

**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Adam Špiesz

**Testy dobré shody pro Cauchyho
rozdělení**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Zdeněk Hlávka, Ph.D.

Studijní program: Finanční matematika

Praha 2025

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

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

V dne

Podpis autora

Rád by som poďakoval môjmu školiteľovi, doc. RNDr. Zdeňkovi Hlávkovi, Ph.D., za jeho ochotu, trpezlivosť a podporu počas písania bakalárskej práce. Jeho cenné pripomienky a včasná spätná väzba mi boli veľkou oporou.

Název práce: Testy dobré shody pro Cauchyho rozdělení

Autor: Adam Špiesz

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Zdeněk Hlávka, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Táto bakalárska práca sa zaoberá testami dobrej zhody pre Cauchyho rozdelenie. Poskytuje prehľad viacerých testov, založených na vzdialenosti od empirickej distribučnej funkcie, vierohodnostnom pomere, empirickej charakteristickej funkcii a Kullbackovej-Leiblerovej divergencii. Podrobnejšie sa venuje testovej štatistike vychádzajúcej z chí-kvadrát testu, ktorá využíva prvú a poslednú poradovú štatistiku. Práca sa tiež zaoberá odhadmi použitými v tejto testovej štatistike a navrhuje jej modifikáciu použitím kvantilového odhadu parametru σ . Následne bola vykonaná simulačná štúdia v programe Mathematica pomocou metódy Monte Carlo, ktorej výsledky ukázali, že navrhnutá úprava testovej štatistiky zvýšila jej silu. Simulácie zároveň potvrdili, že testy založené na Kullbackovej-Leiblerovej divergencii vykazujú celkovo najväčšiu silu.

Klíčová slova: Cauchyho rozdelenie, test dobrej zhody, kryptomeny

Title: Goodness-of-fit tests for the Cauchy distribution

Author: Adam Špiesz

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zdeněk Hlávka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This bachelor's thesis addresses goodness-of-fit tests for the Cauchy distribution. It provides an overview of several tests based on the distance from the empirical distribution function, the likelihood ratio, the empirical characteristic function, and the Kullback–Leibler (KL) divergence. In greater detail, it focuses on a test derived from the chi-squared test, which utilizes the first and last order statistics. The thesis also addresses the estimators used in this test statistic and proposes a modification using a quantile estimator for the parameter σ . A simulation study was then conducted in Mathematica using the Monte Carlo method. Its results demonstrated that the proposed adjustment increases the test's power. The simulations further confirmed that the tests based on KL divergence exhibit the greatest power overall.

Keywords: Cauchy distribution, goodness-of-fit test, cryptocurrencies

Obsah

Úvod	2
1 Cauchyho rozdelenie a test dobrej zhody	3
1.1 Cauchyho rozdelenie	3
1.2 Štatistický test dobrej zhody	5
1.3 Odhad parametrov	5
1.4 Príklad	6
2 Prehľad testov	9
2.1 Testy založené na vzdialenosti od distribučnej funkcie	9
2.2 Testy založené na vierohodnostnom pomere	10
2.3 Testy založené na empirickej charakteristickej funkcii	10
2.4 Testy založené na odhadoch Shannonovej entropie v Kullbackovej– Leiblerovej divergencii	11
2.5 Ebner, Eid, Klar	12
2.6 Zhrnutie	13
3 Rublíkova testová štatistika	14
3.1 Testová štatistika	14
4 Simulácia	17
4.1 Výpočet	17
4.2 Výsledky	19
4.3 Príklad	21
5 Dôkaz	23
Záver	25
Seznam použité literatury	26

Úvod

Štatistické testy dobrej zhody sú nástrojom na overenie, či daný súbor dát pochádza z určitého pravdepodobnostného rozdelenia. Cauchyho rozdelenie predstavuje špecifický typ spojitého rozdelenia, ktoré sa vyznačuje ťažkými chvostami a absenciou strednej hodnoty a rozptylu.

Cielom tejto práce je oboznámiť sa s aktuálnou odbornou literatúrou venujúcou sa testom dobrej zhody pre Cauchyho rozdelenie a podrobne predstaviť vybrané testy. Hlavným zámerom je porovnať dostupné testy na základe ich vlastností pri realistických alternatívach.

Prvá kapitola práce sa venuje základným definíciám a vysvetleniu problematiky. Druhá kapitola sa zameriava na niekoľko druhov štatistických testov, ktoré sa objavujú v literatúre. V tretej kapitole je podrobnejšie predstavený štatistický test navrhnutý Rublík (2001), vrátane jeho matematického ododenia. V štvrtej kapitole sú z popísaných testov vybrané niektoré, pre ktoré sa pomocou metódy Monte Carlo (s $N = 10,000$ opakovaniami) vypočítavajú kritické hodnoty a následne sila testu na hladine $\alpha = 0.05$. Nakoniec budú výsledky zhodnotené.

1. Cauchyho rozdelenie a test dobrej zhody

1.1 Cauchyho rozdelenie

Cauchyho rozdelenie, známe aj ako Cauchyho-Lorentzovo rozdelenie, je spojité rozdelenie určené dvoma parametrami, polohou ($\mu \in \mathbb{R}$) a škálou ($\sigma > 0$). Jeho hustota pravdepodobnosti je definovaná nasledovne:

$$f_0(x; \mu, \sigma) = \frac{1}{\pi\sigma[1 + (\frac{x-\mu}{\sigma})^2]}, \quad x \in \mathbb{R}, \quad (1.1)$$

ďalej budeme značiť

$$f_0(x) = f_0(x; 0, 1) = \frac{1}{\pi(1 + x^2)}, \quad x \in \mathbb{R}. \quad (1.2)$$

Graf hustoty je vykreslený v Obrázku 1.1. Distribučná funkcia Cauchyho rozdelenia je:

$$F_0(x; \mu, \sigma) = \frac{1}{\pi} \arctan\left(\frac{x - \mu}{\sigma}\right) + \frac{1}{2}, \quad x \in \mathbb{R} \quad (1.3)$$

a značíme

$$F_0(x) = F_0(x; 0, 1) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}, \quad x \in \mathbb{R}. \quad (1.4)$$

Graf distribučnej funkcie je vykreslený v Obrázku 1.2. Ďalej ho budeme značiť ako $C(\mu, \sigma)$ (Mahdizadeh a Zamanzade, 2019). Kvantil distribučnej funkcie (1.4) sa dá definovať aj exaktne a to ako

$$\xi_p = \tan((p - 0.5)\pi) \quad (1.5)$$

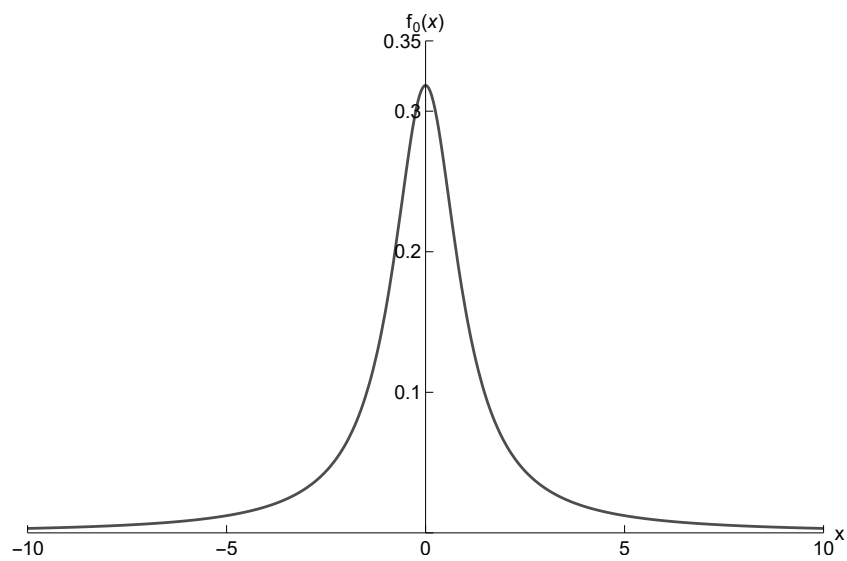
Vlastnosti a použitie

Cauchyho rozdelenie je typickým príkladom rozdelenia, ktoré nemá strednú hodnotu. Má široké využitie v rôznych oblastiach vedy a techniky. Napríklad, opisuje energetické spektrum excitovaného stavu atómu, ako aj rezonančný stav elementárnych častíc. Ďalej sa využíva pri štúdiu rozdelenia kontaktných rezistencií v elektronike, distribúcie hypocentier zemetrasení alebo rozdielov rýchlostí spôsobených rôznymi vírovými štruktúrami (Mahdizadeh a Zamanzade, 2019).

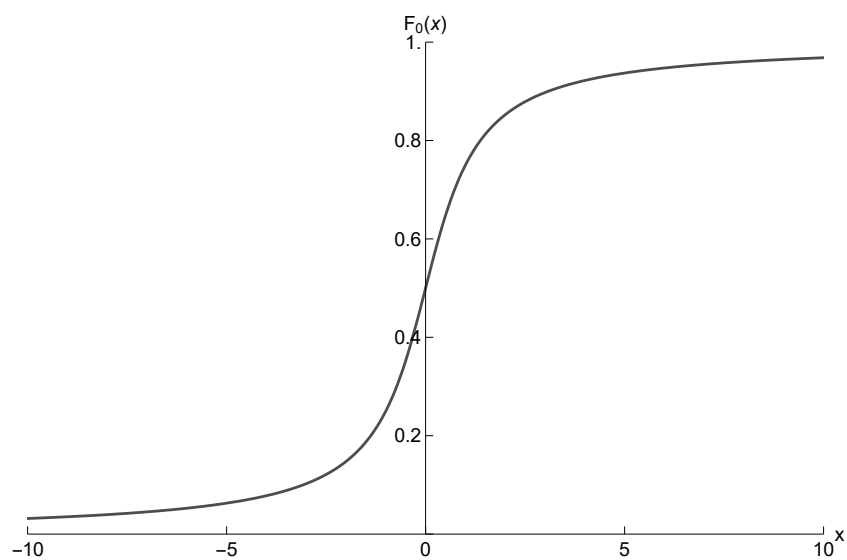
Podľa štúdie od Szczygielski a kol. (2020) sa Cauchyho rozdelenie ukázalo ako najlepšie vyhovujúce pre väčšinu analyzovaných kryptomien a potvrdzuje, že ich návratnosti nie sú normálne rozdelené. Tento výskum zahŕňal najrozsiahlejšiu analýzu návratností kryptomien, pričom testoval až 58 potenciálnych rozdelení na vzorke 15 najväčších kryptomien.

Rozdelenie sa tiež vyznačuje ťažkými chvostami a treba taktiež poznamenať, že

$$X \sim \mathcal{C}(\mu, \sigma) \Leftrightarrow (X - \mu)/\sigma \sim \mathcal{C}(0, 1). \quad (1.6)$$



Obr. 1.1: Hustota štandardného Cauchyho rozdelenia



Obr. 1.2: Distribučná funkcia štandardného Cauchyho rozdelenia

1.2 Štatistický test dobrej zhody

Majme náhodný výber X_1, X_2, \dots, X_n , s neznámou distribučnou funkciou $F(x)$. Testujeme zloženú hypotézu H_0

$$H_0 : F(x) = F_0(x; \mu, \sigma) \quad (1.7)$$

oproti alternatíve

$$H_1 : F(x) \neq F_0(x; \mu, \sigma), \quad (1.8)$$

kde $F_0(x, \mu, \sigma)$ je distribučná funkcia Cauchyho rozdelenia s neznámymi parametrami ($\mu \in \mathbb{R}$) a ($\sigma > 0$).

1.3 Odhad parametrov

Testovanie dobrej zhody s určeným rozdelením, u ktorého dopredu nepoznáme parametre, sa obvykle nezaobíde bez odhadnutia parametrov.

Kvantilové odhady

Cauchyho rozdelenie je špecifické neexistenciou strednej hodnoty. Jednoduchý odhad momentov tu teda neobstojí. Podľa Mahdizadeh a Zamanzade (2019) sú odhady založené na maximálnej vierohodnosti zložité. Testové štatistiky, ktoré spomenieme v ďalšej kapitole si ale vystačia s jednoduchými odhadmi parametrov μ a σ pomocou mediánu a polovičného medzikvartilového rozpätia.

$$\hat{\mu}_n^Q = \begin{cases} (X_n^{(n/2)} + X_n^{(n/2+1)})/2 & \text{ak } n \text{ je párne,} \\ X_n^{((n+1)/2)} & \text{inak.} \end{cases} \quad (1.9)$$

$$\hat{\sigma}_n^Q = \frac{1}{2}(\hat{\xi}_{0,75} - \hat{\xi}_{0,25}), \quad (1.10)$$

kde $\hat{\xi}_p$ je p -tý empirický kvantil, $X_n^{(i)}$ je i -ta poradová štatistika. Medián je prirodzeným odhadom μ , pretože Cauchyho rozdelenie je symetrické okolo $x = \mu$. Polovičný medzikvartilový rozptyl je motivovaný tým, že pri Cauchyho rozdelení sa 1. kvartil rovná $x - \sigma$ a 3. kvartil sa rovná $x + \sigma$. Rozdiel kvartilov sa tým pádom rovná 2σ . Polovica medzikvartilového rozidelu je tým pádom prirodzeným odhadom parametru σ .

Alternatívny odhad

Pri štatistikách, ktoré zvolil Rublík (2001), sa použil rovnaký odhad μ ako v (1.9). Pri parametri σ sa použil iný odhad ako vo zvyšných štatistikách, teda

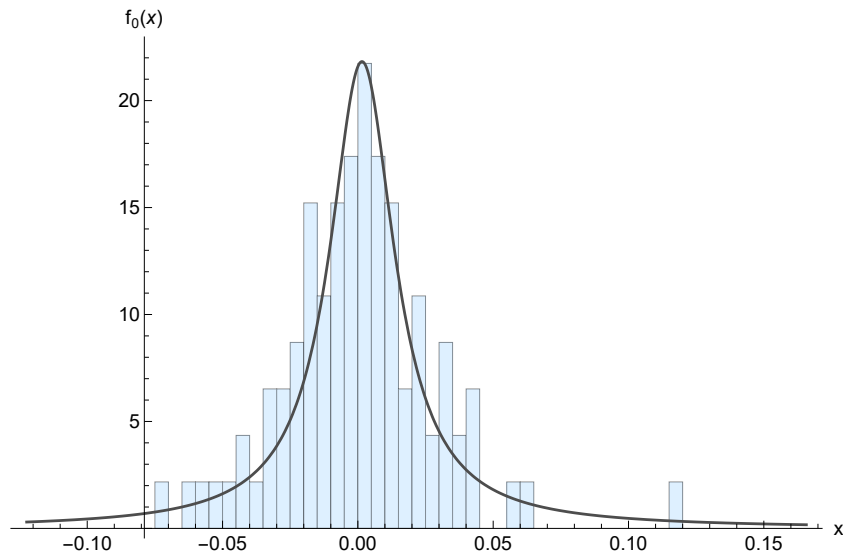
$$\hat{\sigma}_n^R = \sum_{i=1}^n c_{ni} X_n^{(i)}, \quad (1.11)$$

$$c_{ni} = \frac{1}{n} J\left(\frac{i}{n+1}\right), \quad J(u) = \frac{8 \tan(\pi(u-0.5))}{\sec^4(\pi(u-0.5))} = -8 \cos(\pi u) \sin^3(\pi u). \quad (1.12)$$

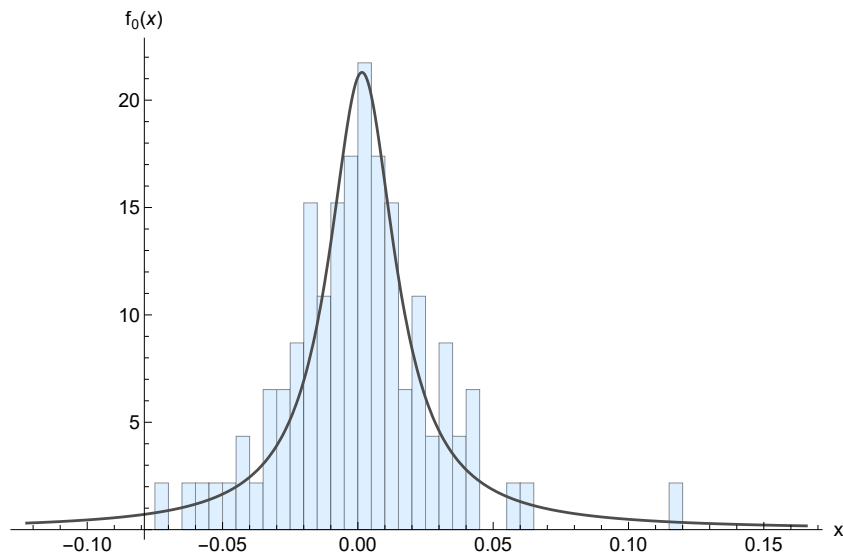
Dôvodom zvolenia tohoto odhadu je dobrá vlastnosť pri konštrukcii testovej štatistiky (3.7) v Tvrdení 3.1.

1.4 Príklad

Pre ilustráciu spomenutých odhadov parametrov a aj samotného Cauchyho rozdelenia uvidíme príklad. Uvažujeme dáta logaritmických výnosov Bitcoinu v časovom období od 1. júla 2024 do 31. septembra 2024, verejne dostupné na platforme www.coingecko.com. Následne uvidíme dva grafy hustôt Cauchyho rozdelenia s príslušnými odhadmi parametrov, spolu s histogramami týchto dát, pričom veľkosť výberu $n = 92$. Vypočítané odhady parametru σ sú uvedené v tabuľkách, zatiaľ čo oboch prípadoch používame (1.9) ako odhad parametra μ . Vypočítaná hodnota odhadu $\hat{\mu}_n^Q = 1.52 \times 10^{-3}$. Grafy aj výpočty pochádzajú z programu Wolfram Mathematica.



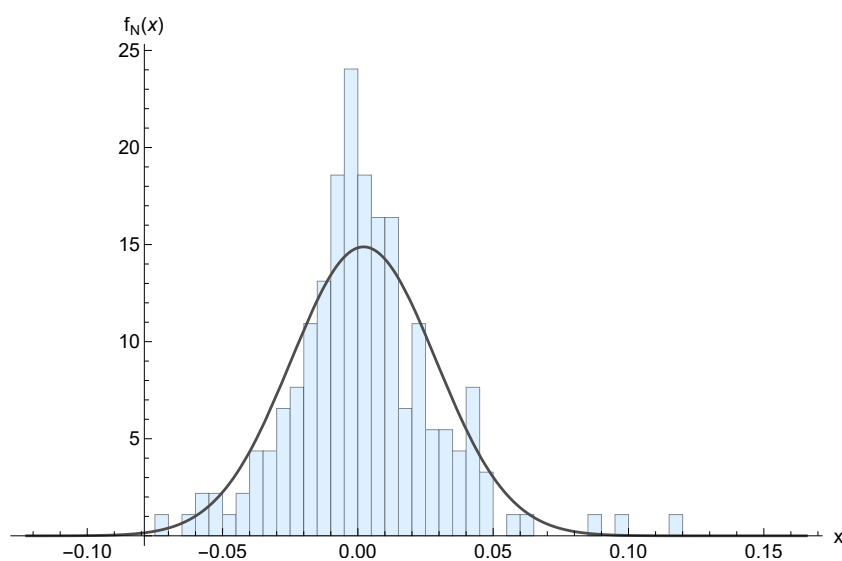
Obr. 1.3: Graf hustoty Cauchyho rozdelenia s polovičným medzikvartilovým rozpätím ako odhad parametru σ , spolu s histogramom logaritmickejých výnosov Bitcoinu. Vypočítaná hodnota $\hat{\sigma}_n^Q = 1.49 \times 10^{-2}$.



Obr. 1.4: Graf hustoty Cauchyho rozdelenia s Rublíkovým odhadom parametru σ spolu s histogramom logaritmickejých výnosov Bitcoinu. Vypočítaná hodnota $\hat{\sigma}_n^R = 1.46 \times 10^{-2}$.

Vhodnosť Cauchyho rozdelenia a normálneho rozdelenia

Výber Bitcoinu je motivovaný štúdiou Szczygielski a kol. (2020), ktorá zdôrazňuje vhodnosť použitia Cauchyho rozdelenia pri modelovaní výnosov kryptomien. Pre ilustráciu uskutočníme test normality. V tomto príklade používame dáta logartnických výnosov v časovom období od 1. júla 2024 do 31. decembra 2024, taktiež z platformy www.coingecko.com. P-hodnota Andersonovho-Darlingovho testu normality pri spomenutom súbore logaritnických výnosov vyšla 1×10^{-4} . Teda zamietame hypotézu, že súbor je z normálneho rozdelenia. Pre ilustráciu nevhodnosti normálneho rozdelenia uvidíme graf histogramu dát spolu hustotou normálneho rozdelenia s odhadnutými parametrami.



Obr. 1.5: Graf hustoty normálneho rozdelenia s maximálne vierohodnými odhadmi s histogramom logaritnických výnosov Bitcoinu.

Z grafu je zrejmé, že normálne rozdelenie nedokáže zachytiť ťažké chvosty pozorované v dátach. Chybné predpokladanie normálneho rozdelenia pri modelovaní výnosov Bitcoinu môže viesť k podhodnoteniu rizika extrémnych pohybov cien. V praxi to napríklad môže spôsobiť nesprávne nastavenie rizikových metrík, ako hodnota pri riziku (VaR), a tým podceňiť možnosť veľkých strát.

2. Prehľad testov

V tejto kapitole predstavíme niekoľko testov dobrej zhody. Pôjde len o prehľad, podrobnejšie rozoberieme testovú štatistiku navrhnutú Rublík (2001) v 3. kapitole. Tu uvedieme niekoľko testových štatistík z prehľadového článku Mahdizadeh a Zamanzade (2019). V tomto článku autori popísali a porovnali 14 testových štatistík založených na rôznych metódach. Vybrané testové štatistiky sme doplnili ešte o zaujímavú skupinu testov z článku Ebner a kol. (2022).

2.1 Testy založené na vzdialenosti od distribučnej funkcie

Kolmogorovov–Smirnovov test, Andersonova–Darlingova štatistika a Cramér–von Misesova štatistika sú založené na váženej vzdialenosti medzi empirickou distribučnou funkciou a skutočnou distribučnou funkciou. Následujú ich definície, tak ako su uvedené v článku Mahdizadeh a Zamanzade (2019).

Kolmogorovov–Smirnovov test

Medzi najznámejšie EDF testy patrí Kolmogorov–Smirnovov test. Jeho testová štatistika je definovaná ako maximálna absolútna odchýlka medzi empirickou distribučnou funkciou $F_n(x)$ a teoretickou distribučnou funkciou $F_0(x)$:

$$\text{KS} = \sup_x |F_n(x) - F_0(x)|. \quad (2.1)$$

Z tohoto výrazu je možné odvodiť explicitnejší výraz. Uvedieme ho tak ako v Mahdizadeh a Zamanzade (2019), kde boli použité aj odhady parametrov

$$\text{KS} = \max_{i=1, \dots, n} \left(\max \left\{ \frac{i}{n} - F_0 \left(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^Q \right), F_0 \left(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^Q \right) - \frac{i-1}{n} \right\} \right). \quad (2.2)$$

Cramér–von Misesova štatistika

Cramér–von Misesova štatistika je definovaná ako

$$W^2 = n \int_{-\infty}^{\infty} \left(F_n(x) - F_0(x) \right)^2 dF_0(x). \quad (2.3)$$

Explicitný výraz s použitím odhadu parametrov pre teoretickú distribučnú funkciu Cauchyho rozdelenia je

$$W^2 = \sum_{i=1}^n \left(F_0 \left(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^Q \right) - \frac{i-0.5}{n} \right)^2 + \frac{1}{12n}. \quad (2.4)$$

Andersonova–Darlingova štatistika

Ďalším silným testom Andersonova–Darlingova štatistika, používa sa najmä pre malé výbery. Je definovaná takto:

$$A^2 = n \int_{-\infty}^{\infty} \frac{\left[F_n(x) - F_0(x) \right]^2}{F_0(x)(1 - F_0(x))} dF_0(x).$$

V tejto rovnici váhová funkcia $\frac{1}{F_0(x)(1-F_0(x))}$ zabezpečuje, že odchýlky v chvostoch majú väčší vplyv na výslednú hodnotu testovej štatistiky. Tak ako pri predošlych dvoch testoch, uvádzame explicitnú formu s odhadmi.

$$A^2 = -\frac{2}{n} \sum_{i=1}^n \left((i-0.5) \log F_0(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^Q) + (n-i+0.5) \log \left(1 - F_0(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^Q) \right) \right) - n. \quad (2.5)$$

Pri všetkých troch testoch hypotézu zamietame pri veľkých hodnotách testových štatistík. Kritické hodnoty Kolmogorovovho-Smirnovovho testu vypočítame pomocou simulácie v štvrtej kapitole.

2.2 Testy založené na vierohodnostnom pomere

Mahdizadeh a Zamanzade (2019) uviedli nasledujúce tri testy. Sú to modifikácie testov navrhnutých Zhang (2002):

$$\begin{aligned} Z_K &= \max_{i=1, \dots, n} \left\{ (i-0.5) \log \left(\frac{i-0.5}{n F_0(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^R)} \right) + (n-i+0.5) \log \left(\frac{n-i+0.5}{n(1-F_0(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^R))} \right) \right\}, \\ Z_A &= -\sum_{i=1}^n \left[\frac{\log \{ F_0(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^R) \}}{n-i+0.5} + \frac{\log \{ 1 - F_0(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^R) \}}{i-0.5} \right], \\ Z_C &= \sum_{i=1}^n \left[\log \left(\frac{1/F_0(X_n^{(i)}; \hat{\mu}_n^Q, \hat{\sigma}_n^R) - 1}{(n-0.5)/(i-0.75) - 1} \right) \right]^2. \end{aligned} \quad (2.6)$$

Tieto štatistiky sú založené na vierohodnostnom pomere (ang. likelihood ratio). Hypotézu zamietame pre veľké hodnoty testových štatistík.

2.3 Testy založené na empirickej charakteristickej funkcii

Gürtler a Henze (2000) navrhli test založený na empirickej charakteristickej funkcii. Najprv definujeme

$$\Psi_n(t) = \frac{1}{n} \sum_{j=1}^n \exp(itY_j), \quad (2.7)$$

kde

$$Y_j = \frac{X_j - \hat{\mu}_n^Q}{\hat{\sigma}_n^Q}, \quad (2.8)$$

pre $j = 1, \dots, n$, sú štandardizované dáta. Testová štatistika

$$D_{n,\lambda} = n \int_{-\infty}^{\infty} |\Psi_n(t) - e^{-|t|}|^2 e^{-\lambda|t|} dt \quad (2.9)$$

predstavuje väženú L^2 -vzdialenosť medzi Ψ_n a charakteristickou funkciou štandardného Cauchyho rozdelenia, kde λ je pevne zvolený kladný váhový parameter. Veľké hodnoty $D_{n,\lambda}$ znamenajú zamietnutie nulovej hypotézy H_0 . Po algebraických úpravách sa odvodila alternatívna reprezentácia $D_{n,\lambda}$ v tvare:

$$D_{n,\lambda} = \frac{2}{n} \sum_{j=1}^n \sum_{k=1}^n \frac{\lambda}{\lambda^2 + (Y_j - Y_k)^2} - 4 \sum_{j=1}^n \frac{1 + \lambda}{(1 + \lambda)^2 + Y_j^2} + \frac{2n}{2 + \lambda}. \quad (2.10)$$

Gürtler a Henze (2000) zistili pomocou simulačnej štúdie, že použitie $\lambda = 5$ vedie k silnému testu.

2.4 Testy založené na odhadoch Shannonovej entropie v Kullbackovej–Leiblerovej divergencii

Ďalšie testy využívajú Kullbackovu–Leiblerovu divergenciu medzi ľubovoľnou funkciou f a f_0 . Tá je daná nasledujúcim výrazom.

$$D(f, f_0; \mu, \sigma) = \int_{-\infty}^{\infty} f(x) \log \left(\frac{f(x)}{f_0(x; \mu, \sigma)} \right) dx \quad (2.11)$$

$$= -H(f) - \int_{-\infty}^{\infty} f(x) \log(f_0(x; \mu, \sigma)) dx, \quad (2.12)$$

kde $H(f)$ je Shannonova entropia f definovaná ako

$$H(f) = - \int_{-\infty}^{\infty} f(x) \log(f(x)) dx \quad (2.13)$$

Test založený na Vašíčkovom odhade

Vašíček (1976) predstavil jednoduchý odhad, ktorý sa často používa pri testoch zhody. Tento odhad je definovaný nasledovne:

$$HV_{m,n} = \frac{1}{n} \sum_{i=1}^n \log \left(\frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right) \quad (2.14)$$

kde m je celé číslo menšie alebo rovné $n/2$ a nazýva sa window size. $X_n^{(1)} \leq \dots \leq X_n^{(n)}$ sú poradové štatistiky založené na náhodnom výbere o rozsahu n , $X_n^{(i)} = X_n^{(1)}$ pre $i < 1$, a $X_n^{(i)} = X_n^{(n)}$ pre $i > n$. Mahdizadeh a Zamanzade (2019) navrhli použiť

$$KL_1 = \exp \left(-HV_{m,n} - \frac{1}{n} \sum_{i=1}^n \log \left(f_0 \left(X_i; \hat{\mu}_n^Q, \hat{\sigma}_n^R \right) \right) \right) \quad (2.15)$$

ako testovú štatistiku. Veľké hodnoty KL_1 znamenajú zamietnutie hypotézy.

Test založený na odhade Ebrahimiho a kolektívu

Ebrahimi a kol. (1994) navrhli použiť tento odhad Shannonovej entropie.

$$HE_{m,n} = \frac{1}{n} \sum_{i=1}^n \log \left(\frac{n}{d_i m} (Y_{(i+m)} - Y_{(i-m)}) \right), \quad (2.16)$$

kde

$$d_i = \begin{cases} 1 + \frac{i+1}{m} - \frac{i}{m^2} & 1 \leq i \leq m \\ 2 & m+1 \leq i \leq n-m, \\ 1 + \frac{n-i}{m+1} & n-m+1 \leq i \leq n \end{cases} \quad (2.17)$$

a $Y_{(i)}$ je definované

$$\begin{cases} Y_{(i-m)} = a + \frac{i-1}{m} (X_n^{(1)} - a) & 1 \leq i \leq m \\ Y_{(i)} = X_n^{(i)} & m+1 \leq i \leq n-m \\ Y_{(i+m)} = b - \frac{n-i}{m} (b - X_n^{(n)}) & n-m+1 \leq i \leq n \end{cases} \quad (2.18)$$

konštanty a, b sú určené tak, že $P(a < X < b) \approx 1$. Napríklad, keď F V tomto prípade volíme $a = \bar{X}_n - ks$ ($b = \bar{X}_n + ks$), kde

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (2.19)$$

a k je vhodné číslo, zhruba od 3 do 5 (v simulácií bolo použité $k = 5$). Mahdizadeh a Zamanzade (2019) použili ako testovú štatistiku výraz

$$KL_2 = \exp \left(-HE_{m,n} - \frac{1}{n} \sum_{i=1}^n \log \left(f_0 \left(X_i; \hat{\mu}_n^Q, \hat{\sigma}_n^R \right) \right) \right). \quad (2.20)$$

Hypotézu zamietame pri veľkej hodnote KL_2 .

2.5 Ebner, Eid, Klar

Ďalšia testová štatistika bola navrhnutá Ebner a kol. (2022). Testová štatistika je definovaná nasledovne

$$E_n = n \int_{-\infty}^{\infty} \left| \frac{1}{n} \sum_{j=1}^n \left(it - \frac{2Y_j}{1+Y_j^2} \right) e^{itY_j} \right|^2 \omega(t) dt \quad (2.21)$$

a je to je vážená L^2 -vzdialenosť od

$$E \left[\left(it - \frac{2Y}{1+Y^2} \right) \exp(itY) \right]$$

k 0. Y_j je definované rovnako ako v (2.8) a i značí imaginárnu jednotku. V tomto prípade $|\cdot|$ značí komplexnú absolútnu hodnotu a $\omega : \mathbb{R} \rightarrow (0, \infty)$ značí kladnú funkciu váh danú

$$\omega(t) = \omega_a(t) = \exp(-a|t|), t \in \mathbb{R}.$$

Treba poznamenať že ω je symetrická okolo počiatku, teda $\omega(t) = \omega(-t)$ platí pre všetky $t \in \mathbb{R}$, a že platí

$$\int_{-\infty}^{\infty} t^6 \omega(t) dt < \infty. \quad (2.22)$$

Numericky vyčísliteľný vzorec bez potreby integrácie je odvodený ako

$$E_n = E_{n,a} = \frac{1}{n} \sum_{j,k=1}^n \left(\frac{8aY_{n,j}Y_{n,k}}{(1+Y_{n,j}^2)(1+Y_{n,k}^2)((Y_{n,j}-Y_{n,k})^2+a^2)} - \frac{16aY_{n,j}(Y_{n,j}-Y_{n,k})}{(1+Y_{n,j}^2)((Y_{n,j}-Y_{n,k})^2+a^2)^2} + \frac{4a^3-12a(Y_{n,j}-Y_{n,k})^2}{((Y_{n,j}-Y_{n,k})^2+a^2)^3} \right). \quad (2.23)$$

Ebner a kol. (2022) uvažuje $a \in \{1, \dots, 6\}$. Je možné odvodiť aj asymptotickú testovú štatistiku keď $a \rightarrow 0$. Táto štatistika má aj známe asymptotické normálne rozdelenie.

2.6 Zhrnutie

Oba články, Mahdizadeh a Zamanzade (2019) a Ebner a kol. (2022), obsahovali samostatné simulačné štúdie. Z nich môžeme doporučiť ako najlepšie testy KL_1 (2.15) a KL_2 (2.20). Tieto zahrnieme aj v našej simulačnej štúdií.

3. Rublíkova testová štatistika

Nasledujúcu testovú štatistiku, ktorú navrhol Rublík (2001) popíšeme podrobnejšie. Autor čerpá z rozsiahlej publikácie od Greenwood a Nikulin (1996), venovanej rôznym metódam chí-kvadrát testovania. Konkrétne ide o chí-kvadrát test pre spojité rozdelenia s parametrami polohy a škály, ktorý umožňuje testovanie len na časti výberového priestoru. Inými slovami, namiesto použitia celého náhodného výberu X_1, \dots, X_n sa pre výpočet testovej štatistiky využije iba jeho podmnožina.

3.1 Testová štatistika

Definujeme empirickú distribučnú funkciu

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(-\infty, t)}(X_i) \quad (3.1)$$

A ďalej definujeme p -ty výberový kvantil ako

$$\hat{\xi}_{p,n} = \inf \{t : F_n(t) \geq p\}$$

Tvrdenie 3.1 (Rublík, 1999). *Majme celé číslo $k \geq 1$ a reálne čísla $0 < p_1 < \dots < p_k < 1$. Predpokladajme, že X_1, \dots, X_n je náhodný výber z Cauchyho rozdelenia $C(\mu, \sigma)$ a nech*

$$\Delta_n = \Delta_n(p_1, \dots, p_k) = \begin{pmatrix} F_0\left(\hat{\xi}_{p_1,n}; \hat{\theta}_n^R\right) - p_1 \\ F_0\left(\hat{\xi}_{p_2,n}; \hat{\theta}_n^R\right) - p_2 \\ \vdots \\ F_0\left(\hat{\xi}_{p_k,n}; \hat{\theta}_n^R\right) - p_k \end{pmatrix}, \quad (3.2)$$

kde $F_0(\cdot; \theta)$ je funkcia definovaná v (1.3) a $\hat{\theta}_n^R = (\hat{\mu}_n^Q, \hat{\sigma}_n^R)'$ je vektor odhadov definovaných vo vzťahoch (1.11) a (1.9)

(I) Vektor rozdielov Δ_n má presné nulové rozdelenie, t.j.,

$$(\Delta_n \mid C(\mu, \sigma)) \stackrel{D}{=} (\Delta_n \mid C(0,1)) \quad (3.3)$$

pre všetky $(\mu, \sigma) \in \Theta$ a pre všetky $n > 1$.

(II) Keď $n \rightarrow \infty$,

$$(\sqrt{n}\Delta_n \mid C(\mu, \sigma)) \stackrel{D}{\rightarrow} N(\mathbf{0}, \Sigma). \quad (3.4)$$

Tu

$$\Sigma = \Sigma(k, p_1, \dots, p_k) = \mathbf{A} + \mathbf{G} \quad (3.5)$$

kde \mathbf{A} je symetrická matica veľkosti $k \times k$ s prvkami $a_{ij} = p_i(1 - p_j)$ pre všetky $i \leq j$, a

$$G_{ij} = \frac{\sin^2(p_i\pi) \sin^2(p_j\pi)}{4} - \frac{\sin^2(p_i\pi)}{2} \min\{p_j, 1 - p_j\} - \frac{\sin^2(p_j\pi)}{2} \min\{p_i, 1 - p_i\} - \frac{\sin(2p_i\pi) \sin(2p_j\pi)}{2\pi^2}. \quad (3.6)$$

Ak $p_i \neq \frac{1}{2}$ pre $i = 1, \dots, k$, potom je matica Σ regulárna a pre

$$Q_n(k, p_1, \dots, p_k) = n \Delta_n(p_1, \dots, p_k)' \Sigma^{-1} \Delta_n(p_1, \dots, p_k) \quad (3.7)$$

platí slabá konvergencia k chí-kvadrát rozdeleniu s k stupňami voľnosti

$$(Q_n \mid C(\mu, \sigma)) \xrightarrow{D} \chi_k^2 \quad (3.8)$$

keď $n \rightarrow \infty$.

Dôkaz sa nachádza v piatej kapitole. Tvrdenie (3.1) nám poskytuje asymptotické rozdelenie testovej štatistiky

$$Q_n(k, p_1, \dots, p_k), \quad (3.9)$$

ako aj fakt, že táto štatistika má presné nulové rozdelenie. Nulovú hypotézu H_0 teda môžeme zamietnuť pri veľkých hodnotách $Q_n(k, p_1, \dots, p_k)$. Na stranách 91–92 v D'Agostino (2017) sa však uvádza, že tento druh testov nevykazuje dostatočnú silu, najmä v porovnaní s testami založenými na distribučných funkciách. Ako uvádza Rublík (2001) (a ako opätovne ukážeme aj pomocou simulácií), tento problém sa vyskytuje aj pri teste pre Cauchyho rozdelenie, konkrétne pre

$$R_F = Q(n, 0.1, 0.9) = n \Delta_n(0.1, 0.9)' (\Sigma(2, 0.1, 0.9))^{-1} \Delta_n(0.1, 0.9). \quad (3.10)$$

Rublík (2001) navyše argumentuje, že okrem slabého asymptotického správania je príčinou nízkej sily testu aj skutočnosť, že hodnoty p_i sú určené fixne. Preto navrhuje iný test, ktorý používa argumenty p_i závislé od veľkosti náhodného výberu. Ako sa napríklad píše na str. 37 v Hájek a Šidák (1967), platí

$$E \left[F_0(X_n^{(i)}; \theta) \right] = \frac{i}{n+1}, \quad (3.11)$$

za predpokladu, že θ je skutočnou hodnotou parametrov Cauchyho rozdelenia. Vzhľadom na ťažké chvosty Cauchyho rozdelenia navrhuje Rublík (2001) použiť testovú štatistiku

$$R_1 = Q_n \left(2, \frac{1}{n+1}, \frac{n}{n+1} \right). \quad (3.12)$$

Táto štatistika využíva prvú a poslednú poradovú štatistiku ako odhady $\frac{1}{n+1}$ -tého a $\frac{n}{n+1}$ -tého kvantilu. Konkrétne,

$$\hat{\xi}_{\frac{1}{n+1}, n} = \inf \left\{ t : F_n(t) \geq \frac{1}{n+1} \right\} = X_n^{(1)}, \quad (3.13)$$

a analogicky pre $p_2 = \frac{n}{n+1}$. Tým pádom je definovaný vektor

$$\Delta_n^R = \begin{pmatrix} F_0(X_n^{(1)}; \hat{\theta}_n^R) - \frac{1}{\frac{n}{n+1}} \\ F_0(X_n^{(n)}; \hat{\theta}_n^R) - \frac{1}{\frac{n}{n+1}} \end{pmatrix}, \quad (3.14)$$

a výsledná testová štatistika je daná vzťahom

$$Q_n \left(2, \frac{1}{n+1}, \frac{n}{n+1} \right) = n (\Delta_n^R)' \left(\Sigma \left(2, \frac{1}{n+1}, \frac{n}{n+1} \right) \right)^{-1} \Delta_n^R. \quad (3.15)$$

Pri kritickej hodnote $q(\alpha, n)$ zamietame nulovú hypotézu, že výber pochádza z Cauchyho rozdelenia, práve vtedy, keď

$$Q_n \left(2, \frac{1}{n+1}, \frac{n}{n+1} \right) > q(\alpha, n). \quad (3.16)$$

Ďalšia testová štatistika založená na odhade σ_n^Q

Štatistické testy v prvej časti tejto kapitoly používajú odhad parametrov Cauchyho rozdelenia (1.9) a (1.11). Kvalitatívny rozdiel medzi odhadmi parametru σ je vidno na porovnaní síl Kolmogorovovho-Smirnovovho testu v Rublík (2001) a v Mahdizadeh a Zamanzade (2019). V prvom spomenutom článku sa používa odhad $\hat{\sigma}_n^R$ a v druhom sa používa odhad $\hat{\sigma}_n^Q$. Avšak v prvom článku simulácia priniesla nižšiu silu ako v druhom. Ak teda tento rozdiel síl spočíva práve v rozdielnom používaní parametrov, použiť odhad $\hat{\sigma}_n^Q$ v Rublíkovej štatistike by mohlo priniesť lepší výsledok.

Preto sme sa rozhodli použiť odhady (1.9) a (1.10) v testovej štatistike (3.10), teda uvažujeme kvantilové odhady

$$\hat{\theta}_n^Q = (\hat{\mu}_n^Q, \hat{\sigma}_n^Q) \quad (3.17)$$

a pomocou

$$\Delta_n^Q = \left(F_0 \left(X_n^{(1)}; \hat{\theta}_n^Q \right) - \frac{1}{\frac{n}{n+1}}, F_0 \left(X_n^{(n)}; \hat{\theta}_n^Q \right) - \frac{n}{n+1} \right)', \quad (3.18)$$

vytvárame novú testovú štatistiku

$$R_2 = n \Delta_n^{Q'} \left(\Sigma \left(2, X_n^{(1)}, X_n^{(n)} \right) \right)^{-1} \Delta_n^Q. \quad (3.19)$$

Dôvodom použitia odhadu $\hat{\sigma}_n^R$ boli dobré vlastnosti v Tvrdení 3.1. Keďže asymptotické rozdelenie takejto testovej štatistiky nepotrebujeme poznať, stačí dokázať, že má presné nulové rozdelenie. To je ale zrejmé pre vlastnosť (1.6) a pretože platí

$$\hat{\sigma}_n^Q > 0, \quad \hat{\sigma}_n^Q(aX_1 + b, \dots, aX_n + b) = |a| \hat{\sigma}_n^Q(X_1, \dots, X_n), \quad a \in \mathbb{R} \quad (3.20)$$

čo plynie z vlastností kvantilov. Matica Σ nezávisí na odhadoch parametrov a preto jej regularita platí naďalej. Zaujímavosťou testovej štatistiky R_2 je, že používa len 5 poradových štatistík z výberu. V simulácii porovnáme silu štatistík R_1 a R_2 .

4. Simulácia

V predošlých kapitolách sme predstavili niekoľko testových štatistík. Pre porovnanie týchto testov je užitočné zistiť silu pri rôznych alternatívach. Pri viacerých testoch nie je možné silu zistiť analyticky, či už pre slabú asymptotiku, alebo pre neznáme rozdelenie testovej štatistiky. Preto volíme prístup simulácie, kde aproximujeme silu testu pri alternatíve. Najprv ale bude potrebná aproximácia kritického oboru. V skratke teda pomocou metódy Monte Carlo s $N = 10000$ opakovaniami vypočítame kritické hodnoty a následne silu na hladine $\alpha = 0.05$.

4.1 Výpočet

Výber testových štatistík

Pre simuláciu sme vybrali tradičný Kolmogorovov-Smirnovov test (2.2) a dva testy (2.15) a (2.20) ktoré navrhli Mahdizadeh a Zamanzade (2019). Tieto testy sú založené na KL vzdialenosti a následnom odhade Shannonovej entropie. Taktiež sme zahrnuli tri verzie štatistiky popísané v Rublík (2001).

Výber alternatív

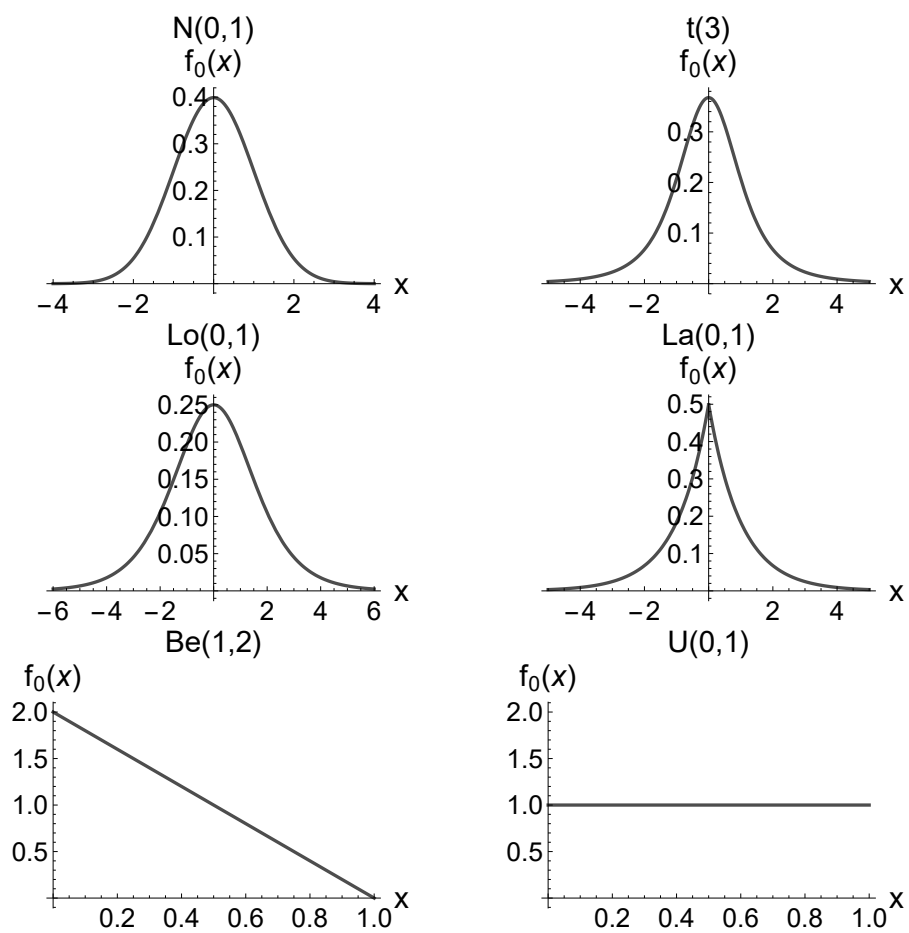
Ako alternatívne rozdelenia sme zvolili:

- Štandardné normálne rozdelenie $N(0,1)$
- Študentovo T-rozdelenie s tromi stupňami voľnosti $St(3)$
- Štandardné logistické rozdelenie $Lo(0,1)$
- Laplaceovo rozdelenie s parametrami 0 a 1 $La(0,1)$
- Beta rozdelenie s parametrami 1 a 2 $Be(1,2)$
- Štandardné rovnomerné rozdelenie $U(0,1)$

Realizácia simulácie

Simuláciu sme uskutočnili v programe Wolfram Mathematica. Prvý krok bol naprogramovať výrazy vybraných testových štatistík ako funkcie v programovacom jazyku Wolfram. Následovné 3 body opíšu proces aproximácie kritických hodnôt vybraných testových štatistík metódou Monte Carlo.

1. Pre každé $n \in \{10, 20, 30, 50, 100\}$ vygenerujeme súbor 10 000 výberov o veľkosti n z $C(0,1)$.
2. Vypočítame testovú štatistiku cez každý z výberov, ako výsledok nám vznikne súbor 10 000 testových štatistík pre každé n .
3. Na požadovanej hladine α vypočítame $1 - \alpha$ kvantil zo spomenutého súboru testových štatistík. Tento kvantil je vypočítanou kritickou hodnotou na hladine α .



Obr. 4.1: Hustota rozdělení použitých v simulacích

Následujúca tabuľka zahŕňa vypočítané kritické hodnoty pre každé n a každú z vybraných testových štatistík

Tabuľka 4.1: Kritické hodnoty

n	KS	KL_1	KL_2	R_1	R_F	R_2
10	0.29	1.29	0.85	10.96	15.9	5.71
20	0.21	1.16	0.74	10.39	13.44	6.91
30	0.17	1.11	0.66	9.27	10.83	7.33
50	0.13	1.04	0.54	8.70	8.71	7.67
100	0.09	0.97	0.38	8.18	7.18	7.76

V prípade každej testovej štatistiky platí, že kritická hodnota je minimom kritického oboru a teda hypotézu zamietame pre veľké hodnoty testovej štatistiky. Ďalšie štyri body zahŕňajú proces výpočtu síl testov na hladine $\alpha = 0.05$.

4. Máme súbor kritických hodnôt pre každú z testových štatistík a pre každé n na hladine 5%.
5. Pre každé n a každú alternatívu vygenerujeme súbor 10 000 výberov z daného alternatívneho rozdelenia.
6. Vypočítame testovú štatistiku cez každý z výberov, ako výsledok nám vznikne súbor 10 000 testových štatistík pre každé n a každú alternatívu.
7. Silu vypočítame ako podiel testových štatistík, ktoré padnú do vypočítaného kritického oboru a všetkých testových štatistík.

4.2 Výsledky

Následuje 5 tabuliek pre každé $n \in \{10, 20, 30, 50, 100\}$ s vypočítanými hodnotami síl pre každú vybranú testovú štatistiku a každú vybranú alternatívu.

Tabuľka 4.2: Výsledky pre $n = 10$

Rozd.	KS	KL₁	KL₂	R₁	R_F	R₂
$N(0,1)$	0.03	0.17	0.19	0.01	0	0.05
$St(3)$	0.03	0.32	0.35	0.01	0.01	0.07
$Lo(0,1)$	0.03	0.25	0.27	0.01	0	0.06
$La(0,1)$	0.03	0.16	0.17	0.01	0	0.05
$Be(1,2)$	0.09	0.49	0.50	0.06	0.01	0.21
$U(0,1)$	0.08	0.62	0.65	0.07	0.01	0.24

Tabuľka 4.3: Výsledky pre $n = 20$

Rozd.	KS	KL₁	KL₂	R₁	R_F	R₂
$N(0,1)$	0.04	0.44	0.45	0.09	0.02	0.16
$St(3)$	0.05	0.81	0.88	0.25	0.03	0.39
$Lo(0,1)$	0.04	0.68	0.73	0.16	0.03	0.27
$La(0,1)$	0.03	0.41	0.46	0.07	0.02	0.13
$Be(1,2)$	0.33	0.97	0.95	0.76	0.12	0.85
$U(0,1)$	0.25	0.99	0.99	0.82	0.13	0.9

Tabuľka 4.4: Výsledky pre $n = 30$

Rozd.	KS	KL₁	KL₂	R₁	R_F	R₂
$N(0,1)$	0.05	0.67	0.64	0.28	0.06	0.33
$St(3)$	0.1	0.98	0.99	0.72	0.13	0.77
$Lo(0,1)$	0.07	0.91	0.94	0.52	0.09	0.58
$La(0,1)$	0.04	0.65	0.73	0.23	0.04	0.29
$Be(1,2)$	0.66	1	1	0.99	0.41	0.99
$U(0,1)$	0.57	1	1	1	0.43	1

Tabuľka 4.5: Výsledky pre $n = 50$

Rozd.	KS	KL₁	KL₂	R₁	R_F	R₂
$N(0,1)$	0.09	0.91	0.84	0.63	0.19	0.65
$St(3)$	0.24	1	1	0.99	0.41	0.99
$Lo(0,1)$	0.15	1	1	0.94	0.29	0.94
$La(0,1)$	0.06	0.94	0.97	0.65	0.09	0.68
$Be(1,2)$	0.97	1	1	1	0.83	1
$U(0,1)$	0.95	1	1	1	0.85	1

Tabuľka 4.6: Výsledky pre $n = 100$

Rozd.	KS	KL₁	KL₂	R₁	R_F	R₂
$N(0,1)$	0.24	0.99	0.96	0.94	0.51	0.94
$St(3)$	0.78	1	1	1	0.87	1
$Lo(0,1)$	0.51	1	1	1	0.73	1
$La(0,1)$	0.14	1	1	0.99	0.24	0.99
$Be(1,2)$	1	1	1	1	1	1
$U(0,1)$	1	1	1	1	1	1

Zhodnotenie

Z výsledkov simulácie vyplýva, že celkovo najväčšiu silu majú štatistiky KL_1 a KL_2 vytvorené z odhadu Shannonovej entropie pre KL-vzdialenosť, pričom najmä pri menšom rozsahu výberu mala väčšiu silu štatistika KL_2 . Z Rublíkových testových štatistík vzišla najlepšie nami upravená štatistika R_2 . Ako tiež vidíme, výmena fixných kvantilov v R_F za pohyblivé kvantily v R_1 a R_2 významne zvýšila silu.

4.3 Príklad

V nasledujúcom príklade aplikujeme testy dobrej zhody pre Cauchyho rozdelenie na reálne dáta. Cieľom príkladu je demonštrovať praktickú využiteľnosť testových štatistík a ukázať, či zvolené dáta zodpovedajú vlastnostiam Cauchyho rozdelenia.

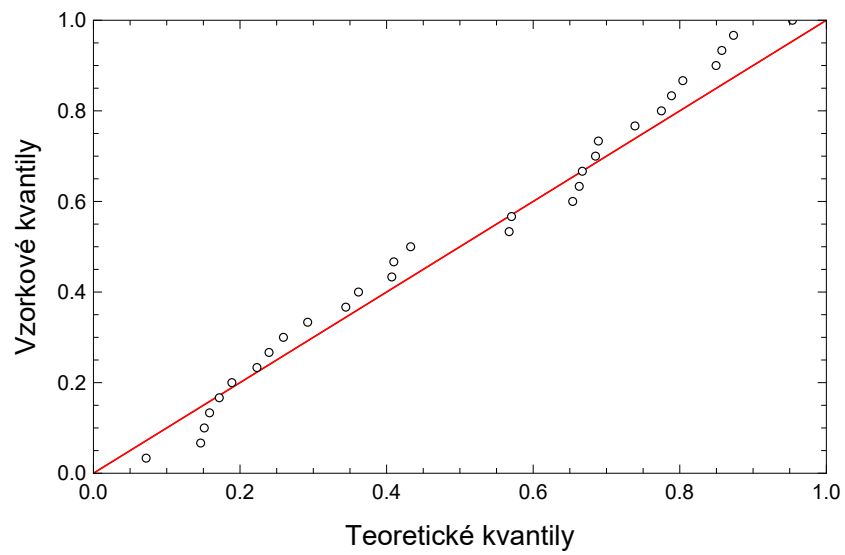
Konkrétne pôjde o denné logaritmické výnosy kryptomien Bitcoin, Ethereum a Ripple, počas prvých 30 dní roka 2025. Dáta pochádzajú z verejne dostupných zdrojov platformy www.coingecko.com. Výber kryptomien je motivovaný štúdiou Szczygielski a kol. (2020) ktorá zdôrazňuje vhodnosť použitia Cauchyho rozdelenia pri modelovaní ťažkých chvostov výnosov kryptomien.

Teoretické rozdelenie testových štatistík nemáme k dispozícii. A na rozhodnutie testu budú slúžiť len kritické hodnoty vypočítané v Tabulke 4.1. Na výpočet p-hodnoty by bola potrebné zachovať nasimulované rozdelenie testových štatistík. Keďže toto empirické rozdelenie nemáme k dispozícii a zároveň iba demonštrujeme použitie testových štatistík, nakoniec len zhodnotíme, či zamietame alebo nezamietame hypotézu, že dataset logaritmických výnosov daných kryptomien je z Cauchyho rozdelenia na hladine $\alpha = 0.05$.

	KS	KL1	KL2	Q1	T	Q2
Krit. hod.	0.29	1.29	0.85	10.96	15.90	5.71
Bitcoin	0.113	1.15	0.648	2.49	4.67	1.96
<i>Výsledok</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>
Ethereum	0.111	1.12	0.708	2.90	2.64	2.95
<i>Výsledok</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>
Ripple	0.113	1.15	0.648	2.49	4.67	1.96
<i>Výsledok</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>	<i>Nez.</i>

Tabuľka 4.7: Výsledky testových štatistík pre $n = 30$

Ako vidíme, žiadna testová štatistika nezamietla ani jednu hypotézu. Pre ilustráciu uvidieme Q-Q graf pre dáta z logaritmických výnosov Bitcoinu.



Obr. 4.2: Q-Q graf vzorkových kvantilov prvých 30 logaritmickejých výnosov Bitcoinu v roku 2025 oproti teoretickým kvantilom Cauchyho rozdelenia.

5. Dôkaz

Dôkaz Vety 3.1 (Rublák, 1999). Nech X_1, \dots, X_n je náhodný výber z $C(0,1)$. Dostávame z Vety 4.2.1 z Jurečková a Sen (1996)

$$\hat{\mu}_n^Q - m = \frac{1}{nf_0(m)} \sum_{i=1}^n \{F_0(m) - \mathbb{I}(X_i \leq m)\} + \mathcal{O}_P(n^{-1/2}), \quad m = 0. \quad (5.1)$$

Po dosadení dostávame výraz

$$\frac{\pi}{n} \sum_{i=1}^n \left\{ \frac{1}{2} - \mathbb{I}(X_i \leq 0) \right\} + \mathcal{O}_P(n^{-1/2}), \quad (5.2)$$

z tohoto dostávame

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\pi}{2} \text{sign}(X_i) \right\} + \mathcal{O}_P(n^{-1/2}). \quad (5.3)$$

Ďalej zo vzťahu (1.16) z (Rublák, 1999) platí

$$\hat{\sigma}_n^R - 1 = \frac{1}{n} \sum_{i=1}^n \psi(X_i) + \mathcal{O}_P(n^{-1}), \quad (5.4)$$

kde

$$\psi = -2 \cos(2 \arctan(x)). \quad (5.5)$$

Aplikáciou (5.1) a (5.4) dostávame

$$\hat{\theta}_n^R - \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{n} \sum_{j=1}^n \mathbf{1}(X_j) + \mathcal{O}_P(n^{-1/2}), \quad \mathbf{1}(x) = \begin{pmatrix} \frac{\pi}{2} \text{sign}(x) \\ -2 \cos(2 \arctan(x)) \end{pmatrix}. \quad (5.6)$$

Teda

$$E(\mathbf{1}(x)) = \mathbf{0}, \quad \mathbf{L} = \text{Var}(\mathbf{1}(x)) = \begin{pmatrix} \frac{\pi^2}{4} & 0 \\ 0 & 2 \end{pmatrix}. \quad (5.7)$$

Zo vzťahu (1.11) z Rublák (1999) platí

$$\hat{\sigma}_n^R > 0, \quad \hat{\sigma}_n^R(aX_1 + b, \dots, aX_n + b) = |a| \hat{\sigma}_n^R(X_1, \dots, X_n). \quad (5.8)$$

a tiež platí

$$\hat{\mu}_n(aZ_1 + b, \dots, aZ_n + b) = a\hat{\mu}_n(Z_1, \dots, Z_n) + b \quad (5.9)$$

pre $a > 0$. To znamená, že predpoklady Vety 2.2 z (Rublák, 1997) sú splnené. Táto veta hovorí o presnej nulovej distribúcií obdobného vektoru rozdielov ako je $\mathbf{\Delta}_n$, avšak s všeobecnými odhadmi parametrov $\hat{\sigma}$ a $\hat{\mu}$ ktoré spĺňajú podmienky spomenuté vyššie. Ďalej nám ukazuje, že obdobná štatistika T_n so spomenutými odhadmi má normálne rozdelenie, teda platí (3.3) a (3.4), pričom

$$\mathbf{\Sigma} = \mathbf{A} + \mathbf{D}\mathbf{\Psi}\mathbf{C}' + \mathbf{C}\mathbf{\Psi}'\mathbf{D} + \mathbf{D}\mathbf{\Psi}\mathbf{L}\mathbf{\Psi}'\mathbf{D} \quad (5.10)$$

kde

$$\mathbf{D} = \text{diag}(f_0(\xi_1), \dots, f_0(\xi_k)), \quad \xi_i = \xi_{p_i} \quad (5.11)$$

$$\mathbf{\Psi} = \begin{pmatrix} 1, & \xi_1 \\ \vdots & \\ 1, & \xi_k \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \text{cov}(\mathbb{I}_{(-\infty, \xi_1)}(x), \mathbf{I}(x, \theta)) \\ \vdots \\ \text{cov}(\mathbb{I}_{(-\infty, \xi_k)}(x), \mathbf{I}(x, \theta)) \end{pmatrix} \quad (5.12)$$

a j -ty riadok matice \mathbf{C} má tvar

$$\mathbf{C}_j = \left(-\frac{\pi}{2} \min\{p_j, 1 - p_j\}, \frac{\sin(2p_j\pi)}{\pi} \right). \quad (5.13)$$

Kombináciou (5.7) - (5.12) získame platnosť vzťahov (3.5) a (3.6). Regularita matice (3.5) vyplýva z III. časti Vety 2.1 z Rublík (1997) s pomocou Vety 2.2 z Rublík (1997). Posledný vzťah (3.8) vyplýva zo základnej vlastnosti mnohonásobného normálneho rozdelenia. \square

Záver

Táto bakalárska práca sa venovala testovaniu dobrej zhody dát s Cauchyho rozdelením, ktoré je špecifické najmä svojimi ťažkými chvostami a neexistenciou konečných momentov.

V prvej časti sme spracovali prehľad testových štatistík navrhovaných v literatúre, pričom sme sa zamerali na rôzne prístupy: testy založené na empirickej distribučnej funkcii, vierohodnostnom pomere, charakteristickej funkcii a entropii. Zdrojom boli práce Mahdizadeh a Zamanzade (2019) a Ebner a kol. (2022).

V ďalšej časti sme sa zamerali na testovú štatistiku predstavenú Rublík (2001), vrátane vysvetlenia jej konštrukcie a tvrdenia o jej vlastnostiach, pričom sme rozviedli dôkaz.

Ďalším prínosom práce je vlastná simulačná štúdia, ktorá porovnáva silu vybraných testov pri rôznych alternatívach. V rámci nej sme zároveň navrhli jednoduchú úpravu Rublíkovej testovej štatistiky – konkrétne sme vymenili pôvodný odhad škálového parametra za kvantilový odhad. Táto modifikácia sa ukázala ako efektívna a viedla k zvýšeniu sily testu.

Z výsledkov taktiež vyplynulo, že testy založené na Kullbackovej-Leiblerovej divergencii majú pri väčšine alternatív najväčšiu silu, čo je v súlade s poznatkami z literatúry.

Práca tak poskytuje prehľad dostupných metód pre testovanie zhody s Cauchyho rozdelením, porovnáva ich výkonnosť a zároveň prináša konkrétne vylepšenie jedného z existujúcich testov.

Seznam použité literatury

- D'AGOSTINO, R. B. (2017). *Goodness-of-fit-techniques*. Routledge.
- EBNER, B., EID, L. a KLAR, B. (2022). Cauchy or not Cauchy? New goodness-of-fit tests for the Cauchy distribution. *Statistical Papers*, pages 1–34.
- EBRAHIMI, N., PFLUGHOEFT, K. a SOOFI, E. S. (1994). Two measures of sample entropy. *Statistics & Probability Letters*, **20**(3), 225–234.
- GREENWOOD, P. E. a NIKULIN, M. S. (1996). *A Guide to Chi-squared Testing*. Wiley & Sons, New York.
- GÜRTLER, N. a HENZE, N. (2000). Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Annals of the Institute of Statistical Mathematics*, **52**, 267–286.
- HÁJEK, J. a ŠIDÁK, Z. (1967). Theory of Rank Tests, Academia Publ. *House, Prague*.
- JUREČKOVÁ, J. a SEN, P. K. (1996). *Robust statistical procedures: asymptotics and interrelations*, volume 311. John Wiley & Sons.
- MAHDIZADEH, M. a ZAMANZADE, E. (2019). Goodness-of-fit testing for the Cauchy distribution with application to financial modeling. *Journal of King Saud University-Science*, **31**(4), 1167–1174.
- RUBLÍK, F. (1997). A quantile goodness-of-fit test applicable to distributions with non-differentiable densities. *Kybernetika*, **33**(5), 505–524.
- RUBLÍK, F. (1999). A Goodness-of-Fit Test for Cauchy Distribution. *Tatra Mountains Mathematical Publications*, **17**, 71–81. Supported by Grant No. 1/4196/97 from VEGA.
- RUBLÍK, F. (2001). A quantile goodness-of-fit test for Cauchy distribution, based on extreme order statistics. *Applications of Mathematics*, **46**, 339–351.
- SZCZYGIELSKI, J. J., KARATHANASOPOULOS, A. a ZAREMBA, A. (2020). One shape fits all? A comprehensive examination of cryptocurrency return distributions. *Applied Economics Letters*, **27**(19), 1567–1573.
- VAŠÍČEK, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **38**(1), 54–59.
- ZHANG, J. (2002). Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(2), 281–294.