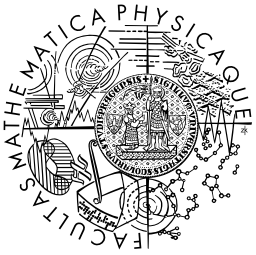


MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

**BAKALÁŘSKÁ PRÁCE**

Tomáš Jurčo

# Testy nezávislosti pro posloupnost veličin s Poissonovým rozdělením

Katedra pravděpodobnosti a matematické statistiky

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

Studijní program: Obecná matematika

Praha 2022

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

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

V ..... dne .....  .....

Podpis autora

Chtěl bych poděkovat své vedoucí práce, RNDr. Šárce Hudecové, Ph.D., za cenné rady, tipy, opravy a trpělivost, ale především za předané pracovní nadšení. Dále děkuji své rodině a přátelům za celkovou psychickou podporu během studia.

Název práce: Testy nezávislosti pro posloupnost veličin s Poissonovým rozdělením

Autor: Tomáš Jurčo

Katedra: Katedra pravděpodobnosti a matematické statistiky

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

Abstrakt: Tato práce se zabývá testováním závislosti v časových řadách stejně rozdělených náhodných veličin s Poissonovým rozdělením. V úvodní části jsou zadefinovány důležité pojmy a definice, zejména autokorelační funkce, její odhady a INAR(1) model. Dále jsou v práci popsány tři druhy testů nezávislosti – testy založené na odhadech autokorelační funkce, test jednoduchých iterací a testy založené na kontingenčních tabulkách. Popsané testy jsou následně porovnány v simulační studii za platnosti nulové hypotézy nezávislosti a za alternativy modelu INAR(1).

Klíčová slova: Poissonovo rozdělení, testy nezávislosti, časové řady

Title: Independence testing for series of Poisson variables

Author: Tomáš Jurčo

Department: Department of Probability and Mathematical Statistics

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

Abstract: This thesis deals with tests of independence for time series of identically distributed Poisson random variables. In the introductory part, important terms and definitions are defined, in particular the autocorrelation function, its estimates and INAR(1) model. Three types of tests of independence are described in the thesis – tests based on estimates of the autocorrelation function, simple runs test and tests based on contingency tables. These tests are compared in a simulation study under the null hypothesis of independence and under the alternative of INAR(1) model.

Keywords: Poisson distribution, tests of independence, time series

# Obsah

# Úvod

Pro popis mnoha náhodných veličin z reálného světa lze využít Poissonovo rozdělení – např. počet dopravních nehod v určitý den na daném území, počty nově narozených v daném měsíci, obecněji počet událostí za jednotku času. Pokud se tyto události v čase opakují, je možné si klást otázku: „Je mezi výskyty událostí nějaký vztah?“ Samotná otázka může mít vícero možných interpretací, my k ní budeme přistupovat ze statistického pohledu, kdy budeme pozorovat počty událostí za jednotku času s cílem zjistit, zda jsou tyto počty, které lze popsat náhodnými veličinami, v čase na sobě nezávislé.

V této práci se budeme zabývat problémem testování nezávislosti posloupnosti náhodných veličin se stejným Poissonovým rozdělením, přičemž se omezíme na situace, kdy parametr $\lambda$ v Poissonovu rozdělení nabývá hodnot blízkých nule, případně se pohybuje nejvýše v řádu jednotek.

V první kapitole práce se seznámíme s některými základními pojmy, definicí autokorelační funkce, možnostmi jejích odhadů a nakonec s modelem využívaným k popisu závislých časových řad.

V další části práce popíšeme testovanou hypotézu a alternativu. Následně zavedeme tři typy testů, které jsou vhodné pro uvažovanou situaci.

Poslední částí práce je simulační studie, ve které prozkoumáme vlastnosti popsaných testů pro různé situace a různé rozsahy dat.

Pro přehlednost práce a komfort čtenáře budeme v průběhu práce zpracovávanou látku ilustrovat na vzorku reálných dat. Konkrétně využijeme údaje o měsíčních počtech nahlášených případů pracovní neschopnosti způsobených popáleninami evidovaných agenturou British Columbia Workers Compensation Board v období od ledna 1987 do prosince 1994, které byly též využity v práci Freeland (1998).

Vlastním přínosem autora v práci je rozepsání některých důkazů, podrobné vysvětlení některých kroků v popisovaných testech a provedení simulační studie.

# 1. Základní pojmy a modely

V této kapitole zavedeme důležité pojmy a definice, které budeme využívat v následujících částech práce. Nejprve si zadefinujeme Poissonovo rozdělení a popíšeme jeho vlastnosti.

**Definice 1.** *Náhodná veličina $X$ má Poissonovo rozdělení s parametrem $\lambda > 0$, píšeme $X \sim Po(\lambda)$, jestliže nabývá pouze hodnot $0, 1, ...,$ a to s pravděpodobnostmi*

$$\mathsf{P}(X = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \qquad k = 0, 1, ...,$$

*kde $\lambda > 0$ je dané číslo.*

**Věta 1.** *Nechť je $X$ náhodná veličina s Poissonovým rozdělením s parametrem $\lambda$. Pak platí, že $\mathsf{E}\ X = \lambda$ a $\mathrm{var}(X) = \lambda$.*

*Důkaz.* Plyne z přímého výpočtu, viz Dupač a Hušková (2003, strana 29). □

V aplikacích zpravidla hodnotu parametru $\lambda$ neznáme, je nutné střední hodnotu, resp. rozptyl, odhadovat. Užitečné jsou konzistentní, případně i nestranné odhady. Při tvorbě odhadů můžeme využít vlastností výběrového průměru a výběrového rozptylu.

**Věta 2.** *Nechť $X_1, ..., X_n$ je náhodný výběr z $Po(\lambda)$, kde $n \geq 2$, s výběrovým průměrem $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ a výběrovým rozptylem $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Pak platí $\mathsf{E}\ \overline{X}_n = \mathsf{E}\ S_n^2 = \lambda$ a současně $\overline{X}_n$, resp. $S_n^2$, jsou konzistentní odhady $\lambda$ pro $n \to \infty$.*

*Důkaz.* Plyne z přímého výpočtu využitím věty 1 a alternativního předpisu pro výběrový rozptyl

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \lambda)^2 - 2\frac{1}{n-1}\sum_{i=1}^n (X_i - \lambda)(\overline{X}_n - \lambda) + \frac{n}{n-1}\sum_{i=1}^n (\overline{X}_n - \lambda)^2 =$$

$$= \frac{1}{n-1}\sum_{i=1}^n (X_i - \lambda)^2 - \frac{n}{n-1}\sum_{i=1}^n (\overline{X}_n - \lambda)^2,$$

kde $\lambda$ je dle věty 1 střední hodnota $X_i$. Konzistence plyne přímým výpočtem ze zákona velkých čísel, viz Dupač a Hušková (2003, strana 95). □

Vzhledem k uvedeným vlastnostem jsou $\overline{X}_n$ a $S_n^2$ nestrannými a konzistentními odhady parametru $\lambda$ pro náhodný výběr z $Po(\lambda)$.

## 1.1 Časové řady

Důležitým stavebním kamenem této práce je pojem časové řady.

**Definice 2.** *Časovou řadou nazveme chronologicky uspořádanou posloupnost náhodných veličin* $\{X_t,\, t \in T\}$ *definovaných na stejném pravděpodobnostním prostoru* $(\Omega, \mathcal{A}, \mathsf{P})$*, kde* $T \neq \emptyset$*,* $T \subset \mathbb{N}$*,* $t \in T$ *je čas.*

Vidíme, že se jedná o speciální typ náhodného procesu s diskrétním časem; náhodné procesy se spojitým časem ($T \subset \mathbb{R}$) nebudou předmětem této práce. Zároveň se omezíme na řady, které jsou označovány jako stacionární.

**Definice 3.** *Časovou řadu* $\{X_t,\, t \in T\}$ *nazýváme striktně stacionární, jestliže pro libovolné* $n \in \mathbb{N}$*,* $h \in \mathbb{Z}$*,* $x_1, x_2, ..., x_n \in \mathbb{R}$ *a* $t_1, t_2, ..., t_n \in T$ *taková, že* $t_k + h \in T$ *pro všechna* $k \in \{1, ..., n\}$*, platí*

$$F_{X_{t_1},...,X_{t_n}}(x_1, ..., x_n) = F_{X_{t_1+h},...,X_{t_n+h}}(x_1, ..., x_n),$$

*kde* $F_{X_{t_1},...,X_{t_n}}$ *je distribuční funkce sdruženého rozdělení náhodného vektoru veličin* $X_{t_1}, ..., X_{t_n}$*.*

*Časovou řadu označíme jako (slabě) stacionární, jestliže* $\mathsf{E}\, X_t = \mu$*,* $\mu \in \mathbb{R}$*,* $\mathrm{var}(X_t) = \sigma^2$*,* $0 < \sigma^2 < \infty$*, pro všechna* $t \in T$*, a pro libovolná* $s, t \in T$ *a* $h \in \mathbb{N}_0$ *taková, že* $s + h$*,* $t + h \in T$ *platí*

$$\mathrm{cov}(X_s, X_t) = \mathrm{cov}(X_{s+h}, X_{t+h}).$$

Z definice vyplývá, že striktně stacionární časová řada s konečnými druhými momenty je zároveň stacionární.

**Definice 4.** *Nechť je* $\{X_t,\, t \in T\}$ *stacionární časová řada, kde* $\mathsf{E}\, X_t = \mu$*,* $\mu \in \mathbb{R}$*,* $\mathrm{var}(X_t) = \sigma^2$*,* $0 < \sigma^2 < \infty$*, pro všechna* $t \in T$*. Pak autokovarianční funkci* $\gamma_k$ *definujeme jako*

$$\gamma_k = \mathsf{E}\,(X_t - \mu)(X_{t-k} - \mu), \qquad k \in \mathbb{Z}.$$

*Obdobně definujeme autokorelační funkci* $\rho_k$ *jako*

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma^2}, \qquad k \in \mathbb{Z}.$$

Autokorelační funkce $\{\rho_k\}$ popisuje strukturu závislosti v časové řadě $\{X_t\}$. Pokud je $\{X_t\}$ posloupnost nezávislých stejně rozdělených náhodných veličin (tuto vlastnost budeme dále značit jako *iid*) s konečným rozptylem, pak $\rho_k = 0$ pro všechna $k \neq 0$, $k \in \mathbb{Z}$. Naopak pokud $\rho_k \neq 0$ pro nějaké $k > 0$, pak

$$\rho_k = \mathsf{E}\,(X_t - \mu)(X_{t-k} - \mu) \neq \mathsf{E}\,(X_t - \mu)\,\mathsf{E}\,(X_{t-k} - \mu),$$

neboť $\mathsf{E}\,(X_t - \mu) = 0$, tudíž náhodné veličiny časové řady $\{X_t\}$ nejsou nezávislé.

Současně pro $k \in \mathbb{Z}$ platí vztah

$$\rho_{-k} = \rho_k, \tag{1.1}$$

tedy stačí uvažovat $\{\rho_k, k \geq 0\}$.

## 1.2 Odhady autokorelační funkce

Máme-li stacionární časovou řadu $\{X_t, t = 1, ..., n\}$, pak můžeme potřebovat odhadnout autokorelační funkci. Ke konstrukci odhadu lze využít výběrového průměru a výběrového rozptylu.

Standardní odhad $\hat{\rho}_k$ autokorelační funkce $\rho_k$ je tvaru

$$\hat{\rho}_k = \frac{\frac{1}{n}\sum_{t=k+1}^{n}(X_t - \overline{X}_n)(X_{t-k} - \overline{X}_n)}{S_n^2},$$

pro $k = 0, 1, ..., n-1$. Tento odhad je pak za jistých předpokladů konzistentní odhad $\rho_k$ pro $n \to \infty$, jak ukazuje tvrzení 3.

V následujících kapitolách se budeme setkávat s časovými řadami $\{X_t, t = 1, ..., n\}$, kde $X_t \sim Po(\lambda)$, $\lambda > 0$, pro všechna $t = 1, ..., n$. Na základě vlastností Poissonova rozdělení uvedených ve větě 1 a větě 2 lze odhad autokorelační funkce $\hat{\rho}_k$ modifikovat na

$$\widetilde{\rho}_k = \frac{\frac{1}{n}\sum_{t=k+1}^{n}(X_t - \overline{X}_n)(X_{t-k} - \overline{X}_n)}{\overline{X}_n}, \tag{1.2}$$

pro hodnoty $k = 1, ..., n-1$. Alternativně můžeme též v odhadu zohlednit počet sčítanců v čitateli

$$\dot{\rho}_k = \frac{(n-k)^{-1}\sum_{t=k+1}^{n}(X_t - \overline{X}_n)(X_{t-k} - \overline{X}_n)}{\overline{X}_n}, \qquad k = 1, ..., n-1,$$

a to za daných předpokladů bez porušení konzistence odhadu, jak ukazuje následující tvrzení.

**Tvrzení 3.** *Nechť $\{X_t, t = 1, ..., n\}$ je stacionární časová řada, kde $X_t \sim Po(\lambda)$ pro všechna $t = 1, ..., n$, autokovarianční funkce $\gamma_k$ splňuje*

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty$$

*a současně pro všechna $s, t = 1, ..., n$*

$$\mathsf{E}\,(X_s X_{s+t} - \mathsf{E}\; X_s X_{s+t})(X_{s+n} X_{s+t+n} - \mathsf{E}\; X_{s+n} X_{s+t+n}) \xrightarrow[n\to\infty]{} 0.$$

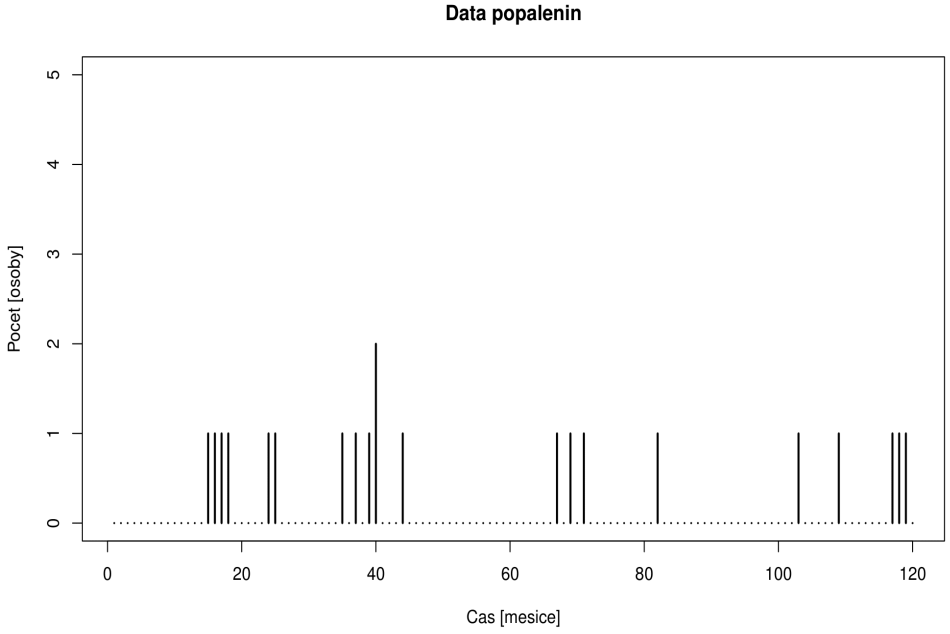
*Pak pro pevné $k$ jsou $\hat{\rho}_k$, $\widetilde{\rho}_k$, $\dot{\rho}_k$ konzistentní odhady autokorelační funkce $\rho_k$ pro $n \to \infty$.*

*Důkaz.* Důkaz konzistence $\hat{\rho}_k$ je uveden např. v Lain (2020). Z důkazu plyne, že za daných předpokladů platí

$$\frac{1}{n}\sum_{t=k+1}^{n}(X_t - \overline{X}_n)(X_{t-k} - \overline{X}_n) \xrightarrow[n\to\infty]{P} \operatorname{cov}(X_t, X_{t-k}), \qquad S_n^2 \xrightarrow[n\to\infty]{P} \lambda.$$

Současně z věty 6 dále plyne, že

$$\overline{X}_n \xrightarrow[n\to\infty]{P} \lambda,$$




Obrázek 1.1: Zaznamenané počty popálenin v jednotlivých měsících.

tedy $\overline{X}_n$ a $S_n^2$ jsou konzistentními odhady střední hodnoty $\lambda$. Pak z Cramérovy-Sluckého věty

$$\widetilde{\rho}_k = \frac{S_n^2}{\overline{X}_n}\hat{\rho}_k \xrightarrow{d} \lambda, \qquad n \to \infty,$$

neboť $S_n^2/\overline{X}_n \xrightarrow[n\to\infty]{P} 1$. Jelikož $\lambda$ je konstanta, $\widetilde{\rho}_k \xrightarrow[n\to\infty]{P} \lambda$, a tedy se jedná o konzistentní odhad parametru $\lambda$.

Obdobným postupem ověříme konzistenci $\dot{\rho}_k$. S použitím Cramérovy-Sluckého věty a konzistence $\widetilde{\rho}_k$

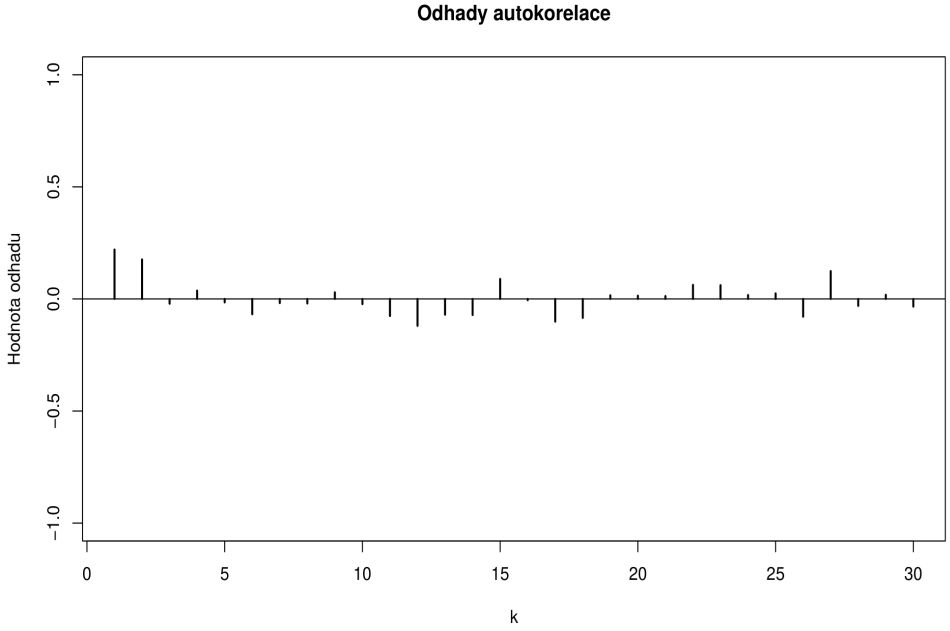
$$\dot{\rho}_k = \frac{nS_n^2}{(n-k)\overline{X}_n}\hat{\rho}_k \xrightarrow{d} \lambda, \qquad n \to \infty,$$

neboť $\frac{n}{n-k} \xrightarrow[n\to\infty]{P} 1$ pro všechna $k = 1, ..., n-1$. Tudíž $\dot{\rho}_k \xrightarrow[n\to\infty]{P} \lambda$, a $\dot{\rho}_k$ je konzistentní odhad $\lambda$.

□

Pro ilustraci porovnáme odhady $\hat{\rho}_k$, $\widetilde{\rho}_k$ a $\dot{\rho}_k$ autokorelační funkce $\rho_k$ na datech hlášených úrazů popsaných v úvodu této práce.

*Příklad.* Uvažujme data počtů nahlášených pracovních neschopností způsobených popáleninami – celkem se jedná o 120 hodnot, které zobrazuje graf 1.1. Následně spočteme odhady, přičemž uvažujeme pouze $k = 1, ..., 30$, viz graf 1.2. Získané výsledky pro $k = 1, 2, 5$ uvádí tabulka 1.1. Pozorujeme, že z použitých odhadů nabývá $\hat{\rho}_k$ hodnot nejvýše o několik setin větších než zbývající odhady $\widetilde{\rho}_k$ a $\dot{\rho}_k$. Mezi $\widetilde{\rho}_k$ a $\dot{\rho}_k$ je pouze zanedbatelný rozdíl, neboť zvolené hodnoty $k$ jsou malé vůči celkovému počtu pozorování.




Obrázek 1.2: Graf odhadů autokorelace $\hat{\rho}_k$ pro $k = 1, ..., 30$.

Pro hodnoty autokorelace $\rho$ platí, že $\rho \in [-1, 1]$, tedy hodnoty uvedené v tabulce 1.1 naznačují, že počty hlášených pracovních neschopností způsobených popáleninami v jednotlivých měsících by mohly být závislé. Tuto ideu ověříme v následující kapitole. Zároveň si povšimneme, že s rostoucím $k$ se hodnoty odhadů autokorelace blíží 0, proto se obecně v této práci zaměříme pouze na závislost dvou po sobě jdoucích náhodných veličin v časové řadě.

## 1.3 Modely pro časové řady závislých náhodných veličin s Poissonovým rozdělením

Existuje nekonečné množství možností, jak mohou být na sobě náhodné veličiny v časové řadě závislé. My v této kapitole ukážeme jeden z možných využívaných modelů – tzv. *PoINAR(1) model*, tedy INAR(1) model s marginálním Poissonovým rozdělením, viz Jung a Tremayne (2003).

Pro účely této sekce uvažujme posloupnost nezávislých, stejně rozdělených náhodných veličin $W_t \sim Po(\beta)$, $t \in \mathbb{N}_0$, $\beta > 0$. Pak PoINAR(1) model časové

| Odhad | $k = 1$ | $k = 2$ | $k = 5$ |
|---|---|---|---|
| $\hat{\rho}_k$ | 0,220 | 0,176 | −0,016 |
| $\widetilde{\rho}_k$ | 0,204 | 0,164 | −0,014 |
| $\dot{\rho}_k$ | 0,206 | 0,167 | −0,015 |

Tabulka 1.1: Srovnání odhadů korelace pro známá data.

řady $\{X_t,\, t \in \mathbb{N}_0\}$ je definován rekurzivně vztahy

$$X_0 \sim Po\left(\frac{\beta}{1-a}\right), \qquad X_t = a \circ X_{t-1} + W_t, \tag{1.3}$$

kde $a \in (0,1)$ je daný parametr a náhodné veličiny $W_t$ a $X_s$ jsou nezávislé pro všechna $s < t$, $s, t \in \mathbb{N}_0$. Operátor $\circ$, nazývaný též jako *thinning operator*, byl poprvé popsán v článku Steutel a Harn (1979) a udává operaci

$$a \circ X_{t-1} = Y_{1,t-1} + Y_{2,t-1} + ... + Y_{X_{t-1},t-1} = \sum_{i=1}^{X_{t-1}} Y_{i,t-1},$$

kde $Y_{i,t-1}$ jsou vzájemně nezávislé, stejně rozdělené náhodné veličiny takové, že $\mathsf{P}(Y_{i,t-1} = 1) = a$ a $\mathsf{P}(Y_{i,t-1} = 0) = 1 - a$; zároveň $Y_{i,t-1}$ jsou nezávislé na $X_s$ pro všechna $s = 1, ..., t-1$ a pro všechna $i$. Pak náhodná veličina $a \circ X_{t-1}$ má za podmínky $X_{t-1}$ pravděpodobnostní rozdělení

$$a \circ X_{t-1} | X_{t-1} \;\sim\; Bi(X_{t-1}, a).$$

Platí, že náhodné veličiny $X_t$ mají marginálně Poissonovo rozdělení, což popisuje následující tvrzení.

**Tvrzení 4.** *Nechť $\{X_t,\, t \in \mathbb{N}_0\}$ je časová řada definovaná rekurzivními vztahy* (1.3)*. Pak jsou náhodné veličiny $X_t$ pro všechna $t \in \mathbb{N}_0$ stejně rozdělené a platí*

$$X_t \;\sim\; Po(\lambda), \qquad t \in \mathbb{N}_0,$$

*kde $\lambda = \frac{\beta}{1-a}$.*

Tuto vlastnost můžeme ukázat, před provedením samotného důkazu si však pro úplnost připomeneme definici vytvořující funkce.

**Definice 5.** *Nechť $X$ je náhodná veličina s diskrétním rozdělením a s hodnotami v $\mathbb{N}_0$. Pak zavádíme vytvořující funkci $P_X$ náhodné veličiny $X$ jako*

$$P_X(s) = \mathsf{E}\,[s^X], \qquad s \in [-1, 1].$$

*Důkaz.* Označme $\lambda := \frac{\beta}{1-a}$. Ukážeme za pomoci indukce, že $X_t \sim Po(\lambda)$ pro všechna $t \in \mathbb{N}_0$. Z předpokladů vztah platí pro $t = 0$, předpokládejme, že platí i pro všechna $t = 0, ..., n-1$.

Označme $S_{n-1} := a \circ X_{n-1}$. Pak pro $t = n$ můžeme, díky nezávislosti $W_n$ a $S_{n-1}$, psát:

$$P_{X_n}(s) = P_{S_{n-1}}(s) P_{W_n}(s).$$

Jelikož $W_n \sim Po(\beta)$, můžeme určit $P_{W_n}(s)$:

$$P_{W_n}(s) = \mathsf{E}\,[s^W] = \sum_{t=0}^{\infty} \frac{e^{-\beta}\beta^n}{n!} s^n = e^{-\beta} \sum_{t=0}^{\infty} \frac{(\beta s)^n}{n!} = e^{\beta(s-1)},$$

přičemž obdobným přímým výpočtem získáme, že $P_{Y_{1,n-1}}(s) = 1 + a(s-1)$. Současně je $S_{n-1}$ náhodným součtem stejně rozdělených, nezávislých náhodných veličin $Y_{i,n-1}$, tudíž je vytvořující funkce $S_{n-1}$ dána vztahem

$$P_{S_{n-1}}(s) = P_{X_{n-1}}(P_{Y_{1,n-1}}(s)).$$

Pak platí

$$\begin{aligned} P_{X_n}(s) &= P_{X_{n-1}}(P_{Y_{1,n-1}}(s))\ \mathrm{e}^{\beta(s-1)} = \\ &= P_{X_{n-1}}(1 + a(s-1))\ \mathrm{e}^{\beta(s-1)} = \\ &= \mathrm{e}^{\frac{\beta}{1-a}(a(s-1))}\ \mathrm{e}^{\beta(s-1)} = \mathrm{e}^{(s-1)(\beta + a\frac{\beta}{1-a})} \\ &= e^{\frac{\beta}{1-a}(s-1)}, \end{aligned}$$

kde jsme ve třetí rovnosti využili indukční předpoklad. Tudíž $X_t \sim Po(\lambda)$, $\lambda = \frac{\beta}{1-a}$, pro všechna $t \in \mathbb{N}_0$.

□

Z konstrukce dále plyne, že $\{X_t\}$ je striktně stacionární homogenní Markovův řetězec a platí vztah

$$\mathsf{E}\,(X_t | X_{t-1}) = aX_{t-1} + \lambda.$$

Odtud přímým výpočtem plyne vztah pro autokorelaci

$$\rho_k = \mathrm{cor}(X_t, X_{t-k}) = a^k, \qquad k = 1, ..., n-1,$$

viz např. Grunwald a kol. (2000). S pomocí tohoto vztahu můžeme snadno popsat lineární závislost mezi dvěma libovolnými náhodnými veličinami $X_n$ a $X_m$, $n, m \in \mathbb{N}_0$, $n \neq m$ výše popsané časové řady. Povšimněme si, že $\rho_k$ nabývá v tomto modelu pouze nezáporných hodnot.

# 2. Testy nezávislosti

V minulé kapitole jsme ukázali, jaký model lze uvažovat při práci s časovou řadou závislých náhodných veličin s Poissonovým rozdělením. V aplikacích však obvykle předem nevíme, zda lze náhodné veličiny považovat za nezávislé, či nikoli.

Pro veškeré uváděné testy budeme uvažovat časovou řadu $\{X_t,\ t = 1, ..., n\}$, $n \in \mathbb{N}$. Následně provedeme test platnosti hypotézy

$$H_0 : X_t \ \sim \ iid, X_t \ \sim \ Po(\lambda) \qquad t \in \{1, ..., n\},$$

oproti alternativě

$H_1$: $\{X_t\}$ je časová řada stejně rozdělených náhodných veličin, $X_t \ \sim \ Po(\lambda)$ pro všechna $t \in \{1, ..., n\}$, avšak $\exists\ s, t \in \{1, ..., n\}, s \neq t$: $X_s$ a $X_t$ jsou závislé.

Současně, jak jsme stanovili v úvodu této práce, se zaměříme na situace, kdy je $\lambda$ malé a budeme očekávat, že řadu $\{X_t\}$ lze za alternativy $H_1$ popsat pomocí modelu PoINAR(1).

Testovaná hypotéza $H_0$ je specifickým případem obecné hypotézy $\widetilde{H_0}$:

$$\widetilde{H_0} : X_t \ \sim \ iid, \qquad t \in \{1, ..., n\},$$

kterou lze testovat oproti zcela obecné alternativě

$$\widetilde{H_1} : \exists\ s, t \in T\text{: } X_s \text{ a } X_t \text{ jsou závislé, nemusí být stejně rozdělené.}$$

K vyřešení otázky nezávislosti pro obecné náhodné veličiny můžeme použít několik běžně používaných testů, popsaných například v knize Cipra (2008, kapitola 9.7). Vzhledem k zaměření této práce na časové řady náhodných veličin s Poissonovým rozdělením, navíc za předpokladu malých hodnot parametru $\lambda$, mnohé běžně využívané obecné testy mohou dosahovat méně přesných výsledků, případně nebudou zcela fungovat bez jistých úprav. Pro tento specifický typ časových řad budeme testy, případně hlavní idey úprav běžně užívaných testů, čerpat z článku Jung a Tremayne (2003), u kterých k výše uvedeným problémům nedochází.

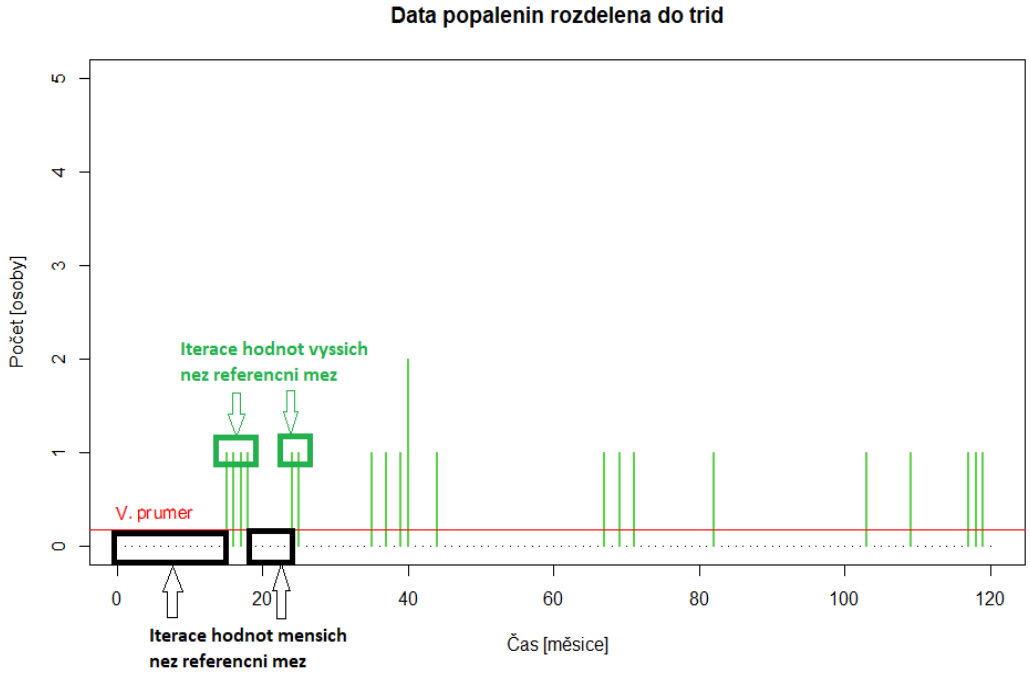
## 2.1 Test jednoduchých iterací

Jednu z nejjednodušších myšlenek reprezentuje *test jednoduchých iterací (angl. simple runs test)*, založený na porovnávání dat vůči předem zvolené referenční mezi – pro potřeby testu označené jako $M$.

Současně, v kontextu tohoto testu, budeme *iterací (angl. run)* rozumět skupinu po sobě jdoucích pozorování takových, že všechna pozorování obsažená v dané skupině mají vyšší, resp. menší, hodnotu, než je hodnota zvolené referenční meze. Celkový počet iterací v testovaných datech označme $R$.

Mějme časovou řadu $\{X_t,\ t = 1, ..., n\}$ a zvolme vhodnou referenční mez. Před provedením samotného testu je třeba pozorovaná data rozdělit do dvou tříd – první bude obsahovat data s hodnotami vyššími než referenční mez $M$, druhá bude tvořena hodnotami menšími než referenční mez $M$. Označme dále

$$k = \sum_{t=1}^{n} \mathbf{1}\{X_t > M\}, \qquad l = \sum_{t=1}^{n} \mathbf{1}\{X_t < M\}. \tag{2.1}$$




Obrázek 2.1: Příklady iterací vytvořené podle upravené referenční meze.

Obvykle je jako referenční mez uvažována hodnota výběrového mediánu $M_n$ pozorovaných dat, viz Cipra (2008, strana 324). V tomto případě pak očekáváme, že nad i pod referenční mezí se nachází shodný počet pozorování, tedy $k = l$. Pokud $k = l \pm 1$, pak zvolíme libovolný nezařazený prvek, nabývající hodnoty výběrového mediánu, a zařadíme jej do menší ze tříd. Veškerá dosud nezařazená pozorování, tedy pozorování nabývající hodnoty výběrového mediánu, vyloučíme. V takto upraveném souboru dat pak spočítáme počet iterací $R$, jak ilustruje obrázek 2.1.

Testová statistika $Z$ je pak dána předpisem

$$Z = \frac{R - (k+1)}{\sqrt{\frac{k(k+1)}{2k-1}}},$$

pro kterou dle knihy Gibbons a Chakraborti (2003, sekce 3.2) za nulové hypotézy $\widetilde{H_0}$ platí, že

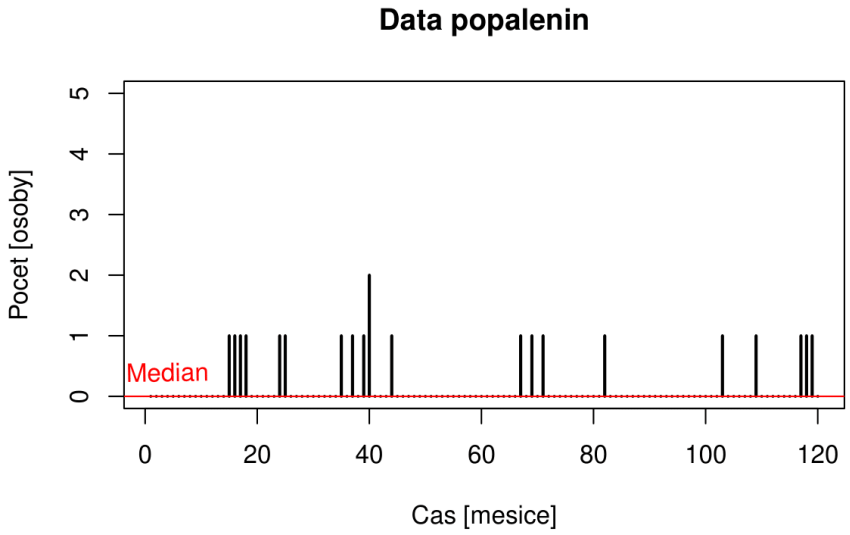
$$Z \xrightarrow{d} N(0,1), \qquad n \to \infty.$$

Popsaný test je oboustranný a asymptotický, přičemž jeho kritický obor je dán vztahem

$$|Z| \geq u_{1-\alpha/2},$$

kde $u_{1-\alpha/2}$ je $(1-\alpha/2)$-kvantil normálního rozdělení $N(0,1)$. Tedy čím je vyšší absolutní hodnota rozdílu počtu iterací a počtu prvků v jedné ze tříd, tím spíše pochází pozorovaná data z časové řady závislých náhodných veličin.

Tato obecná verze testu jednoduchých iterací však není vhodná například právě pro testování nezávislosti pro časové řady náhodných veličin s Poissonovým rozdělením $Po(\lambda)$, kde $\lambda$ je malé, jak ukážeme na příkladu.




Obrázek 2.2: Referenční mez obecného testu jednoduchých iterací.

*Příklad.* Uvažujme data popálenin. Pak medián těchto pozorovaných dat je roven nule (viz graf 2.2). Současně ze 120 pozorování je pouze 20 pozorovaných hodnot nenulových, přičemž, již z podstaty dat, nebyly pozorovány žádné záporné hodnoty, tedy hodnoty menší než zvolená referenční mez. Nelze proto test provést výše popsaným postupem.

Vhodnou úpravu testu zmiňují např. Jung a Tremayne (2006). Mějme časovou řadu $\{X_t,\, t = 1, ..., n\}$ stejně rozdělených náhodných veličin s Poissonovým rozdělením. Oproti výše popsanému postupu, za referenční mez $M$ nyní zvolíme hodnotu výběrového průměru. Testová statistika $Z^*$ je pak dána předpisem

$$Z^* = \frac{R - 1 - \frac{2kl}{N}}{\sqrt{2kl\frac{2lk-N}{N^2(N-1)}}}, \qquad N = k + l,$$

kde $k$, $l$ jsou dány vztahy (2.1). Zároveň Wald a Wolfowitz (1940) ukázali, že za nulové hypotézy $H_0$ pro rozdělení testové statistiky platí

$$Z^* \xrightarrow{d} N(0, 1), \qquad n \to \infty.$$

Popsaný test je jednostranný, neboť předpokládáme, že za alternativy $H_1$ lze řadu $\{X_t\}$ popsat pomocí modelu PoINAR(1), kde

$$\rho_k = a^k > 0, \qquad a \in (0, 1),\, k = 1, ..., n - 1.$$

Pak nulovou hypotézu nezávislosti zamítáme pro malé hodnoty testové statistiky:

$$H_0 \text{ zamítáme právě tehdy, když } Z < u_\alpha,$$

kde $u_\alpha$ je $\alpha$-kvantil normálního rozdělení $N(0, 1)$.

*Příklad.* Uvažujme data popálenin, přičemž za referenční mez zvolíme výběrový průměr $\overline{X}_n$, kde $\overline{X}_n = 0{,}175$. Pak data rozdělíme do dvou tříd, viz graf 2.1, kde $k = 20$ a $l = 100$. Spočteme hodnotu testové statistiky $Z^* = -2{,}437$ a příslušnou p-hodnotu $p(Z^*) = \mathsf{F}(Z^*) = 0.007$, kde $\mathsf{F}$ je distribuční funkce normovaného normálního rozdělení. Tudíž pro zvolenou hladinu spolehlivosti $\alpha = 0{,}05$ zamítáme hypotézu ve prospěch alternativy, neboť platí $Z^* < u_\alpha = -1{,}645$.

## 2.2 Testy založené na odhadu autokorelace

Další test můžeme založit na vlastnostech autokorelační funkce a jejích odhadů, které jsme popsali v kapitole 1.2.

Mějme časovou řadu $\{X_t,\ t = 1, ..., n\}$ stejně rozdělených náhodných veličin s Poissonovým rozdělením s parametrem $\lambda$. Za nulové hypotézy $H_0$, zadefinované v úvodu kapitoly, platí

$$\rho_1 = 0,$$

tudíž i hodnoty odhadu $\widetilde{\rho_1}$, definovaného vztahem (1.2), by měly být za $H_0$ blízké nule. Této myšlenky je využito k sestrojení testové statistiky $S_n$ dané předpisem

$$S_n = \widetilde{\rho_1}\sqrt{n} = (\overline{X}_n\sqrt{n})^{-1}\sum_{t=2}^{n}(X_{t-1} - \overline{X}_n)(X_t - \overline{X}_n),$$

kde $\overline{X}_n$ je výběrový průměr. Shodné testové statistiky využívá i *skórový test* odvozený pro PoINAR(1) model ze sekce 1.3, viz Freeland (1998).

Vlastnosti testové statistiky pro časovou řadu náhodných veličin s Poissonovým rozdělením shrnuje následující věta.

**Věta 5.** *Nechť $\{X_t,\ t = 1, ..., n\}$ je časová řada vzájemně nezávislých, stejně rozdělených náhodných veličin, kde $X_t \sim Po(\lambda)$ pro všechna $t = 1, ..., n$. Pak testová statistika $S_n$ konverguje v distribuci*

$$S_n \xrightarrow{d} N(0,1), \qquad \text{pro } n \to \infty.$$

*Důkaz.* Nejprve si upravíme předpis pro testovou statistiku $S_n$:

$$\begin{aligned}
S_n &= (\overline{X}_n\sqrt{n})^{-1}\sum_{t=2}^{n}(X_{t-1} - \overline{X}_n)(X_t - \overline{X}_n) = \\
&= \frac{1}{\overline{X}_n\sqrt{n}}\sum_{t=2}^{n}\Big[(X_{t-1} - \lambda) + (\lambda - \overline{X}_n)\Big]\Big[(X_t - \lambda) + (\lambda - \overline{X}_n)\Big] = \\
&= \frac{1}{\overline{X}_n\sqrt{n}}\left[\sum_{t=2}^{n}(X_{t-1} - \lambda)(X_t - \lambda) + \sum_{t=2}^{n}(X_{t-1} - \lambda)(\lambda - \overline{X}_n) + \right. \\
&+ \left.\sum_{t=2}^{n}(X_t - \lambda)(\lambda - \overline{X}_n) + \sum_{t=2}^{n}(\lambda - \overline{X}_n)^2\right] = \\
&= \frac{1}{\overline{X}_n\sqrt{n}}\left[\sum_{t=2}^{n}(X_{t-1} - \lambda)(X_t - \lambda) + 2(\lambda - \overline{X}_n)\sum_{t=2}^{n}(X_t - \lambda) + \right. \\
&+ \left.(\lambda - \overline{X}_n)(X_1 + X_n) + \sum_{t=2}^{n}(\lambda - \overline{X}_n)^2\right].
\end{aligned}$$

K ukázání konvergence výsledného výrazu využijeme vlastnosti Poissonova rozdělení (viz věta 1). Z centrální limitní věty víme, že platí

$$\sqrt{n}(\overline{X}_n - \lambda) \xrightarrow{d} N(0, \lambda), \qquad \text{pro } n \to \infty. \tag{2.2}$$

Dále ze silného zákona velkých čísel získáme

$$\overline{X}_n \xrightarrow{P} \lambda, \qquad \text{pro } n \to \infty. \tag{2.3}$$

Pak označme

$$A_n := \frac{1}{\overline{X}_n \sqrt{n}} \sum_{t=2}^{n} (X_{t-1} - \lambda)(X_t - \lambda).$$

Ukážeme, že výraz $S_n - A$ konverguje k nule v pravděpodobnosti.

$$\begin{aligned} S_n - A_n = \frac{1}{\overline{X}_n \sqrt{n}} \Bigg[ & 2(\lambda - \overline{X}_n) \sum_{t=2}^{n} (X_t - \lambda) + \\ & + (\lambda - \overline{X}_n)(X_1 + X_n) + \sum_{t=2}^{n} (\lambda - \overline{X}_n)^2 \Bigg] = \\ & = \frac{2(\lambda - \overline{X}_n)}{\overline{X}_n} \sqrt{n}(\overline{X}_n - \lambda) + \frac{\lambda - \overline{X}_n}{\overline{X}_n \sqrt{n}} (-X_1 + 2\lambda + X_n) - \\ & - \frac{\lambda - \overline{X}_n}{\overline{X}_n} \frac{n-1}{\sqrt{n}} (\overline{X}_n - \lambda). \end{aligned}$$

Pak poslední výraz se zjevně pro $n \to \infty$ chová stejně jako výraz

$$\frac{\lambda - \overline{X}_n}{\overline{X}_n} \sqrt{n}(\overline{X}_n - \lambda) + \frac{\lambda - \overline{X}_n}{\overline{X}_n \sqrt{n}} (-X_1 + 2\lambda + X_n).$$

Jelikož $(-X_1 + 2\lambda + X_n)$ je konečná náhodná veličina, pak ze vztahu 2.3 vyplývá, že

$$\frac{\lambda - \overline{X}_n}{\overline{X}_n \sqrt{n}} (-X_1 + 2\lambda + X_n) \xrightarrow{P} 0, \qquad \text{pro } n \to \infty.$$

Obdobně ze vztahů 2.2, 2.3, Cramérovy-Sluckého věty a vztahů konvergencí v pravděpodobnosti a v distribuci plyne

$$\frac{\lambda - \overline{X}_n}{\overline{X}_n} \sqrt{n}(\overline{X}_n - \lambda) \xrightarrow{P} 0, \qquad \text{pro } n \to \infty.$$

Tudíž

$$S_n - A_n \xrightarrow{P} 0, \qquad \text{pro } n \to \infty. \tag{2.4}$$

Zbývá ukázat konvergenci výrazu $A_n$. Jelikož z nezávislosti veličin $\{X_t\}$ platí

$$\mathsf{E}\left[(X_{t-1} - \lambda)(X_t - \lambda)\right] = \mathsf{E}\,(X_{t-1} - \lambda)\,\mathsf{E}\,(X_t - \lambda) = (\lambda - \lambda)(\lambda - \lambda) = 0,$$

je součet v $A_n$ součtem náhodných veličin $Y_t := (X_{t-1} - \lambda)(X_t - \lambda)$, které tvoří striktně stacionární řadu. Veličiny $Y_t$ jsou zároveň 1-závislé, viz definice 6. Z centrální limitní věty pro posloupnost $m$-závislých centrovaných náhodných veličin (věta 7) platí

$$\frac{1}{\sqrt{n}} \sum_{t=2}^{n} Y_t \xrightarrow[n \to \infty]{d} N(0, \Delta^2),$$

přičemž

$$n \operatorname{var} \overline{Y}_n \xrightarrow[n\to\infty]{d} \Delta^2.$$

Počítejme

$$n \operatorname{var} \overline{Y}_n = \frac{1}{n} \operatorname{var} \left( \sum_{t=2}^{n} Y_t \right) = \frac{1}{n} \left( \sum_{t=2}^{n} \operatorname{var} Y_t + 2 \sum_{t=2}^{n-1} \operatorname{cov}(Y_t, Y_{t-1}) \right), \tag{2.5}$$

kde z nezávislosti $\{X_t\}$ plyne

$$\begin{aligned} \operatorname{cov}(Y_t, Y_{t-1}) &= \mathsf{E}\,(Y_t Y_{t-1}) - \mathsf{E}\; Y_t\, \mathsf{E}\; Y_{t-1} = \\ &= \mathsf{E}\; \left[ (X_{t-1} - \lambda)(X_t - \lambda)^2 (X_{t+1} - \lambda) \right] - 0 = \\ &= \mathsf{E}\,(X_{t-1} - \lambda)\, \mathsf{E}\,(X_t - \lambda)^2\, \mathsf{E}\,(X_{t+1} - \lambda) = \\ &= 0, \end{aligned}$$

neboť $\{Y_t\}$ a $\{X_t - \lambda\}$ jsou posloupnosti centrovaných náhodných veličin. Současně $\{Y_t\}$ je posloupnost stejně rozdělených veličin, neboť $\{X_t\}$ jsou stejně rozdělené. Pak

$$\begin{aligned} \operatorname{var} Y_t &= \operatorname{var} \left[ (X_{t-1} - \lambda)(X_t - \lambda) \right] = \\ &= \operatorname{var} X_{t-1} X_t + \lambda^2 \operatorname{var} X_{t-1} + \lambda^2 \operatorname{var} X_t + 2 \operatorname{cov}(X_{t-1} X_t, -\lambda X_{t-1}) + \\ &+ 2 \operatorname{cov}(X_{t-1} X_t, -\lambda X_t). \end{aligned}$$

Jelikož $\{X_t\}$ jsou stejně rozdělené a nezávislé náhodné veličiny, lze předpis rozptylu $Y_t$ dále upravit:

$$\operatorname{var} Y_t = \operatorname{var} X_{t-1} X_t + 2\lambda^2 \operatorname{var} X_t - 4\lambda \operatorname{cov}(X_{t-1} X_t, X_t). \tag{2.6}$$

Vypočítáme jednotlivé členy výrazu 2.6. Platí

$$\begin{aligned} \operatorname{var} X_{t-1} X_t &= \mathsf{E}\,(X_{t-1} X_t)^2 - (\mathsf{E}\; X_{t-1} X_t)^2 = \\ &= (\mathsf{E}\; X_1^2)^2 - (\mathsf{E}\; X_1)^2 = \\ &= \lambda^2 (\lambda + 1)^2 - \lambda^4 = \\ &= \lambda^2 (2\lambda + 1), \end{aligned}$$

dále z vlastností Poissonova rozdělení $\operatorname{var} X_t = \lambda$ a nakonec

$$\begin{aligned} \operatorname{cov}(X_{t-1} X_t, X_t) &= \mathsf{E}\; X_{t-1} X_t^2 - \mathsf{E}\; (X_{t-1} X_t)\, \mathsf{E}\; X_t = \\ &= \lambda^2 (\lambda + 1) - \lambda^3 = \\ &= \lambda^2. \end{aligned}$$

Dosazením do výrazu 2.6 získáme

$$\operatorname{var} Y_t = \lambda^2 (2\lambda + 1) + 2\lambda^3 - 4\lambda^3 = \lambda^2.$$

Tudíž pro výraz 2.5 platí

$$n \operatorname{var} \overline{Y}_n = \frac{n-1}{n} (\lambda^2 - 0) \xrightarrow[n\to\infty]{} \lambda^2.$$

Pak použitím vztahu 2.3 a Cramérovy-Sluckého věty získáme

$$A_n \xrightarrow[n\to\infty]{d} N(0,1),$$

tudíž díky vztahu 2.4 konverguje i testová statistika $S_n$

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

□

Popsaný test má asymptotickou přesnost, přičemž je obvykle používán jako test jednostranný. Na základě věty 5 a předpokladu, že za alternativy $H_1$ lze řadu $\{X_t\}$ popsat pomocí modelu PoINAR(1), kde

$$\rho_k = a^k > 0, \qquad a \in (0,1),\ k = 1, ..., n-1,$$

můžeme formalizovat pravidlo rozhodování testu. Pro zvolenou testovou hladinu $\alpha$ hypotézu zamítneme ve prospěch alternativy pro velké hodnoty testové statistiky:

$$H_0 \text{ zamítáme právě tehdy, když } S_n > u_{1-\alpha},$$

kde $u_\alpha$ je $\alpha$-kvantil normálního rozdělení $N(0,1)$.

Alternativně lze také využít testových statistik konstruovaných na základě odhadů autokorelační funkce $\hat{\rho}_k$ a $\dot{\rho}_k$

$$\hat{S}_n = \sqrt{n}\hat{\rho}_1, \qquad \dot{S}_n = \sqrt{n}\dot{\rho}_1,$$

neboť pro tyto statistiky za nulové hypotézy $H_0$ za použití věty 2 a Cramérovy-Sluckého věty platí

$$\hat{S}_n \xrightarrow[n\to\infty]{d} N(0,1), \qquad \dot{S}_n \xrightarrow[n\to\infty]{d} N(0,1).$$

Tudíž také zamítáme $H_0$ právě tehdy, když $\hat{S}_n > u_{1-\alpha}$, resp. $\dot{S}_n > u_{1-\alpha}$.

*Příklad.* Uvažujme data popálenin. Položme $\alpha = 0{,}05$. Pak s využitím tabulky 1.1 víme, že

$$S_{120} = \sqrt{120}\widetilde{\rho_1} = 2{,}240.$$

A obdobně

$$\hat{S}_{120} = 2{,}414, \qquad \dot{S}_{120} = 2{,}259.$$

Jelikož $u_{1-\alpha} = 1.645$, pro všechny použité testové statistiky zamítáme nulovou hypotézu ve prospěch alternativy, neboli se jedná o data pocházející z časové řady závislých náhodných veličin. P-hodnoty těchto testů udává předpis $p(t) = 1-\mathsf{F}(t)$, kde $\mathsf{F}(t)$ je distribuční funkce normovaného normálního rozdělení a $t$ je realizovaná hodnota příslušné testové statistiky. Pak $p(S_{120}) = 0{,}013$, $p(\hat{S}_{120}) = 0{,}008$ a $p(S_{120}) = 0{,}012$.

## 2.3 Kontingenční tabulky

Poslední skupina uvažovaných testů využívá popisu testovaných řad pomocí kontingenčních tabulek.

Pro časovou řadu $\{X_t,\, t = 1, ..., n\}$ náhodných veličin s Poissonovým rozdělením s parametrem $\lambda > 0$ uvažujme posloupnost $\{Y_t,\, t = 1, ..., n\}$, kde

$$Y_t = \mathbf{1}_{[X_t>0]}.$$

Pak

$$Y_t \sim Alt(p),$$

kde $p = \mathsf{P}(X_1 = 0) = \mathrm{e}^{-\lambda}$.

Neboť $Y_t$ je funkcí pouze $X_t$ pro všechna $t \in \{1, ..., n\}$, lze testování platnosti hypotézy $H_0$ oproti alternativě $H_1$ pro časovou řadu $\{X_t\}$ provést alternativním způsobem. Položme pro posloupnost $\{Y_t\}$ upravenou hypotézu

$$\dot{H}_0 : Y_t \ \sim \ iid$$

a upravenou alternativu

$\dot{H}_1$: $\{Y_t\}$ jsou stejně rozdělené, ale nejsou závislé náhodné veličiny.

Pak provedeme test platnosti hypotézy $\dot{H}_0$ proti alternativě $\dot{H}_1$ pro posloupnost $\{Y_t\}$.

Položme

$$n_{jk} = \sum_{t=1}^{n-1} \mathbf{1}\{Y_t = j, Y_{t+1} = k\} \qquad j, k \in \{0, 1\},$$

které udává počet všech dvojic takových, že $Y_t = j$ a $Y_{t+1} = k$ pro všechna $t = 1, ..., n-1$. Pak pro posloupnost náhodných veličin $\{Y_t\}$ za hypotézy $\dot{H}_0$ platí, že

$$\mathsf{P}(Y_t = j) = \mathsf{P}(Y_t = j|Y_{t-1} = 0) = \mathsf{P}(Y_t = j|Y_{t-1} = 1), \qquad j \in \{0, 1\},$$

neboli

$$\mathsf{P}(Y_t = j, Y_{t+1} = k) = \mathsf{P}(Y_t = j)\,\mathsf{P}(Y_{t+1} = k), \qquad j, k \in \{0, 1\}. \tag{2.7}$$

Pak za předpokladu, že $\{Y_t\}$ tvoří homogenní Markovův řetězec, můžeme sestrojit test poměrem věrohodnosti. Sdružené rozdělení $Y_1, ..., Y_n$ je dáno

$$\mathsf{P}\left(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n\right) = \mathsf{P}\left(Y_1 = y_1\right) \prod_{j,k} p_{jk}^{n_{jk}},$$

kde $p_{jk} = \mathsf{P}(Y_t = j, Y_{t+1} = k)$. Označme $p = (p_{00}, p_{01}, p_{10}, p_{11})$ a uvažujme logaritmickou věrohodnost podmíněnou počátečním stavem, tedy

$$l_n(p) = \sum_{j=0}^{1} \sum_{k=0}^{1} n_{jk} \log p_{jk}.$$

| | $Y_{t+1}=0$ | $Y_{t+1}=1$ | $\sum$ |
|---|---|---|---|
| $Y_t=0$ | $p_{00}$ | $p_{01}$ | $p_{0+}$ |
| $Y_t=1$ | $p_{10}$ | $p_{11}$ | $p_{1+}$ |
| $\sum$ | $p_{+0}$ | $p_{+1}$ | 1 |

Tabulka 2.1: Kontingenční tabulka pravděpodobností.

Jelikož platí $p_{00}+p_{01}+p_{10}+p_{11}=1$, spočteme odhad $p_{jk}$ pomocí metody Lagrangeových multiplikátorů. Položme

$$f(p) = l_n(p) - \gamma\left(p_{00}+p_{01}+p_{10}+p_{11}-1\right),$$

pak

$$\frac{\partial f(p)}{\partial p_{jk}} = \frac{n_{jk}}{p_{jk}} - \gamma.$$

Řešením rovnice $\frac{\partial f(p)}{\partial p_{jk}}=0$ získáváme odhad hodnoty $p_{jk}$ ve tvaru

$$\hat{p}_{jk} = \frac{n_{jk}}{\gamma},$$

přičemž díky vlastnosti součtu pravděpodobností $p_{jk}$, viz tabulka 2.1, musí platit

$$1 = \sum_{j,k}\hat{p}_{jk} = \sum_{j,k}\frac{n_{jk}}{\gamma},$$

tudíž $\gamma = \sum_{j,k} n_{jk} = n-1$. Výsledný odhad je

$$\hat{p}_{jk} = \frac{n_{jk}}{n-1}, \qquad \hat{p} = \left(\hat{p}_{00},\hat{p}_{01},\hat{p}_{10},\hat{p}_{11}\right).$$

Obdobným postupem za nulové hypotézy $H_0$ získáme s pomocí vztahu (2.7) odhad

$$\widetilde{p}_{jk} = \frac{n_{j+}n_{+k}}{(n-1)^2}, \qquad \widetilde{p} = \left(\widetilde{p}_{00},\widetilde{p}_{01},\widetilde{p}_{10},\widetilde{p}_{11}\right),$$

kde

$$n_{j+} = \sum_{k=0}^{1} n_{jk}, \qquad n_{+k} = \sum_{j=0}^{1} n_{jk}.$$

Pak test poměrem věrohodností je dán testovou statistikou

$$\begin{aligned} LR_n &= 2\left(l_n(\hat{p}) - l_n(\widetilde{p})\right) = \\ &= 2\left(\sum_{j,k} n_{jk}\log\left(\frac{n_{jk}}{n-1}\right) - \sum_{j,k} n_{jk}\log\left(\frac{n_{j+}n_{+k}}{(n-1)^2}\right)\right) = \\ &= 2\sum_{j=0}^{1}\sum_{k=0}^{1} n_{jk}\log\left(\frac{n_{jk}(n-1)}{n_{j+}n_{+k}}\right), \end{aligned} \tag{2.8}$$

| | $Y_{t+1}=0$ | $Y_{t+1}=1$ | $\sum$ |
|---|---|---|---|
| $Y_t=0$ | $n_{00}$ | $n_{01}$ | $n_{0+}$ |
| $Y_t=1$ | $n_{10}$ | $n_{11}$ | $n_{1+}$ |
| $\sum$ | $n_{+0}$ | $n_{+1}$ | $n-1$ |

Tabulka 2.2: Kontingenční tabulka.

která má za platnosti nulové hypotézy $\dot{H}_0$ asymptotické rozdělení $\chi^2$ s $3-2=1$ stupněm volnosti, viz Anděl (2007, strana 179). Jedná se o jednostranný test, kdy zamítáme nulovou hypotézu, pokud platí

$$LR_n \geq \chi^2_1(1-\alpha),$$

kde $\chi^2_1(1-\alpha)$ je $(1-\alpha)$-kvantil $\chi^2$-rozdělení o 1 stupni volnosti.

Alternativně je možné testovat platnost hypotézy $\dot{H}_0$ pomocí $\chi^2$-*testu nezávislosti v kontingenční tabulce*, viz tabulka 2.2. Testová statistika $\chi^2$ je dána předpisem

$$\chi^2 = \sum_{j=0}^{1}\sum_{k=0}^{1} \frac{(n_{jk} - \frac{n_{j+}n_{+k}}{n-1})^2}{\frac{n_{j+}n_{+k}}{n-1}},$$

viz Anděl (2007, kapitola 13).

Za platnosti hypotézy $\dot{H}_0$ má testová statistika asymptotické rozdělení $\chi^2$ s 1 stupněm volnosti, viz Billingsley (1961, strana 26). Jedná se o jednostranný test, hypotézu $\dot{H}_0$, a tedy i hypotézu $H_0$, zamítáme pro velké hodnoty testové statistiky, tedy pokud pro danou testovou hladinu $\alpha$ platí

$$\chi^2 \geq \chi^2_1(1-\alpha),$$

kde $\chi^2_1(1-\alpha)$ je $(1-\alpha)$-kvantil $\chi^2$-rozdělení o 1 stupni volnosti.

*Příklad.* Uvažujme data popálenin. Sestavíme kontingenční tabulku, viz tabulka 2.3. Následně vypočteme testové statistiky $LR_n$ a $\chi^2$. Pro tato data vychází $LR_n = 4{,}873$, $\chi^2 = 5{,}691$ s p-hodnotami $p(LR_n) = 1 - \mathsf{F}(LR_n) = 0.027$ a $p(\chi^2) = 0.017$, kde $\mathsf{F}(t)$ je distribuční funkce $\chi^2$-rozdělení o 1 stupni volnosti v realizované hodnotě testové statiky $t$. Za zvolené testové hladiny $\alpha = 0{,}05$ dále platí, že $\chi^2_1(1-\alpha) = 3{,}841$, tudíž i oba tyto testy zamítají $H_0$ ve prospěch alternativy. Počty pracovních neschopností způsobených popáleninami tedy na sobě v po sobě jdoucích měsících závisí.

| | $Y_{t+1}=0$ | $Y_{t+1}=1$ | $\sum$ |
|---|---|---|---|
| $Y_t=0$ | 86 | 13 | 99 |
| $Y_t=1$ | 13 | 7 | 20 |
| $\sum$ | 99 | 20 | 119 |

Tabulka 2.3: Kontingenční tabulka pro data popálenin.

# 3. Simulační studie

V této kapitole porovnáme chování testů popsaných v předchozí kapitole ve dvou situacích – za nulové hypotézy $H_0$ a za alternativy $H_1$ popsané modelem PoINAR(1). Následně na základě dosažených výsledků provedeme diskuzi o vhodnosti zkoumaných testů pro jednotlivé uvažované situace. Pro přehlednost jsou jednotlivé testy v tabulkách výsledků označovány názvy svých testových statistik zavedených v předchozí kapitole – test jednoduchých iterací je reprezentován statistikou $Z^*$, testy založené na odhadu autokorelace popisují statistiky $S_n$, $\hat{S}_n$ a $\dot{S}_n$ a testy založené na kontingenční tabulce zastupují statistiky $\chi^2$ ($\chi^2$-test) a $LR_n$ (test poměrem věrohodnosti).

Veškeré testy v této kapitole byly provedeny za nastavení *set.seed(0505)* a jsou prováděny na hladině $\alpha = 0{,}05$. Bylo provedeno 1000 opakování simulací pro každé nastavení parametrů modelu a pro každou volbu $n$, kde $n$ je počet hodnot v dané časové řadě. Uvažujeme $n \in \{50, 100, 200, 500\}$. Hodnoty uvedené v tabulkách byly vypočteny z celkového počtu úspěšně proběhlých testů pro danou simulaci.

## 3.1 Situace za nulové hypotézy

Uvažujme $\{X_t\}$ posloupnost nezávislých stejně rozdělených náhodných veličin s Poissonovým rozdělením s parametrem $\lambda$. Uvažujeme $\lambda \in \{0{,}2, 0{,}5, 1{,}0\}$.

Jak ukazují tabulky 3.1, 3.2 a 3.3, veškeré uvažované testy přibližně dosahují požadované testové hladiny $\alpha$, přičemž sledujeme, že s rostoucím $n$ se hladiny testů blíží k předepsané hladině $\alpha = 0{,}05$. Typicky nejvíce konzervativních hodnot dosahují testy založené na odhadu autokorelace pro všechny možnosti testových statistik $S_n$, $\hat{S}_n$ a $\dot{S}_n$. Dále pozorujeme, že pro malé $n$ je skutečná hladina testu nejblíže předepsané hladině $\alpha$ pro test založený na iteracích.

Naopak pro $\lambda \in \{0{,}5, 1{,}0\}$ dosáhly nejvyšších antikonzervativních výsledků testy založené na kontingenční tabulce s testovými statistikami $\chi^2$ a $LR_n$ pro $n = 100$, $\lambda = 1{,}0$. Pro hodnotu parametru $\lambda = 0{,}2$ dosáhl nejvyšších antikonzervativních výsledků test založený na iteracích.

Současně až na $LR_n$-test veškeré uvažované testy úspěsně proběhly ve všech simulacích. $LR_n$-test dle předpisu (2.8) své testové statistiky $LR_n$, neproběhne, pokud libovolné $n_{jk} = 0$. Toto nejsnáze nastává pro nejkratší časové řady ($n = 50$), konkrétně neproběhl třikrát pro volbu $\lambda = 1{,}0$ a jedenkrát pro $\lambda = 0{,}5$. K výraznému nárůstu chybovosti došlo pro volbu $\lambda = 0{,}2$, kdy pro $n = 50$ neproběhlo 270 z 1000 testů, dále 56 pro $n = 100$ a 3 pro $n = 200$.

| $n$ | $Z^*$ | $S_n$ | $\hat{S}_n$ | $\dot{S}_n$ | $\chi^2$ | $LR_n$ |
|---|---|---|---|---|---|---|
| 50 | 0,052 | 0,034 | 0,026 | 0,040 | 0,038 | 0,043 |
| 100 | 0,053 | 0,038 | 0,036 | 0,039 | 0,061 | 0,065 |
| 200 | 0,049 | 0,039 | 0,037 | 0,039 | 0,049 | 0,050 |
| 500 | 0,048 | 0,045 | 0,039 | 0,046 | 0,042 | 0,042 |

Tabulka 3.1: Tabulka hladin testů za nulové hypotézy $H_0$ pro $\lambda = 1{,}0$.

| $n$ | $Z^*$ | $S_n$ | $\hat{S}_n$ | $\dot{S}_n$ | $\chi^2$ | $LR_n$ |
|---|---|---|---|---|---|---|
| 50 | 0,045 | 0,031 | 0,030 | 0,036 | 0,039 | 0,041 |
| 100 | 0,047 | 0,032 | 0,026 | 0,033 | 0,044 | 0,045 |
| 200 | 0,041 | 0,036 | 0,036 | 0,036 | 0,043 | 0,042 |
| 500 | 0,053 | 0,050 | 0,053 | 0,051 | 0,047 | 0,047 |

Tabulka 3.2: Tabulka hladin testů za nulové hypotézy $H_0$ pro $\lambda = 0{,}5$.

| $n$ | $Z^*$ | $S_n$ | $\hat{S}_n$ | $\dot{S}_n$ | $\chi^2$ | $LR_n$ |
|---|---|---|---|---|---|---|
| 50 | 0,067 | 0,039 | 0,039 | 0,042 | 0,032 | 0,025 |
| 100 | 0,053 | 0,037 | 0,044 | 0,037 | 0,036 | 0,029 |
| 200 | 0,037 | 0,036 | 0,037 | 0,036 | 0,040 | 0,054 |
| 500 | 0,062 | 0,057 | 0,054 | 0,057 | 0,058 | 0,060 |

Tabulka 3.3: Tabulka hladin testů za nulové hypotézy $H_0$ pro $\lambda = 0{,}2$.

## 3.2 Situace za alternativy modelu PoINAR(1)

Uvažujme časovou řadu $\{X_t\}$ generovanou modelem PoINAR(1), viz sekce 1.3. Simulace provedeme pro hodnoty parametru $a \in \{0{,}2, 0{,}5, 0{,}7\}$, přičemž $X_t \sim Po(\lambda)$, kde, obdobně jako za nulové hypotézy $H_0$, $\lambda \in \{0{,}2, 0{,}5, 1{,}0\}$. Pro každou simulaci pak na základě tvrzení 4 hodnotu parametru $\beta$ vypočítáme.

Pro dané $\lambda$ a dané $a$ v tabulkách 3.4, 3.5 a 3.6 pozorujeme růst síly testů s rostoucím $n$. Pro $\lambda \in \{0{,}5, 1{,}0\}$ tvoří nejsilnější skupinu testů testy založené na odhadu autokorelace, přičemž mezi nimi vyniká $\hat{S}_n$-test. Následovány jsou testem jednoduchých iterací s testovou statistikou $Z^*$, přičemž, jak naznačuje tabulka 3.6, pro velmi nízké hodnoty parametru $\lambda$ bude $Z^*$-test silnější než testy založené na odhadech autokorelace. Naopak z uvažovaných testů jsou slabší testy založené na kontingenčních tabulkách, a to zejména pro krátké časové řady.

Z provedených simulací vyplývá, že sílu testu výrazně ovlivňuje hodnota parametru $a$, přičemž tato pozorování se mezi jednotlivými uvažovanými hodnotami $\lambda$ výrazně neliší. Zatímco pro $a$ blízké 1 všechny uvažované testy vykazují vysokou sílu i pro krátké časové řady ($n = 50$), pro $a = 0{,}2$ a délku řad $n = 50$ síla testu nepřesáhla hodnotu 0,4 pro žádný z testů, u testů v kontingenčních tabulkách dokonce nedosáhla ani hodnoty 0,25. Právě pravděpodobnost chyby druhého druhu za $a = 0{,}2$ většina testů výrazně sníží až pro časové řady délky $n = 500$.

Obdobně jako za nulové hypotézy $H_0$, i za alternativy $H_1$ v určitých situacích některé testy neproběhnou. Pro $\lambda \in \{0{,}5, 1{,}0\}$ tyto komplikace pozorujeme pouze pro $n = 50$ v jednotkách případů u testů založených na kontingenčních tabulkách s testovými statistikami $\chi^2$ a $LR_n$, přičemž $\chi^2$-test lze označit jako odolnější vůči selhání než $LR_n$-test, neboť neproběhne právě tehdy, když $n_{j+} = 0$ nebo $n_{+k} = 0$. Se snižováním parametru $\lambda$ pozorujeme nárůst počtu neproběhlých testů, neboť žádný z uvažovaných testů neproběhne, pokud je testovaná časová řada tvořena pouze posloupností nulových hodnot. Dále pro dané $\lambda$ tento počet roste s parametrem závislosti $a$ (viz sekce 1.3), jak ukazuje tabulka 3.7.

| $a$ | $n$ | $Z^*$ | $S_n$ | $\hat{S}_n$ | $\dot{S}_n$ | $\chi^2$ | $LR_n$ |
|---|---|---|---|---|---|---|---|
| 0,2 | 50 | 0,237 | 0,313 | 0,317 | 0,322 | 0,103 | 0,103 |
| | 100 | 0,366 | 0,525 | 0,551 | 0,532 | 0,234 | 0,225 |
| | 200 | 0,581 | 0,818 | 0,837 | 0,818 | 0,416 | 0,416 |
| | 500 | 0,897 | 0,992 | 0,992 | 0,992 | 0,796 | 0,794 |
| 0,5 | 50 | 0,806 | 0,830 | 0,918 | 0,837 | 0,671 | 0,667 |
| | 100 | 0,953 | 0,993 | 0,999 | 0,994 | 0,956 | 0,951 |
| | 200 | 0,987 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| | 500 | 0,987 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| 0,7 | 50 | 0,971 | 0,954 | 0,997 | 0,960 | 0,965 | 0,958 |
| | 100 | 0,984 | 1,000 | 1,000 | 1,000 | 0,998 | 0,998 |
| | 200 | 0,991 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| | 500 | 0,993 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |

Tabulka 3.4: Souhrnná tabulka síly testů za alternativy $H_1$ posloupnosti náhodných veličin z $Po(1{,}0)$.

| $a$ | $n$ | $Z^*$ | $S_n$ | $\hat{S}_n$ | $\dot{S}_n$ | $\chi^2$ | $LR_n$ |
|---|---|---|---|---|---|---|---|
| 0,2 | 50 | 0,292 | 0,294 | 0,305 | 0,303 | 0,167 | 0,167 |
| | 100 | 0,489 | 0,536 | 0,550 | 0,541 | 0,335 | 0,330 |
| | 200 | 0,688 | 0,780 | 0,796 | 0,781 | 0,547 | 0,543 |
| | 500 | 0,974 | 0,989 | 0,991 | 0,990 | 0,937 | 0,937 |
| 0,5 | 50 | 0,903 | 0,826 | 0,872 | 0,834 | 0,814 | 0,811 |
| | 100 | 0,998 | 0,991 | 0,996 | 0,992 | 0,992 | 0,991 |
| | 200 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| | 500 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| 0,7 | 50 | 0,986 | 0,955 | 0,984 | 0,962 | 0,981 | 0,975 |
| | 100 | 1,000 | 0,999 | 0,999 | 0,999 | 0,999 | 0,999 |
| | 200 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| | 500 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |

Tabulka 3.5: Souhrnná tabulka síly testů za alternativy $H_1$ posloupnosti náhodných veličin z $Po(0{,}5)$.

| $a$ | $n$ | $Z^*$ | $S_n$ | $\hat{S}_n$ | $\dot{S}_n$ | $\chi^2$ | $LR_n$ |
|---|---|---|---|---|---|---|---|
| 0,2 | 50 | 0,378 | 0,269 | 0,309 | 0,292 | 0,217 | 0,192 |
| | 100 | 0,539 | 0,505 | 0,523 | 0,505 | 0,401 | 0,359 |
| | 200 | 0,749 | 0,746 | 0,772 | 0,750 | 0,637 | 0,597 |
| | 500 | 0,972 | 0,976 | 0,977 | 0,976 | 0,952 | 0,949 |
| 0,5 | 50 | 0,854 | 0,773 | 0,824 | 0,793 | 0,788 | 0,764 |
| | 100 | 0,973 | 0,954 | 0,960 | 0,954 | 0,953 | 0,939 |
| | 200 | 1,000 | 1,000 | 1,000 | 1,000 | 0,998 | 0,996 |
| | 500 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| 0,7 | 50 | 0,935 | 0,912 | 0,919 | 0,915 | 0,920 | 0,926 |
| | 100 | 0,990 | 0,986 | 0,989 | 0,986 | 0,985 | 0,985 |
| | 200 | 0,999 | 0,999 | 0,999 | 0,999 | 0,999 | 0,998 |
| | 500 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |

Tabulka 3.6: Souhrnná tabulka síly testů za alternativy $H_1$ posloupnosti náhodných veličin z $Po(0{,}2)$.

| $a$ | $n$ | $Z^*$ | $S_n$ | $\hat{S}_n$ | $\dot{S}_n$ | $\chi^2$ | $LR_n$ |
|---|---|---|---|---|---|---|---|
| 0,2 | 50 | | | | | | 92 |
| | 100 | | | | | | 9 |
| 0,5 | 50 | 4 | 4 | 4 | 4 | 5 | 61 |
| | 100 | | | | | 1 | 6 |
| 0,7 | 50 | 52 | 52 | 52 | 52 | 54 | 131 |
| | 100 | 3 | 3 | 3 | 3 | 3 | 14 |

Tabulka 3.7: Tabulka počtů selhání testů za alternativy $H_1$ pro posloupnosti náhodných veličin z $Po(0{,}2)$.

# Závěr

V této práci jsme se zabývali testováním nezávislosti v časových řadách stejně rozdělených náhodných veličin s Poissonovým rozdělením a popsali jsme test jednoduchých iterací, testy založené na odhadu autokorelace pro tři různé odhady a testy založené na kontingenční tabulce – $\chi^2$-test a test poměrem věrohodnosti. Ilustraci jsme provedli na příkladu měsíčních počtů nahlášených případů pracovní neschopnosti způsobených popáleninami evidovaných agenturou British Columbia Workers Compensation Board v období od ledna 1987 do prosince 1994.

Popsané testy jsme následně porovnali v simulační studii, ve které jsme dospěli k následujícím poznatkům. Pro hypotézu $H_0$ i alternativu $H_1$ a libovolné hodnoty $n$ lze jako nejvíce univerzální test doporučit test založený na odhadu autokorelace, přičemž mezi jednotlivými testovými statistikami $S_n$, $\hat{S}_n$ a $\dot{S}_n$ jsou pouze zanedbatelné rozdíly. Z této trojice testů vychází test daný testovou statistikou $\hat{S}_n$ jako nejvíce konzervativní za $H_0$ a nejsilnější za $H_1$.

Naopak nejméně vhodné, především pro krátké časové řady nebo řady náhodných veličin s Poissonovým rozdělením, jehož parametr se blíží nule, se za hypotézy $H_0$ i alternativy $H_1$ ukázaly být testy založené na kontingenčních tabulkách, neboť z popisovaných testů nejčastěji neproběhnou a mají znatelně nižší sílu.

# Seznam použité literatury


Anděl, J. (2007). *Základy matematické statistiky*. Matfyzpress, Praha, 2. opravené vydání. ISBN 80-7378-001-1.

Billingsley, P. (1961). Statistical methods in markov chains. *The Annals of mathematical statistics*, **32**(1), 12–40. ISSN 0003-4851.

Cipra, T. (2008). *Finanční ekonometrie*. Ekopress, Praha, 1. vydání. ISBN 978-80-86929-43-9.

Dupač, V. a Hušková, M. (2003). *Pravděpodobnost a matematická statistika*. Karolinum, Praha, 1. vydání. ISBN 80-246-0009-9.

Freeland, R. K. (1998). *Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data*. PhD thesis, University of British Columbia. URL `https://open.library.ubc.ca/collections/ubctheses/831/items/1.0088709`.

Gibbons, J. D. a Chakraborti, S. (2003). *Nonparametric statistical inference*. Dekker, New York, 4. opravené a rozšířené vydání. ISBN 0-8247-4052-1.

Grunwald, G. K., Hyndman, R. J., Tedesco, L. a Tweedie, R. L. (2000). Non-gaussian conditional linear AR(1) models. *Australian and New Zealand journal of statistics*, **42**(4), 479–495. ISSN 1369-1473.

Jung, R. C. a Tremayne, A. R. (2003). Testing for serial dependence in time series models of counts. *Journal of time series analysis*, **24**(1), 65–84. ISSN 0143-9782.

Jung, R. C. a Tremayne, A. R. (2006). Binomial thinning models for integer time series. *Statistical modelling*, **6**(2), 81–96. ISSN 1471-082X.

Lain, M. (2020). *Robustní odhady autokorelační funkce*. Diplomová práce, Univerzita Karlova, Matematicko-fyzikální fakulta, Katedra pravděpodobnosti a matematické statistiky, Praha.

Prášková, Z. (2007). *Základy náhodných procesů. II.* Učební texty Univerzity Karlovy v Praze. Karolinum, Praha, 1. vydání. ISBN 978-80-246-0971-3.

Steutel, F. a Harn, van, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of probability*, **7**(5), 893–899. ISSN 0091-1798.

Wald, A. a Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of mathematical statistics*, **11**(2), 147–162. ISSN 0003-4851.

# Seznam obrázků

# Seznam tabulek

# A. Dodatky

V této kapitole uvedeme pro úplnost některé definice a tvrzení, které jsou využity v důkazech v průběhu této práce, avšak samy o sobě nejsou podstatné pro probíraná témata.

V důkazu tvrzení 3 využíváme následující větu, viz Prášková (2007, strana 80).

**Věta 6.** *Nechť* $\{X_t, t \in \mathbb{Z}\}$ *je reálná stacionární posloupnost se střední hodnotou* $\mu$ *a autokovarianční funkcí* $\gamma$*, pro kterou* $\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$*. Pak pro* $n \to \infty$ *platí*

$$\overline{X}_n \xrightarrow[n\to\infty]{P} \mu.$$

V důkazu věty 5 využíváme následující definici a větu, převzaté z Prášková (2007, strana 87).

**Definice 6.** *Nechť* $\{X_t, t = 1, ..., n\}$ *je časová řada. Pak řekneme, že náhodné veličiny této časové řady jsou* $m$*-závislé, kde* $m \in \mathbb{N}_0$ *je dané číslo, jestliže pro všechna* $t = 1, ..., n$ *jsou náhodné vektory* $(..., X_{t-1}, X_t)$ *a* $(X_{t+m+1}, X_{t+m+2}, ...)$ *nezávislé.*

**Věta 7.** *Nechť* $\{X_t, t = 1, ..., n\}$ *je časová řada reálných, centrovaných, striktně stacionárních,* $m$*-závislých náhodných veličin s konečnými druhými momenty a autokovarianční funkcí* $\gamma$*, pro kterou* $\Delta_m^2 = \sum_{k=-m}^{m} \gamma_k \neq 0$*.*

*Potom pro* $n \to \infty$

$$n \operatorname{var} \overline{X}_n \to \Delta_m^2, \tag{A.1}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \xrightarrow{d} N(0, \Delta_m^2). \tag{A.2}$$