MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Filip Trusina

Testy Poissonova rozdělení

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce:	doc. RNDr. Zbyněk Pawlas, Ph.D.
Studijní program:	Matematika
Studijní obor:	Obecná matematika

Praha 2019

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

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

V Praze dne 16. 5. 2019 Podpis autora Trusina Filip

Chtěl bych zde poděkovat svému školiteli Zbyňku Pawlasovi za ochotu, vstřícnost a odborné i lidské zázemí při vedení mé bakalářské práce.

Název práce: Testy Poissonova rozdělení

Autor: Filip Trusina

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Zbyněk Pawlas, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V této práci se zabýváme otázkou, zda posloupnost nezávislých stejně rozdělených náhodných veličin pochází z Poissonova rozdělení. Pro tento problém představíme dva různé postupy a několik testů pro každý postup. První postup je založen na asymptotické aproximaci rozdělení testových statistik. Druhý postup vychází z generování testovacích vzorků. Dále na základě námi provedených simulačních studií diskutujeme sílu, výhody a nevýhody jednotlivých testů.

Klíčová slova: Poissonovo pravděpodobnostní rozdělení, test dobré shody, test metodou testovacích vzorků

Title: Tests for the Poisson distribution

Author: Filip Trusina

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zbyněk Pawlas, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In this work we deal with the question whether a sequence of independent identically distributed random variables comes from the Poisson distribution. For this task we present two different approaches and couple of tests for each approach. The first approach is based on the asymptotic approximation of distribution of test statistics. The second approach uses generation of test samples. Based on simulations done by us, we discuss the power of individual tests and their advantages and disadvantages.

Keywords: Poisson distribution, goodness-of-fit test, bootstrap sample test

Obsah

Kapitola 1

Představení problému

Mějme posloupnost nezáporných celočíselných nezávislých stejně rozdělených náhodných veličin $X_1, X_2, \ldots, X_n$. Zajímá nás, zda tato posloupnost je generována z Poissonova pravděpodobnostního rozdělení s neznámým parametrem $\lambda > 0$, to jest rozdělení dané předpisem

$$P(X = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \quad \lambda > 0, \quad k = 0, 1, 2, \ldots.$$

Toto rozdělení budeme značit $Po(\lambda)$. O této posloupnosti můžeme uvažovat jako o náhodném výběru z rozdělení. Pro řešení této otázky představíme dva různé postupy a několik testů pro každý postup. Pro každý test provedeme simulační studii za účelem určení empirického odhadu síly testu. V reálném světě by se toto rozdělení dalo použít pro modelování počtu aut, které projedou danou ulicí za den, nebo například počtu telefonátů přijatých telefonním centrem za den. Tedy v našem případě bychom měli posloupnost čísel $X_1, X_2, \ldots, X_n$, která by reprezentovala počet aut, které projedou danou ulicí, a my bychom chtěli zjistit, zda počet aut má Poissonovo rozdělení s neznámým parametrem $\lambda > 0$.

1.1 Historie

Poissonovo rozdělení je pojmenované po francouzském matematiku Siméonu Denisu Poissonovi, který žil v letech 1781 až 1840 a který toto rozdělení představil společně se svým pojetím teorie pravděpodobnosti v roce 1837 v práci [7]. Nicméně už v roce 1711 přišel francouzský matematik Abraham de Moivre (1667–1754) s podobnými výsledky ohledně tohoto rozdělení.

1.2 Formulace hypotézy

Pro náš problém předpokládáme nulovou hypotézu

$$H_0 : X \sim Po(\lambda), \quad \lambda > 0,$$

kterou budeme testovat proti alternativě

$$H_1 : X \text{ není z } Po(\lambda), \quad \lambda > 0.$$

Neboli předpokládáme H_0, že náhodný výběr $X_1, X_2, \ldots, X_n$ pochází z Poissonova rozdělení s neznámým parametrem $\lambda > 0$. Jako alternativu volíme H_1 ve tvaru, že náhodný výběr $X_1, X_2, \ldots, X_n$ nepochází z Poissonova rozdělení.

1.3 Definice

Definice 1. Náhodný výběr $X_1, \ldots, X_n$ *o velikosti* n *je uspořádaná* n*-tice nezávislých, stejně rozdělených náhodných veličin* X_i*, kde* $i \in \{1, \ldots, n\}$.

Definice 2. Střední hodnota *náhodné veličiny* X *je definována předpisem*

$$\mathrm{E}X = \int_{\mathbb{R}} x \, \mathrm{d}P(x),$$

kde P *je pravděpodobnostní míra určující rozdělení náhodné veličiny* X.

Definice 3. Rozptyl *náhodné veličiny* X *je definován předpisem*

$$\operatorname{var} X = \mathrm{E}X^2 - (\mathrm{E}X)^2.$$

Definice 4. Disperze *náhodné veličiny* X *je definována předpisem*

$$\mathrm{D}X = \frac{\operatorname{var} X}{\mathrm{E}X}.$$

1.4 Značení

V této práci budeme z důvodu přehlednosti používat některá značení.

$Po(\lambda)$	Poissonovo rozdělení s parametrem λ
χ^2_{n-1}	chí-kvadrát rozdělení o $n-1$ stupních volnosti
$\chi^2_{n-1}(1-\alpha)$	$(1-\alpha)$-kvantil chí-kvadrát rozdělení o $n-1$ stupních volnosti
$\Phi^{-1}(1-\alpha)$	$(1-\alpha)$-kvantil standardního normálního rozdělení
$\overline{X} = \overline{X}_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$	průměr náhodného výběru X_1, X_2, ..., X_n
$S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$	výběrový rozptyl
$Bi(n,p)$	binomické rozdělení s parametry n, p
$NB(n,p)$	negativně binomické rozdělení s parametry n, p
$PoM(\lambda_1, \lambda_2, p)$	směs Poissonových rozdělení s parametry λ_1, λ_2 a koeficientem směsi p
$PoZM(\lambda, p)$	nulou modifikované Poissonovo rozdělení s parametry λ a p
$F_n(k)$	empirická distribuční funkce náhodného výběru daná předpisem $F_n(k) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{I}\{X_i \leq k\}$

Kapitola 2

Testy založené na aproximaci rozdělením

2.1 Aproximace rozdělením

V tomto postupu využíváme vlastností rozdělení, které má asymptoticky daná statistika za platnosti nulové hypotézy. Jelikož některé testové statistiky v této kategorii by pro náhodný výběr obsahující samé nuly neměly smysl, například testová statistika z podkapitoly 2.3, budeme, za účelem dodržení hladiny testu $\alpha = 0{,}05$, postupovat následujícím způsobem. Pokud nastane případ, že náhodný výběr bude obsahovat samé nuly, zvolíme náhodné číslo z množiny $\{1, 2, \ldots, 20\}$. Pokud jsme zvolili číslo 20, zamítneme nulovou hypotézu H_0. Podobnou úpravu provedeme i v testech uvedených v kapitole 4. Nyní představíme testy a jim příslušné testové statistiky, které budeme v tomto typu testování používat.

2.2 Podmíněná chí-kvadrát statistika

Jako první test představíme test uveden v [2, kap. 12.5]. K jeho odvození využijeme následující lemma 1.

Lemma 1. *Nechť X_1, X_2, ..., X_n je náhodný výběr z $Po(\lambda)$, kde $\lambda > 0$. Pak podmíněné rozdělení náhodného vektoru $(X_1, X_2, \ldots, X_n)^T$, za podmínky $\sum\limits_{i=1}^{n} X_i = n\overline{X}$, je multinomické $M(n\overline{X}, (1/n, \ldots, 1/n))$.*

Důkaz. Sdružené rozdělení veličin X_1, X_2, ..., X_n je

$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{\lambda^{\sum x_i} \mathrm{e}^{-n\lambda}}{x_1! \ldots x_n!}.$$

Rozdělení náhodné veličiny $Y = \sum_{i=1}^{n} X_i$ je

$$P(Y = y) = \frac{(n\lambda)^y \mathrm{e}^{-n\lambda}}{y!},$$

to jest Poissonovo rozdělení s parametrem $n\lambda$. Podmíněné rozdělení náhodných veličin $X_1, \ldots, X_n$ za podmínky $Y = y$ pak bude dáno pravděpodobnostmi

$$\frac{y!}{x_1! \ldots x_n!}(1/n)^y, \quad \text{kde } \sum_{i=1}^{n} x_i = y,$$

což odpovídá multinomickému rozdělení s n třídami a s pravděpodobnostmi $(1/n, \ldots, 1/n)^T$. $\square$

Na základě lemmatu 1 získáme testovou statistiku

$$Q_A = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\overline{X}} = \frac{(n-1)S^2}{\overline{X}}, \tag{2.1}$$

kde S^2 je výběrový rozptyl náhodného výběru. Pak dle [2, věta 12.5] testová statistika Q_A má za platnosti nulové hypotézy H_0 přibližně chí-kvadrát rozdělení s $n-1$ stupni volnosti. Tedy budeme zamítat nulovou hypotézu H_0, pokud

$$Q_A \leq \chi^2_{n-1}(\alpha/2) \quad \text{nebo} \quad Q_A \geq \chi^2_{n-1}(1-\alpha/2).$$[1]

Jak vidíme, jedná se o oboustranný test. V článku [4, str. 616] se uvažuje jiná alternativní hypotéza, která vede na jednostrannou verzi tohoto testu. V tom případě je kritérium zamítnutí H_0 ve tvaru $Q_A > \chi^2_{n-1}(1-\alpha)$. Jednostranný test by ovšem selhával v situaci, kdy je výběrový rozptyl menší než výběrový průměr.

2.3 Neymanova–Scottova statistika

Další test je popsán v [4, str. 616–617]. My zde budeme používat oboustrannou verzi tohoto testu. Testová statistika tohoto testu je založená na vztahu (2.1) pro Q_A. Z předchozí kapitoly víme, že Q_A má za platnosti nulové hypotézy H_0 přibližně chí-kvadrát rozdělení s $n-1$ stupni volnosti. Využijeme znalosti střední hodnoty a rozptylu chí-kvadrát rozdělení a toho, že pro velká n ho lze aproximovat normálním rozdělením. Proto

$$T_{NS} = \frac{Q_A - (n-1)}{\sqrt{2(n-1)}} = \sqrt{\frac{n-1}{2}} \left(\frac{S^2}{\overline{X}} - 1 \right) \tag{2.2}$$

má pro $n \to \infty$ asymptoticky standardní normální rozdělení. Tedy budeme zamítat nulovou hypotézu H_0, pokud

$$T_{NS} \leq \Phi^{-1}(\alpha/2) \quad \text{nebo} \quad T_{NS} \geq \Phi^{-1}(1-\alpha/2).$$

V kapitole 3 je ukázáno, že pro menší velikost n náhodného výběru není tento test spolehlivý. Asymptotická aproximace funguje až pro velké hodnoty n, což je pozorováno v [4].

2.4 Anscombeho statistika

Poslední test v této kategorii je test představený v [4, str. 614–616], my zde budeme používat jeho oboustrannou verzi. Anscombe v roce 1948 v [3] ukázal, že pokud náhodná veličina $N \sim Po(\lambda)$, pak

$$\operatorname{var} \sqrt{N + \frac{3}{8}} = \frac{1}{4} + O\left(\frac{1}{\lambda}\right), \quad \lambda \to \infty, \tag{2.3}$$

[1]Značení v [2] pro kvantily chí-kvadrát rozdělení je opačné, to jest kvantil na levé straně jest $(1 - \frac{\alpha}{2})$ a na pravé $(\frac{\alpha}{2})$

kde člen $O(\frac{1}{\lambda})$ znamená, že existuje $K > 0$ takové, že $O(\frac{1}{\lambda}) \leq K\frac{1}{\lambda}$. Na základě tohoto faktu je možné definovat $Y_i = \sqrt{X_i + 3/8}$. Rovnice (2.3) naznačuje, že Y_i má pro velké λ přibližně normální rozdělení s rozptylem 1/4 a střední hodnotou $\mathrm{E}\sqrt{N + 3/8}$, kde $N \sim Po(\lambda)$. Proto testová statistika

$$T_{An} = 4\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \tag{2.4}$$

má za platnosti nulové hypotézy H_0 přibližně chí-kvadrát rozdělení o $n-1$ stupních volnosti. Výsledky představené v kapitole 3 ukazují, že je tato úvaha správná i pro poměrně malá λ. Nicméně pro velmi malá λ je test založený na této testové statistice (2.4) již dost nepřesný. Nulovou hypotézu budeme zamítat, pokud

$$T_{An} < \chi^2_{n-1}(\alpha/2) \quad \text{nebo} \quad T_{An} > \chi^2_{n-1}(1-\alpha/2).$$

Kapitola 3

Výsledky testů založených na aproximaci rozdělením

3.1 Empirické výsledky za platnosti H_0

V této části uvedeme výsledky testů založených na testových statistikách Q_A, T_{NS}, T_{An} daných vzorci (2.1), (2.2) a (2.4). Tabulka 1 obsahuje výsledky na základě námi provedených simulací Poissonova rozdělení, to jest za platnosti nulové hypotézy H_0. Dává nám informaci o pravděpodobnosti chyby prvního druhu[1] pro jednotlivé testové statistiky. Testy jsme prováděli na 1 000 000 náhodných výběrech z $Po(\lambda)$, na hladině α=0,05, pro různé velikosti náhodného výběru, značeno n, a různé hodnoty λ. Hodnoty v tabulce 1 značí relativní počet zamítnutí nulové hypotézy.

λ	n	Test		
		Q_A	T_{NS}	T_{An}
1	n=50	0,0443	0,0455	0,2657
1	n=25	0,0412	0,0391	0,0964
1	n=10	0,0337	0,0346	0,0281
1	n=5	0,0296	0,0246	0,0272
3	n=50	0,0482	0,0465	0,0497
3	n=25	0,0458	0,0452	0,0498
3	n=10	0,0470	0,0435	0,0465
3	n=5	0,0455	0,0433	0,0530
5	n=50	0,0493	0,0472	0,0610
5	n=25	0,0487	0,0454	0,0598
5	n=10	0,0469	0,0425	0,0597
5	n=5	0,0435	0,0440	0,0568
10	n=50	0,0493	0,0473	0,0563
10	n=25	0,0496	0,0456	0,0558
10	n=10	0,0483	0,0430	0,0545
10	n=5	0,0463	0,0473	0,0530

Tabulka 1: Relativní četnost chyb prvního druhu ze 1 000 000 simulací

[1] Zamítnutí platné nulové hypotézy

Z tabulky 1 vidíme, že testy poměrně dobře dodržují stanovenou hladinu testů pro různé volby λ a n s výjimkou, pokud $\lambda = 1$, kde testy jsou poměrně nepřesné. V [2] je doporučeno používat chí-kvadrát test dobré shody, pokud je splněno $\overline{X} \geq 5$. Naše simulace potvrzují, že pro $\lambda \geq 5$ test založený na Q_A zamítá platnou hypotézu přibližně v 5 % případů. Pro pro větší velikost n náhodného výběru dodržuje tento test lépe hladinu testu. Přesnost testu T_{NS} roste s počtem n, což nám potvrzuje asymptotickou normalitu. Dále vidíme, že asymptotická aproximace pro testovou statistiku T_{An} pro $\lambda \to \infty$ je správná, nicméně pro malá λ je test založen na této statistice velmi nepřesný. Z tabulky 1 můžeme utvořit závěr, že za platnosti nulové hypotézy nejlépe funguje test založený na testové statistice Q_A.

3.2 Empirické výsledky za platnosti H_1

V této části uvedeme výsledky testů založených na Q_A, T_{NS}, T_{An} za platnosti alternativní hypotézy H_1. Budeme provádět tyto testy na pravděpodobnostních rozděleních různých od Poissonova. Jako ukazatel, zda budeme nějaké rozdělení zamítat lépe, nám poslouží disperze. Poissonovo rozdělení má disperzi rovnou 1, čím vzdálenější bude hodnota disperze nějakého rozdělení od 1, tím spíše budou testy na tomto rozdělení zamítat nulovou hypotézu.

Jako první rozdělení zde uvedeme binomické rozdělení $Bi(m,p)$[2], které popisuje četnost výskytu náhodného jevu, kde parametr m udává počet pokusů a $p \in (0,1)$ pravděpodobnost daného jevu. Toto rozdělení je dáno předpisem

$$P(X = k) = \binom{m}{k} p^k (1-p)^{m-k}, \quad k = 0, 1, 2, \ldots$$

Pro střední hodnotu a rozptyl tohoto rozdělení platí

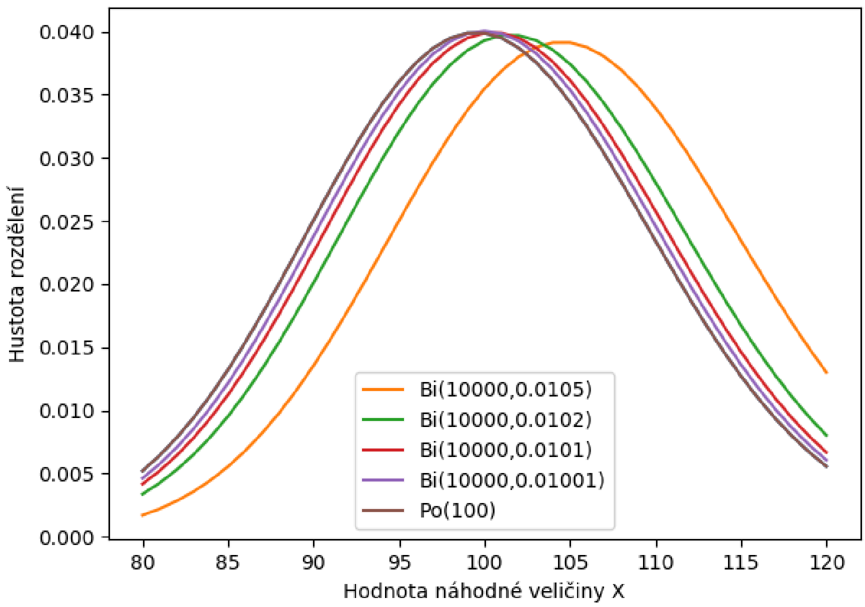
$$\mathrm{E}X = mp \quad \text{a} \quad \operatorname{var} X = mp(1-p)$$

Koeficient disperze pro toto rozdělení je roven

$$\mathrm{D}X = \tfrac{\operatorname{var} X}{\mathrm{E}X} = 1 - p.$$

Na tomto rozdělení je z našeho pohledu zajímavé, že binomické rozdělení se chová podobně jako Poissonovo v případě, že m je dostatečně velké a p dostatečně malé, kde pak $p = \frac{\lambda}{m}$, což je ukázáno například v [1]. Výsledky v tabulce 2 ukazují tuto vlastnost na několika případech. Tato vlastnost je také ukázána na obrázku 1.

[2] za účelem přehlednosti budeme používat jako parametr m místo obvyklého n, kterým značíme velikost náhodného výběru



Obrázek 1: Podobnost Poissonova rozdělení a binomického rozdělení, pro přehlednost jsou jednotlivé pravděpodobnosti jevů proloženy spojitou křivkou

Jako druhé rozdělení budeme brát v úvahu negativně binomické rozdělení $NB(m,p)$, to jest rozdělení dané předpisem

$$P(X=k)=\binom{m+k-1}{m-1}p^m(1-p)^k, \quad k=0,1,2,\dots,$$

kde $m \geq 1$ a $0 < p < 1$ je pravděpodobnost úspěchu. Jde o rozdělení počtu neúspěchů před dosažením m-tého úspěchu. Dále, pokud $X \sim NB(m,p)$, pak

$$\mathrm{E}X = m\frac{1-p}{p} \quad a \quad \operatorname{var} X = m\frac{1-p}{p^2}.$$

Koeficient disperze pro toto rozdělení je roven

$$\mathrm{D}X = \frac{\operatorname{var} X}{\mathrm{E}X} = \frac{1}{p}.^3$$

Další rozdělení, na kterém budeme testovat nulovou hypotézu, bude směs Poissonových rozdělení. To jest rozdělení $PoM(\lambda_1,\lambda_2,p)$, které je směs dvou nezávislých Poissonových rozdělení s parametry $\lambda_1 > 0, \lambda_2 > 0$ a pravděpodobností $p \in (0,1)$ výběru druhého z nich, to jest $PoM(\lambda_1,\lambda_2,p) = (1-p)Po(\lambda_1)+pPo(\lambda_2)$. Z linearity střední hodnoty je zřejmé, že pro střední hodnotu tohoto rozdělení bude platit, že

$$\mathrm{E}X = (1-p)\lambda_1 + p\lambda_2.$$

Druhý moment je

$$\mathrm{E}X^2 = (1-p)(\lambda_1^2+\lambda_1) + p(\lambda_2^2+\lambda_2).$$

[3] budeme testovat toto rozdělení pouze pro $p \geq 0{,}5$, jelikož pro menší p by byla disperze úměrně vzdálená 1 a testy by zamítaly hypotézu skoro vždy

Pak pro rozptyl bude platit

$$\begin{aligned}\operatorname{var} X = \mathrm{E}X^2 - (\mathrm{E}X)^2 &= (1-p)(\lambda_1^2+\lambda_1) + p(\lambda_2^2+\lambda_2) \\ &\quad - (1-p)^2\lambda_1^2 - p^2\lambda_2^2 - 2p(1-p)\lambda_1\lambda_2 \\ &= (1-p)\lambda_1 + p\lambda_2 + p(1-p)(\lambda_1-\lambda_2)^2.\end{aligned}$$

Koeficient disperze pro toto rozdělení je roven

$$\mathrm{D}X = \frac{\operatorname{var} X}{\mathrm{E}X} = 1 + \frac{p(1-p)(\lambda_1-\lambda_2)^2}{(1-p)\lambda_1 + p\lambda_2}.$$

Další rozdělení bude nulou modifikované Poissonovo rozdělení[4], které budeme značit $PoZM(\lambda, p)$, kde s pravděpodobností $p \in (0,1)$ vybíráme 0 a s pravděpodobností $(1-p)$ volíme z $Po(\lambda)$, kde $\lambda > 0$, tedy je to rozdělení dané pravděpodobnostmi

$$P(X=k) = \begin{cases} p + (1-p)\mathrm{e}^{-\lambda}, & \text{pokud} \quad k = 0, \\ (1-p)\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}, & \text{pokud} \quad k = 1, 2, \dots. \end{cases}$$

Střední hodnotu a rozptyl tohoto rozdělení můžeme určit jako speciální případ vzorců pro $PoZM(\lambda_1, \lambda_2, p)$, kde bereme $\lambda_1 = \lambda$ a $\lambda_2 = 0$. Odtud vidíme, že

$$\mathrm{E}X = (1-p)\lambda \quad \text{a} \quad \operatorname{var} X = (1-p)\lambda(1+p\lambda).$$

Koeficient disperze pro toto rozdělení je roven

$$\mathrm{D}X = \tfrac{\operatorname{var} X}{\mathrm{E}X} = 1 + p\lambda.$$

Následující tabulka 2 obsahuje výsledky testů za platnosti výše vypsaných alternativ H_1, testy provádíme na stejné hladině $\alpha = 0{,}05$ a pro velikost vzorku $n = 50$. Abychom získali pravděpodobnost volby v rozděleních PoM a $PoZM$, tak při generování vzorků volíme náhodné číslo v intervalu $(0,1)$ a srovnáním s p určíme, z jakého rozdělení budeme generovat realizaci X_i. Jak vidíme, tak naše úvaha o vztahu disperze a zamítání nulové hypotézy byla správná. Pro binomické rozdělení vidíme, že T_{An} je silnější nežli ostatní testy. Pro negativně binomické rozdělení je nejsilnější test T_{NS}. Z tabulky 2 je patrné, že test založený na testové statistice T_{An} si vede o poznání hůře, oproti zbylým testovým statistikám, v případě, že střední hodnota daného rozdělení je malá, toto je dáno tím, že přesnost testové statistiky T_{An} je malá pro malé střední hodnoty, což je důsledek závislosti asymptotičnosti rozdělení na parametru λ, kde víme, že pro $\lambda \to \infty$ přesnost aproximace T_{An} asymptotický rozdělením roste. Nicméně vidíme, že všechny testy si vedou poměrně dobře proti testovaným alternativám, pokud disperze není velmi blízká 1.

Na následujícím obrázku 2 je ukázána pravděpodobnost zamítnutí nulové hypotézy H_0 v závislosti na rostoucí pravděpodobnosti p výběru v rozdělení $PoM(5, 10, p)$ pro testy Q_A, T_{NS}, T_{An}, kde pro $p = 0$ se jedná vlastně o výběr z $Po(5)$. Pro zvětšující se pravděpodobnost p se zvyšuje i pravděpodobnost zamítnutí nulové hypotézy H_0. Po dosažení přibližně $p = 0{,}5$ začne pravděpodobnost zamítnutí nulové hypotézy H_0 klesat, pro $p = 1$ se jedná o výběr z $Po(10)$.

[4] v angličtině zero-modified Poisson distribution

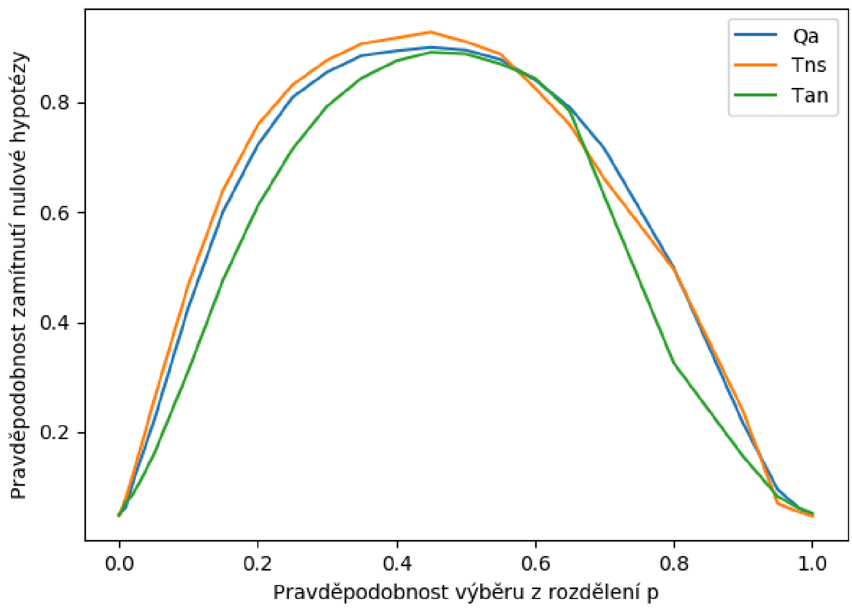
Rozdělení	EX	Disperze	Test		
			Q_A	T_{NS}	T_{An}
$Bi(1000, 0.01)$	10	0,99	0,0485	0,0447	0,0559
$Bi(100, 0.1)$	10	0,9	0,0704	0,0439	0,0778
$Bi(50, 0.3)$	15	0,7	0,3677	0,2568	0,3679
$Bi(30, 0.6)$	18	0,4	0,9957	0,9880	0,9945
$NB(3, \frac{3}{4})$	1	1,33	0,3103	0,3574	0,0709
$NB(2, \frac{2}{3})$	1	1,5	0,4979	0,5470	0,0497
$NB(1, \frac{1}{2})$	1	2	0,8407	0,8678	0,1067
$NB(5, \frac{1}{2})$	5	2	0,9154	0,9313	0,9078
$NB(10, \frac{2}{3})$	5	1,5	0,5472	0,5946	0,5299
$NB(15, \frac{3}{4})$	5	1,33	0,3302	0,3777	0,3219
$PoM(1, 5, \frac{1}{2})$	3	2,33	0,9987	0,9983	0,9952
$PoM(2, 5, \frac{1}{2})$	3,5	1,64	0,7404	0,7811	0,7141
$PoM(3, 5, \frac{1}{2})$	4	1,25	0,2254	0,2613	0,2106
$PoM(1, 2, \frac{1}{2})$	1,5	1,16	0,1384	0,1637	0,0202
$PoM(1, 3, \frac{1}{2})$	2	1,5	0,5575	0,6129	0,2682
$PoM(1, 4, \frac{1}{2})$	2,5	1,9	0,9273	0,9452	0,8708
$PoM(1, 5, \frac{1}{4})$	2	1,5	0,9915	0,9918	0,8677
$PoM(1, 5, \frac{1}{10})$	1,4	2,03	0,8172	0,8386	0,2494
$PoM(5, 10, \frac{1}{2})$	7,5	1,83	0,8941	0,9157	0,8967
$PoM(1, 10, \frac{1}{4})$	3,25	3,83	1,0000	1,0000	0,9997
$PoZM(10, \frac{1}{20})$	9,5	1,5	0,5534	0,5930	0,8454
$PoZM(10, \frac{1}{10})$	9	2	0,8924	0,9129	0,9847
$PoZM(3, \frac{1}{10})$	2,7	1,3	0,2843	0,3347	0,3845
$PoZM(2, \frac{1}{10})$	1,8	1,2	0,1535	0,1911	0,0352
$PoZM(1, \frac{1}{10})$	0,9	1,1	0,0722	0,0938	0,2370

Tabulka 2: Relativní četnost zamítnutí nulové hypotézy za platnosti alternativ rozdělení ze 1 000 000 simulací

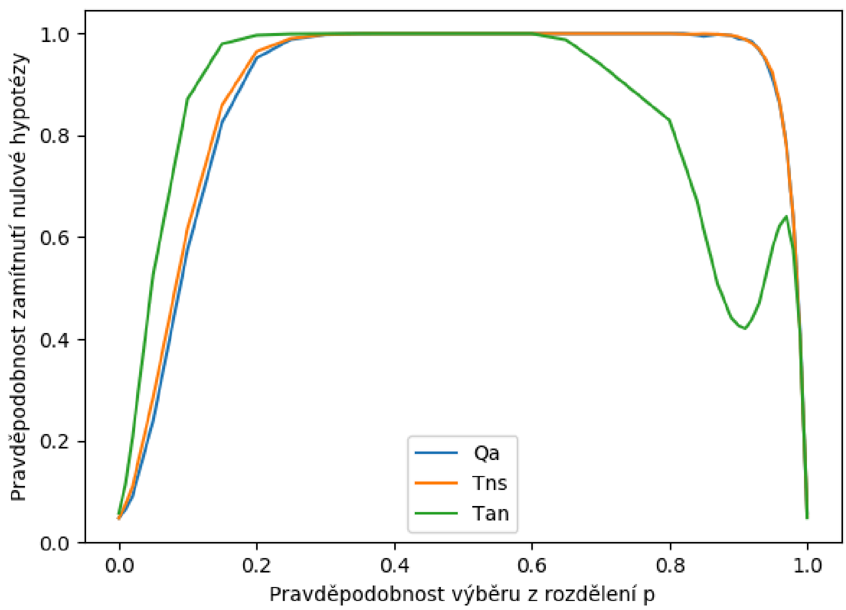
Vidíme, že test T_{NS} je z těchto tří testů nejsilnější pro tuto alternativu rozdělení. Z obrázku 2 je poměrně dobře vidět, že testová statistika Q_A je slabší než ostatní testové statistiky při testování alternativního rozdělení $PoM(5, 10, p)$. A to i o několik procent, tento rozdíl je nejvíce vidět v oblasti od $p = 0{,}5$ do $p = 1$.

Na obrázku 3 je ukázána pravděpodobnost zamítnutí nulové hypotézy H_0 v závislosti na rostoucí pravděpodobnosti p výběru v rozdělení $PoZM(5, p)$. Tedy pro $p = 0$ se jedná o výběr z $Po(5)$[5]. Z obrázku 3 vidíme, že test T_{An} je proti této alternativě $PoZM(5, p)$ silnější, nežli ostatní testy pro malá p. Musíme si uvědomit, že pro větší hodnoty p žádný z testů nedodržuje hladinu $\alpha = 0{,}05$. Podobný závěr platí pro menší hodnotu parametru λ, kde se ještě znatelněji ukáže, že test založený na T_{An} nedodržuje hladinu. Vidíme, že testy Q_A a T_{NS} dávají při této alternativě velmi podobné výsledky.

[5]Pro $p = 1$ bychom získali posloupnost samých 0, což naše oprava u testů na tento případ nám zaručí dodržení hladiny testu



Obrázek 2: Pravděpodobnost zamítnutí nulové hypotézy ve směsi Poissonových rozdělení s parametry $\lambda_1 = 5, \lambda_2 = 10$ v závislosti na pravděpodobnosti výběru druhého z nich



Obrázek 3: Pravděpodobnost zamítnutí nulové hypotézy v nulou modifikovaném Poissonově rozdělení s parametrem $\lambda = 5$ v závislosti na pravděpodobnosti vložení nuly

Kapitola 4

Test založen na testovacích vzorcích

Použijme totožný postup pro testování, jaký je použit v článku [9]. Postup je převzat z [8], kde jsou podrobněji vysvětleny jeho vlastnosti. Jedná se o princip parametrického Monte Carlo testování, který nespoléhá na asymptotické rozdělení testových statistik, ale je numericky náročnější než postup popsaný v kapitole 2. Princip testu spočívá ve srovnávání pozorovaných dat s testovacími vzorky nagenerovanými za platnosti nulové hypotézy. Budeme testovat na hladině $\alpha = 0{,}05$. Mějme náhodný výběr X_1, X_2, $\ldots$, X_n. Spočteme průměr náhodného výběru, to jest $\overline{X} = \hat{\lambda}$, a zvolenou testovou statistiku $T = T(X_1, \ldots, X_n)$. Nyní vygenerujeme testovací vzorek[1] z Poissonova rozdělení s parametrem $\hat{\lambda}$ o stejné velikosti jako je velikost n náhodného výběru. Těchto testovacích vzorků vygenerujeme $b = 99$. Spočítáme příslušnou testovou statistiku našich testovacích vzorků, čímž dostaneme $T_1, \ldots, T_b$. Za nulové hypotézy je jakékoli uspořádání hodnot $T, T_1, \ldots, T_b$ stejně pravděpodobné, proto je pravděpodobnost, že T překročí alespoň k z hodnot $T_1, \ldots, T_b$, rovna $\frac{b+1-k}{b+1}$. To nás vede k tomu, že zamítáme nulovou hypotézu H_0 ve prospěch alternativy H_1, pokud hodnota testové statistiky spočtená z dat překročí hodnotu alespoň k testových statistik testovacích vzorků. Přitom k je určeno tak, aby $\alpha = \frac{b+1-k}{b+1}$ neboli $k = (b+1)(1-\alpha)$. V našem případě $b = 99$ a $\alpha = 0{,}05$ je $k = 95$. Tímto způsobem se zajistí, že daný postup dodržuje teoreticky hladinu testu za platnosti nulové hypotézy. Nyní představíme několik vhodných voleb testové statistiky T, které pak budeme porovnávat v následující kapitole.

4.1 M-testová statistika

Testová statistika, kterou budeme používat v tomto testu a budeme ji značit M, je testová statistika představena v [9, str. 242–244], ve kterém je také plné odvození této testové statistiky. Proto tady ukážeme pouze nejdůležitější kroky. Mějme náhodný výběr $X_1, X_2, \ldots, X_n$ z testovaného rozdělení, nyní označme posloupnost

$$\hat{m}_k = \frac{1}{n}\sum_{j=1}^{n}|k - X_j|, \quad k = 0, 1, 2, \ldots.$$

[1] v angličtině bootstrap sample

Položme výběrový průměr $\overline{X} = \hat{\lambda}$, vidíme, že platí $\hat{\lambda} = \hat{m}_0$. Dále položme $\hat{F}(0) = \hat{p}(0) = (\hat{m}_1 + 1 - \hat{\lambda})/2$, $\hat{p}(k)$ a $\hat{F}(k)$ jsou určeny rekursivním předpisem

$$\hat{p}(k) = \frac{\hat{m}_{k+1} - (k+1-\hat{\lambda})(2\hat{F}(k-1)-1)}{2(k+1)}, \quad k = 1, 2, 3, \ldots,$$

kde $\hat{p}(k)$ je odhad pravděpodobnosti $P(X = k)$ rozdělení náhodného výběru. Hodnota $\hat{F}(k)$ je odhad hodnoty distribuční funkce skutečného rozdělení v k. Následující tabulka 3 ukazuje, že tento odhad je přesnější, nežli běžná empirická distribuční funkce $F_n(k)$. Tabulka 3 obsahuje průměr hodnot odhadů distribuční funkce $F(k;1)$ Poissonova rozdělení v bodě k ze 100 realizací náhodného výběru o velikosti $n = 50$ z Poissonova rozdělení s parametrem $\lambda = 1$.

k	0	1	2	3	4	5	6	7
$F(k;1)$	0,3679	0,7357	0,9197	0,9810	0,9963	0,9994	0,9999	0,9999
$\hat{F}(k)$	0,3684	0,7309	0,9164	0,9794	0,9957	0,9992	0,9998	0,9999
$F_n(k)$	0,3684	0,7390	0,9314	0,9824	0,9970	0,9996	1,0000	1,0000

Tabulka 3: Porovnání odhadů distribuční funkce se skutečnou distribuční funkcí Poissonova rozdělení s parametrem $\lambda = 1$, zprůměrování 100 pokusů

Z tabulky 3 vidíme, že odhad $\hat{F}(k)$ je přesnější nežli odhad daný empirickou distribuční funkcí. Toto motivuje testovou statistiku

$$M = n \sum_{j=0}^{\infty} (\hat{F}(j) - F(j;\hat{\lambda}))^2 p(j;\hat{\lambda}), \tag{4.1}$$

kde $\hat{F}(j)$ je výše diskutovaný odhad distribuční funkce v j, $F(j;\hat{\lambda})$ je hodnota distribuční funkce Poissonova rozdělení s parametrem $\hat{\lambda} = \overline{X}$ v j, a $p(j;\hat{\lambda})$ je pravděpodobnost $P(X = j)$ v Poissonově rozdělení s parametrem $\hat{\lambda} = \overline{X}$, n je velikost vzorku.

4.2 Modifikovaná M-testová statistika

Jako druhý test představíme test založený na modifikaci M-testové statistiky, viz (4.1). Naše testová statistika MM bude podobná, s tím rozdílem, že hustotu v M odhadneme za pomocí rozdílu odhadů distribučních funkcí, podobně v testové statistice v podkapitole 4.4. Tedy využijeme odhady popsané v části o M-testové statistice. Naše nová modifikovaná MM testová statistika bude ve tvaru

$$MM = n \sum_{j=0}^{\infty} (\hat{F}(j) - F(j;\hat{\lambda}))^2 (\hat{F}(j) - \hat{F}(j-1)), \tag{4.2}$$

kde $\hat{F}(j)$ je odhad distribuční funkce z podkapitoly 4.1 a $F(j;\hat{\lambda})$ je distribuční funkce Poissonova rozdělení s parametrem $\hat{\lambda}$, n je velikost náhodného výběru.

4.3 Cramérova–von Misesova statistika

Testová statistika, kterou budeme využívat v tomto testu, je pojmenována po švédském matematiku Haraldu Cramérovi a rakouském matematiku Richardu von Misesovi. Tato testová statistika byla představena v [5]. Testová statistika W^2 je ve tvaru

$$W^2 = n \sum_{j=0}^{\infty} (F_n(j) - F(j; \hat{\lambda}))^2 p(j; \hat{\lambda}), \tag{4.3}$$

kde $F_n(j)$ je empirická distribuční funkce náhodného výběru, $F(j; \hat{\lambda})$ je distribuční funkce Poissonova rozdělení s parametrem $\hat{\lambda}$ v j, dále $p(j; \hat{\lambda})$ je pravděpodobnost $P(X = j)$ v Poissonově rozdělení s parametrem $\hat{\lambda}$ a n je velikost náhodného výběru. Vidíme, že tato testová statistika je velmi podobná testové statistice M, jen s tím rozdílem, že jako odhad distribuční funkce je použita empirická distribuční funkce.

4.4 Modifikovaná Cramérova–von Misesova statistika

Testová statistika CM je založena na následujícím principu. Při určování testové statistiky W^2 je potřeba znát Poissonovo pravděpodobnostní rozdělení s parametrem $\hat{\lambda}$, zatímco testová statistika CM tuto znalost nepotřebuje, hodnotu pravděpodobnosti $P(X = j)$ nahradíme díky znalosti náhodného výběru za pomocí rozdílu dvou empirických distribučních funkcí, a to v j a $j - 1$. To nám dává testovou statistiku ve tvaru

$$CM = n \sum_{j=0}^{\infty} (F_n(j) - F(j; \hat{\lambda}))^2 (F_n(j) - (F_n(j-1)), \tag{4.4}$$

kde $F_n(j)$ je empirická distribuční funkce náhodného výběru, $F(j; \hat{\lambda})$ je distribuční funkce Poissonova rozdělení s parametrem $\hat{\lambda}$ v j a n je velikost výběru. Následující tabulka 4 ukazuje průměr ze 100 pokusů rozdílu mezi pravděpodobností $P(X = k)$ v Poissonově rozdělení s parametrem 2 a rozdílem empirických distribučních funkcí z náhodného výběru o velikost $n = 50$ z $Po(2)$.

k	0	1	2	3	4	5	6	7
$p(k; \lambda)$	0,135	0,270	0,270	0,180	0,090	0,036	0,012	0,003
$F_n(k) - F_n(k-1)$	0,140	0,267	0,273	0,170	0,096	0,036	0,012	0,002

Tabulka 4: Porovnání pravděpodobnosti $P(X = k)$ a rozdílu hodnoty empirické distribuční funkce v j-tém a $(j-1)$-tém kroku Poissonova rozdělení s parametrem $\lambda = 2$, zprůměrování 100 pokusů

4.5 Kolmogorovova–Smirnovova statistika

Kolmogorovův–Smirnovův test založen na této testové statistice je popsán v [6][2]. V tomto případě budeme využívat jeho jednovýběrovou testovou statistiku, jelikož testujeme, zda náhodný výběr pochází z nějakého známého rozdělení, v našem případě z $Po(\lambda), \lambda > 0$. Testová statistika je ve tvaru

$$KS = \sqrt{n} \sup_{0 \leq j} |F_n(j) - F(j; \hat{\lambda})|, \tag{4.5}$$

kde $F_n(j)$ je empirická distribuční funkce v j, $F(j; \hat{\lambda})$ je hodnota distribuční funkce Poissonova rozdělení s parametrem $\hat{\lambda}$ a n velikost výběru.

4.6 Fisherova disperze

Další testová statistika, kterou budeme v této části používat, je Fisherova disperze použitá také v [9, str. 245] daná předpisem

$$FD = \left(\frac{1}{2\sqrt{n}} \left[\frac{(n-1)S^2}{\overline{X}} - n \right] \right)^2, \tag{4.6}$$

kde S^2 je výběrový rozptyl, $\overline{X}$ je výběrový průměr a n je velikost výběru. Výhoda této testové statistiky spočívá v tom, že nemusíme pro výpočet počítat distribuční funkci nebo pravděpodobnost $P(X = k)$ v Poissonově rozdělení, které jsou v ostatních testech uvedených v této části potřeba.

[2]my zde nepoužíváme tento test, pouze jeho testovou statistiku

Kapitola 5

Výsledky testů založených na testovacích vzorcích

5.1 Zrychlení algoritmu

Některé testové statistiky představené v kapitole 4 pracují s nekonečnou sumou, což pro přesnost výsledku je pěkné, nicméně z hlediska výpočetní náročnosti prakticky nedosažitelné. Tento problém nastává především pro testové statistiky M a W^2. Proto při výpočtu statistik uděláme menší úpravu. Například pokud generujeme náhodný výběr z Poissonova rozdělení s parametrem $\lambda = 3$, pak pravděpodobnost, že nagenerujeme číslo větší nežli 15, je dána vztahem

$$P(X \geq 15) = \sum_{i=15}^{\infty} \frac{3^i}{i!}\mathrm{e}^{-3} = 1 - \sum_{0}^{i=14} \frac{3^i}{i!}\mathrm{e}^{-3} = 6{,}7 \cdot 10^{-7}.$$

Proto hodnota, kterou bychom v dané sumě přičítali k testové statistice pro i mnohonásobně větší nežli parametr λ, je zanedbatelná vůči dosažené hodnotě testové statistiky. Je rozumné proto omezit horní hranici v nekonečné řadě například vztahem $(\lambda + 5) \cdot 2$, což nám dává 12 pro $\lambda = 1$, 16 pro $\lambda = 3$, 20 pro $\lambda = 5$ a 30 pro $\lambda = 10$. V případě, že parametr λ neznáme, použijeme ve vzorci místo λ horní celou část výběrového průměru $\overline{X}$.

5.2 Empirické výsledky za platnosti H_0

V této části uvedeme výsledky testů založených na testových statistikách M, MM, W^2, CM, KS, FD, viz (4.1)–(4.6), za platnosti nulové hypotézy H_0. Tyto výsledky jsou obsaženy v tabulce 5. Tabulka 5 nám dává informaci o pravděpodobnosti chyby prvního druhu pro jednotlivé testové statistiky. Test jsme prováděli na hladině $\alpha = 0{,}05$ pro 10 000 vzorků, pro různé velikosti náhodného výběru n, různé hodnoty λ a pro počet testovacích vzorků $b = 99$.

λ	Velikost n	Test					
		M	MM	W^2	CM	KS	FD
1	n=50	0,051	0,050	0,050	0,051	0,050	0,051
1	n=25	0,051	0,050	0,050	0,050	0,051	0,051
1	n=10	0,052	0,048	0,050	0,049	0,052	0,051
1	n=5	0,052	0,047	0,049	0,048	0,053	0,053
3	n=50	0,050	0,050	0,050	0,051	0,050	0,049
3	n=25	0,050	0,050	0,050	0,050	0,051	0,050
3	n=10	0,051	0,052	0,051	0,051	0,051	0,051
3	n=5	0,054	0,052	0,052	0,052	0,053	0,053
5	n=50	0,050	0,050	0,049	0,050	0,051	0,051
5	n=25	0,050	0,050	0,050	0,050	0,051	0,051
5	n=10	0,051	0,051	0,051	0,050	0,051	0,051
5	n=5	0,054	0,052	0,053	0,051	0,053	0,053
10	n=50	0,050	0,050	0,049	0,050	0,050	0,050
10	n=25	0,049	0,050	0,049	0,050	0,050	0,050
10	n=10	0,051	0,050	0,050	0,050	0,051	0,051
10	n=5	0,052	0,052	0,052	0,051	0,053	0,052

Tabulka 5: Relativní četnost chyby prvního druhu ze 10 000 simulací

Z tabulky 5 vidíme, že testy založené na všech testových statistikách přibližně dodržují hladinu testu α, a to pro všechny volby parametru λ. Nicméně po snížení velikosti vzorku na $n = 5$ vidíme, že testy jsou méně efektivní a nedodržují procento zamítnutí vzorků, jak by za platnosti H_0 měly. Pro parametr $\lambda = 1$ vidíme, že testy pro n menší jsou více nepřesné, nežli pro vyšší λ.

5.3 Empirické výsledky za platnosti H_1

Následující tabulka 6 nám ukazuje výsledky testů za platnosti alternativ popsaných v části 3.2. Z tabulky 6 vidíme, že pro binomické rozdělení jsou testy založené na testových statistikách M a MM slabší nežli zbylé testy. Dále je vidět, že test založen na FD je lepší než ostatní. U negativně binomického rozdělení vidíme, že test založen na testové statistice MM si vede mnohem lépe vůči ostatním statistikám, s výjimkou FD, nežli v binomickém případě. K tomuto závěru docházíme i u směsi Poissonových rozdělení a nulou modifikovaného Poissonova rozdělení. Tedy můžeme říci, že z našich testových statistik si pro nejvíce alternativ nejlépe vedla FD. Mezi testovými statistikami založenými na hodnotách empirických distribučních funkcí si nejlépe vede MM, která i skoro ve všech případech byla lepší, nežli její základní verze M.

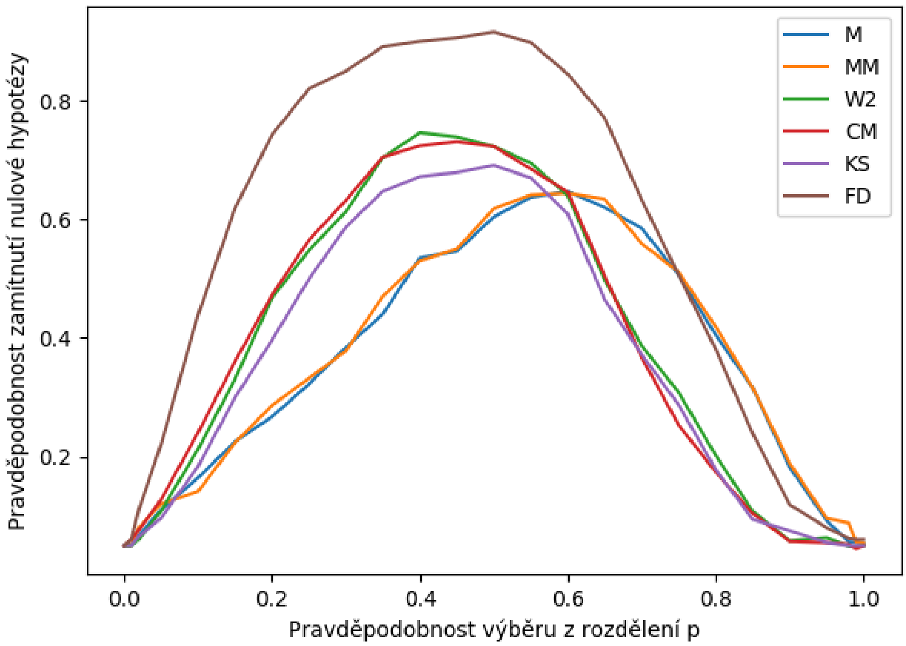
Na následujícím obrázku 4 je ukázána pravděpodobnost zamítnutí nulové hypotézy H_0 v závislosti na rostoucí pravděpodobnosti p výběru v rozdělení $PoM(5, 10, p)$. Tedy pro $p = 0$ se jedná o výběr z $Po(5)$ a pro $p = 1$ se jedná o výběr z $Po(10)$. Z obrázku 4 vidíme, že test založený na testové statistice FD je silnější nežli zbylé testy. Vidíme, že modifikované verze testů MM a CM jsou velmi podobně úspěšné jako jejich základní verze M a W^2. Všechny testy překročí 60 % zamítnutí nulové hypotézy H_0 pro $p = 0{,}6$.

Rozdělení	Test					
	M	MM	W^2	CM	KS	FD
$Bi(1000, 0.01)$	0,050	0,049	0,048	0,049	0,049	0,050
$Bi(100, 0.1)$	0,034	0,045	0,064	0,058	0,054	0,063
$Bi(50, 0.3)$	0,097	0,095	0,228	0,198	0,172	0,311
$Bi(30, 0.6)$	0,575	0,539	0,953	0,933	0,842	0,988
$NB(3, \frac{3}{4})$	0,257	0,310	0,192	0,226	0,211	0,331
$NB(2, \frac{2}{3})$	0,439	0,451	0,342	0,377	0,359	0,515
$NB(1, \frac{1}{2})$	0,819	0,839	0,726	0,774	0,743	0,858
$NB(5, \frac{1}{2})$	0,799	0,797	0,679	0,689	0,658	0,925
$NB(10, \frac{2}{3})$	0,397	0,415	0,286	0,308	0,254	0,575
$NB(15, \frac{3}{4})$	0,245	0,258	0,166	0,175	0,155	0,338
$PoM(1, 5, \frac{1}{2})$	0,998	0,997	0,986	0,989	0,984	0,997
$PoM(2, 5, \frac{1}{2})$	0,692	0,701	0,543	0,564	0,515	0,754
$PoM(3, 5, \frac{1}{2})$	0,191	0,210	0,129	0,136	0,129	0,241
$PoM(1, 2, \frac{1}{2})$	0,121	0,131	0,085	0,099	0,101	0,153
$PoM(1, 3, \frac{1}{2})$	0,542	0,567	0,383	0,409	0,407	0,575
$PoM(1, 4, \frac{1}{2})$	0,933	0,941	0,831	0,855	0,827	0,938
$PoM(1, 5, \frac{1}{4})$	0,967	0,971	0,948	0,956	0,931	0,993
$PoM(1, 5, \frac{1}{10})$	0,625	0,662	0,558	0,589	0,545	0,826
$PoM(5, 10, \frac{1}{2})$	0,593	0,596	0,726	0,737	0,662	0,898
$PoM(1, 10, \frac{1}{4})$	1,000	1,000	1,000	1,000	0,999	1,000
$PoZM(10, \frac{1}{20})$	0,926	0,927	0,226	0,173	0,213	0,570
$PoZM(10, \frac{1}{10})$	0,994	0,995	0,647	0,609	0,611	0,896
$PoZM(3, \frac{1}{10})$	0,523	0,587	0,207	0,260	0,312	0,307
$PoZM(2, \frac{1}{10})$	0,213	0,267	0,140	0,169	0,186	0,176
$PoZM(1, \frac{1}{10})$	0,080	0,098	0,076	0,083	0,078	0,087

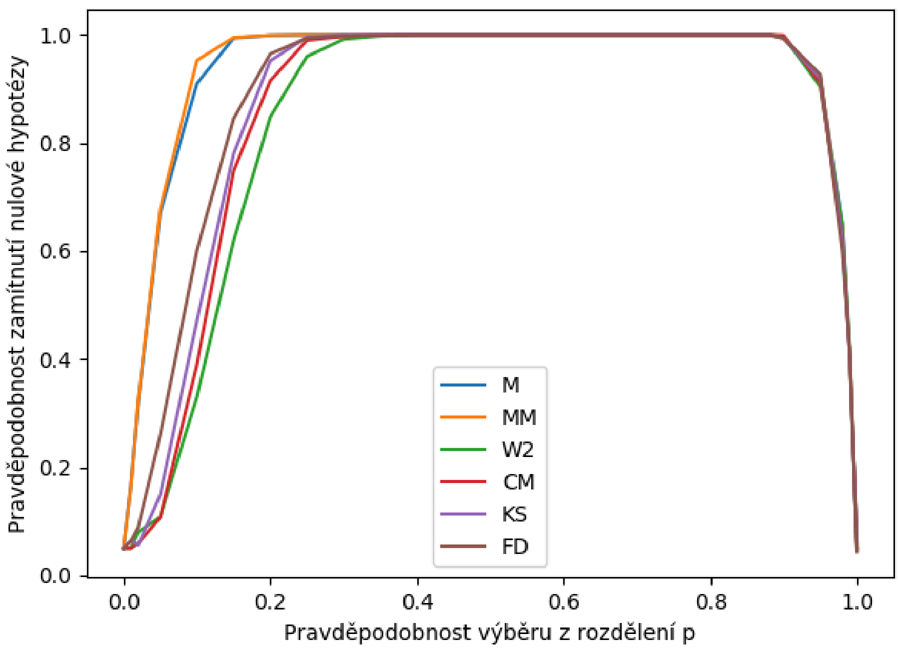
Tabulka 6: Relativní četnost zamítnutí nulové hypotézy za platnosti alternativ rozdělení ze 10 000 simulací

Na následujícím obrázku 5 je ukázána pravděpodobnost zamítnutí nulové hypotézy H_0 v závislosti na rostoucí pravděpodobnosti p výběru v rozdělení $PoZM(5, p)$. Tedy pro $p = 0$ se jedná o výběr z $Po(5)$[1]. Z obrázku 5 vidíme, že testy založené na testových statistikách M, MM jsou silnější nežli zbylé testy a také mají velmi podobnou úspěšnost zamítnutí nulové hypotézy, i když se jeví, že test založen na M je slabší nežli jeho modifikovaná verze MM. Všechny testy dosahují 100 % zamítnutí nulové hypotézy H_0 pro velké p, takže jsou poměrně dobré. Pro p jdoucí k 1 testovaný náhodný výběr začne obsahovat samé nuly a pravděpodobnost testů klesne kvůli naší úpravě na 5 %.

[1] Pro $p = 1$ bychom získali posloupnost samých 0, což naše oprava u testů na tento případ nám bude dodržovat hladinu testu



Obrázek 4: Pravděpodobnost zamítnutí nulové hypotézy ve směsi Poissonových rozdělení s parametry $\lambda_1 = 5$ a $\lambda_2 = 10$ v závislosti na pravděpodobnosti výběru druhého z nich



Obrázek 5: Pravděpodobnost zamítnutí nulové hypotézy v nulou modifikovaném Poissonově rozdělení s parametrem $\lambda = 5$ v závislosti na rostoucí pravděpodobnosti výběru nuly

Závěr

V této práci jsme se zabývali testy Poissonova rozdělení. Uvedli jsme dva postupy a několik testů pro každý z nich.

V rámci našich simulací jsme došli k podobným výsledkům a závěrům, diskutovaným v kapitolách 3 a 5, ke kterým došli jejich autoři v článcích [4] a [9].

V této práci jsme na rozdíl od [2] ukázali i empirické výsledky, které jsme podložili námi udělanými simulacemi. Pro některé testy použité v článku [4] jsme se zabývali v kapitole 3 i otázkou síly těchto testů proti alternativním rozdělením. Ukázali jsme, že testy si vedou poměrně dobře vůči testům z kapitoly 5, ba dokonce v některých situacích i srovnatelně s nejsilnějším testem založeným na testové statistice FD dané vztahem (4.6).

Literatura

[1] Anderson T. W., Samuels S. M. (1965): Some inequalities among binomial and Poisson probabilities, *in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. **1**, 1–12.

[2] Anděl J. (2005): *Základy matematické statistiky.* Praha: Matfyzpress, **1.** vydání, ISBN 80-86732-40-1.

[3] Anscombe F. J. (1948): The transformation of Poisson, binomial and negative-binomial data, *Biometrika*, Vol. **35**, No. **3/4**, 246–254.

[4] Brown L. D., Zhao L. H. (2002): A test for the Poisson distribution, *Sankhyā Ser. A*, Vol. **64**, No. **3**, 611–625.

[5] Cramér H. (1928): On the composition of elementary errors, *Scand. Actuar. J.*, Vol. **1**, 13–74.

[6] Massey Jr. F. J. (1951): The Kolmogorov–Smirnov test for goodness of fit, *J. Amer. Statist. Assoc.*, Vol. **46**, No. **253**, 68–78.

[7] Poisson S. D. (1837): *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilitiés*, Paříž: Bachelier.

[8] Stute W., Manteiga W., Quindimil M. (1993): Bootstrap based goodness-of-fit tests, *Metrika*, Vol. **40**, 243–256.

[9] Székely G. J., Rizzo M. L. (2004): Mean distance test of Poisson distribution, *Statist. Probab. Letter.*, Vol. **67**, 241–247.

Přílohy

Zdrojový kód programu v Pythonu verze 3.7.1 je dostupný jako elektronická příloha v SISu.