



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

## **BAKALÁŘSKÁ PRÁCE**

Tomáš Laube

# **Testy dobré shody s Poissonovým rozdělením založené na charakterizaci Steinova typu**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Šárka Hudecová, Ph.D.

Studijní program: Obecná matematika

Praha 2024

Prohlašuji, že jsem tuto bakalářskou práci vypracoval 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

Na tomto místě bych chtěl poděkovat vedoucí mé práce RNDr. Šárce Hudecové, Ph.D., za pomoc, vstřícnost a cenné rady při psaní této práce a také mé rodině za podporu během celého studia.

Název práce: Testy dobré shody s Poissonovým rozdělením založené na charakterizaci Steinova typu

Autor: Tomáš Laube

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Šárka Hudecová, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato bakalářská práce se zabývá testováním dobré shody s Poissonovým rozdělením metodou založenou na Steinově-Chenově identitě. Steinova-Chenova identita charakterizuje Poissonovo rozdělení na základě střední hodnoty transformace. V práci jsou představeny dva typy testových statistik a následně jsou odvozeny jejich asymptotické vlastnosti za nulové hypotézy. V závěrečné části jsou statistické testy porovnány v simulační studii.

Klíčová slova: Poissonovo rozdělení, Steinova-Chenova identita, testy dobré shody

Title: Goodness-of-fit tests for Poisson distribution based on Stein-type characterization

Author: Tomáš Laube

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Šárka Hudecová, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This bachelor's thesis deals with the goodness-of-fit tests for Poisson distribution based on the Stein-Chen identity. Stein-Chen identity characterizes the Poisson distribution on the basis of expected value of transformation. In this thesis, two types of statistics are presented. Their asymptotic properties are derived. In the last part, presented statistical tests are compared in a simulation study.

Keywords: Poisson distribution, Stein-Chen identity, goodness-of-fit tests

# Obsah

<b>Úvod</b>	<b>2</b>
<b>1 Základní definice a tvrzení</b>	<b>3</b>
1.1 Poissonovo rozdělení . . . . .	3
1.2 Steinova-Chenova identita . . . . .	3
1.3 Pomocné věty . . . . .	4
<b>2 Testy dobré shody</b>	<b>5</b>
2.1 Testová statistika jako podíl . . . . .	7
2.1.1 Asymptotické rozdělení pro funkce v mocninném tvaru . . . . .	8
2.1.2 Konstrukce statistického testu . . . . .	10
2.2 Testová statistika jako rozdíl . . . . .	13
<b>3 Simulační studie</b>	<b>16</b>
3.1 Hladina testů . . . . .	16
3.2 Síla testů . . . . .	19
<b>Závěr</b>	<b>22</b>
<b>Seznam použité literatury</b>	<b>23</b>
<b>Seznam tabulek</b>	<b>24</b>

# Úvod

Poissonovo rozdělení je základní rozdělení pro modelování počtů událostí za jednotku času. Tématem této práce je testování dobré shody s Poissonovým rozdělením. Zjednodušeně řečeno testujeme, zda náhodný výběr pochází z Poissonova rozdělení. Existuje celá řada takových testů. V této práci popíšeme testové statistiky založené na Steinově-Chenově identitě.

Hlavním zdrojem této práce je článek Weiß, Puig a Aleksandrov (2023). V článku je dokázána zmíněná Steinova-Chenova identita pro Poissonovo a binomické rozdělení. Tato identita charakterizuje rozdělení na základě střední hodnoty vhodné třídy transformací. Cílem práce tedy není pouze odvodit testovou statistiku, ale též najít vhodnou transformační funkci, pro kterou bude mít test dobré vlastnosti. V předlohovém článku je dále pro Poissonovo rozdělení sestrojena testová statistika v podílovém tvaru. V této práci kromě této podílové statistiky sestrojíme ještě statistiku rozdílovou. Pomocí centrální limitní věty odvodíme asymptotické rozdělení obou typů statistik za nulové hypotézy a v závěrečné simulační studii porovnáme vlastnosti testů pro praktické použití.

V první kapitole vyslovíme a dokážeme Steinovu-Chenovu identitu pro Poissonovo rozdělení a uvedeme základní definice a známá tvrzení.

Ve druhé kapitole z Steinovy-Chenovy identity odvodíme oba typy testových statistik a pomocí centrální limitní věty a delta-věty odvodíme jejich asymptotické rozdělení. Také zde uvádíme obecný postup pro odhad rozptylu testových statistik, který dále použijeme ke konstrukci kritického oboru. V závěrečné části této kapitoly se zaměříme na konkrétní třídu funkcí – funkce ve tvaru  $f(x) = s^{x-1}$ ,  $s \in (0,1)$ , souhrnně je budeme nazývat mocninné funkce. Pro mocninné funkce vyjádříme analytický tvar asymptotického rozptylu a pomocí tohoto vyjádření navrhneme alternativní způsob odhadu rozptylu.

V závěrečné kapitole uvádíme simulační studii. V první části zkoumáme statistické testy podle dodržování hladiny. Dále zde také porovnáme testy s mocninnými funkcemi dle volby odhadu rozptylu. Ve druhé podkapitole zkoumáme sílu testů proti různým alternativám.

Vlastním přínosem této práce je podrobné rozepsání odvození asymptotického rozdělení testových statistik a důkazu Steinovy-Chenovy identity. Dále pak alternativní odhad rozptylu pro testové statistiky s mocninnými funkcemi. Vlastním přínosem je též simulační studie v závěrečné kapitole.

# 1. Základní definice a tvrzení

## 1.1 Poissonovo rozdělení

V této části připomeneme definici Poissonova rozdělení a jeho základní vlastnosti.

**Definice 1.** *Náhodná veličina  $X$  má Poissonovo rozdělení s parametrem  $\lambda > 0$ , jestliže*

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \forall k \in \mathbb{N}_0.$$

Značíme  $X \sim \text{Po}(\lambda)$ .

Poissonovo rozdělení mají náhodné veličiny vyjadřující počet událostí za jednotku času. Poissonovým rozdělením bychom mohli modelovat například počet zákazníků obchodu za jeden den, počet pojistných událostí za měsíc nebo počet aut projíždějících ulicí Sokolovská za týden.

V dalších částech budeme potřebovat znalost střední hodnoty, rozptylu a vytvořující funkce Poissonova rozdělení. Nejprve spočteme vytvořující funkci. Nechť  $X \sim \text{Po}(\lambda)$  a  $s \in (0,1)$ , pak

$$P_X(s) = \sum_{n=0}^{\infty} s^n P(X = n) = \sum_{n=0}^{\infty} s^n \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda s)^n}{n!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}.$$

Se znalostí vytvořující funkce jednoduše spočteme střední hodnotu

$$E X = P'_X(1) = e^{\lambda} \lambda = \lambda$$

a rozptyl

$$\text{var}(X) = P''_X(1) + P'_X(1) - (P_X(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Vzorce jsou odvozeny například v Johnson, Kemp a Kotz (2005, strana 59).

## 1.2 Steinova-Chenova identita

Steinova-Chenova identita je hlavní teoretické tvrzení, o které se opírá tato práce. Charakterizuje Poissonovo rozdělení na základě střední hodnoty transformace. Větu s důkazem jsme převzali z článku Weiß a kol. (2023, strana 2) a pečlivě jsme zdůvodnili všechny kroky v důkazu.

**Věta 1** (Steinova-Chenova identita). *Nechť  $\lambda > 0$  a  $X$  je náhodná veličina. Pak  $X \sim \text{Po}(\lambda)$  právě tehdy, když pro každou omezenou funkci  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  platí*

$$E[Xf(X)] = \lambda E[f(X+1)]. \quad (1.1)$$

*Důkaz.* Necht  $X \sim \text{Po}(\lambda)$ . Pak z definice střední hodnoty dostáváme:

$$\begin{aligned} \mathbb{E}[Xf(X)] &= \sum_{n=1}^{\infty} nf(n) \frac{\lambda^n}{n!} e^{-\lambda} = \lambda \sum_{n=1}^{\infty} f(n) \frac{\lambda^{n-1}}{(n-1)!} e^{-\lambda} = \lambda \sum_{n=0}^{\infty} f(n+1) \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= \lambda \mathbb{E}[f(X+1)]. \end{aligned}$$

Obráceně, ať  $\epsilon > 0$ . Uvažujme funkce  $f_s : \mathbb{N}_0 \rightarrow \mathbb{R}$ ,  $f_s(x) = s^{x-1}$ , pro  $s \in [\epsilon, 1]$ . Případ  $s = 0$  musíme vyloučit, neboť pro  $x = 0$  a  $s = 0$  není výraz  $f_s(x)$  definován. Funkce  $f_s$  jsou omezené na množině  $\mathbb{N}_0$ , neboť  $\forall x \in \mathbb{N}_0, \forall s \in [\epsilon, 1] : f_s(x) \leq 1$ . Vezměme vytvořující funkci náhodné veličiny  $X$

$$P_X(s) = \mathbb{E}[s^X] = \mathbb{E}[f_s(X+1)]$$

a její derivaci

$$P'_X(s) = \mathbb{E}[Xs^{X-1}] = \mathbb{E}[Xf_s(X)].$$

Dosazením do předpokladu (1.1) dostáváme diferenciální rovnici

$$P'_X(s) = \lambda P_X(s).$$

Jejím jediným řešením je funkce  $P_X(s) = ke^{\lambda s}$ ,  $k \in \mathbb{R}$ . Z podmínky pro vytvořující funkci nutně  $\lim_{s \rightarrow 1^-} P_X(s) = 1$ . Tedy  $P_X(s) = e^{\lambda(s-1)}$ , což je přesně vytvořující funkce Poissonova rozdělení s parametrem  $\lambda$ .

□

### 1.3 Pomocné věty

V této kapitole uvedeme ještě 2 známá tvrzení, jejichž důkazy lze nalézt například v knize van der Vaart (1998, strana 16 a 26).

**Věta 2** (Mnohorozměrná Lindebergova centrální limitní věta). *Necht  $\mathbf{X}_1, \mathbf{X}_2, \dots$  jsou  $k$ -rozměrné, nezávislé, stejně rozdělené náhodné vektory. Dále necht střední hodnota  $\mathbb{E} \mathbf{X}_1 = \boldsymbol{\mu}$  a rozptylová matice  $\text{var} \mathbf{X}_1 = \boldsymbol{\Sigma}$  je konečná. Pak platí*

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i - \boldsymbol{\mu} \right) \xrightarrow[n \rightarrow \infty]{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

**Věta 3** ( $\Delta$ -věta). *Necht  $\{\mathbf{T}_n\}_{n=1}^{\infty}$  splňuje*

$$\sqrt{n} (\mathbf{T}_n - \boldsymbol{\mu}) \xrightarrow[n \rightarrow \infty]{d} N(\mathbf{0}, \boldsymbol{\Sigma})$$

pro nějaké  $\boldsymbol{\mu} \in \mathbb{R}^k$  a matici  $\boldsymbol{\Sigma}$ . Necht  $g : \mathbb{R}^k \rightarrow \mathbb{R}^p$  je funkce, která je spojitě diferencovatelná na okolí bodu  $\boldsymbol{\mu}$ . Označme  $\nabla g$  gradient funkce  $g$ . Pak platí

$$\sqrt{n} (g(\mathbf{T}_n) - g(\boldsymbol{\mu})) \xrightarrow[n \rightarrow \infty]{d} N(\mathbf{0}, \nabla g(\boldsymbol{\mu}) \boldsymbol{\Sigma} \nabla g(\boldsymbol{\mu})^T).$$



## 2. Testy dobré shody

V této části přiblížíme problém testování dobré shody. Pro test máme naporozovaná data  $X_1, X_2, \dots, X_n$ . Předpokládáme, že jde o náhodný výběr, tedy že pozorování jsou nezávislá a že pochází ze stejného rozdělení. V našem případě chceme testovat shodu s Poissonovým rozdělením, tedy se můžeme omezit na model všech diskrétních rozdělení, jejichž nosičem je množina  $\mathbb{N}_0$  (stejně jako u Poissonova rozdělení). Testovat budeme hypotézu  $H_0$  proti alternativě  $H_1$ , kde

$$\begin{aligned} H_0 &: \exists \lambda \in (0, \infty), \text{ že } X_i \sim \text{Po}(\lambda), \\ H_1 &: \nexists \lambda \in (0, \infty), \text{ že } X_i \sim \text{Po}(\lambda). \end{aligned} \quad (2.1)$$

Za platnosti nulové hypotézy má náhodný výběr Poissonovo rozdělení s parametrem  $\lambda$ . Zamítnutím nulové hypotézy prokazujeme, že náhodný výběr neodpovídá tomuto pravděpodobnostnímu rozdělení.

V další části za pomoci Steinovy-Chenovy identity sestrojíme testovou statistiku pro testování  $H_0$  a odvodíme její asymptotické rozdělení. Asymptotické rozdělení testové statistiky v podílovém tvaru je stručně odvozeno v článku Weiß a kol. (2023, strana 5). Zde vše pečlivě rozepíšeme a dále odvodíme alternativu v podobě rozdílové statistiky. Dále ještě spočteme přesný rozptyl testových statistik za nulové hypotézy pro konkrétní třídu funkcí.

Nechť je dána funkce  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$ , která splňuje předpoklady věty 1, tedy je omezená. Za platnosti nulové hypotézy platí rovnost (1.1). Můžeme ji upravit do tvaru rozdílu

$$\mathbb{E}[Xf(X)] - \lambda \mathbb{E}[f(X+1)] = 0, \quad (2.2)$$

nebo podílu

$$\frac{\mathbb{E}[Xf(X)]}{\lambda \mathbb{E}[f(X+1)]} = 1. \quad (2.3)$$

Zde navíc potřebujeme předpokládat, že  $\mathbb{E}[f(X+1)] \neq 0$ .

Označme

$$\mathbf{Z}_n = \begin{pmatrix} Z_{n,1} \\ Z_{n,2} \\ Z_{n,3} \end{pmatrix} = \begin{pmatrix} \bar{X}_n \\ \frac{1}{n} \sum_{i=1}^n X_i f(X_i) \\ \frac{1}{n} \sum_{i=1}^n f(X_i + 1) \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \mathbb{E} X \\ \mathbb{E}[Xf(X)] \\ \mathbb{E}[f(X+1)] \end{pmatrix}. \quad (2.4)$$

Nyní ukážeme asymptotické vlastnosti náhodného vektoru  $\mathbf{Z}_n$ . V celé kapitole budeme uvažovat  $X$  jako obecnou náhodnou veličinu, která je stejně rozdělená jako náhodný výběr  $X_1, \dots, X_n$ .

**Lemma 4.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z diskrétního rozdělení na  $\mathbb{N}_0$  s konečnou střední hodnotou a  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce. Pak  $\mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{P} \boldsymbol{\mu}$ .*

*Důkaz.* Použijeme zákon velkých čísel na jednotlivé složky vektoru  $\mathbf{Z}_n$ . Z předpokladu platí  $\mathbb{E} X < \infty$  a  $f(x)$  je omezená, tedy i  $\mathbb{E} |Xf(X)| < \infty$  a  $\mathbb{E} |f(X+1)| < \infty$ . Proto

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{P} \mathbb{E} X,$$

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n X_i f(X_i) &\xrightarrow[n \rightarrow \infty]{P} \mathbf{E}[X f(X)], \\ \frac{1}{n} \sum_{i=1}^n f(X_i + 1) &\xrightarrow[n \rightarrow \infty]{P} \mathbf{E}[f(X + 1)].\end{aligned}$$

Konvergence po složkách implikuje konvergenci vektoru, tedy  $\mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{P} \boldsymbol{\mu}$ . □

**Lemma 5.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z diskrétního rozdělení na  $\mathbb{N}_0$  s konečným druhým momentem a  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce. Pak platí*

$$\sqrt{n} (\mathbf{Z}_n - \boldsymbol{\mu}) \xrightarrow[n \rightarrow \infty]{d} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

kde  $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1,2,3}$  je varianční matice s hodnotami

$$\begin{aligned}\sigma_{11} &= \mathbf{E}[X^2] - \mu_1^2, \\ \sigma_{22} &= \mathbf{E}[(X f(X))^2] - \mu_2^2, \\ \sigma_{33} &= \mathbf{E}[(f(X + 1))^2] - \mu_3^2, \\ \sigma_{12} &= \sigma_{21} = \mathbf{E}[X^2 f(X)] - \mu_1 \mu_2, \\ \sigma_{13} &= \sigma_{31} = \mathbf{E}[X f(X + 1)] - \mu_1 \mu_3, \\ \sigma_{23} &= \sigma_{32} = \mathbf{E}[X f(X) f(X + 1)] - \mu_2 \mu_3.\end{aligned}$$

*Důkaz.* Plyne přímo z centrální limitní věty (věta 2). Vektor  $\mathbf{Z}_n$  je součtem nezávislých, stejně rozdělených vektorů:

$$\mathbf{Z}_n = \begin{pmatrix} \bar{X}_n \\ \frac{1}{n} \sum_{i=1}^n X_i f(X_i) \\ \frac{1}{n} \sum_{i=1}^n f(X_i + 1) \end{pmatrix} = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} X_i \\ X_i f(X_i) \\ f(X_i + 1) \end{pmatrix}.$$

Spočteme prvky varianční matice vektoru

$$\begin{pmatrix} X_i \\ X_i f(X_i) \\ f(X_i + 1) \end{pmatrix}$$

a ověříme, že je konečná:

$$\begin{aligned}\sigma_{11} &= \text{var } X = \mathbf{E}[X^2] - \mu_1^2, \\ \sigma_{22} &= \text{var}(X f(X)) = \mathbf{E}[(X f(X))^2] - \mu_2^2, \\ \sigma_{33} &= \text{var}(f(X + 1)) = \mathbf{E}[(f(X + 1))^2] - \mu_3^2, \\ \sigma_{12} &= \sigma_{21} = \text{cov}(X, X f(X)) = \mathbf{E}[X^2 f(X)] - \mu_1 \mu_2, \\ \sigma_{13} &= \sigma_{31} = \text{cov}(X, f(X + 1)) = \mathbf{E}[X f(X + 1)] - \mu_1 \mu_3, \\ \sigma_{23} &= \sigma_{32} = \text{cov}(X f(X), f(X + 1)) = \mathbf{E}[X f(X) f(X + 1)] - \mu_2 \mu_3.\end{aligned}$$

V důkazu lemmatu 4 jsme diskutovali konečnost vektoru  $\boldsymbol{\mu}$ . Dále z předpokladu konečného druhého momentu a omezenosti funkce  $f$  plyne konečnost všech ostatních entit v matici  $\boldsymbol{\Sigma}$ . Ověřili jsme předpoklady centrální limitní věty a tedy

$$\sqrt{n} (\mathbf{Z}_n - \boldsymbol{\mu}) \xrightarrow[n \rightarrow \infty]{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

□

## 2.1 Testová statistika jako podíl

Zaměříme se nejprve na tvar (2.3). Naším cílem je nyní odvodit testovou statistiku a její asymptotické rozdělení. Využijeme vlastnosti, že výběrový průměr je nestranným a konzistentním odhadem střední hodnoty.

Pomocí značení (2.4) položíme statistiku

$$T_n = \frac{Z_{n,2}}{Z_{n,1}Z_{n,3}}.$$

Nyní odvodíme asymptotické vlastnosti statistiky  $T_n$ .

**Lemma 6.** *Nechť  $X_1, \dots, X_n$  je náhodný vektor s diskrétním rozdělením na  $\mathbb{N}_0$  s konečnou nenulovou střední hodnotou a  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce splňující  $E[f(X+1)] \neq 0$ . Pak*

$$T_n \xrightarrow[n \rightarrow \infty]{P} \frac{\mu_2}{\mu_1\mu_3}.$$

*Důkaz.* Z lemmatu 4 máme  $\mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{P} \boldsymbol{\mu}$ . Užitím věty o spojitě transformaci s funkcí  $g(x,y,z) = y/(xz)$  dostáváme

$$T_n \xrightarrow[n \rightarrow \infty]{P} \frac{\mu_2}{\mu_1\mu_3}.$$

□

**Věta 7.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z diskrétního rozdělení na  $\mathbb{N}_0$  s konečným druhým momentem a  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce splňující  $E[f(X+1)] \neq 0$ . Pak*

$$\sqrt{n} \left( T_n - \frac{\mu_2}{\mu_1\mu_3} \right) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_T^2),$$

kde

$$\sigma_T^2 = \left( \frac{\mu_2}{\mu_1\mu_3} \right)^2 \left[ \frac{\sigma_{11}}{\mu_1^2} - 2 \frac{\sigma_{12}}{\mu_1\mu_2} + 2 \frac{\sigma_{13}}{\mu_1\mu_3} + \frac{\sigma_{22}}{\mu_2^2} - 2 \frac{\sigma_{23}}{\mu_2\mu_3} + \frac{\sigma_{33}}{\mu_3^2} \right]$$

a  $\sigma_{i,j}$  jsou prvky matice  $\boldsymbol{\Sigma}$  z lemmatu 5.

*Důkaz.* Z lemmatu 5 máme

$$\sqrt{n} (\mathbf{Z}_n - \boldsymbol{\mu}) \xrightarrow[n \rightarrow \infty]{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

Dále použijeme  $\Delta$ -větu (věta 3) s funkcí  $g(x,y,z) = y/(xz)$ . Gradient funkce  $g$  je

$$\nabla g(x,y,z) = \begin{pmatrix} -y & 1 & -y \\ x^2z & xz & xz^2 \end{pmatrix}.$$

Asymptotický rozptyl potom vychází

$$\begin{aligned} \sigma_T^2 &= \nabla g(\mu_1, \mu_2, \mu_3) \boldsymbol{\Sigma} \nabla g(\mu_1, \mu_2, \mu_3)^T \\ &= \begin{pmatrix} -\mu_2 & 1 & -\mu_2 \\ \mu_1^2\mu_3 & \mu_1\mu_3 & \mu_1\mu_3^2 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \begin{pmatrix} -\mu_2 \\ \mu_1^2\mu_3 \\ 1 \\ \mu_1\mu_3 \\ -\mu_2 \\ \mu_1\mu_3^2 \end{pmatrix} \\ &= \frac{1}{\mu_1^2\mu_3^2} \left[ \frac{\mu_2^2\sigma_{11}}{\mu_1^2} - 2 \frac{\mu_2\sigma_{12}}{\mu_1} + 2 \frac{\mu_2^2\sigma_{13}}{\mu_1\mu_3} + \sigma_{22} - 2 \frac{\mu_2\sigma_{23}}{\mu_3} + \frac{\mu_2^2\sigma_{33}}{\mu_3^2} \right] \\ &= \left( \frac{\mu_2}{\mu_1\mu_3} \right)^2 \left[ \frac{\sigma_{11}}{\mu_1^2} - 2 \frac{\sigma_{12}}{\mu_1\mu_2} + 2 \frac{\sigma_{13}}{\mu_1\mu_3} + \frac{\sigma_{22}}{\mu_2^2} - 2 \frac{\sigma_{23}}{\mu_2\mu_3} + \frac{\sigma_{33}}{\mu_3^2} \right]. \end{aligned}$$

Dostáváme tedy

$$\sqrt{n} \left( T_n - \frac{\mu_2}{\mu_1 \mu_3} \right) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_T^2).$$

□

Nyní můžeme právě dokázanou větu zesílit předpokladem Poissonova rozdělení (tj. za platnosti nulové hypotézy).

**Věta 8.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z  $Po(\lambda)$  a  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce splňující  $E[f(X+1)] \neq 0$ . Pak*

$$\sqrt{n} (T_n - 1) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_T^2),$$

kde

$$\sigma_T^2 = \frac{\sigma_{11}}{\mu_1^2} - 2 \frac{\sigma_{12}}{\mu_1 \mu_2} + 2 \frac{\sigma_{13}}{\mu_1 \mu_3} + \frac{\sigma_{22}}{\mu_2^2} - 2 \frac{\sigma_{23}}{\mu_2 \mu_3} + \frac{\sigma_{33}}{\mu_3^2} \quad (2.5)$$

a  $\sigma_{i,j}$  jsou prvky matice  $\Sigma$  z lemmatu 5.

*Důkaz.* Z věty 7 máme

$$\sqrt{n} \left( T_n - \frac{\mu_2}{\mu_1 \mu_3} \right) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_T^2),$$

kde

$$\sigma_T^2 = \left( \frac{\mu_2}{\mu_1 \mu_3} \right)^2 \left[ \frac{\sigma_{11}}{\mu_1^2} - 2 \frac{\sigma_{12}}{\mu_1 \mu_2} + 2 \frac{\sigma_{13}}{\mu_1 \mu_3} + \frac{\sigma_{22}}{\mu_2^2} - 2 \frac{\sigma_{23}}{\mu_2 \mu_3} + \frac{\sigma_{33}}{\mu_3^2} \right].$$

Úpravou rovnosti (1.1) z věty 1 platí  $\frac{\mu_2}{\mu_1 \mu_3} = 1$ . Dosazením do výsledku výše dostáváme

$$\sigma_T^2 = \frac{\sigma_{11}}{\mu_1^2} - 2 \frac{\sigma_{12}}{\mu_1 \mu_2} + 2 \frac{\sigma_{13}}{\mu_1 \mu_3} + \frac{\sigma_{22}}{\mu_2^2} - 2 \frac{\sigma_{23}}{\mu_2 \mu_3} + \frac{\sigma_{33}}{\mu_3^2}.$$

□

Nyní máme v rukou nástroj pro výpočet asymptotického rozdělení testové statistiky. Pro obecnou funkci  $f$  ale může být problém přesně spočítat střední hodnoty  $E[Xf(X)]$ ,  $E[f(X+1)]$  a tím pádem též asymptotický rozptyl. Zde se potom musí použít numerický software, nebo lze střední hodnoty konzistentně odhadnout. V následujícím případě ukážeme konkrétní třídu funkcí, pro které lze asymptotický rozptyl dále upravovat, a pro tyto funkce tak najdeme různé konzistentní odhady rozptylu.

### 2.1.1 Asymptotické rozdělení pro funkce v mocninném tvaru

Nechť  $X \sim Po(\lambda)$  je náhodná veličina,  $s \in (0,1)$  a  $f_s : \mathbb{N}_0 \rightarrow \mathbb{R}$  funkce s předpisem  $f_s(x) = s^{x-1}$ . Podle věty 8 spočteme asymptotické rozdělení statistiky  $T_n$ .

Pro náhodnou veličinu  $Y$  budeme značit její vytvořující funkci jako  $P_Y(s)$ . Nejprve spočteme derivace vytvořujících funkcí, které dále použijeme k výpočtu.

$$\begin{aligned} P_X(s) &= \mathbb{E}[s^X], \\ P'_X(s) &= \mathbb{E}[Xs^{X-1}], \\ P_{X+1}(s) &= \mathbb{E}[s^{X+1}], \\ P'_{X+1}(s) &= \mathbb{E}[(X+1)s^X], \\ P''_{X+1}(s) &= \mathbb{E}[(X^2+X)s^{X-1}], \\ P_{2X}(s) &= \mathbb{E}[s^{2X}], \\ P'_{2X}(s) &= \mathbb{E}[2Xs^{2X-1}], \\ P''_{2X}(s) &= \mathbb{E}[(4X^2-2X)s^{2X}]. \end{aligned}$$

Nyní pomocí vytvořujících funkcí šikovně vyjádříme střední hodnoty, kterými jsou vyjádřeny prvky matice  $\Sigma$ :

$$\begin{aligned} \mathbb{E}[Xf(X)] &= \mathbb{E}[Xs^{X-1}] = P'_X(s), \\ \mathbb{E}[f(X+1)] &= \mathbb{E}[s^X] = P_X(s), \\ \mathbb{E}[(Xf(X))^2] &= \mathbb{E}[X^2s^{2(X-1)}] = \frac{\mathbb{E}[2Xs^{2(X-1)}]}{4} + \frac{\mathbb{E}[(4X^2-2X)s^{2(X-1)}]}{4} \\ &= \frac{P''_{2X}(s)}{4} + \frac{\mathbb{E}[2Xs^{2X-1}]}{4s} = \frac{P''_{2X}(s)}{4} + \frac{P'_{2X}(s)}{4s}, \\ \mathbb{E}[(f(X+1))^2] &= \mathbb{E}[s^{2X}] = P_{2X}(s), \\ \mathbb{E}[X^2f(X)] &= \mathbb{E}[X^2s^{X-1}] = \mathbb{E}[(X^2+X)s^{X-1}] - \mathbb{E}[Xs^{X-1}] \\ &= P''_{X+1}(s) - P'_X(s), \\ \mathbb{E}[Xf(X+1)] &= \mathbb{E}[Xs^X] = \mathbb{E}[(X+1)s^X] - \mathbb{E}[s^X] = P'_{X+1}(s) - P_X(s), \\ \mathbb{E}[Xf(X)f(X+1)] &= \mathbb{E}[Xs^{X-1}s^X] = \frac{1}{2}\mathbb{E}[2Xs^{2X-1}] = \frac{P'_{2X}(s)}{2}. \end{aligned}$$

Nyní použijeme předpoklad Poissonova rozdělení k přesnému výpočtu vytvořujících funkcí. Již dříve jsme ukázali, že

$$P_X(s) = e^{\lambda(s-1)}.$$

Dále počítáme

$$\begin{aligned} P'_X(s) &= \lambda e^{\lambda(s-1)}, \\ P_{X+1}(s) &= \sum_{n=0}^{\infty} s^n \mathbb{P}(X+1=n) = \sum_{n=0}^{\infty} s^n \mathbb{P}(X=n-1) = \sum_{n=-1}^{\infty} s^{n+1} \mathbb{P}(X=n) \\ &= 0 + \sum_{n=0}^{\infty} s^{n+1} \mathbb{P}(X=n) = \sum_{n=0}^{\infty} s^{n+1} \frac{\lambda^n}{n!} e^{-\lambda} = se^{-\lambda} \sum_{n=0}^{\infty} \frac{(s\lambda)^n}{n!} = se^{\lambda(s-1)}, \\ P'_{X+1}(s) &= e^{\lambda(s-1)} + s\lambda e^{\lambda(s-1)} = e^{\lambda(s-1)}(1+s\lambda), \\ P''_{X+1}(s) &= \lambda e^{\lambda(s-1)} + \lambda e^{\lambda(s-1)}(1+s\lambda) = \lambda e^{\lambda(s-1)}(2+s\lambda), \end{aligned}$$

$$\begin{aligned}
P_{2X}(s) &= \sum_{n=0}^{\infty} s^n \mathbf{P}(2X = n) = \sum_{n=0}^{\infty} s^n \mathbf{P}(X = \frac{n}{2}) = \sum_{k=0}^{\infty} s^{2k} \mathbf{P}(X = k) \\
&= \sum_{k=0}^{\infty} s^{2k} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s^2 \lambda)^k}{k!} = e^{-\lambda} e^{s^2 \lambda} = e^{\lambda(s^2-1)}, \\
P'_{2X}(s) &= e^{\lambda(s^2-1)} 2\lambda s, \\
P''_{2X}(s) &= e^{\lambda(s^2-1)} (2\lambda s)^2 + e^{\lambda(s^2-1)} 2\lambda = 2\lambda e^{\lambda(s^2-1)} (2\lambda s^2 + 1).
\end{aligned}$$

Teď už můžeme vyjádřit všechny prvky matice  $\Sigma$  a vektoru  $\mu$ :

$$\begin{aligned}
\sigma_{11} &= \text{var}(X) = \lambda, \\
\sigma_{22} &= \mathbf{E}[(Xf(X))^2] - (\mathbf{E}[Xf(X)])^2 = \lambda e^{\lambda(s^2-1)}(1 + \lambda s^2) - \lambda^2 e^{2\lambda(s-1)}, \\
\sigma_{33} &= \mathbf{E}[(f(X+1))^2] - (\mathbf{E}[f(X+1)])^2 = e^{\lambda(s^2-1)} - e^{2\lambda(s-1)}, \\
\sigma_{12} &= \mathbf{E}[X^2 f(X)] - \mathbf{E}[X] \mathbf{E}[Xf(X)] = \lambda e^{\lambda(s-1)}(1 + s\lambda) - \lambda^2 e^{\lambda(s-1)} \\
&= \lambda e^{\lambda(s-1)}(1 + s\lambda - \lambda), \\
\sigma_{13} &= \mathbf{E}[Xf(X+1)] - \mathbf{E}[X] \mathbf{E}[f(X+1)] = s\lambda e^{\lambda(s-1)} - \lambda e^{\lambda(s-1)} \\
&= \lambda e^{\lambda(s-1)}(s - 1), \\
\sigma_{23} &= \mathbf{E}[Xf(X)f(X+1)] - \mathbf{E}[Xf(X)] \mathbf{E}[f(X+1)] = \lambda s e^{\lambda(s^2-1)} - \lambda e^{2\lambda(s-1)}, \\
\mu_1 &= \mathbf{E}[X] = \lambda, \\
\mu_2 &= \mathbf{E}[Xf(X)] = \lambda e^{\lambda(s-1)}, \\
\mu_3 &= \mathbf{E}[f(X+1)] = e^{\lambda(s-1)}.
\end{aligned} \tag{2.6}$$

Nyní můžeme použít větu 8 k spočtení asymptotického rozdělení  $T_n$ :

$$\sqrt{n}(T_n - 1) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_T^2),$$

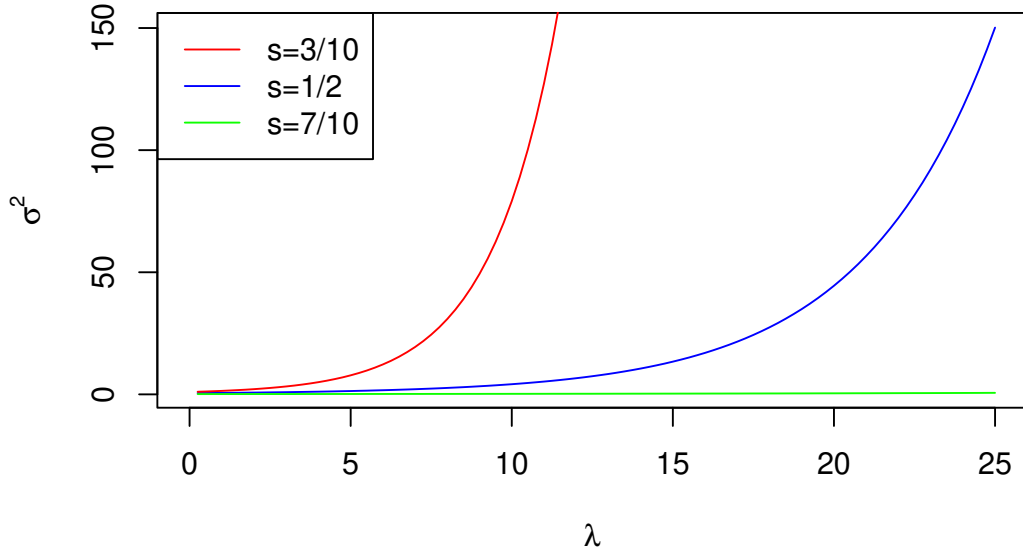
kde

$$\begin{aligned}
\sigma_T^2 &= \frac{\sigma_{11}}{\mu_1^2} - 2 \frac{\sigma_{12}}{\mu_1 \mu_2} + 2 \frac{\sigma_{13}}{\mu_1 \mu_3} + \frac{\sigma_{22}}{\mu_2^2} - 2 \frac{\sigma_{23}}{\mu_2 \mu_3} + \frac{\sigma_{33}}{\mu_3^2} \\
&= \frac{\lambda}{\lambda^2} - 2 \frac{\lambda e^{\lambda(s-1)}(1 + s\lambda - \lambda)}{\lambda^2 e^{\lambda(s-1)}} + 2 \frac{\lambda e^{\lambda(s-1)}(s - 1)}{\lambda e^{\lambda(s-1)}} + \frac{e^{\lambda(s^2-1)} - e^{2\lambda(s-1)}}{e^{2\lambda(s-1)}} \\
&\quad - 2 \frac{\lambda s e^{\lambda(s^2-1)} - \lambda e^{2\lambda(s-1)}}{\lambda e^{\lambda(s-1)} e^{\lambda(s-1)}} + \frac{\lambda e^{\lambda(s^2-1)}(1 + s^2 \lambda) - \lambda^2 e^{2\lambda(s-1)}}{\lambda^2 e^{2\lambda(s-1)}} \\
&= \frac{1 - 2 - 2\lambda(s - 1) + 2\lambda(s - 1)}{\lambda} + \frac{e^{\lambda(s^2-1)}}{e^{2\lambda(s-1)}} \left( \frac{1 + \lambda s^2}{\lambda} - 2s + 1 \right) \\
&= -\frac{1}{\lambda} + \frac{e^{\lambda(s^2-1)}}{e^{2\lambda(s-1)}} \frac{1 + \lambda(s - 1)^2}{\lambda} = \frac{1}{\lambda} \left( e^{\lambda(s-1)^2} - 1 \right) + e^{\lambda(s-1)^2} (s - 1)^2.
\end{aligned}$$

V grafu 2.1 jsou zaneseny hodnoty  $\sigma_T^2$  v závislosti na  $\lambda$  pro fixní  $s = 0,3; 0,5$  a  $0,7$ .

## 2.1.2 Konstrukce statistického testu

Nyní můžeme zkonstruovat statistický test hypotézy  $H_0$ . K tomu potřebujeme konzistentní odhad pro  $\sigma_T^2$  z věty 8. Uvažujme znovu obecnou (omezenou)



Obrázek 2.1: Graf závislosti  $\sigma_T^2$  na  $\lambda$  pro funkci  $f_s$ .

funkci  $f$ . Rozptyl  $\sigma_T^2$  je pouze kombinací součtů, součinů a podílů různých středních hodnot, které můžeme konzistentně odhadnout výběrovými průměry. Prvky  $\mu_1, \mu_2, \mu_3$  jsme již odhadli v lemmatu 4. Zbylé prvky odhadneme následovně:

$$\hat{\sigma}_{12} = \frac{1}{n} \sum_{i=1}^n X_i^2 f(X_i) - Z_{n,1} Z_{n,2} \xrightarrow[n \rightarrow \infty]{P} \mathbb{E}[X^2 f(X)] - \mathbb{E} X \mathbb{E}[X f(X)] = \sigma_{12},$$

$$\hat{\sigma}_{13} = \frac{1}{n} \sum_{i=1}^n X_i f(X_i + 1) - Z_{n,1} Z_{n,3} \xrightarrow[n \rightarrow \infty]{P} \mathbb{E}[X f(X + 1)] - \mathbb{E} X \mathbb{E} f(X + 1) = \sigma_{13},$$

$$\hat{\sigma}_{23} = \frac{1}{n} \sum_{i=1}^n X_i f(X_i) f(X_i + 1) - Z_{n,2} Z_{n,3} \xrightarrow[n \rightarrow \infty]{P} \mathbb{E}[X f(X) f(X + 1)] - \mathbb{E}[X f(X)] \mathbb{E}[f(X + 1)] = \sigma_{23},$$

$$\hat{\sigma}_{11} = \frac{1}{n-1} \sum_{i=1}^n (X_i - Z_{n,1})^2 \xrightarrow[n \rightarrow \infty]{P} \text{var } X = \sigma_{11},$$

$$\hat{\sigma}_{22} = \frac{1}{n-1} \sum_{i=1}^n (X_i f(X_i) - Z_{n,2})^2 \xrightarrow[n \rightarrow \infty]{P} \text{var } X f(X) = \sigma_{22},$$

$$\hat{\sigma}_{33} = \frac{1}{n-1} \sum_{i=1}^n (f(X_i + 1) - Z_{n,3})^2 \xrightarrow[n \rightarrow \infty]{P} \text{var } f(X + 1) = \sigma_{33}.$$

(2.7)

„Dosazením“ těchto odhadů do předpisu pro  $\sigma_T^2$  (2.5) dostáváme, použitím věty o spojitě transformaci, požadovaný konzistentní odhad rozptylu:

$$\hat{\sigma}_n^2 = \frac{\hat{\sigma}_{11}}{\hat{\mu}_1^2} - 2 \frac{\hat{\sigma}_{12}}{\hat{\mu}_1 \hat{\mu}_2} + 2 \frac{\hat{\sigma}_{13}}{\hat{\mu}_1 \hat{\mu}_3} + \frac{\hat{\sigma}_{22}}{\hat{\mu}_2^2} - 2 \frac{\hat{\sigma}_{23}}{\hat{\mu}_2 \hat{\mu}_3} + \frac{\hat{\sigma}_{33}}{\hat{\mu}_3^2} \xrightarrow[n \rightarrow \infty]{P} \sigma_T^2. \quad (2.8)$$

Pro funkce  $f_s(x) = s^{x-1}$  jsme v části 2.1.1 rozptyl  $\sigma_T^2$  vyjádřili jako funkci proměnných  $s$  a  $\lambda$ . Zde je jediné neznámé  $\lambda$ , které odhadneme výběrovým průměrem. Z věty o spojitě transformaci pak

$$\hat{\sigma}_n^2 = \frac{1}{\bar{X}_n} \left( e^{\bar{X}_n(s-1)^2} - 1 \right) + e^{\bar{X}_n(s-1)^2} (s-1)^2 \xrightarrow[n \rightarrow \infty]{P} \sigma_T^2. \quad (2.9)$$

**Věta 9.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr s rozdělením  $Po(\lambda)$ ,  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce a  $\hat{\sigma}_n^2$  konzistentní odhad rozptylu  $\sigma_T^2$ . Pak*

$$\sqrt{n} \left( \frac{T_n - 1}{\sqrt{\hat{\sigma}_n^2}} \right) \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

*Důkaz.* Z předpokladu platí

$$\hat{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{P} \sigma_T^2.$$

Úpravou věty 8 máme

$$\sqrt{n} \left( \frac{T_n - 1}{\sqrt{\sigma_T^2}} \right) \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

Z Cramérový-Sluckého věty pak dostáváme

$$\sqrt{n} \left( \frac{T_n - 1}{\sqrt{\hat{\sigma}_n^2}} \right) \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

□

Teď již můžeme konstruovat kritický obor statistického testu. Označme  $u_\alpha$   $\alpha$ -kvantil  $N(0,1)$  rozdělení. Z věty 9 plyne

$$P \left( -u_{1-\alpha/2} < \sqrt{n} \frac{T_n - 1}{\sqrt{\hat{\sigma}_n^2}} < u_{1-\alpha/2} \right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha,$$

tedy

$$P \left( 1 - \frac{\sqrt{\hat{\sigma}_n^2} u_{1-\alpha/2}}{\sqrt{n}} < T_n < 1 + \frac{\sqrt{\hat{\sigma}_n^2} u_{1-\alpha/2}}{\sqrt{n}} \right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha.$$

Odtud dostáváme kritický obor s asymptotickou hladinou  $\alpha$

$$C(\alpha) = \left( -\infty, 1 - \frac{\sqrt{\hat{\sigma}_n^2} u_{1-\alpha/2}}{\sqrt{n}} \right) \cup \left( 1 + \frac{\sqrt{\hat{\sigma}_n^2} u_{1-\alpha/2}}{\sqrt{n}}, \infty \right).$$

Hypotézu  $H_0$  tedy zamítáme právě tehdy, když  $T_n \in C(\alpha)$ .



## 2.2 Testová statistika jako rozdíl

V této kapitole se zaměříme na tvar rozdílu (2.2). Analogicky ke kapitole 2.1 odvodíme testovou statistiku, její asymptotické rozdělení a na závěr kritický obor statistického testu. Položíme statistiku

$$V_n = Z_{n,2} - Z_{n,1}Z_{n,3}.$$

Nejprve obecně odvodíme asymptotické vlastnosti.

**Lemma 10.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z diskrétního rozdělení na  $\mathbb{N}_0$  s konečnou střední hodnotou a  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce. Pak*

$$V_n \xrightarrow[n \rightarrow \infty]{P} \mu_2 - \mu_1\mu_3.$$

*Důkaz.* Z lemmatu 4 máme  $\mathbf{Z}_n \xrightarrow[n \rightarrow \infty]{P} \boldsymbol{\mu}$ . Užitím věty o spojitě transformaci s funkcí  $g(x, y, z) = y - xz$  na vektor  $\mathbf{Z}_n$  dostáváme

$$Z_{n,2} - Z_{n,1}Z_{n,3} \xrightarrow[n \rightarrow \infty]{P} \mu_2 - \mu_1\mu_3.$$

□

**Věta 11.** *Nechť  $X_1, \dots, X_n$  je náhodný výběr z diskrétního rozdělení na  $\mathbb{N}_0$  s konečným druhým momentem a  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce. Pak*

$$\sqrt{n}(V_n - (\mu_2 - \mu_1\mu_3)) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_V^2),$$

kde

$$\sigma_V^2 = \mu_3^2\sigma_{11} - 2\mu_3\sigma_{12} + 2\mu_1\mu_3\sigma_{13} + \sigma_{22} - 2\mu_1\sigma_{23} + \mu_1^2\sigma_{33} \quad (2.10)$$

a prvky  $\mu_i, \sigma_{i,j}$  jsme definovali v (2.4) a v lemmatu 5.

*Důkaz.* Z lemmatu 5 máme

$$\sqrt{n}(\mathbf{Z}_n - \boldsymbol{\mu}) \xrightarrow[n \rightarrow \infty]{d} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

Použijeme  $\Delta$ -větu (věta 3) s funkcí  $g(x, y, z) = y - xz$ . Gradient funkce  $g$  je

$$\nabla g(x, y, z) = \begin{pmatrix} -z & 1 & -x \end{pmatrix}.$$

Asymptotický rozptyl vychází

$$\begin{aligned} \sigma_V^2 &= \begin{pmatrix} -\mu_3 & 1 & -\mu_1 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \begin{pmatrix} -\mu_3 \\ 1 \\ -\mu_1 \end{pmatrix} \\ &= \mu_3^2\sigma_{11} - 2\mu_3\sigma_{12} + 2\mu_1\mu_3\sigma_{13} + \sigma_{22} - 2\mu_1\sigma_{23} + \mu_1^2\sigma_{33}. \end{aligned}$$

Dostáváme tedy

$$\sqrt{n}(V_n - (\mu_2 - \mu_1\mu_3)) \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_V^2).$$

□

K předpokladům věty 11 přidáme ještě předpoklad Poissonova rozdělení, budeme tak mít asymptotické rozdělení testové statistiky za nulové hypotézy.

**Věta 12.** Necht  $X_1, \dots, X_n$  je náhodný výběr z  $Po(\lambda)$  a  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce. Pak

$$\sqrt{n}V_n \xrightarrow[n \rightarrow \infty]{d} N(0, \sigma_V^2),$$

kde  $\sigma_V^2$  je z (2.10).

*Důkaz.* Náhodné veličiny  $X_1, \dots, X_n$  jsou z Poissonova rozdělení, tedy platí rovnost (2.2), neboli  $\mu_2 - \mu_1\mu_3 = 0$ . Dosazením do věty 11 dostáváme požadovanou konvergenci. □

Ke konstrukci kritického oboru potřebujeme konzistentní odhad  $\tilde{\sigma}_n^2$  rozptylu  $\sigma_V^2$ . Použijeme stejné odhady (2.7) jako v kapitole 2.1.2 s tím rozdílem, že nyní „dosadíme“ do předpisu (2.10):

$$\tilde{\sigma}_n^2 = \hat{\mu}_3^2 \hat{\sigma}_{11} - 2\hat{\mu}_3 \hat{\sigma}_{12} + 2\hat{\mu}_1 \hat{\mu}_3 \hat{\sigma}_{13} + \hat{\sigma}_{22} - 2\hat{\mu}_1 \hat{\sigma}_{23} + \hat{\mu}_1^2 \hat{\sigma}_{33}. \quad (2.11)$$

Z věty o spojitě transformaci platí

$$\tilde{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{P} \sigma_V^2$$

**Věta 13.** Necht  $X_1, \dots, X_n$  je náhodný výběr z  $Po(\lambda)$ ,  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  je omezená funkce a  $\tilde{\sigma}_n^2$  je konzistentní odhad rozptylu  $\sigma_V^2$ . Pak

$$\sqrt{n} \frac{V_n}{\sqrt{\tilde{\sigma}_n^2}} \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

*Důkaz.* Postupujeme zcela analogicky k důkazu věty 9. Z předpokladu plyne  $\tilde{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{P} \sigma_V^2$ . Úpravou věty 12 máme

$$\sqrt{n} \frac{V_n}{\sqrt{\sigma_V^2}} \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

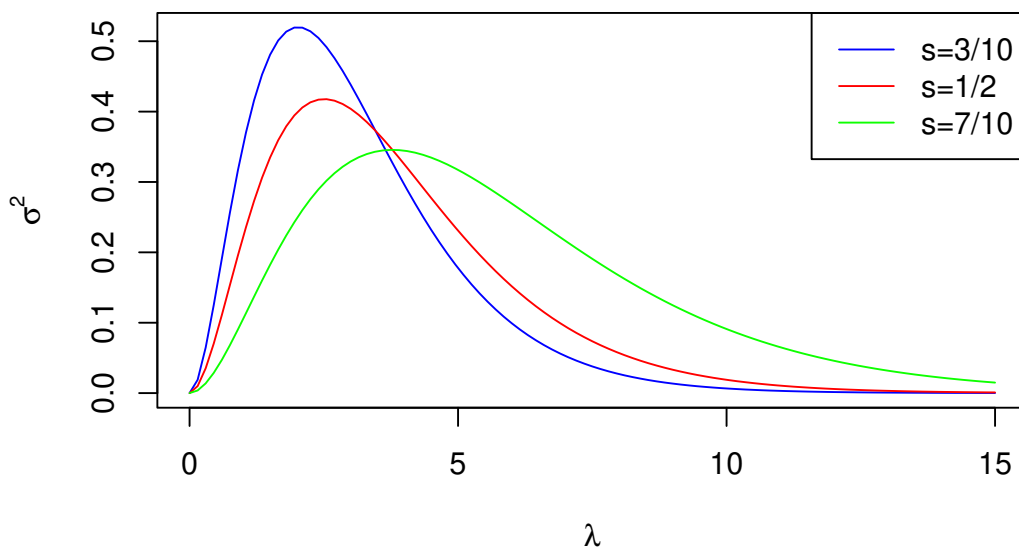
Z Cramérovoy-Sluckého věty dostáváme

$$\sqrt{n} \frac{V_n}{\sqrt{\tilde{\sigma}_n^2}} \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

□

Nyní již samotný kritický obor s asymptotickou hladinou  $\alpha$ . Z platnosti

$$\begin{aligned} & \mathbf{P} \left( -u_{1-\alpha/2} < \sqrt{n} \frac{V_n}{\sqrt{\tilde{\sigma}_n^2}} < u_{1-\alpha/2} \right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha, \\ & \mathbf{P} \left( -\frac{u_{1-\alpha/2} \sqrt{\tilde{\sigma}_n^2}}{\sqrt{n}} < V_n < \frac{u_{1-\alpha/2} \sqrt{\tilde{\sigma}_n^2}}{\sqrt{n}} \right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha \end{aligned}$$



Obrázek 2.2: Graf závislosti  $\sigma_V^2$  na  $\lambda$  pro funkci  $f_s$ .

dostáváme

$$C(\alpha) = \left( -\infty, -\frac{u_{1-\alpha/2}\sqrt{\tilde{\sigma}_n^2}}{\sqrt{n}} \right) \cup \left( \frac{u_{1-\alpha/2}\sqrt{\tilde{\sigma}_n^2}}{\sqrt{n}}, \infty \right).$$

Hypotézu  $H_0$  zamítáme právě tehdy, když  $V_n \in C(\alpha)$ .

Na závěr uvedeme ještě výpočet  $\sigma_V^2$  pro  $f$  v mocninném tvaru. Postup je stejný jako v kapitole 2.1.1, mění se pouze dosazení do vzorce. Aplikujeme hodnoty z (2.6) na vzorec z věty 11. Hodnota  $\sigma_V^2$  pak pro pevné  $s \in (0,1)$  a funkci  $f_s(x) = s^{x-1}$  vychází

$$\begin{aligned} \sigma_V^2 &= \mu_3^2\sigma_{11} - 2\mu_3\sigma_{12} + 2\mu_1\mu_3\sigma_{13} + \sigma_{22} - 2\mu_1\sigma_{23} + \mu_1^2\sigma_{33} \\ &= e^{2\lambda(s-1)}\lambda - 2e^{\lambda(s-1)}(\lambda e^{\lambda(s-1)})(1 + s\lambda - \lambda) + 2e^{\lambda(s-1)}\lambda e^{\lambda(s-1)}\lambda(s-1) \\ &\quad - 2\lambda^2(se^{\lambda(s^2-1)} - e^{2\lambda(s-1)}) + \lambda e^{\lambda(s^2-1)}(1 + \lambda s^2) - \lambda^2 e^{2\lambda(s-1)} \\ &\quad - 2\lambda(\lambda s e^{\lambda(s^2-1)} - \lambda e^{2\lambda(s-1)}) + \lambda^2(e^{\lambda(s^2-1)} - e^{2\lambda(s-1)}) \\ &= e^{2\lambda(s-1)}(-\lambda - 2\lambda(s\lambda - \lambda) + 2\lambda^2(s-1)) + e^{\lambda(s^2-1)}(-2\lambda^2s + \lambda + \lambda^2s^2 + \lambda^2) \\ &= -\lambda e^{2\lambda(s-1)} + \lambda e^{\lambda(s^2-1)}(1 + \lambda(s-1)^2). \end{aligned}$$

V grafu (2.2) jsou zaneseny hodnoty  $\sigma_V^2$  v závislosti na  $\lambda$  pro pevné  $s = 1/4$ ,  $1/2$  a  $3/4$ .

Využitím výpočtu výše můžeme pro mocninné funkce odhadnout rozptyl  $\tilde{\sigma}_n^2$  jako

$$\tilde{\sigma}_n^2 = -\bar{X}_n e^{2\bar{X}_n(s-1)} + \bar{X}_n e^{\bar{X}_n(s^2-1)}(1 + \bar{X}_n(s-1)^2). \quad (2.12)$$

Z věty o spojitě transformaci plyne

$$\tilde{\sigma}_n^2 \xrightarrow[n \rightarrow \infty]{P} \sigma_V^2.$$

## 3. Simulační studie

V předchozí kapitole jsme odvodili několik testů hypotézy  $H_0$  (2.1). V této kapitole provedeme simulace, abychom získali vhled do chování testů pro konečné  $n$ . Jako referenční test budeme používat test založený na indexu disperze, který je popsán například v Johnson a kol. (2005, strana 179). Testová statistika je

$$D_n = \sqrt{\frac{n}{2}} \left( \frac{S_n^2}{\bar{X}_n} - 1 \right)$$

a za nulové hypotézy platí

$$D_n \xrightarrow[n \rightarrow \infty]{d} N(0,1).$$

Simulace budeme provádět v programu R Core Team (2023). Implementaci testů jsme prováděli samostatně a nepoužívali jsme žádné balíčky.

### 3.1 Hladina testů

Všechny námi popsané testy jsou asymptotické. Nejprve se proto podíváme, jak dodržují hladinu pro různé rozsahy výběru. Zajímá nás poměr počtu zamítnutí platné nulové hypotézy k celkovému počtu opakování.

Zvolme pevné  $\lambda$  a rozsah výběru  $n$ . Nagenerujeme náhodný výběr  $X_1, \dots, X_n$  z rozdělení  $\text{Po}(\lambda)$  a provedeme na něj zvolený test. Takto nagenerujeme 1000  $n$ -tic a spočteme počet zamítnutí nulové hypotézy, označme ho  $N$ . Odhad hladiny testu je pak

$$\frac{N}{1000}.$$

Nejprve se zaměříme na testy s obecnou funkcí  $f$ . Jediný předpoklad, který musíme dodržet, je omezenost  $f$  na  $\mathbb{N}_0$ . Zvolme funkce

$$\begin{aligned} f_1(x) &= (x^2 + 1)^{-1}, \\ f_2(x) &= \cos(x), \\ f_3(x) &= e^{-x}. \end{aligned}$$

Simulaci provedeme postupně s rozsahy výběru 20, 50, 100, 250, 500 a 2000. Parametr  $\lambda$  budeme volit 1, 3 a 5.

Označme testy odvozené v kapitole 2 následovně:  $T_{i,n}$  test založený na podílové statistice s funkcí  $f_i$ .  $V_{i,n}$  test založený na rozdílové statistice s funkcí  $f_i$ ,  $i \in \{1, 2, 3\}$ . Všechny testy uvažujeme na hladině  $\alpha = 0,05$ .

Z tabulek 3.1–3.3 vidíme, že pro rostoucí  $n$  se pozorovaná hladina blíží hodnotě 0,05. Pro nízké rozsahy výběru a volby  $\lambda = 3$  a 5 empirická hladina výrazně převyšuje cílovou hodnotu 0,05. Můžeme říct, že pro  $\lambda \leq 3$  je dostatečný rozsah výběru 250. Pro vyšší hodnoty je potřeba brát  $n$  ještě vyšší. Pro  $\lambda = 5$  z tabulky 3.3 vidíme, že pro funkce  $f_1$  a  $f_3$  dává i rozsah  $n = 500$  stále příliš vysoké hodnoty hladiny. Pro rozsah  $n = 2000$  všechny testy odpovídají zvolené hladině  $\alpha$ .

$n$	$T_{1,n}$	$V_{1,n}$	$T_{2,n}$	$V_{2,n}$	$T_{3,n}$	$V_{3,n}$	$D_n$
20	0,091	0,053	0,115	0,066	0,094	0,056	0,042
50	0,058	0,049	0,095	0,049	0,058	0,050	0,044
100	0,057	0,042	0,077	0,045	0,058	0,042	0,038
250	0,057	0,053	0,053	0,043	0,056	0,054	0,053
500	0,046	0,046	0,052	0,043	0,045	0,046	0,051
2000	0,049	0,051	0,044	0,043	0,049	0,050	0,051

Tabulka 3.1: Odhad hladiny pro  $\lambda = 1$

$n$	$T_{1,n}$	$V_{1,n}$	$T_{2,n}$	$V_{2,n}$	$T_{3,n}$	$V_{3,n}$	$D_n$
20	0,303	0,297	0,079	0,069	0,307	0,289	0,050
50	0,116	0,119	0,055	0,040	0,117	0,117	0,050
100	0,089	0,090	0,072	0,066	0,093	0,087	0,048
250	0,072	0,064	0,058	0,061	0,070	0,064	0,039
500	0,053	0,053	0,048	0,039	0,055	0,055	0,050
2000	0,057	0,056	0,047	0,049	0,059	0,058	0,051

Tabulka 3.2: Odhad hladiny pro  $\lambda = 3$

$n$	$T_{1,n}$	$V_{1,n}$	$T_{2,n}$	$V_{2,n}$	$T_{3,n}$	$V_{3,n}$	$D_n$
20	0,315	0,291	0,037	0,080	0,408	0,359	0,052
50	0,243	0,238	0,023	0,048	0,256	0,262	0,054
100	0,253	0,255	0,042	0,071	0,265	0,278	0,061
250	0,137	0,137	0,027	0,045	0,140	0,142	0,036
500	0,095	0,095	0,048	0,047	0,101	0,102	0,055
2000	0,063	0,062	0,044	0,041	0,056	0,056	0,054

Tabulka 3.3: Odhad hladiny pro  $\lambda = 5$

Dále provedeme simulace s funkcemi v mocninném tvaru. Za  $s$  budeme volit  $3/10$ ,  $1/2$  a  $7/10$ , tedy funkce v testových statistikách budou

$$f_4(x) = \left(\frac{3}{10}\right)^{x-1},$$

$$f_5(x) = \left(\frac{1}{2}\right)^{x-1},$$

$$f_6(x) = \left(\frac{7}{10}\right)^{x-1}.$$

Značení pro testové statistiky je stejné jako výše. Budeme uvažovat odhad rozptylu metodou odvozenou v kapitole 2.1.1 využívající vytvořující funkce, tj. vyjádření (2.9) pro statistiku  $T$ , resp. (2.12) pro statistiku  $V$ . Všechny testy opět uvažujeme na hladině  $\alpha = 0,05$ .

Z tabulek 3.4–3.6 vidíme, že testy s mocninnými funkcemi dodržují hladinu až na výjimky dobře. Navíc pro rostoucí  $\lambda$  nepozorujeme výrazné zhoršení pozorované hladiny. Dále vidíme, že podílové a rozdílové statistiky dodržují hladinu srovnatelně. Zaměříme-li se na jednotlivé řádky tabulek, vidíme, že test  $T_{4,n}$  dodržuje hladinu hůře, než testy  $T_{5,n}$  a  $T_{6,n}$ .

Celkově lze říct, že testy založené na podílové statistice můžeme označit jako o trochu konzervativnější. Dále testy založené na mocninné funkci jsou z hlediska

$n$	$T_{4,n}$	$V_{4,n}$	$T_{5,n}$	$V_{5,n}$	$T_{6,n}$	$V_{6,n}$	$D_n$
20	0,068	0,050	0,055	0,052	0,045	0,042	0,042
50	0,060	0,051	0,052	0,051	0,050	0,050	0,044
100	0,048	0,042	0,042	0,041	0,037	0,040	0,038
250	0,053	0,049	0,052	0,050	0,047	0,047	0,053
500	0,050	0,041	0,044	0,044	0,040	0,038	0,051
2000	0,050	0,049	0,047	0,047	0,048	0,048	0,051

Tabulka 3.4: Odhad hladiny pro  $\lambda = 1$  a funkce v mocninném tvaru

$n$	$T_{4,n}$	$V_{4,n}$	$T_{5,n}$	$V_{5,n}$	$T_{6,n}$	$V_{6,n}$	$D_n$
20	0,108	0,038	0,063	0,036	0,045	0,030	0,050
50	0,079	0,043	0,061	0,052	0,059	0,049	0,050
100	0,064	0,041	0,051	0,042	0,056	0,055	0,048
250	0,056	0,043	0,041	0,037	0,035	0,035	0,039
500	0,049	0,053	0,044	0,048	0,052	0,052	0,050
2000	0,054	0,050	0,053	0,052	0,062	0,059	0,051

Tabulka 3.5: Odhad hladiny pro  $\lambda = 3$  a funkce v mocninném tvaru

$n$	$T_{4,n}$	$V_{4,n}$	$T_{5,n}$	$V_{5,n}$	$T_{6,n}$	$V_{6,n}$	$D_n$
20	0,033	0,057	0,028	0,050	0,045	0,033	0,052
50	0,047	0,056	0,051	0,045	0,050	0,048	0,054
100	0,047	0,037	0,051	0,039	0,059	0,054	0,061
250	0,052	0,042	0,046	0,044	0,034	0,038	0,036
500	0,055	0,048	0,058	0,056	0,049	0,049	0,055
2000	0,047	0,051	0,047	0,045	0,049	0,047	0,054

Tabulka 3.6: Odhad hladiny pro  $\lambda = 5$  a funkce v mocninném tvaru

dodržování hladiny lepší než testy s obecnými funkcemi. To může být způsobeno i tím, že u obecných testů je potřeba odhadnout více entit, tedy bez dostatečného rozsahu výběru dochází k větším nepřesnostem než u testů s mocninnými funkcemi, u kterých odhadujeme pouze parametr  $\lambda$ . Z testů odvozených v této práci můžeme jako nejlepší z hlediska dodržování hladiny označit testy  $V_{5,n}$ ,  $T_{6,n}$  a  $V_{6,n}$ .

V této části ještě krátce srovnáme hladinu testů založených na funkci  $f_6$ , změníme-li odhad rozptylu. Označme  $T'_{6,n}$  test založený na podílové statistice s funkcí  $f_6$ , kde pro odhad rozptylu použijeme (2.8), tedy postup využitý pro obecné funkce. Analogicky označme  $V'_{6,n}$  test založený na rozdílové statistice s funkcí  $f_6$  a s odhadem rozptylu z (2.11).

Z tabulky 3.7 vidíme, že testy, ve kterých odhadujeme rozptyl obecnou metodou, mají pro malé rozsahy značně vyšší pozorovanou hladinu. Je to způsobeno zřejmě tím, jak jsme již zmínili v odstavci výše, že v obecném případě odhadujeme více entit, tedy konvergence probíhá pomaleji. Pro testy založené na mocninných funkcích tedy můžeme doporučit využít odhad rozptylu založený na vytvořujících funkcích, tj. (2.9) pro podílové statistiky a (2.12) pro rozdílové statistiky.

n	$T_{6,n}$	$T'_{6,n}$	$V_{6,n}$	$V'_{6,n}$
20	0,045	0,150	0,033	0,138
50	0,050	0,100	0,048	0,097
100	0,059	0,107	0,054	0,102
250	0,034	0,058	0,038	0,053
500	0,049	0,060	0,049	0,060
2000	0,049	0,054	0,047	0,054

Tabulka 3.7: Srovnání hladiny testů podle různých odhadů rozptylu pro  $\lambda = 5$ .

## 3.2 Síla testů

Nyní se zaměříme na sílu jednotlivých testů proti různým alternativám. Síla testu za dané alternativy je pravděpodobnost, že zamítneme (neplatnou) nulovou hypotézu. Postupovat budeme stejně jako v kapitole výše, pouze budeme data generovat z jiného rozdělení než z Poissonova.

Jako alternativu budeme volit negativně binomické rozdělení s různými parametry. Pro náhodnou veličinu  $X$  z negativně binomického rozdělení s parametry  $p, r$ , kde  $p > 0$  a  $r > 0$  platí

$$P(X = x) = \binom{r+x-1}{r-1} \left(\frac{1}{p+1}\right)^r \left(\frac{p}{p+1}\right)^x, \quad x \in \mathbb{N}_0.$$

Pro střední hodnotu a rozptyl platí  $E X = rp$ ,  $\text{var } X = rp(1+p)$  (Johnson a kol. (2005), strana 208). V dalším textu budeme používat alternativní definici, kterou z definice výše dostaneme substitucí  $p = \mu/r$ . Protože  $p > 0$  a  $r > 0$ , pak nutně  $\mu > 0$ . Pro náhodnou veličinu  $X \sim \text{NB}(\mu, r)$  dostáváme parametrizaci

$$P(X = x) = \binom{r+x-1}{r-1} \left(\frac{r}{\mu+r}\right)^r \left(\frac{\mu}{\mu+r}\right)^x, \quad x \in \mathbb{N}_0,$$

a pro střední hodnotu a rozptyl platí

$$E X = \mu, \quad \text{var } X = \mu + \frac{\mu^2}{r}.$$

Je dobré si uvědomit, že negativně binomické rozdělení má stejný nosič jako Poissonova rozdělení. Dále pro náhodnou veličinu  $X_r \sim \text{NB}(\mu, r)$  platí

$$\lim_{r \rightarrow \infty} \text{var } X_r = \mu = E X_r,$$

což je vlastnost charakterizující Poissonovo rozdělení. Lze též dokázat, že pro rozdělení  $\text{NB}(\mu, r)$  platí, že čím větší je parametr  $r$ , tím více se toto rozdělení blíží k rozdělení  $\text{Po}(\mu)$ .

Zaměříme se nejprve na testy s obecnou funkcí ( $T_1$  až  $V_3$ ). Z tabulek 3.8–3.11 vidíme, že největší sílu proti vybraným alternativám má test  $D_n$ . Pro malá  $n$  je rozdíl v síle testu  $D_n$  oproti ostatním testům velmi výrazný. V tabulce 3.10 vidíme, že síla testu  $T_{2,n}$  proti alternativě  $\text{NB}(5,3)$  s rostoucím  $n$  klesá k 0, tedy tento test není konzistentní, a proto ho rozhodně nemůžeme doporučit k testování dobré shody s Poissonovým rozdělením. Způsobeno je to zřejmě tím, že funkce  $\cos$  je periodická s oborem hodnot  $[-1,1]$ . V odhadu  $\hat{\sigma}_n^2$  pak jsou složky  $\hat{\mu}_2$  a  $\hat{\mu}_3$  blízké 0, což způsobuje, že je  $\hat{\sigma}_n^2$  velké.

Nyní se zaměříme na testy s mocninnými funkcemi. Používali jsme stejný odhad rozptylu (2.9), resp. (2.12) jako v předchozí kapitole. Pohledem do tabulek 3.12–3.15 vidíme, že síla všech testů proti všem alternativám se s rostoucím rozsahem výběru  $n$  blíží k 1. Můžeme tedy tyto testy považovat za konzistentní. Pro všechny testy (s výjimkou  $T_{6,20}$ ) a všechny zkoumané alternativy je síla rozdílového testu vyšší než síla podílového testu, což bychom před provedením simulace čekali spíše obráceně. Porovnáme-li jednotlivé testy podle parametru  $s$  mocninné funkce, vidíme, že s rostoucím  $s$  roste síla proti vybraným alternativám. Z toho můžeme usoudit, že nejvhodnější k praktickému použití je volit parametr  $s$  co nejblíže k 1. Zvláštní je téměř nulová síla testu  $T_{4,n}$  pro  $n \leq 50$  proti NB(5,10). To může být způsobeno velkým rozptylem statistiky  $T_{4,n}$  pro  $\lambda \geq 5$  (vidno z obrázku 2.1) v kombinaci s tím, že rozdělení NB(5,10) a Po(5) jsou si blízké. Při porovnání síly testu  $D_n$  s testy s mocninnými funkcemi pozorujeme, že ve všech případech je síla testu  $D_n$  viditelně vyšší než u ostatních testů.

Celkově můžeme říct, že ze sledovaných testů je pro testování dobré shody s Poissonovým rozdělením nejvhodnější test  $D_n$ . Z testů odvozených v této práci se jako nejlepší jeví test  $V_{6,n}$ .

$n$	$T_{1,n}$	$V_{1,n}$	$T_{2,n}$	$V_{2,n}$	$T_{3,n}$	$V_{3,n}$	$D_n$
20	0,024	0,079	0,216	0,062	0,021	0,087	0,217
50	0,074	0,185	0,294	0,116	0,055	0,187	0,361
100	0,227	0,341	0,397	0,161	0,211	0,342	0,571
250	0,665	0,727	0,599	0,350	0,656	0,730	0,870
500	0,920	0,939	0,759	0,588	0,922	0,941	0,987
2000	1	1	0,999	0,996	1	1	1

Tabulka 3.8: Odhad síly testů proti alternativě NB(1,3).

$n$	$T_{1,n}$	$V_{1,n}$	$T_{2,n}$	$V_{2,n}$	$T_{3,n}$	$V_{3,n}$	$D_n$
20	0,049	0,052	0,155	0,053	0,047	0,054	0,094
50	0,049	0,065	0,148	0,056	0,043	0,069	0,091
100	0,046	0,075	0,150	0,057	0,044	0,074	0,144
250	0,092	0,125	0,193	0,086	0,083	0,132	0,245
500	0,206	0,244	0,253	0,163	0,201	0,244	0,357
2000	0,686	0,697	0,522	0,419	0,688	0,702	0,862

Tabulka 3.9: Odhad síly testů proti alternativě NB(1,10).

$n$	$T_{1,n}$	$V_{1,n}$	$T_{2,n}$	$V_{2,n}$	$T_{3,n}$	$V_{3,n}$	$D_n$
20	0,024	0,102	0,044	0,054	0,019	0,086	0,882
50	0,486	0,633	0,054	0,068	0,201	0,459	0,998
100	0,961	0,975	0,037	0,102	0,854	0,905	1
250	1	1	0,023	0,171	1	1	1
500	1	1	0,010	0,301	1	1	1
2000	1	1	0	0,815	1	1	1

Tabulka 3.10: Odhad síly testů proti alternativě NB(5,3).



$n$	$T_{1,n}$	$V_{1,n}$	$T_{2,n}$	$V_{2,n}$	$T_{3,n}$	$V_{3,n}$	$D_n$
20	0,057	0,049	0,036	0,056	0,133	0,099	0,363
50	0,035	0,052	0,032	0,057	0,043	0,056	0,571
100	0,110	0,137	0,025	0,054	0,048	0,085	0,849
250	0,597	0,624	0,034	0,079	0,362	0,406	0,995
500	0,944	0,949	0,028	0,103	0,800	0,813	1
2000	1	1	0,032	0,309	1	1	1

Tabulka 3.11: Odhad síly testů proti alternativě NB(5,10).

$n$	$T_{4,n}$	$V_{4,n}$	$T_{5,n}$	$V_{5,n}$	$T_{6,n}$	$V_{6,n}$	$D_n$
20	0,063	0,129	0,112	0,144	0,155	0,161	0,217
50	0,166	0,234	0,237	0,272	0,285	0,287	0,361
100	0,328	0,399	0,441	0,460	0,504	0,511	0,571
250	0,726	0,754	0,802	0,818	0,855	0,857	0,870
500	0,933	0,944	0,971	0,974	0,985	0,986	0,987
2000	1	1	1	1	1	1	1

Tabulka 3.12: Odhad síly testů s mocninnými funkcemi proti NB(1,3).

$n$	$T_{4,n}$	$V_{4,n}$	$T_{5,n}$	$V_{5,n}$	$T_{6,n}$	$V_{6,n}$	$D_n$
20	0,055	0,063	0,058	0,064	0,071	0,069	0,094
50	0,061	0,070	0,067	0,070	0,067	0,070	0,091
100	0,067	0,079	0,088	0,097	0,110	0,112	0,144
250	0,118	0,144	0,155	0,164	0,200	0,203	0,245
500	0,228	0,250	0,290	0,302	0,322	0,327	0,357
2000	0,684	0,694	0,769	0,775	0,829	0,829	0,862

Tabulka 3.13: Odhad síly testů s mocninnými funkcemi proti NB(1,10).

$n$	$T_{4,n}$	$V_{4,n}$	$T_{5,n}$	$V_{5,n}$	$T_{6,n}$	$V_{6,n}$	$D_n$
20	0,001	0,561	0,311	0,682	0,748	0,811	0,882
50	0,034	0,823	0,803	0,936	0,980	0,986	0,998
100	0,446	0,972	0,993	0,998	1	1	1
250	0,984	1	1	1	1	1	1
500	1	1	1	1	1	1	1
2000	1	1	1	1	1	1	1

Tabulka 3.14: Odhad síly testů s mocninnými funkcemi proti NB(5,3).

$n$	$T_{4,n}$	$V_{4,n}$	$T_{5,n}$	$V_{5,n}$	$T_{6,n}$	$V_{6,n}$	$D_n$
20	0,001	0,168	0,037	0,187	0,159	0,225	0,363
50	0,006	0,244	0,157	0,339	0,387	0,436	0,571
100	0,027	0,356	0,362	0,535	0,704	0,744	0,849
250	0,204	0,569	0,769	0,852	0,966	0,971	0,995
500	0,566	0,826	0,979	0,986	0,999	0,999	1
2000	0,999	1	1	1	1	1	1

Tabulka 3.15: Odhad síly testů s mocninnými funkcemi proti NB(5,10).

# Závěr

V této práci jsme se zabývali testováním dobré shody s Poissonovým rozdělením založeným na Steinově-Chenově identitě. V první kapitole jsme uvedli základní definice a tvrzení. Také jsme zde dokázali Steinovu-Chenovu identitu. Ve druhé kapitole jsme sestrojili testové statistiky a odvodili jsme jejich asymptotická rozdělení. Dále jsme přesněji vyjádřili rozptyl testových statistik s mocninnými funkcemi, což nám umožňuje konstruovat přesnější testy. Ve třetí kapitole jsme uvedli simulační studii.

Statistických testů zabývajících se problémem testování dobré shody s Poissonovým rozdělením existuje celá řada. Nejpoužívanější z nich je test založený na indexu disperze, který jsme použili jako referenční test v simulační studii v kapitole 3. Testů založených na Steinově-Chenově identitě lze sestřit mnoho. V simulační studii jsme viděli, že volba funkce  $f$  výrazně ovlivňuje vlastnosti testu. Pro funkci  $f(x) = \cos(x)$  dokonce příslušný podílový test není konzistentní. Naopak pro volbu  $f(x) = (0,7)^{x-1}$  je hladina rozdílového testu i síla proti uvedeným alternativám blízko hodnotám referenčního testu. Ze všech zkoumaných testů má nejlepší vlastnosti test založený na indexu disperze. Z dalších testů bychom dle simulační studie doporučili test  $V_{n,6}$ .

Za svůj vlastní přínos považuji hlavně výpočet rozptylu testových statistik za nulové hypotézy popsany v kapitole 2.1.1 a celou simulační studii, zejména pak srovnání podílových a rozdílových statistik. Vlastním přínosem je též podrobné odvození asymptotického rozdělení testových statistik.

# Seznam použité literatury

JOHNSON, N. L., KEMP, A. W. a KOTZ, S. (2005). *Univariate Discrete Distribution*. 3. vydání. Wiley-interscience. ISBN 0-471-27246-9.

R CORE TEAM (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

VAN DER VAART, A. W. (1998). *Asymptotic statistics*. Cambridge University Press. ISBN 0-521-49603-9.

WEISS, C. H., PUIG, P. a ALEKSANDROV, B. (2023). Optimal Stein-type goodness-of-fit tests for count data. *Biometrical Journal*, **65**, 2200073. URL <https://doi.org/10.1002/bimj.202200073>.

# Seznam tabulek

3.1	Odhad hladiny pro $\lambda = 1$ . . . . .	17
3.2	Odhad hladiny pro $\lambda = 3$ . . . . .	17
3.3	Odhad hladiny pro $\lambda = 5$ . . . . .	17
3.4	Odhad hladiny pro $\lambda = 1$ a funkce v mocninném tvaru . . . . .	18
3.5	Odhad hladiny pro $\lambda = 3$ a funkce v mocninném tvaru . . . . .	18
3.6	Odhad hladiny pro $\lambda = 5$ a funkce v mocninném tvaru . . . . .	18
3.7	Srovnání hladiny testů podle různých odhadů rozptylu pro $\lambda = 5$ .	19
3.8	Odhad síly testů proti alternativě <b>NB(1,3)</b> . . . . .	20
3.9	Odhad síly testů proti alternativě <b>NB(1,10)</b> . . . . .	20
3.10	Odhad síly testů proti alternativě <b>NB(5,3)</b> . . . . .	20
3.11	Odhad síly testů proti alternativě <b>NB(5,10)</b> . . . . .	21
3.12	Odhad síly testů s mocninnými funkcemi proti <b>NB(1,3)</b> . . . . .	21
3.13	Odhad síly testů s mocninnými funkcemi proti <b>NB(1,10)</b> . . . . .	21
3.14	Odhad síly testů s mocninnými funkcemi proti <b>NB(5,3)</b> . . . . .	21
3.15	Odhad síly testů s mocninnými funkcemi proti <b>NB(5,10)</b> . . . . .	21