu = 0.3$										
$T_{n_1 n_2}^{(1)}$	12.605	10.612	6.080	4.925	2.538	1.955	0.821	0.601	0.370	0.262
$T_{n_1 n_2}^{(2)}$	5.536	4.264	3.383	2.528	1.865	1.352	0.917	0.645	0.582	0.402
$\mu = 0.4$										
$T_{n_1 n_2}^{(1)}$	12.615	10.632	6.073	4.915	2.539	1.964	0.840	0.615	0.369	0.262
$T_{n_1 n_2}^{(2)}$	5.564	4.287	3.398	2.541	1.890	1.357	0.942	0.658	0.578	0.401
$\mu = 0.5$										
$T_{n_1 n_2}^{(1)}$	12.648	10.642	6.062	4.926	2.563	1.978	0.847	0.615	0.375	0.267
$T_{n_1 n_2}^{(2)}$	5.572	4.307	3.426	2.555	1.885	1.367	0.949	0.663	0.592	0.411
$\mu = 0.6$										
$T_{n_1 n_2}^{(1)}$	12.613	10.629	6.104	4.932	2.567	1.988	0.851	0.624	0.382	0.272
$T_{n_1 n_2}^{(2)}$	5.600	4.310	3.451	2.580	1.919	1.394	0.955	0.673	0.607	0.419
$\mu = 0.8$										
$T_{n_1 n_2}^{(1)}$	12.627	10.656	6.124	4.972	2.603	2.019	0.877	0.642	0.399	0.283
$T_{n_1 n_2}^{(2)}$	5.680	4.386	3.508	2.641	1.971	1.430	0.993	0.705	0.633	0.439
$\mu = 1.0$										
$T_{n_1 n_2}^{(1)}$	12.658	10.702	6.154	5.010	2.648	2.060	0.904	0.666	0.416	0.297
$T_{n_1 n_2}^{(2)}$	5.735	4.444	3.581	2.699	2.030	1.489	1.042	0.738	0.667	0.465

Tabulka 4.7: Aproximace kritických hodnot pro  $N(0,1)$  při změně parametru  $\mu$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

$C(0,1)$	a=0.25		a=0.5		a=1		a=2		a=3	
$C(\mu,1)$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\mu = 0.0$										
$T_{n_1 n_2}^{(1)}$	12.181	10.568	6.082	5.151	2.860	2.366	1.236	0.989	0.709	0.559
$T_{n_1 n_2}^{(2)}$	5.726	4.697	3.872	3.141	2.547	2.024	1.623	1.267	1.217	0.945
$\mu = 0.1$										
$T_{n_1 n_2}^{(1)}$	12.207	10.570	6.100	5.152	2.861	2.359	1.236	0.986	0.708	0.560
$T_{n_1 n_2}^{(2)}$	5.719	4.702	3.879	3.138	2.536	2.021	1.628	1.265	1.215	0.947
$\mu = 0.2$										
$T_{n_1 n_2}^{(1)}$	12.165	10.572	6.088	5.173	2.866	2.368	1.239	0.992	0.713	0.563
$T_{n_1 n_2}^{(2)}$	5.737	4.698	3.886	3.142	2.552	2.030	1.623	1.279	1.231	0.950
$\mu = 0.3$										
$T_{n_1 n_2}^{(1)}$	12.199	10.570	6.105	5.157	2.869	2.367	1.235	0.989	0.716	0.558
$T_{n_1 n_2}^{(2)}$	5.748	4.716	3.873	3.141	2.548	2.030	1.628	1.274	1.226	0.945
$\mu = 0.4$										
$T_{n_1 n_2}^{(1)}$	12.204	10.568	6.099	5.165	2.880	2.371	1.234	0.990	0.720	0.563
$T_{n_1 n_2}^{(2)}$	5.754	4.717	3.894	3.149	2.551	2.029	1.621	1.274	1.233	0.955
$\mu = 0.5$										
$T_{n_1 n_2}^{(1)}$	12.186	10.578	6.102	5.168	2.879	2.379	1.234	0.994	0.718	0.566
$T_{n_1 n_2}^{(2)}$	5.741	4.706	3.903	3.159	2.566	2.041	1.628	1.277	1.233	0.960
$\mu = 0.6$										
$T_{n_1 n_2}^{(1)}$	12.197	10.580	6.126	5.173	2.884	2.382	1.240	0.992	0.717	0.565
$T_{n_1 n_2}^{(2)}$	5.754	4.725	3.900	3.149	2.572	2.049	1.634	1.279	1.233	0.960
$\mu = 0.8$										
$T_{n_1 n_2}^{(1)}$	12.181	10.582	6.135	5.193	2.895	2.397	1.246	1.004	0.725	0.567
$T_{n_1 n_2}^{(2)}$	5.782	4.744	3.942	3.183	2.589	2.058	1.650	1.290	1.244	0.961
$\mu = 1.0$										
$T_{n_1 n_2}^{(1)}$	12.191	10.597	6.120	5.204	2.914	2.406	1.253	1.010	0.729	0.573
$T_{n_1 n_2}^{(2)}$	5.817	4.775	3.952	3.204	2.595	2.082	1.653	1.300	1.251	0.971

Tabulka 4.8: Aproximace kritických hodnot pro  $C(0,1)$  při změně parametru  $\mu$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

DEx(0,1)	a=0.25		a=0.5		a=1		a=2		a=3	
DEx( $\mu,1$ )	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\mu = 0.0$										
$T_{n_1 n_2}^{(1)}$	12.441	10.582	6.029	4.971	2.625	2.083	0.962	0.726	0.468	0.346
$T_{n_1 n_2}^{(2)}$	5.588	4.417	3.556	2.760	2.124	1.596	1.178	0.861	0.788	0.569
$\mu = 0.1$										
$T_{n_1 n_2}^{(1)}$	12.459	10.587	6.024	4.966	2.633	2.079	0.951	0.723	0.466	0.341
$T_{n_1 n_2}^{(2)}$	5.617	4.433	3.561	2.755	2.111	1.599	1.165	0.850	0.786	0.562
$\mu = 0.2$										
$T_{n_1 n_2}^{(1)}$	12.444	10.568	6.023	4.973	2.644	2.085	0.955	0.725	0.465	0.342
$T_{n_1 n_2}^{(2)}$	5.585	4.437	3.549	2.761	2.130	1.599	1.166	0.859	0.780	0.564
$\mu = 0.3$										
$T_{n_1 n_2}^{(1)}$	12.436	10.584	6.048	4.985	2.642	2.095	0.963	0.730	0.471	0.350
$T_{n_1 n_2}^{(2)}$	5.579	4.423	3.556	2.768	2.131	1.610	1.179	0.864	0.797	0.578
$\mu = 0.4$										
$T_{n_1 n_2}^{(1)}$	12.427	10.599	6.051	4.985	2.661	2.100	0.957	0.732	0.469	0.347
$T_{n_1 n_2}^{(2)}$	5.612	4.433	3.580	2.783	2.150	1.614	1.179	0.866	0.793	0.574
$\mu = 0.5$										
$T_{n_1 n_2}^{(1)}$	12.434	10.583	6.080	5.000	2.649	2.101	0.969	0.732	0.474	0.352
$T_{n_1 n_2}^{(2)}$	5.620	4.453	3.609	2.797	2.157	1.626	1.184	0.871	0.799	0.581
$\mu = 0.6$										
$T_{n_1 n_2}^{(1)}$	12.486	10.617	6.074	5.017	2.663	2.118	0.969	0.738	0.482	0.355
$T_{n_1 n_2}^{(2)}$	5.660	4.482	3.611	2.810	2.166	1.640	1.197	0.881	0.813	0.588
$\mu = 0.8$										
$T_{n_1 n_2}^{(1)}$	12.452	10.638	6.123	5.039	2.698	2.134	1.000	0.756	0.492	0.362
$T_{n_1 n_2}^{(2)}$	5.696	4.541	3.675	2.835	2.201	1.659	1.230	0.903	0.828	0.598
$\mu = 1.0$										
$T_{n_1 n_2}^{(1)}$	12.466	10.664	6.159	5.079	2.731	2.172	1.016	0.767	0.499	0.371
$T_{n_1 n_2}^{(2)}$	5.766	4.573	3.733	2.897	2.250	1.705	1.246	0.920	0.851	0.614

Tabulka 4.9: Aproximace kritických hodnot pro DEx(0,1) při změně parametru  $\mu$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

N(0,1) N(0, $\sigma^2$ )	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\sigma^2 = 1.0$										
$T_{n_1 n_2}^{(1)}$	12.660	10.623	6.071	4.886	2.538	1.939	0.816	0.596	0.367	0.260
$T_{n_1 n_2}^{(2)}$	5.559	4.255	3.365	2.498	1.846	1.336	0.905	0.633	0.575	0.397
$\sigma^2 = 1.1$										
$T_{n_1 n_2}^{(1)}$	12.633	10.644	6.102	4.947	2.583	1.999	0.851	0.624	0.390	0.276
$T_{n_1 n_2}^{(2)}$	5.608	4.330	3.456	2.580	1.937	1.404	0.959	0.675	0.616	0.425
$\sigma^2 = 1.2$										
$T_{n_1 n_2}^{(1)}$	12.594	10.643	6.133	4.980	2.615	2.028	0.884	0.650	0.404	0.288
$T_{n_1 n_2}^{(2)}$	5.659	4.406	3.494	2.648	1.991	1.458	1.012	0.713	0.648	0.448
$\sigma^2 = 1.3$										
$T_{n_1 n_2}^{(1)}$	12.602	10.669	6.157	5.016	2.639	2.060	0.919	0.681	0.425	0.305
$T_{n_1 n_2}^{(2)}$	5.715	4.463	3.606	2.723	2.048	1.509	1.067	0.759	0.687	0.483
$\sigma^2 = 1.4$										
$T_{n_1 n_2}^{(1)}$	12.609	10.684	6.161	5.045	2.681	2.098	0.936	0.696	0.444	0.319
$T_{n_1 n_2}^{(2)}$	5.772	4.500	3.650	2.782	2.119	1.567	1.101	0.792	0.726	0.508
$\sigma^2 = 1.5$										
$T_{n_1 n_2}^{(1)}$	12.567	10.684	6.176	5.065	2.712	2.123	0.968	0.719	0.465	0.336
$T_{n_1 n_2}^{(2)}$	5.766	4.532	3.688	2.825	2.171	1.608	1.157	0.829	0.765	0.541
$\sigma^2 = 1.6$										
$T_{n_1 n_2}^{(1)}$	12.533	10.681	6.197	5.083	2.738	2.161	0.986	0.740	0.474	0.344
$T_{n_1 n_2}^{(2)}$	5.812	4.575	3.748	2.876	2.211	1.658	1.195	0.864	0.787	0.556
$\sigma^2 = 1.8$										
$T_{n_1 n_2}^{(1)}$	12.497	10.679	6.198	5.117	2.764	2.200	1.025	0.771	0.507	0.373
$T_{n_1 n_2}^{(2)}$	5.844	4.633	3.794	2.943	2.294	1.734	1.262	0.921	0.859	0.614
$\sigma^2 = 2.0$										
$T_{n_1 n_2}^{(1)}$	12.466	10.684	6.202	5.144	2.811	2.238	1.061	0.810	0.535	0.397
$T_{n_1 n_2}^{(2)}$	5.873	4.668	3.869	3.001	2.368	1.802	1.340	0.988	0.912	0.660

Tabulka 4.10: Aproximace kritických hodnot pro N(0,1) při změně parametru  $\sigma^2$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

$C(0,1)$	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\sigma^2 = 1.0$										
$T_{n_1 n_2}^{(1)}$	12.181	10.568	6.082	5.151	2.860	2.366	1.236	0.989	0.709	0.559
$T_{n_1 n_2}^{(2)}$	5.726	4.697	3.872	3.141	2.547	2.024	1.623	1.267	1.217	0.945
$\sigma^2 = 1.1$										
$T_{n_1 n_2}^{(1)}$	12.187	10.570	6.102	5.178	2.890	2.386	1.248	1.002	0.724	0.571
$T_{n_1 n_2}^{(2)}$	5.744	4.724	3.906	3.174	2.570	2.059	1.642	1.291	1.242	0.966
$\sigma^2 = 1.2$										
$T_{n_1 n_2}^{(1)}$	12.150	10.576	6.112	5.191	2.906	2.408	1.268	1.022	0.743	0.587
$T_{n_1 n_2}^{(2)}$	5.757	4.749	3.943	3.198	2.621	2.093	1.680	1.321	1.272	0.992
$\sigma^2 = 1.3$										
$T_{n_1 n_2}^{(1)}$	12.142	10.574	6.107	5.201	2.927	2.423	1.274	1.029	0.750	0.596
$T_{n_1 n_2}^{(2)}$	5.773	4.759	3.945	3.214	2.637	2.107	1.687	1.335	1.288	1.010
$\sigma^2 = 1.4$										
$T_{n_1 n_2}^{(1)}$	12.177	10.586	6.128	5.210	2.931	2.435	1.294	1.043	0.761	0.603
$T_{n_1 n_2}^{(2)}$	5.791	4.789	3.973	3.231	2.648	2.127	1.723	1.349	1.309	1.026
$\sigma^2 = 1.5$										
$T_{n_1 n_2}^{(1)}$	12.078	10.575	6.122	5.219	2.943	2.452	1.301	1.055	0.772	0.614
$T_{n_1 n_2}^{(2)}$	5.795	4.796	3.978	3.259	2.672	2.151	1.730	1.375	1.331	1.045
$\sigma^2 = 1.6$										
$T_{n_1 n_2}^{(1)}$	12.091	10.566	6.120	5.222	2.950	2.459	1.312	1.063	0.786	0.624
$T_{n_1 n_2}^{(2)}$	5.814	4.820	3.985	3.262	2.684	2.166	1.746	1.384	1.349	1.061
$\sigma^2 = 1.8$										
$T_{n_1 n_2}^{(1)}$	12.045	10.552	6.119	5.229	2.973	2.479	1.335	1.088	0.796	0.638
$T_{n_1 n_2}^{(2)}$	5.814	4.832	4.007	3.289	2.728	2.199	1.777	1.422	1.375	1.082
$\sigma^2 = 2.0$										
$T_{n_1 n_2}^{(1)}$	12.023	10.544	6.120	5.244	2.983	2.501	1.355	1.097	0.809	0.648
$T_{n_1 n_2}^{(2)}$	5.816	4.836	4.041	3.318	2.746	2.228	1.810	1.442	1.389	1.098

Tabulka 4.11: Aproximace kritických hodnot pro  $C(0,1)$  při změně parametru  $\sigma^2$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

DEx(0,1)	a=0.25		a=0.5		a=1		a=2		a=3	
DEx(0, $\sigma^2$ )	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
$\sigma^2 = 1.0$										
$T_{n_1 n_2}^{(1)}$	12.441	10.582	6.029	4.971	2.625	2.083	0.962	0.726	0.468	0.346
$T_{n_1 n_2}^{(2)}$	5.588	4.417	3.556	2.760	2.124	1.596	1.178	0.861	0.788	0.569
$\sigma^2 = 1.1$										
$T_{n_1 n_2}^{(1)}$	12.427	10.578	6.048	5.005	2.665	2.121	0.975	0.746	0.488	0.360
$T_{n_1 n_2}^{(2)}$	5.656	4.468	3.630	2.810	2.181	1.652	1.212	0.895	0.829	0.599
$\sigma^2 = 1.2$										
$T_{n_1 n_2}^{(1)}$	12.394	10.599	6.085	5.028	2.693	2.151	1.000	0.768	0.504	0.376
$T_{n_1 n_2}^{(2)}$	5.667	4.507	3.671	2.855	2.232	1.699	1.243	0.927	0.859	0.628
$\sigma^2 = 1.3$										
$T_{n_1 n_2}^{(1)}$	12.412	10.617	6.101	5.055	2.733	2.183	1.033	0.791	0.520	0.391
$T_{n_1 n_2}^{(2)}$	5.721	4.571	3.710	2.906	2.283	1.742	1.294	0.961	0.893	0.657
$\sigma^2 = 1.4$										
$T_{n_1 n_2}^{(1)}$	12.403	10.613	6.102	5.071	2.754	2.214	1.043	0.803	0.540	0.403
$T_{n_1 n_2}^{(2)}$	5.717	4.588	3.747	2.945	2.330	1.782	1.321	0.987	0.926	0.683
$\sigma^2 = 1.5$										
$T_{n_1 n_2}^{(1)}$	12.381	10.609	6.127	5.109	2.795	2.236	1.081	0.829	0.557	0.421
$T_{n_1 n_2}^{(2)}$	5.746	4.619	3.825	2.986	2.382	1.819	1.380	1.034	0.965	0.717
$\sigma^2 = 1.6$										
$T_{n_1 n_2}^{(1)}$	12.388	10.615	6.143	5.114	2.794	2.252	1.085	0.840	0.571	0.430
$T_{n_1 n_2}^{(2)}$	5.780	4.643	3.833	3.022	2.402	1.848	1.388	1.048	0.990	0.732
$\sigma^2 = 1.8$										
$T_{n_1 n_2}^{(1)}$	12.343	10.633	6.130	5.139	2.835	2.293	1.128	0.879	0.601	0.452
$T_{n_1 n_2}^{(2)}$	5.807	4.684	3.878	3.066	2.460	1.914	1.465	1.110	1.049	0.778
$\sigma^2 = 2.0$										
$T_{n_1 n_2}^{(1)}$	12.285	10.619	6.146	5.165	2.853	2.311	1.156	0.903	0.618	0.471
$T_{n_1 n_2}^{(2)}$	5.820	4.720	3.920	3.111	2.516	1.950	1.515	1.153	1.084	0.810

Tabulka 4.12: Aproximace kritických hodnot pro DEx(0,1) při změně parametru  $\sigma^2$ . Testové statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_2 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

## 4.2 Aplikace testu na diskrétní rozdělení

Přejdeme však konečně k aplikaci testu na situace, pro které byl tento test primárně konstruován-pro testování hypotézy, zda dva náhodné výběry pocházejí z rozdělení se stejnou distribuční funkcí  $F_0$ , není-li však tato spojitá. Postupně zde testuji různé kombinace dvojic pro následující typy a tvary rozdělení: '**Bi-**' značí binomické rozdělení  $\text{Bi}(20, 0.5)$ , '**Ge-**' pak geometrické  $\text{Ge}(0.1)$  a '**Po-**' Poissonovo rozdělení  $\text{Po}(10)$ . Parametry těchto tří rozdělení jsou voleny tak, aby měly společnou střední hodnotu(=10), mají však různé rozptyly. Ke skupině těchto tří přidáme ta samá rozdělení, změníme však parametry tak, aby výběry měly pro změnu stejný rozptyl(=8.3), ale různé střední hodnoty. Jsou to '**Bi+**' pro  $\text{Bi}(33,0.5)$ , '**Ge+**' pro  $\text{Ge}((\sqrt{34}-1)/16.5)$  a '**Po+**' pro  $\text{Po}(8.25)$ .

Těchto šest vzorků mezi sebou vzájemně porovnáme, daný test tedy aplikujeme pro celkem 21 různých dvojic výběrů. Pro každou dvojici pak spočítáme pravděpodobnosti zamítnutí hypotézy a opět se pokusíme aproximovat dané kritické hodnoty. Oboje počítáme stejným způsobem jako v předcházejícím odstavci.

Pravděpodobnosti zamítnutí v tabulkách (4.13)-(4.14) ukazují citlivost testu pro dané dvojice rozdělení. Zajímavější výsledky nám však ukazují odhadnuté kritické hodnoty v (4.15)-(4.16)-zdá se, že aproximace kritických hodnot pro konkrétní dvojici výběrů si jsou opět velice podobné. Podívejme se nejprve na statistiku  $T_{n_1 n_2}^{(1)}$ . Pro malé hodnoty  $a$ , konkrétně pro  $a = 0.25$  se aproximované hodnoty liší nejvíce-pohybují se zde v intervalu (12.8,14.4). Pro  $a = 1$  a  $a = 3$  se je však rozdíl v kritických hodnotách maximálně dvě desetiny pro všechny možné dvojice námi generovaných výběrů-znamená to tedy, že v tomto případě dokážeme aproximovat kritické hodnoty na velice dobré úrovni i bez znalosti hypotetického rozdělení. Kritická hodnota spočítaná pro statistiku  $T_{n_1 n_2}^{(2)}$  se liší víceméně pro všechny parametry  $a$  stejně a to o zhruba pět desetin pro všechny uvažované dvojice.

	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
Bi- x Bi-										
$T_{n_1 n_2}^{(1)}$	4,8	10,3	4,1	9,7	3,6	8,0	5,6	10,9	5,0	10,6
$T_{n_1 n_2}^{(2)}$	4,8	10,4	5,0	9,5	3,9	8,3	5,6	10,9	5,1	10,6
Bi- x Ge-										
$T_{n_1 n_2}^{(1)}$	96,8	99,0	99,6	99,8	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	99,6	99,8	99,9	*	*	*	*	*	*	*
Bi- x Po-										
$T_{n_1 n_2}^{(1)}$	8,4	16,6	11,7	20,6	16,2	26,8	16,3	26,2	19,6	31,8
$T_{n_1 n_2}^{(2)}$	11,7	20,3	16,1	25,8	19,5	29,5	17,8	28,5	20,7	32,8
Bi- x Bi+										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
Bi- x Ge+										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
Bi- x Po+										
$T_{n_1 n_2}^{(1)}$	34,1	49,1	42,7	57,1	59,1	71,2	71,7	80,3	76,1	84,4
$T_{n_1 n_2}^{(2)}$	48,7	63,7	57,2	70,0	65,3	76,5	72,3	82,1	76,2	84,8
Ge- x Ge-										
$T_{n_1 n_2}^{(1)}$	4,0	9,2	4,3	8,5	4,6	9,2	4,9	10,0	5,3	10,4
$T_{n_1 n_2}^{(2)}$	4,6	9,3	4,9	10,0	4,7	9,9	4,9	10,0	5,4	10,3
Ge- x Po-										
$T_{n_1 n_2}^{(1)}$	83,6	91,4	92,3	97,2	98,0	99,2	99,3	99,8	99,6	99,9
$T_{n_1 n_2}^{(2)}$	94,0	96,9	97,1	98,9	98,8	99,4	99,4	99,8	99,6	99,8
Ge- x Bi+										
$T_{n_1 n_2}^{(1)}$	99,9	99,9	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
Ge- x Ge+										
$T_{n_1 n_2}^{(1)}$	59,6	74,2	71,5	83,4	84,9	91,3	92,7	96,7	93,3	97,2
$T_{n_1 n_2}^{(2)}$	74,7	84,1	82,5	89,7	87,6	92,8	92,3	96,1	91,8	95,8
Ge- x Po+										
$T_{n_1 n_2}^{(1)}$	77,5	87,5	85,9	92,7	96,0	98,1	97,2	99,0	99,0	99,7
$T_{n_1 n_2}^{(2)}$	90,8	95,5	93,3	96,5	97,0	98,8	97,3	99,0	98,7	99,6

Tabulka 4.13: Pravděpodobnosti zamítnutí(v %) pro diskrétní rozdělení. Statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_1 = 10\%$ , při  $n_1 = n_2 = 30$ . Symbol '\*' znamená, že hypotéza byla zamítnuta na hladině 100%.



	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
Po- x Po-										
$T_{n_1 n_2}^{(1)}$	4,6	9,9	4,2	8,5	5,2	9,8	5,2	10,4	4,8	9,7
$T_{n_1 n_2}^{(2)}$	4,4	9,4	4,3	9,1	5,2	9,4	5,1	10,2	4,9	9,6
Po- x Bi+										
$T_{n_1 n_2}^{(1)}$	99,8	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
Po- x Ge+										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
Po- x Po+										
$T_{n_1 n_2}^{(1)}$	14,2	23,6	20,2	31,4	32,5	45,7	42,2	55,6	47,0	61,0
$T_{n_1 n_2}^{(2)}$	20,2	32,4	27,1	40,5	36,4	48,7	42,5	55,4	45,1	59,5
Bi+ x Bi+										
$T_{n_1 n_2}^{(1)}$	4,6	9,9	4,2	8,5	5,2	9,8	5,2	10,4	4,8	9,7
$T_{n_1 n_2}^{(2)}$	4,4	9,4	4,3	9,1	5,2	9,4	5,1	10,2	4,9	9,6
Bi+ x Ge+										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
Bi+ x Po+										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
Ge+ x Ge+										
$T_{n_1 n_2}^{(1)}$	5,7	11,2	5,2	9,8	4,7	9,7	3,9	9,3	5,0	10,1
$T_{n_1 n_2}^{(2)}$	5,3	11,0	4,6	9,6	5,1	9,4	3,6	9,4	5,0	9,9
Ge+ x Po+										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
Po+ x Po+										
$T_{n_1 n_2}^{(1)}$	4,6	9,4	5,6	10,5	4,7	9,2	4,8	10,1	5,0	10,1
$T_{n_1 n_2}^{(2)}$	4,6	10,4	5,2	10,7	5,1	10,2	5,0	10,1	5,2	10,1

Tabulka 4.14: Pravděpodobnosti zamítnutí(v %) pro diskrétní rozdělení. Statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_1 = 10\%$ , při  $n_1 = n_2 = 30$ . Symbol '\*' znamená, že hypotéza byla zamítnuta na hladině 100%.

	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
Bi- x Bi-										
$T_{n_1 n_2}^{(1)}$	14.442	11.971	6.842	5.673	3.110	2.530	1.319	1.024	0.729	0.547
$T_{n_1 n_2}^{(2)}$	6.143	5.021	4.265	3.399	2.862	2.226	1.788	1.343	1.301	0.960
Ge										
$T_{n_1 n_2}^{(1)}$	13.219	11.400	6.451	5.567	3.175	2.685	1.543	1.259	0.971	0.776
$T_{n_1 n_2}^{(2)}$	5.843	4.995	4.207	3.541	3.025	2.482	2.122	1.696	1.698	1.342
Po-										
$T_{n_1 n_2}^{(1)}$	14.113	11.839	6.738	5.657	3.145	2.588	1.399	1.101	0.799	0.613
$T_{n_1 n_2}^{(2)}$	6.089	5.038	4.293	3.476	2.948	2.329	1.926	1.470	1.434	1.082
Bi+										
$T_{n_1 n_2}^{(1)}$	13.359	11.508	6.533	5.611	3.212	2.703	1.570	1.258	0.981	0.763
$T_{n_1 n_2}^{(2)}$	5.945	5.054	4.314	3.593	3.099	2.512	2.175	1.712	1.742	1.346
Ge+										
$T_{n_1 n_2}^{(1)}$	13.714	11.675	6.716	5.705	3.329	2.758	1.669	1.308	1.071	0.809
$T_{n_1 n_2}^{(2)}$	6.074	5.109	4.403	3.629	3.172	2.555	2.289	1.767	1.860	1.411
Po+										
$T_{n_1 n_2}^{(1)}$	14.091	11.846	6.745	5.671	3.151	2.595	1.409	1.107	0.810	0.619
$T_{n_1 n_2}^{(2)}$	6.097	5.053	4.315	3.496	2.966	2.341	1.934	1.482	1.449	1.096
Ge x Ge										
$T_{n_1 n_2}^{(1)}$	13.026	11.295	6.387	5.534	3.161	2.702	1.582	1.308	1.030	0.833
$T_{n_1 n_2}^{(2)}$	5.780	4.963	4.176	3.532	3.007	2.500	2.139	1.747	1.740	1.403
Po-										
$T_{n_1 n_2}^{(1)}$	12.973	11.309	6.396	5.537	3.171	2.694	1.566	1.287	0.990	0.797
$T_{n_1 n_2}^{(2)}$	5.794	4.981	4.208	3.558	3.039	2.510	2.157	1.743	1.729	1.379
Bi+										
$T_{n_1 n_2}^{(1)}$	12.831	11.206	6.315	5.516	3.164	2.711	1.618	1.332	1.059	0.849
$T_{n_1 n_2}^{(2)}$	5.714	4.955	4.160	3.544	3.012	2.525	2.180	1.780	1.777	1.432
Ge+										
$T_{n_1 n_2}^{(1)}$	13.884	11.645	6.687	5.623	3.182	2.654	1.511	1.220	0.948	0.748
$T_{n_1 n_2}^{(2)}$	5.991	4.999	4.230	3.480	2.980	2.403	2.041	1.620	1.627	1.274
Po+										
$T_{n_1 n_2}^{(1)}$	13.172	11.412	6.439	5.558	3.168	2.685	1.535	1.258	0.964	0.774
$T_{n_1 n_2}^{(2)}$	5.855	5.008	4.230	3.548	3.030	2.492	2.112	1.699	1.694	1.338

Tabulka 4.15: Aproximace kritických hodnot pro diskrétní rozdělení. Statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_1 = 10\%$ , při  $n_1 = n_2 = 30$ .

	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
Po- x Po-										
$T_{n_1 n_2}^{(1)}$	13.837	11.743	6.677	5.645	3.170	2.630	1.448	1.153	0.861	0.664
$T_{n_1 n_2}^{(2)}$	6.039	5.061	4.305	3.527	3.005	2.400	2.009	1.559	1.550	1.180
Bi-+										
$T_{n_1 n_2}^{(1)}$	13.200	11.441	6.479	5.590	3.203	2.706	1.579	1.274	0.989	0.778
$T_{n_1 n_2}^{(2)}$	5.890	5.038	4.298	3.593	3.093	2.526	2.187	1.736	1.748	1.362
Ge+										
$T_{n_1 n_2}^{(1)}$	13.441	11.552	6.628	5.661	3.284	2.746	1.644	1.310	1.048	0.809
$T_{n_1 n_2}^{(2)}$	5.985	5.065	4.352	3.610	3.158	2.553	2.264	1.775	1.829	1.409
Po-+										
$T_{n_1 n_2}^{(1)}$	13.881	11.745	6.679	5.651	3.174	2.631	1.458	1.154	0.856	0.662
$T_{n_1 n_2}^{(2)}$	6.065	5.063	4.310	3.526	3.014	2.399	2.013	1.561	1.539	1.175
Bi+ x Bi+										
$T_{n_1 n_2}^{(1)}$	13.996	11.790	6.720	5.669	3.174	2.613	1.433	1.128	0.825	0.635
$T_{n_1 n_2}^{(2)}$	6.080	5.061	4.323	3.513	2.993	2.370	1.965	1.516	1.485	1.127
Ge+										
$T_{n_1 n_2}^{(1)}$	13.424	11.478	6.588	5.640	3.300	2.765	1.708	1.361	1.148	0.886
$T_{n_1 n_2}^{(2)}$	5.926	5.032	4.302	3.584	3.124	2.547	2.265	1.794	1.876	1.458
Po-+										
$T_{n_1 n_2}^{(1)}$	13.053	11.354	6.417	5.573	3.217	2.722	1.612	1.304	1.036	0.812
$T_{n_1 n_2}^{(2)}$	5.836	5.014	4.254	3.585	3.105	2.550	2.221	1.772	1.797	1.411
Ge+ x Ge+										
$T_{n_1 n_2}^{(1)}$	14.774	11.928	6.931	5.639	3.119	2.493	1.330	1.024	0.752	0.563
$T_{n_1 n_2}^{(2)}$	6.119	4.919	4.196	3.304	2.792	2.144	1.786	1.339	1.335	0.980
Po-+										
$T_{n_1 n_2}^{(1)}$	13.680	11.672	6.701	5.691	3.301	2.738	1.604	1.269	0.996	0.767
$T_{n_1 n_2}^{(2)}$	6.071	5.111	4.400	3.614	3.167	2.538	2.221	1.724	1.772	1.351
Po-+ x Po-+										
$T_{n_1 n_2}^{(1)}$	14.019	11.808	6.716	5.660	3.158	2.601	1.419	1.123	0.827	0.633
$T_{n_1 n_2}^{(2)}$	6.084	5.056	4.323	3.507	2.979	2.358	1.956	1.509	1.491	1.121

Tabulka 4.16: Aproximace kritických hodnot pro diskrétní rozdělení. Statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_1 = 10\%$ , při  $n_1 = n_2 = 30$ .

### 4.3 Aplikace testu na spojitá rozdělení

Dále jsme zkoumali, jak se naše testová statistika chová pro případ, že oba výběry pocházejí z rozdělení se spojitou distribuční funkcí. Výběry bereme ze šesti rozdělení-z již zkoumaných normálního(N01), Cauchyho(C01) a Laplaceovo(L01), a dále přidáme logaritmické(Log), logaritmicko normální(LN) a exponenciální(Exp)-všechny ve standartizované formě, tj. s parametrem  $\mu=0$  pro střední hodnotu a  $\sigma^2 = 1$  pro rozptyl, v závorkách uvádíme zde použité značení. Opět porovnáváme každé ze jmenovaných rozdělení s každým-máme tedy 21 dvojic, pro které spočítáme pravděpodobnosti zamítnutí  $H_0$  a odhadneme kritické hodnoty. Způsob získání těchto výsledků zůstává stejný jako v předchozích simulacích.

V tabulkách pro pravděpodobnosti zamítnutí(4.17)-(4.18) si můžeme zkontrolovat, že obě testové statistiky přibližně drží hladinu  $\alpha_1$  a  $\alpha_2$ , na které byly testy prováděny. Také zde vidíme, jak je test různě citlivý pro různé dvojice výběrů-zatímco například dvojici (Exp x LN) zamítáme s pravděpodobností v rozsahu přibližně od osmi procent do třiceti tří(pro různé hladiny  $\alpha$  a různá  $a$ ), většinu jiných dvojic zamítáme minimálně v osmdesáti procentech případů.

Co se týče aproximace kritických hodnot uvedených v (4.19)-(4.20), dostáváme zde zatím nejhorší výsledek-resp. tyto aproximace se od sebe pro malá  $a$  liší poměrně dost-pro  $T_{n_1 n_2}^{(1)}$  o maximálně dvanáct desetin, pro  $T_{n_1 n_2}^{(2)}$  už pak až o dvacet pět desetin. S rostoucí  $a$  tyto rozdíly mírně klesají-a pro  $a=3$  dosahují pro  $T_{n_1 n_2}^{(1)}$  pět desetin, pro  $T_{n_1 n_2}^{(2)}$  osm desetin. Pokud tedy neznáme hypotetické rozdělení daného výběru, doporučovala bych z daných možností určitě použít raději  $T_{n_1 n_2}^{(1)}$  pro velké hodnoty  $a$ .

	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
Exp x Exp										
$T_{n_1 n_2}^{(1)}$	4,8	10,3	4,1	9,7	3,6	8,0	5,6	10,9	5,0	10,6
$T_{n_1 n_2}^{(2)}$	4,8	10,4	5,0	9,5	3,9	8,3	5,6	10,9	5,1	10,6
Log										
$T_{n_1 n_2}^{(1)}$	96,8	99,0	99,6	99,8	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	99,6	99,8	99,9	*	*	*	*	*	*	*
LN										
$T_{n_1 n_2}^{(1)}$	8,4	16,6	11,7	20,6	16,2	26,8	16,3	26,2	19,6	31,8
$T_{n_1 n_2}^{(2)}$	11,7	20,3	16,1	25,8	19,5	29,5	17,8	28,5	20,7	32,8
C01										
$T_{n_1 n_2}^{(1)}$	99,6	99,9	*	*	*	*	99,9	*	99,9	*
$T_{n_1 n_2}^{(2)}$	99,9	*	*	*	99,8	*	99,8	99,9	99,6	99,9
N01										
$T_{n_1 n_2}^{(1)}$	98,4	99,3	98,9	99,8	98,8	99,9	96,1	98,5	92,3	97,5
$T_{n_1 n_2}^{(2)}$	98,4	99,5	98,2	99,6	95,0	98,1	87,9	95,2	83,4	92,2
L01										
$T_{n_1 n_2}^{(1)}$	98,8	99,7	99,4	99,8	98,9	99,6	98,3	99,5	96,5	98,8
$T_{n_1 n_2}^{(2)}$	99,6	99,9	99,0	99,8	97,2	99,4	94,7	98,2	91,1	97,0
Log x Log										
$T_{n_1 n_2}^{(1)}$	4,0	9,2	4,3	8,5	4,6	9,2	4,9	10,0	5,3	10,4
$T_{n_1 n_2}^{(2)}$	4,6	9,3	4,9	10,0	4,7	9,9	4,9	10,0	5,4	10,3
LN										
$T_{n_1 n_2}^{(1)}$	83,6	91,4	92,3	97,2	98,0	99,2	99,3	99,8	99,6	99,9
$T_{n_1 n_2}^{(2)}$	94,0	96,9	97,1	98,9	98,8	99,4	99,4	99,8	99,6	99,8
C01										
$T_{n_1 n_2}^{(1)}$	6,5	13,9	9,1	17,4	8,4	17,0	8,8	17,1	9,0	19,0
$T_{n_1 n_2}^{(2)}$	9,4	15,8	9,3	18,0	7,7	14,5	7,9	15,6	9,5	19,2
N01										
$T_{n_1 n_2}^{(1)}$	18,0	29,6	27,1	41,1	35,8	49,7	36,7	53,8	33,1	52,2
$T_{n_1 n_2}^{(2)}$	31,5	44,5	37,6	51,3	41,3	56,3	33,8	52,7	29,9	49,0
L01										
$T_{n_1 n_2}^{(1)}$	16,2	26,0	21,4	33,4	21,2	31,9	19,3	29,6	14,5	26,1
$T_{n_1 n_2}^{(2)}$	20,9	31,7	22,8	34,6	19,0	30,1	15,3	26,1	12,6	22,8

Tabulka 4.17: Pravděpodobnosti zamítnutí(v %) pro spojitá rozdělení. Statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_1 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ . Symbol '\*' znamená, že hypotéza byla zamítnuta na hladině 100%.

	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
LN x LN										
$T_{n_1 n_2}^{(1)}$	4,6	9,9	4,2	8,5	5,2	9,8	5,2	10,4	4,8	9,7
$T_{n_1 n_2}^{(2)}$	4,4	9,4	4,3	9,1	5,2	9,4	5,1	10,2	4,9	9,6
C01										
$T_{n_1 n_2}^{(1)}$	91,8	96,9	97,0	99,0	98,9	99,6	98,7	99,6	98,3	99,6
$T_{n_1 n_2}^{(2)}$	97,1	98,9	98,1	99,5	98,4	99,6	98,1	99,5	97,7	99,3
N01										
$T_{n_1 n_2}^{(1)}$	93,5	97,4	98,2	99,3	99,2	99,9	99,8	99,9	*	*
$T_{n_1 n_2}^{(2)}$	98,9	99,8	99,3	99,9	99,6	99,9	99,9	99,9	99,9	*
L01										
$T_{n_1 n_2}^{(1)}$	92,8	97,8	97,4	99,2	98,8	99,7	99,1	99,7	99,3	99,8
$T_{n_1 n_2}^{(2)}$	98,4	99,1	98,4	99,5	99,1	99,9	98,7	99,7	99,2	99,8
C01 x C01										
$T_{n_1 n_2}^{(1)}$	4,6	9,9	4,2	8,5	5,2	9,8	5,2	10,4	4,8	9,7
$T_{n_1 n_2}^{(2)}$	4,4	9,4	4,3	9,1	5,2	9,4	5,1	10,2	4,9	9,6
N01										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
L01										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
N01 x N01										
$T_{n_1 n_2}^{(1)}$	5,7	11,2	5,2	9,8	4,7	9,7	3,9	9,3	5,0	10,1
$T_{n_1 n_2}^{(2)}$	5,3	11,0	4,6	9,6	5,1	9,4	3,6	9,4	5,0	9,9
L01										
$T_{n_1 n_2}^{(1)}$	*	*	*	*	*	*	*	*	*	*
$T_{n_1 n_2}^{(2)}$	*	*	*	*	*	*	*	*	*	*
L01 x L01										
$T_{n_1 n_2}^{(1)}$	4,6	9,4	5,6	10,5	4,7	9,2	4,8	10,1	5,0	10,1
$T_{n_1 n_2}^{(2)}$	4,6	10,4	5,2	10,7	5,1	10,2	5,0	10,1	5,2	10,1

Tabulka 4.18: Pravděpodobnosti zamítnutí(v %) pro spojitá rozdělení. Statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_1 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ . Symbol '\*' znamená, že hypotéza byla zamítnuta na hladině 100%.

	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
Exp x Exp										
$T_{n_1 n_2}^{(1)}$	11.549	9.1174	4.611	3.476	1.514	1.080	0.375	0.257	0.145	0.097
$T_{n_1 n_2}^{(2)}$	3.321	2.364	1.694	1.166	0.818	0.545	0.352	0.232	0.205	0.134
Log										
$T_{n_1 n_2}^{(1)}$	12.303	10.324	5.780	4.708	2.450	1.919	0.865	0.655	0.422	0.312
$T_{n_1 n_2}^{(2)}$	5.174	4.062	3.261	2.512	1.927	1.454	1.053	0.774	0.713	0.516
LN										
$T_{n_1 n_2}^{(1)}$	12.301	10.073	5.541	4.373	2.222	1.683	0.768	0.560	0.391	0.279
$T_{n_1 n_2}^{(2)}$	4.651	3.541	2.822	2.100	1.661	1.197	0.927	0.654	0.657	0.460
C01										
$T_{n_1 n_2}^{(1)}$	12.171	10.206	5.663	4.634	2.396	1.923	0.923	0.716	0.512	0.384
$T_{n_1 n_2}^{(2)}$	4.970	3.951	3.119	2.439	1.942	1.496	1.152	0.876	0.865	0.639
N01										
$T_{n_1 n_2}^{(1)}$	12.217	10.141	5.532	4.417	2.128	1.628	0.642	0.468	0.275	0.194
$T_{n_1 n_2}^{(2)}$	4.715	3.582	2.702	2.018	1.421	1.025	0.685	0.477	0.424	0.290
L01										
$T_{n_1 n_2}^{(1)}$	12.123	10.082	5.524	4.452	2.192	1.711	0.715	0.537	0.336	0.244
$T_{n_1 n_2}^{(2)}$	4.761	3.675	2.845	2.162	1.599	1.191	0.838	0.601	0.550	0.391
Log x Log										
$T_{n_1 n_2}^{(1)}$	12.403	10.707	6.246	5.223	2.921	2.350	1.167	0.893	0.605	0.450
$T_{n_1 n_2}^{(2)}$	5.995	4.816	4.032	3.170	2.575	1.972	1.515	1.122	1.048	0.767
LN										
$T_{n_1 n_2}^{(1)}$	12.345	10.590	6.107	5.083	2.790	2.245	1.102	0.852	0.569	0.431
$T_{n_1 n_2}^{(2)}$	5.722	4.608	3.815	3.001	2.399	1.847	1.420	1.069	0.988	0.737
C01										
$T_{n_1 n_2}^{(1)}$	12.255	10.620	6.156	5.194	2.886	2.360	1.194	0.939	0.650	0.502
$T_{n_1 n_2}^{(2)}$	5.852	4.755	3.950	3.145	2.545	2.001	1.560	1.191	1.123	0.853
N01										
$T_{n_1 n_2}^{(1)}$	12.484	10.666	6.168	5.095	2.740	2.172	1.000	0.759	0.493	0.364
$T_{n_1 n_2}^{(2)}$	5.757	4.580	3.726	2.888	2.248	1.695	1.235	0.903	0.827	0.601
L01										
$T_{n_1 n_2}^{(1)}$	12.422	10.650	6.187	5.139	2.784	2.230	1.066	0.816	0.535	0.400
$T_{n_1 n_2}^{(2)}$	5.820	4.648	3.850	2.999	2.353	1.791	1.354	1.002	0.916	0.674

Tabulka 4.19: Aproximace kritických hodnot pro spojitá rozdělení. Statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_1 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .

	a=0.25		a=0.5		a=1		a=2		a=3	
	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$	$\alpha_1$	$\alpha_2$
LN x LN										
$T_{n_1 n_2}^{(1)}$	12.487	10.438	5.975	4.809	2.586	2.000	0.972	0.718	0.508	0.365
$T_{n_1 n_2}^{(2)}$	5.352	4.137	3.414	2.582	2.102	1.546	1.226	0.872	0.863	0.613
C01										
$T_{n_1 n_2}^{(1)}$	12.286	10.564	6.046	5.065	2.765	2.253	1.132	0.894	0.632	0.491
$T_{n_1 n_2}^{(2)}$	5.626	4.568	3.749	2.989	2.395	1.882	1.476	1.140	1.091	0.830
N01										
$T_{n_1 n_2}^{(1)}$	12.420	10.542	6.002	4.942	2.596	2.063	0.942	0.715	0.471	0.350
$T_{n_1 n_2}^{(2)}$	5.509	4.355	3.506	2.708	2.066	1.563	1.153	0.851	0.795	0.574
L01										
$T_{n_1 n_2}^{(1)}$	12.422	10.650	6.187	5.139	2.784	2.230	1.066	0.816	0.535	0.400
$T_{n_1 n_2}^{(2)}$	5.820	4.648	3.850	2.999	2.353	1.791	1.354	1.002	0.916	0.674
C01 x C01										
$T_{n_1 n_2}^{(1)}$	12.172	10.569	6.085	5.162	2.866	2.372	1.238	0.991	0.715	0.564
$T_{n_1 n_2}^{(2)}$	5.713	4.700	3.884	3.145	2.557	2.039	1.626	1.276	1.230	0.952
N01										
$T_{n_1 n_2}^{(1)}$	12.410	10.603	6.080	5.039	2.690	2.173	1.042	0.813	0.556	0.428
$T_{n_1 n_2}^{(2)}$	5.611	4.496	3.637	2.853	2.215	1.714	1.298	0.992	0.934	0.708
L01										
$T_{n_1 n_2}^{(1)}$	12.303	10.573	6.082	5.068	2.751	2.230	1.098	0.860	0.595	0.461
$T_{n_1 n_2}^{(2)}$	5.665	4.577	3.730	2.961	2.333	1.826	1.399	1.074	1.014	0.771
N01 x N01										
$T_{n_1 n_2}^{(1)}$	12.627	10.623	6.057	4.900	2.524	1.942	0.821	0.594	0.366	0.260
$T_{n_1 n_2}^{(2)}$	5.524	4.249	3.373	2.520	1.837	1.332	0.902	0.631	0.571	0.398
L01										
$T_{n_1 n_2}^{(1)}$	12.521	10.598	6.036	4.935	2.570	2.014	0.879	0.658	0.415	0.299
$T_{n_1 n_2}^{(2)}$	5.520	4.331	3.453	2.631	1.978	1.469	1.027	0.745	0.679	0.480
L01 x L01										
$T_{n_1 n_2}^{(1)}$	12.433	10.576	6.040	4.971	2.626	2.086	0.960	0.725	0.467	0.343
$T_{n_1 n_2}^{(2)}$	5.569	4.402	3.545	2.747	2.128	1.604	1.177	0.860	0.787	0.566

Tabulka 4.20: Aproximace kritických hodnot pro spojitá rozdělení. Statistiky  $T_{n_1 n_2}^{(1)}$  a  $T_{n_1 n_2}^{(2)}$  jsou počítány na hladinách  $\alpha_1 = 5\%$  a  $\alpha_1 = 10\%$ , při velikostech výběrů  $n_1 = n_2 = 30$ .



# Kapitola 5

## Závěr

Tato diplomová práce začala úvahou, kterak vhodně využít platnosti Glivenko-Cantelliho věty, ale nedojít přitom k testům typu Kolmogorov-Smirnov, které jsou sice konzistentní, ale jejichž aplikace je omezena pouze na spojité distribuční funkce. Využili jsme vzájemně jednoznačného vztahu mezi distribuční funkcí a charakteristickou funkcí dané náhodné veličiny a sestrojili testovou statistiku založenou na váženém funkcionálu rozdílu empirické charakteristické a hypotetické charakteristické funkce. Odpadl nám tak problém s případnou nespojitostí, neboť charakteristická funkce existuje a je spojitá pro všechna  $t \in \mathbb{R}$ , rozšířili jsme tedy možnost použití testů založených na této větě i pro výběry pocházející z rozdělení s diskrétní distribuční funkcí. I my jsme dospěli ke konzistentním testům, nicméně přesto má naše třída testů i své nedostatky.

Závislost limitního rozdělení testové statistiky na rozdělení hypotetické charakteristické funkce je nepříjemná a je základní nevýhodou testů založených na empirických charakteristických funkcích. Přesto jsme pomocí tzv. resampling metod (připomínám, že my jsme použili konkrétně bootstrap s vracením) schopni velice uspokojivě aproximovat kritické hodnoty pro daný výběr, což tento nedostatek příjemně kompenzuje. Ukázali jsme si dokonce, že v některých případech nám tato nevýhoda ani vadit nemusí, protože aproximované kritické hodnoty pro různá rozdělení jsou si tak blízké, že hypotetické rozdělení ani uvažovat nemusíme, jak se nám třeba stalo u vybraných diskrétních rozdělení v kapitole 4.2. Asi by stálo za to podrobněji prozkoumat, pro které situace a případy závislost na limitním rozdělení "mizí" a v kterých případech s ní bohužel musíme počítat.

Podotkněme ještě, že limitní chování daných statistik jsme zde diskutovali pouze pro  $t \in \mathbb{R}^1$ , všechny námi uvedená tvrzení se však dají prokázat i pro  $t \in \mathbb{R}^d$ ,  $d > 1$ . Tomuto rozšíření na mnohorozměrná rozdělení se podrobně věnuje článek [6].

# Dodatek A

## Příloha

### A.1 Citace continuous mapping theorem

**Věta 10** (Continuous mapping theorem). *Bud'  $h : \mathbb{R} \mapsto \mathbb{R}$  borelovsky měřitelná funkce s množinou  $D_h$  všech svých bodů nespojitosti takovou, že  $\mu(D_h) = 0$ . Pokud  $\mu_n \Rightarrow \mu$ , pak rovněž  $\mu_n h^{-1} \Rightarrow \mu h^{-1}$ . Symbol  $\Rightarrow$  zde značí slabou konvergenci.*

**Důsledek 1.** *Je-li funkce  $h$  z předešlé věty spojitá a  $X_n \xrightarrow{D} X$ , pak platí, že  $h(X_n) \xrightarrow{D} h(X)$ .*

*Důkaz.* Věta je i s důsledkem převzata z [3], v této knize jsou také obě tvrzení dokázána. □

### A.2 Zdrojový kód programu

Veškeré výpočty použité při této diplomové práci byly porovány programem R, verze 2.2.1 pro Windows. Tento software je volně šiřitelný na internetu, internetová adresa domovské stránky je <http://www.r-project.org>.

Pravděpodobnosti zamítnutí

```
N1<-30 #velikost 1.vyberu
N2<-30 #velikost 2.vyberu
N<-N1+N2
vektora<-c(0.25,0.5,1,2,3) #vektor parametru "a"
B<-2000 #bootstrap, pocet opakovani
alfa1<-0.05;alfa2<-0.1;
```

```

fce1<-function(x,y,a){vysl<-1/(a^2+(x-y)^2);return(vysl)}
fce2<-function(x,y,a){vysl<-exp(-(x-y)^2/(4*a));return(vysl)}
laplace<-function(n,a,b){ifelse(runif(n)>0.5,1,-1)*rexp(n,1/b)+a}

####statistiky T1 a T2 ####
stat<-function(data,a){
x <- data[1:N1];
y <- data[(N1+1):N];
T1 <- 2*a*N1*N2/N*(sum(outer(x, x, fce1, a = a))/N1^2
+ sum(outer(y, y, fce1, a = a))/N2^2
- 2*sum(outer(x, y, fce1, a = a))/(N1*N2))
T2 <- sqrt(pi/a)*N1*N2/N*(sum(outer(x, x, fce2, a = a))/N1^2
+ sum(outer(y, y, fce2, a = a))/N2^2
- 2*sum(outer(x, y, fce2, a = a))/(N1*N2))
return(c(T1,T2))}

####vlastni telo programu####
pocetOpak<-2000

for(j in 1:length(vektora)){
a<-vektora[j]
RR1 <- matrix(data=0,nrow=pocetOpak,ncol=2)
RR2 <- matrix(data=0,nrow=pocetOpak,ncol=2)
r1<-c(0,0)
r2<-c(0,0)
for (z in 1:pocetOpak){
#vygeneruj danou dvojici vyberu
vyber1<-rpois(30,10)
vyber2<-rpois(30,8.25)
zvyber<-c(vyber1,vyber2) #sdruzeny vyber

#1/bootstrap

bootT1<-NULL;bootT2<-NULL;
for (k in 1:B){
actual<-sample(1:(N1+N2),replace=T)
bootvyber<-zvyber[actual]
bstat<-stat(bootvyber,a);
bootT1<-c(bootT1,bstat[1]);
bootT2<-c(bootT2,bstat[2]);
}
}

```

```

sort1<-sort(bootT1);
sort2<-sort(bootT2);
c11<-sort1[B*(1-alfa1)];
c12<-sort1[B*(1-alfa2)];
c21<-sort2[B*(1-alfa1)];
c22<-sort2[B*(1-alfa2)];

#2/porovnej skut.hodnotu s kritickymi
stat1<-stat(zvyber, a); #statistiky spoctene z puvodnich hodnot
if(stat1[1] >= c11){RR1[z,1]<-1}
  if(stat1[1] >= c12){RR1[z,2]<-1}
if(stat1[2] >= c21){RR2[z,1]<-1}
  if(stat1[2] >= c22){RR2[z,2]<-1}
r1<-r1+c(RR1[z,1],RR1[z,2])
r2<-r2+c(RR2[z,1],RR2[z,2])}

#3/vypis a a psti zamitnuti na danych hladinach pro konkretni statistiku
print(paste("a=",a));
rr1<-r1/pocetOpak
rr2<-r2/pocetOpak
print("Psti zamitnuti u T1:");print(rr1)
print("Psti zamitnuti u T2:");print(rr2)}

```

## Aproximovace kritických hodnot

```

####vlastni telo programu####
pocetOpak<-2000

for(j in 1:length(vektora)){
a<-vektora[j]
bootc11<-NULL; bootc12<-NULL;bootc21<-NULL; bootc22<-NULL
for (z in 1:pocetOpak){

#0/vygeneruje danou dvojici výběrů
vyber1<-rgeom(30,(sqrt(34)-1)/16.5)
vyber2<-rgeom(30,(sqrt(34)-1)/16.5)
zvyber<-c(vyber1,vyber2) #sdruzeny vyber

#1/bootstrap

```

```

bootT1<-NULL;bootT2<-NULL;
  for (k in 1:B){
    actual<-sample(1:(N1+N2),replace=T)
    bootvyber<-zvyber[actual]
bstat<-stat(bootvyber,a);
bootT1<-c(bootT1,bstat[1]);
bootT2<-c(bootT2,bstat[2]);
}
sort1<-sort(bootT1);
sort2<-sort(bootT2);
c11<-sort1[B*(1-alfa1)];
c12<-sort1[B*(1-alfa2)];
c21<-sort2[B*(1-alfa1)];
c22<-sort2[B*(1-alfa2)];

bootc11<-c(bootc11,c11);bootc21<-c(bootc21,c21);
bootc12<-c(bootc12,c12);bootc22<-c(bootc22,c22)}

sort11<-sort(bootc11);sort12<-sort(bootc12);
sort21<-sort(bootc21);sort22<-sort(bootc22);
kr11<-sort11[pocetOpak*(1-alfa1)];
kr12<-sort12[pocetOpak*(1-alfa2)];
kr21<-sort21[pocetOpak*(1-alfa1)];
kr22<-sort22[pocetOpak*(1-alfa2)];

print(paste("a= ",a));
print(paste("Kritické hodnoty pro T1 při alfa1: ",kr11, ",
při alfa2: ", kr12));
print(paste("Kritické hodnoty pro T1 při alfa1: ",kr11, ",
při alfa2: ", kr12))}

```

# Literatura

- [1] Anděl, Jiří. (2002). *Základy matematické analýzy*. preprint, Univerzita Karlova v Praze, 2002, Praha.
- [2] Hušková, Marie; Meintanis, Simos G. (2006) *Change point analysis based on empirical characteristic functions*. *Metrika* 63, No. 2, 145-168.
- [3] Davidson, James. (1994). *Stochastic Limit Theory - Advanced Texts in Econometrics* Oxford University Press: 1994, Oxford.
- [4] Hušková, Marie; Meintanis, Simos G. *Testing procedures based on empirical characteristic functions I: Goodness-of-fit, testing for symmetry and independence* přijato k otištění v tatra Mt. math. Publ.
- [5] Hušková, Marie; Meintanis, Simos G. *Testing procedures based on empirical characteristic functions II: k-sample problem, change point problem* přijato k otištění v tatra Mt. math. Publ.
- [6] Hušková, Marie; Meintanis, Simos G. *Tests for the multivariate k-sample problem based on the empirical characteristic function* přijato k otištění do časopisu *Statistic*.
- [7] Lehman, E.L. Davidson, James. (1994). *Stochastic Limit Theory - Advanced Texts in Econometrics* Oxford University Press: 1994, Oxford.
- [8] Madurkayová B. *Tests based on U-statistics*. Master thesis, Charles University in Prague, Czech Republic, 2006.
- [9] Meintanis, Simos G. (2005) *Permutation tests for homogeneity based on the empirical characteristic function*. *J. Nonparametric Stat.* 17, No. 5, 583-592.
- [10] Ushakov, Nikolai G. (1999). *Selected topics in characteristic functions*. VSP: Utrecht.

**pozice nalezené chyby:**

str.10, 5. řádek

**nesprávně je uvedeno:**

Pak ale z centrální limitní věty pro Hilbertovy prostory (přesné znění této věty najdeme v [3] na str.450 pod názvem 'Functional central limit theorem') dostáváme, že  $Z_n \xrightarrow{D} \mathcal{Z}$  pro  $n \rightarrow \infty$ , kde  $\mathcal{Z}$  je gaussovský proces s kovarianční funkcí tím jsme dokázali tvrzení (1.5).

**správně má být:**

Pak ale z centrální limitní věty pro Hilbertovy prostory (přesné znění této věty najdeme v [3] na str.450 pod názvem 'Functional central limit theorem') dostáváme, že  $Z_n \xrightarrow{D} \mathcal{Z}$  pro  $n \rightarrow \infty$ , kde  $\mathcal{Z}$  je gaussovský proces s kovarianční funkcí

$$K(s, t) = E\mathcal{Z}(s)\mathcal{Z}(t) = C(s - t) + S(s + t) - \\ - C(s)C(t) - C(s)S(t) - S(s)C(t) - S(s)S(t),$$

tím jsme dokázali tvrzení (1.5).