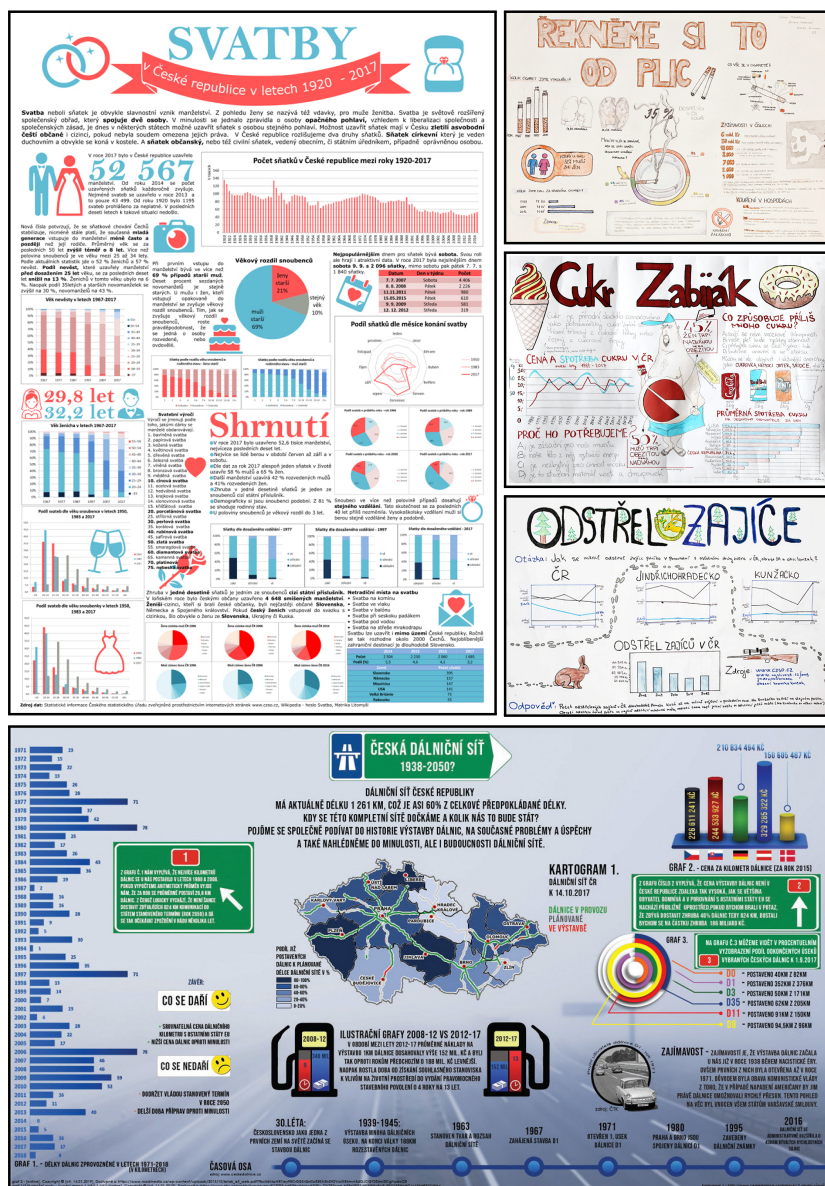




VYUŽITÍ χ^2 TESTŮ V PROSTOROVÉ STATISTICE χ^2 TESTS IN SPATIAL STATISTICS

Martina Petráková, Jiří Dvořák

Adresa: Matematicko-fyzikální fakulta Univerzity Karlovy, Sokolovská 83, 186 75 Praha 8

E-mail: martinapetrakova@centrum.cz, dvorak@karlin.mff.cuni.cz

Abstrakt: V tomto článku představíme několik možností uplatnění známého χ^2 testu při řešení problémů z prostorové statistiky. Nejprve zavedeme pojem bodového procesu a jeho funkce intenzity a představíme dva příklady bodových procesů – binomický a Poissonův bodový proces. Budeme se zabývat testováním, zda pozorovaná data pochází z Poissonova procesu, zda je pro Poissonův proces funkce intenzity konstantní, zda závisí na dané kovariátě, zda je v předem určeném tvaru, závislém pouze na konečném počtu parametrů, či zda je v součinném tvaru. Použití vybraných testů předvedeme pro vhodná reálná data.

Klíčová slova: prostorová statistika, bodový proces, Poissonův proces, binomický proces, funkce intenzity, χ^2 test.

Abstract: We present several ways of using the well-known χ^2 test to solve problems in spatial statistics. First we introduce the notion of a point processes and its intensity function and we provide two examples of point processes – binomial and Poisson point process. We consider the problem of testing the hypotheses that the observed data comes from a Poisson process, that for a Poisson process the intensity function is constant, that it depends on a given covariate, that it has a specified form, depending only on finitely many parameters, or that it is in the product form. For some of the tests we illustrate their use for appropriate real datasets.

Keywords: spatial statistics, point process, Poisson process, binomial process, intensity function, χ^2 test.

1. Úvod

Chceme-li zkoumat data, která zachycují výskyt nějakých jevů v čase (např. příchod zákazníků), v prostoru (např. pozice hnízd daného druhu ptáků) nebo časoprostoru (např. zemětřesení), můžeme využít teorii bodových procesů a odpovídající postupy prostorové statistiky. Bodové procesy popisují náhodné množiny bodů a své využití proto mají v řadě oborů, jmenovitě například v epidemiologii (výskyt nakažených jedinců), ekologii (pozice stromů

daného druhu), medicíně (prostorové uspořádání neuronů), materiálových vědách (výskyt poruch v krystalové mřížce) a mnoha dalších. Pozorovaná data pak tvoří pozice všech bodů procesu, které se vyskytují v zadaném kompaktním pozorovacím okně. Jak poloha pozorovaných bodů tak jejich počet jsou náhodné a je obvyklé, že mezi body dochází k interakcím – body se mohou přitahovat nebo naopak odpuzovat. Tyto vlastnosti (náhodný počet pozorovaných bodů, závislosti mezi jednotlivými body) vedou k tomu, že typicky není možné pro modelování takových dat použít náhodný výběr a k analýze využít postupy klasické statistiky. V tomto článku se zaměříme na případ, kdy tyto postupy využít umíme.

Nejprve představíme dva příklady bodových procesů – *binomický proces* a *Poissonův bodový proces* – a jednu z jejich základních charakteristik, *funkci intenzity*, která popisuje očekávaný počet pozorovaných bodů v dané oblasti. Budeme se zabývat otázkou, co na základě pozorovaných dat umíme říci o funkci intenzity, například zda je možné předpokládat, že je konstantní a bodový proces je tedy tzv. *homogenní*. Jako základní model přitom budeme předpokládat právě Poissonův bodový proces, který má tu důležitou vlastnost, že mezi jeho body nejsou žádné interakce. Přítomnost bodu v daném místě tedy nijak neovlivňuje přítomnost dalších bodů v jeho okolí. Tato vlastnost bude stěžejní k tomu, abychom mohli použít výsledky odvozené pro náhodné výběry a využít známého postupu χ^2 testu na vybrané problémy. Klíčovou roli zde hraje podmíněné rozdělení Poissonova procesu při daném počtu pozorovaných bodů, které je popsáno ve větě 1.

Ve statistice bodových procesů je běžné, že máme k dispozici pouze jednu pozorovanou realizaci, tj. jednu množinu bodů pozorovanou v zadaném pozorovacím okně. Pro tuto situaci v sekci 3 sestavíme formální testy hypotéz, že homogenní bodový proces (tedy proces s konstantní funkcí intenzity) je Poissonův (sekce 3.1.), že je Poissonův bodový proces homogenní (sekce 3.2.), že rozdělení Poissonova procesu závisí na dané kovariátě (sekce 3.3.), že je funkce intenzity Poissonova procesu v daném tvaru závislejícím na neznámých parametrech (sekce 3.4.), nebo že je funkce intenzity Poissonova procesu v součinném tvaru (sekce 3.5.). Některé z těchto testů si také ukážeme na jednoduchých reálných příkladech.

2. Bodové procesy

Bodový proces $\mathbb{X}$ je náhodná lokálně konečná množina bodů v $\mathbb{R}^d$, $d \in \mathbb{N}$. Formálně jde o zobrazení mezi pravděpodobnostním prostorem a prostorem lokálně konečných množin vybaveným vhodnou σ -algebrou. Pro úplnost dodáváme, že pro lokálně konečnou množinu má její průnik s libovolným kompak-

MEZINÁRODNÍ SOUTĚŽ O NEJLEPŠÍ STATISTICKÝ PLAKÁT ISLP POSTER COMPETITION

Michal Novotný, Jakub Fischer

O soutěži

Mezinárodní **soutěž o nejlepší statistický plakát** je vyhlášena v rámci Mezinárodního projektu statistické gramotnosti (International Statistical Literacy Project).

Posláním tohoto projektu je podporovat, vytvářet a spolupodílet se na aktivitách podporujících statistickou gramotnost po celém světě. **Statistickou gramotnost** lze chápat jako znalosti a dovednosti, díky nimž můžeme zpracovávat data, analyzovat je a porozumět statistickým údajům.

Při vytváření statistického plakátu si studenti osvojí sběr statistických dat, způsoby a možnosti jejich zpracování a prezentaci výsledků výzkumu.

Učitelé mohou oživit své vyučovací hodiny a studenti získat zábavným způsobem nové znalosti a dovednosti, pochopit význam statistiky a její praktické využití v každodenním životě. Soutěž může oživit výuku **matematiky, informatiky, výtvarné výchovy a českého či cizího jazyka**, rozvíjí práci v týmu a zájem dětí o svět kolem nás.

V České republice jsou organizátory národního kola soutěže Vysoká škola ekonomická v Praze ve spolupráci s Českým statistickým úřadem.

Mezinárodní projekt statistické gramotnosti pořádá Mezinárodní asociace pro statistické vzdělávání (Int. Association for Statistical Education).

Soutěž o nejlepší statistický plakát (video):

<https://www.youtube.com/watch?v=BsvgqCk303c>

Jak vytvořit statistický plakát (video):

<https://www.youtube.com/watch?v=cP0S3t8aTPo>

Kontaktní informace

Adresa pro zasílání plakátů: Michal Novotný, Český statistický úřad, Na padesátém 81, 100 82 Praha 10 nebo na: statistickyplakat@czso.cz

Informace o soutěži: Michal Novotný, Český statistický úřad, e-mail: michal.novotny@czso.cz

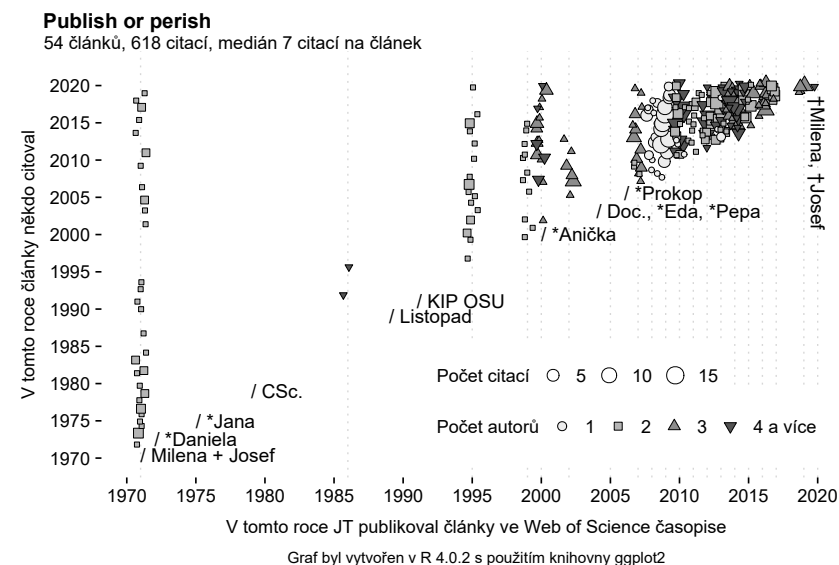
Národní koordinátor soutěže: Jakub Fischer, VŠE v Praze, e-mail: fischerj@vse.cz

ŽIVOT PODLE JOSEFA TVRDÍKA THE LIFE ACCORDING TO JOSEF TVRDÍK

Patrik Galeta

E-mail: galeta@ksa.zcu.cz

Josefa Tvrdíka jsem poznal jako otce jeho dcer, dědu jeho vnoučat a majitele chaty, kam jsme občas jezdili. Až později vyšlo najevo, že oba vedeme, i když na různých úrovních obtížnosti, kurzy (bio)statistiky. Když mi jeho dcery poslaly zprávu, že ztratily otce, napadlo mě odpovědět grafem, jehož upravenou verzi můžete vidět níže. Formou jemu vlastní jsem chtěl uctít člověka pevného základu, jakým je osa X, člověka, který byl rovný a stál vzpřímeně jako osa Y, a to i v dobách, kdy mu v tom bránil režim nebo horšící se zdraví. Graf by možná nevyhověl přísným zásadám vhodné prezentace výsledků (srov. jeho příspěvek v IB 4/2007, str. 6–20). S odkazem na jeho přemýšlivou povahu a smysl pro humor jsem ale doufal, že rozptýlí jeho dcery soustředěné na poslední rozloučení a pod slzami se na chvíli usmějí. Graf je vizuálním podobenstvím životní cesty člověka, během které se proluly osobní radosti s profesními úspěchy (omezenými zde na impaktované články). Dne 23. 11. 2020 ztratil graf osy, Josef Tvrdík odešel. Zanechal ale odkaz, který dovolí graf ještě dlouho doplňovat a aktualizovat.



tem pouze konečný počet bodů a můžeme si ji tedy představit jako množinu izolovaných bodů v prostoru. Ukázky realizací různých bodových procesů v rovině nabízí obrázky 1, 2 a 3.

Zde se omezuje na bodové procesy v euklidovském prostoru. Náhodná je nejen poloha bodů procesu, ale také jejich počet v podmnožinách $\mathbb{R}^d$. Pozorujeme-li tedy nějakou oblast $W \subset \mathbb{R}^d$ (pozorovací okno), pak počet bodů procesu $\mathbb{X}$ v množině W je náhodná veličina, kterou označíme $\mathbb{X}(W)$. V článku se budeme věnovat převážně bodovým procesům v rovině ($d = 2$) a k procesům ve vyšší dimenzi se dostaneme v části 3.5. Všechny zde uvažované množiny budou borelovsky měřitelné.

Jednou ze základních popisných charakteristik bodového procesu je jeho *míra intenzity* Λ , která popisuje očekávaný počet bodů procesu v dané množině. Pro množinu $B \subset \mathbb{R}^d$ pak platí $\Lambda(B) = \mathbb{E}\mathbb{X}(B)$. Pokud je míra intenzity absolutně spojitá vzhledem k d -rozměrné Lebesgueově míře, říkáme její hustotě $\lambda(u)$, $u \in \mathbb{R}^d$, *funkce intenzity* bodového procesu. Platí pak

$$\Lambda(B) = \int_B \lambda(u) du.$$

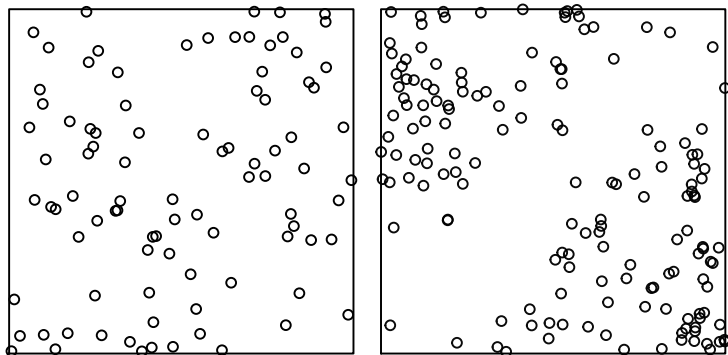
Je-li funkce $\lambda(u)$ konstantní, t.j. $\lambda(u) = \lambda$ pro všechna $u \in \mathbb{R}^d$, hovoříme o *homogenním* bodovém procesu a konstantu λ nazýváme *intenzita* (pro míru intenzity pak platí $\Lambda(B) = \lambda|B|$, kde $|B|$ je Lebesgueova míra množiny B). V opačném případě hovoříme o *nehomogenním* bodovém procesu. Obrázek 1 ukazuje příklad realizace homogenního a nehomogenního bodového procesu a vliv funkce intenzity na rozmístění bodů.

2.1. Binomický bodový proces

Mějme nyní pozorovací okno $W \subset \mathbb{R}^d$ a n nezávislých, stejně rozdělených náhodných vektorů $X_1, \dots, X_n$ s hodnotami v tomto okně. Tyto vektory budou reprezentovat jednotlivé body procesu. Náhodná množina bodů $\{X_1, \dots, X_n\}$ pak představuje (v mnoha ohledech nejjednodušší) příklad bodového procesu, tzv. *binomický proces*. Pro počet bodů $\mathbb{X}(B)$ binomického procesu $\mathbb{X}$ v nějaké množině $B \subset W$ platí

$$\mathbb{X}(B) = \sum_{i=1}^n \mathbb{1}\{X_i \in B\},$$

kde $\mathbb{1}$ značí indikátorovou funkci. Protože X_i jsou nezávislé, stejně rozdělené náhodné vektory, jsou $\mathbb{1}\{X_i \in B\}$ také nezávislé, stejně rozdělené náhodné veličiny s alternativním rozdělením. Proto $\mathbb{X}(B)$ má binomické rozdělení s parametry n a $\Lambda(B)/\Lambda(W)$.



Obrázek 1: Ukázka realizace homogenního (vlevo) a nehomogenního bodového procesu v $\mathbb{R}^2$ (vpravo) v pozorovacím okně $W = [0, 1]^2$. Vlevo vidíme realizaci Poissonova procesu (viz část 2.2.) s intenzitou $\lambda = 100$, vpravo pak realizaci Poissonova procesu s funkcí intenzity $\lambda(x, y) = 500 \cdot |x - y|$.

2.2. Poissonův bodový proces

Dalším, pro tento článek stěžejním příkladem bodového procesu je *Poissonův proces*. Bodový proces $\mathbb{X}$ je Poissonův proces s mírou intenzity Λ , pokud má následující vlastnosti:

- P1: Počet bodů $\mathbb{X}(B)$ v množině $B \subset \mathbb{R}^d$ je náhodná veličina s Poissonovým rozdělením s parametrem $\Lambda(B)$.
- P2: Pro množiny $B_1, \dots, B_m$, $m \in \mathbb{N}$, po dvou disjunktní, jsou náhodné veličiny $\mathbb{X}(B_1), \dots, \mathbb{X}(B_m)$ nezávislé.

V Poissonově procesu nejsou žádné interakce mezi body, ani přitažlivé, ani odpudivé, což je důsledkem nezávislosti počtu bodů v disjunktních množinách. Naopak v binomickém procesu jsou jednotlivé body umísťovány nezávisle na sobě, vzhledem k pevnému celkovému počtu bodů v zadaném okně ale nejsou počty v disjunktních podmnožinách tohoto okna nezávislé. Volně řečeno, pokud už je bod umístěn v jedné podmnožině, nemůže být umístěn v jiné, s ní disjunktní.

Příklad realizace Poissonova a binomického procesu se stejnou intenzitou ukazuje obrázek 2. Následující věta popisuje vztah mezi těmito dvěma bodovými procesy. Stručně řečeno věta říká, že pokud se u Poissonova procesu

rismatu a dílem nutnosti užívání logického myšlení během výuky. Proto jsem bez váhání přistoupil na téma doktorského studia pod jeho taktovkou. Tato příležitost mi nejenže umožnila zkoumat zajímavý obor evolučních algoritmů a prohlubovat základy statistiky, ale rovněž jsem měl možnost intenzivní spolupráce s člověkem, který byl něčím víc než školitelem či později kolegou. Jeho přehled a nadhled pomáhal usměrňovat mé divoké nápady v rámci výuky a výzkumu coby doktoranda. Měl vždy po ruce cenné rady, které přesahovaly meze spolupráce, a pomohly mnoha lidem v různých životních situacích.

Domnívám se, že Josef byl člověk se smyslem pro spravedlnost a čest snad i proto, že většinu života aktivně sportoval. Ke sportu vedl kromě svých dětí i studenty. Osobně vzpomínám, že mi byl schopen tolerovat určité výzkumné nedostatky, pokud jsem je hájil náročným sportovním nasazením. Odsuzoval všechny formy lenosti, od psychické až po fyzickou, což dokazoval neskutečnou vůlí při zdolávání někdy až krutých životních údělů. Měl rád logické úlohy, kterými se s radostí zabýval, a pokořiv je jimi následně častoval své studenty a některé kolegy. Udivoval mě skvělou pamětí, kdy si byl například schopen zapamatovat křestní jména všech svých studentů několika předmětů po prvním týdnu semestru, což já mnohdy nejsem schopen ani na jeho konci. Jako pedagog a školitel byl přísný, což vtipně podtrhoval heslem „Tvrdý k sobě, nemilosrdný k ostatním“. Většina studentů však zpětně vyšší požadavky oceňila, zejména pak ti, kteří dále v životě statistiku využívali v praxi. Svého hesla se držel také tím, že si téměř nikdy na nic nestěžoval, ačkoli k tomu měl s přibývajícím věkem často patřičné důvody.

Byl oblíbený ve společnosti jednak proto, že měl rád zdravý humor typu Járy Cimrmana, a také proto, že měl vzácnou schopnost žertovat na svůj účet. Toto jsem mohl sledovat na sklonku jeho života, který trávil v bytě, a příchozí návštěvy hostil s nadšením a jen stěží je propouštěl. Je nutno zmínit, že byl člověkem pracovitým. V práci trávil většinu dne, což dokazuje především velké množství vysoce hodnocených vědeckých publikací. Ochotně pomáhal kolegům, a snad nikdy neodmítl, když někdo nezkušený potřeboval aplikovat a objasnit výsledky metod statistiky. Byl to velice čestný člověk, který odsuzoval všechny formy podvádění. K lidem se choval zdvořile, k ženám galantně, a to i když už byl odkázán na invalidní vozík. Za sebe mohl říct, že tak jako Josef na své doktorandy nahlížel jako na své „další děti“, tak i já jsem jej vnímal, jako „dalšího rodiče“, který se mě za roky strávené nad společnými cíli pokusil obohatit o zdravý úsudek a moudra, která mě dodnes denně obklopují. Nejen rodině, ale i lidem, se kterými blíže spolupracoval, zůstalo v jejich životech po Josefově odchodu prázdné místo.

morální metr a dokonale ovládal umění sebereflexe. Věděl, jaký vklad do života dostal od svých rodičů, věděl, jakým člověkem chtěl být i jaké lidi chtěl vychovat ze svých dvou dcer. Odtud závěr, i když ještě s odbočkou.

Vraceli jsme se jednou spolu s Pepou a Ivanem Křivým z (nevím už kterého) Compstatu a dostali jsme se s letadlem do velkých turbulencí. Na letuškách byla poznat nervozita, smýkalo to s nimi v uličce sem a tam. Mě ještě nepřešla chuť na hloupé žerty, i obrátil jsem se na Pepu s Ivanem, kteří seděli za mnou, se slovy „Co říkáte, kluci, že stejně byl ten život hezkej?“ Nevím, čím to, že jsem se tehdy nedočkal odpovědi.

Dnes, když se za Pepou ohlídím, bych se ho rád zeptal, tentokrát už vážně, na to samé. Představuji si, že kdyby mi mohl odpovědět, řekl by něco jako „Jasně, ale nepřijde to samo, musíš se, hochu, starat a snažit.“

Pepo, kamaráde, byl jsi skvělý chlapík, budeš tu chybět. A díky, že tu byl.

VZPOMÍNKA NA DOC. JOSEFA TVRDÍKA

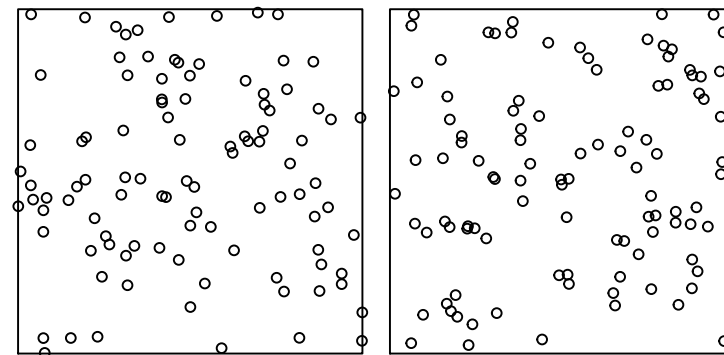
A MEMORY OF DOC. JOSEF TVRDÍK

Petr Bujok

E-mail: petr.bujok@osu.cz

Josef Tvrđík se narodil 28. června 1943 ve Zruči nad Sázavou. Vystudoval gymnázium v Ledči, kde se seznámil se základy chemie, které pak zúročil při studiu na VŠCHT. Tam se později stal též kandidátem věd teorie technické chemie. V rámci postgraduálního studia na Univerzitě Karlově se začal hlouběji zajímat o statistiku, která se mu měla stát jeho celoživotní cestou. Po studiích pracoval nejprve jako programátor. Přitom postupně začal aplikovat a rozvíjet nabyté znalosti statistiky. Své programátorské a statistické dovednosti zúročil při výuce studentů Katedry informatiky a počítačů tehdy nově vznikající Ostravské univerzity. U studentů měl patřičný respekt, neboť sebemenší porce aplikované matematiky vyvolává informatikům zpravidla stavy mdlob. Mimo to, že na katedře zavedl třístupňovou výuku statistiky (od základů až po vícerozměrné metody), započal zde výzkum v oblasti stochastických metod pro řešení problému globální optimalizace. V této oblasti aplikované matematiky později úspěšně habilitoval na Katedře matematiky Ostravské univerzity. Byl aktivním členem několika společností zaměřených na aplikovanou matematiku a účastnil se mnoha statistických konferencí.

Osobně jsem Josefa poznal nejprve jako kantora při studiích statistiky, která mi čněla nad ostatními pilíři informatiky. To bylo z části dílem jeho cha-



Obrázek 2: Ukázka realizace binomického (vlevo) a Poissonova bodového procesu (vpravo) v okně $W = [0, 1]^2$ s intenzitou $\lambda = 100$.

omezíme na pozorovací okno W a podmíníme celkovým počtem bodů procesu v tomto okně, dostáváme právě binomický proces.

Věta 1. *Nechť $n \in \mathbb{N}$, $W \subset \mathbb{R}^d$ je omezená množina a $\mathbb{X}$ je Poissonův bodový proces s funkcí intenzity $\lambda(u)$, $u \in \mathbb{R}^d$. Pak podmíněně při $\mathbb{X}(W) = n$ má restrikce procesu $\mathbb{X}$ na okno W stejné rozdělení jako binomický proces $\mathbb{Y} = \{Y_1, \dots, Y_n\}$, kde Y_i jsou nezávislé náhodné vektory s hustotou*

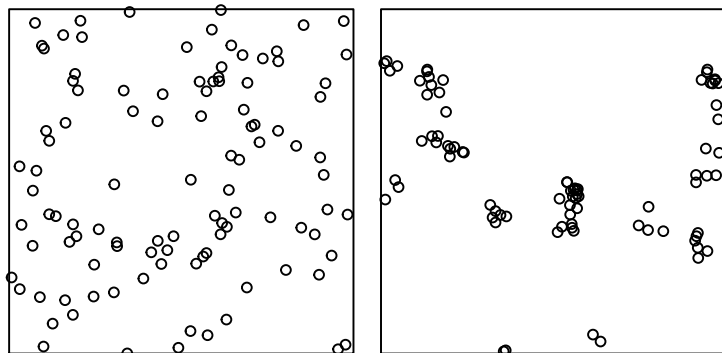
$$f(u) = \frac{\lambda(u)}{\int_W \lambda(v) dv} \cdot \mathbb{1}\{u \in W\}, \quad u \in \mathbb{R}^d.$$

Důkaz můžeme nalézt v práci [4] či jiných zdrojích. Poznamenejme, že v tomto případě shodou rozdělení bodových procesů $\mathbb{X}$ a $\mathbb{Y}$ myslíme shodu rozdělení náhodných vektorů $(\mathbb{X}(B_1), \dots, \mathbb{X}(B_k))$ a $(\mathbb{Y}(B_1), \dots, \mathbb{Y}(B_k))$ pro $k \in \mathbb{N}$ a borelovské množiny $B_1, \dots, B_k \subset W$.

Jak bylo řečeno výše, v Poissonově procesu nejsou žádné interakce mezi body. Důsledkem je, že realizace procesu nemají žádnou strukturu, jak můžeme vidět na obrázku 3, kde porovnáváme realizaci homogenního Poissonova procesu s realizací homogenního bodového procesu s přitažlivými interakcemi mezi body.

3. Testy χ^2 pro bodové procesy

Na následujících stranách představíme několik možností použití χ^2 testu k testování různých hypotéz o bodovém procesu. Vždy budeme předpokládat,



Obrázek 3: Ukázka realizace homogenního Poissonova procesu v okně $W = [0, 1]^2$ s intenzitou $\lambda = 100$ (vlevo) a realizace homogenního bodového procesu ve stejném okně se stejnou intenzitou, ale s přitažlivými interakcemi mezi body (vpravo).

že pozorujeme jednu realizaci bodového procesu $\mathbb{X}$ v omezeném pozorovacím okně $W \subset \mathbb{R}^d$. Naším pozorováním tedy není náhodný výběr, ale jedna (lokálně konečná) množina. To pak vyžaduje vhodnou adaptaci standardních postupů.

3.1. Test Poissonova rozdělení homogenního procesu

Předpokládejme nyní, že bodový proces $\mathbb{X}$ je homogenní s intenzitou $\lambda > 0$. Zajímá nás, zda je tento proces Poissonův. Testovat tedy budeme hypotézu, že je pozorovaná realizace homogenního procesu v okně W realizací Poissonova procesu, proti alternativě, že je realizací nějakého jiného homogenního procesu. Jak naznačíme dále, můžeme k tomu použít test dobré shody s Poissonovým rozdělením, viz [2, sekce 10.4.2.]. Obdobný postup v klasickém kontextu uvádí také [1, sekce 12.5.].

Z definice Poissonova procesu v sekci 2.2. víme, že počet bodů $\mathbb{X}(A)$ v nějaké množině $A \subset W$ je náhodná veličina s Poissonovým rozdělením, a navíc pro disjunktní oblasti jsou tyto počty nezávislé. Rozdělíme-li tedy okno W na m disjunktních podmnožin $A_1, \dots, A_m$, se stejnou Lebesgueovou mírou $|A_1| = \dots = |A_m|$, pak počty bodů v těchto oblastech $\mathbb{X}(A_1), \dots, \mathbb{X}(A_m)$ tvoří náhodný výběr z Poissonova rozdělení se střední hodnotou $\lambda|A_1|$.

odřekli pěší horskou túru, která už nebyla pro ně, a místo toho se v akvaparku na tobogánu vyblbli jako malí kluci).

Řekl bych o Pepovi málo, kdybych mluvil jen o tom, co dělal, a ne, jaký byl. Z mého pohledu na něm byla nejobdivuhodnější jeho úžasné vlídná, přátelská povaha. Kam vkročil, všude byl oblíbený, aniž by k tomu potřeboval vynikat ve flámování nebo generování historek. Byl gentleman každým ze svých mnoha coulů. Nevzpomínám si, že bych ho kdy viděl rozčileného, že by se hádal nebo na někoho zlobil. Až k nevěře mi to přišlo třeba v r. 2006, kdy ho během Compstatu v Římě okradli místní kapsáři. Pepa už tehdy chodil o francouzských holích (prý důsledek dřívějšího sportování pro zdraví – páter poničená basketbalem), a právě proto si ho zřejmě ty hyeny v tramvaji vyhlídly jako snadnou oběť. „Už z dálky jsem viděl, že jdou na mě,“ líčil to Pepa beze zloby v hlase, spíš až trochu pobaveně. V Pepově blízkosti člověk nutně měl pocit, že spřátelit se a přáteli zůstat je ta nejsnazší věc na světě. V obsáhlých, téměř 150stránkových zápiscích nazvaných *Trochu delší životopis*¹, které sepsal na sklonku života, mluví o zástupcích přátel, které v životě získal, jako o jedné ze svých důležitých životních výher.

Na podzim 2011 jsme se oba zúčastnili dvojkonference Stakan/TeXperience v Železném Rudě a oba jsme absolvovali (mimo oficiální program) s několika kolegy a kolegyněmi výlet do Bavorska na Velký Javor (1456 m n. m.). Nahoru se Pepa pochopitelně dostal lanovkou, ale tam si i o berlich užil špetku horské turistiky. Nevím to jistě, ale myslím, že po Železném Rudě jsme se už osobně nesetkali. Přestal jezdit na Robust i na další akce, od roku 2013 už nekandidoval do výboru ČStS. Proslýchalo se, že kromě potíží s chůzí na tom ani jinak není zdravotně dobře. Jeho chorobopis neznám, ale nedivil bych se, kdyby se na něm podepsala Pepova kuřácká vášeň. (Vypráví se, že dosti dobře obstál v testu hloubky této vášně při pobytu ve Švédsku na univerzitě v Lulea blízko polárního kruhu, kde se muselo chodit kouřit ven do 40stupňového mrazu. Pravda, měl přeci jen o něco slabší výsledky než Ivan Křivý, který i za těchto podmínek dokázal vykouřit cigaretu až do konce.) Graf, v němž Patrik Galeta ve zkratce shrnul Pepův život (viz str. 28 v tomto čísle IB), svědčí názorně o tom, že Pepa ani v těžkých časech nerezignoval na práci a téměř doposledka úspěšně a s dobrým ohlasem publikoval. V IB vyšel Pepův článek naposled v č. 2/2017.

Pepa byl statečný chlap. Už zmíněné zápisky, které vznikaly v době pro něj asi nejtemnější, kdy pochoval svou milovanou ženu Milenu (byli jsme si souzeni, napsal o ní) a sám musel tušit svůj blížící se konec, nepůsobí vůbec bolestinsky. A ještě něco se z nich dá vyčíst: Měl sám na sebe velmi přísný

¹https://www.statapol.cz/wp-content/uploads/2021/03/tvrdik_vsechno_2020.pdf

ale mělo seznámení s Tomášem Havránkem. Tomáš v r. 1978 spolu se svým šéfem Metodějem K. Chytilým na ostravském semináři o nových přístupech v užití počítačů propagoval metodu GUHA (automatizované generování hypotéz, předchůdce dnešního data-miningu) a Pepa se „chytil drápkem“. Ještě v témže roce se zúčastnil semináře o této metodě v Alšovicích a v letech 1979 až 1982 pak absolvoval, opět pod Tomášovou taktovkou, postgraduální kurs analýzy dat konaný v Matematickém středisku biologických ústavů ČSAV v Praze – Krči.

Od r. 1984 potom pravidelně utužoval své vazby se statistickou komunitou na konferencích Robust. Jeho první příspěvky na Robustu se týkaly statistického softwaru. Krátce po Listopadu 1989 začal pracovat na katedře informatiky Přírodovědecké fakulty nově zakládané Ostravské univerzity. Našel si zde svou statistickou „niku“ ve stochastických optimalizačních algoritmech a jejich použití v těžkých regresních úlohách. A také přítele a nejbližšího spolupracovníka v osobě Ivana Křivého, spoluzakladatele fakulty a jejího druhého děkana. Jejich spolupráce trvala čtvrt století, až do Ivanovy smrti v r. 2016. (Viz nekrolog z Pepova pera v IB 1/2017.) Na univerzitě působil téměř 30 let a odešel odtud, bohužel nakrátko, až do důchodu.

Pepa byl (stejně jako já) jedním z těch, které Tomáš Havránek přitáhl k výpočetní statistice a členství v IASC (Int. Assoc. for Statistical Computing). V zahraničí jsme se nejčastěji potkávali na konferencích Compstat pořádaných Evropskou regionální sekci IASC, v jejímž výboru si Pepa odsloužil jedno volební období. V tuzemsku jsme se vídali na Robustech, kde exceloval jak v oficiálním programu, tak v zábavě. Kromě anekdoty citované v úvodu si vzpomínám třeba na večírek v r. 1990 v Liblicích, kde Pepa s kolegou sehrál scénku, jak nemá vypadat, ale mnohdy vypadá konzultace statistika se zadavatelem („Takže, pane doktore, pohlaví bude v pátém sloupečku?“ „Ano, ano, v pátém. Někdy v šestém.“) nebo na jeho statistickou analýzou podložený „objev“ z r. 2000, kdy se bilancovalo prvních 20 let Robustu, že pravidelní účastníci Robustu stárnou o třetinu pomaleji než zbytek české populace.

Do naší odborné společnosti Pepa vstoupil hned po jejím založení v r. 1990. Dlouhá léta (cca 20 let!) byl členem výboru a v letech 2005 – 2007 místopředsedou společnosti. A byl velmi pilným přispěvatelem našeho Informačního bulletinu. Nejspíš vůbec nejpilnějším – napočítal jsem 14 článků na „volná“ témata (plus dva překlady do anglických čísel IB), 8 zpráv o akcích či jejich oznámení a 3 nekrology. I po letech je radost jeho texty číst – psal krásnou češtinou, živě, zábavně. Platí to mnohdy i pro zprávy (např. text o pražském Compstatu v IB 4/2004 je malé literární dílko), ba dokonce i nekrology (třeba když v IB 4/2015 vzpomíná, jak si s Ing. Zdeňkem Rothem na Robustu 2008

Označme n pozorovaný počet bodů. Střední hodnoty náhodných veličin $\mathbb{X}(A_1), \dots, \mathbb{X}(A_m)$ (střední počet bodů pozorovaných v jednotlivých množinách) můžeme odhadnout jako

$$\frac{\sum_{i=1}^m \mathbb{X}(A_i)}{m} = \frac{n}{m}.$$

Z počtů bodů v jednotlivých množinách sestavíme vektor $(U_0, \dots, U_k)$, kde náhodná veličina U_j udává, v kolika množinách jsme napozorovali přesně j bodů (pro poslední hodnotu je to k a nebo více bodů – hodnota k je volena buď jako největší hodnota z pozorovaných počtů bodů v jednotlivých množinách nebo tak, aby byla hodnota U_k dostatečně vysoká pro použití χ^2 testu). Vektor $(U_0, \dots, U_k)$ má poté za hypotézy, že $\mathbb{X}$ je Poissonův proces, multinomické rozdělení s pravděpodobnostmi

$$p_j = e^{-n/m} \frac{(n/m)^j}{j!}, \quad j \in \{0, \dots, k-1\} \quad \text{a} \quad p_k = 1 - \sum_{j=0}^{k-1} p_j.$$

Testová statistika je pak tvaru

$$\sum_{j=0}^k \frac{(U_j - np_j)^2}{np_j}$$

a má za hypotézy asymptoticky χ^2 -rozdělení s k stupni volnosti. Hypotézu zamítáme pro vysoké hodnoty testové statistiky. Ilustraci použití testu na konkrétních datech zde neuvádíme s ohledem na nutnost použít vysoký počet množin m a z toho plynoucí nepřehlednost vizualizace problému.

3.2. Homogenita Poissonova procesu

V předchozí části jsme předpokládali homogenitu bodového procesu a hypotézou byla skutečnost, že se jedná o Poissonův proces. Nyní budeme naopak předpokládat, že pozorovaná realizace pochází z Poissonova bodového procesu, a bude nás zajímat, zda je tento proces homogenní. Hypotézou je tedy konstantnost funkce intenzity na pozorovacím okně W , alternativou je její nekonstantnost [2, sekce 6.4.2.].

Díky větě 1 víme, že podmíněně při daném počtu pozorovaných bodů v okně W se na realizaci Poissonova procesu $\mathbb{X}$ můžeme dívat jako na realizaci binomického procesu $\mathbb{Y} = \{Y_1, \dots, Y_n\}$. Pokud je navíc Poissonův proces homogenní s intenzitou $\lambda > 0$, pak náhodné vektory Y_i , udávající polohu jednotlivých bodů, mají rovnoměrné rozdělení v pozorovacím okně W .

Rozdělíme-li toto okno na disjunktní podmnožiny $A_1, \dots, A_m$, jsou očekávané počty bodů v jednotlivých množinách rovny $\lambda \cdot |A_i|$. Pak podmíněně při $\mathbb{X}(W) = n$ má náhodný vektor $(\mathbb{X}(A_1), \dots, \mathbb{X}(A_m))$ multinomické rozdělení $M(n; p_1, \dots, p_m)$, kde $p_j = |A_j|/|W|$.

Testová statistika je v tomto případě tvaru

$$\sum_{j=1}^m \frac{(\mathbb{X}(A_j) - np_j)^2}{np_j}.$$

Speciálně, pokud volíme množiny $A_1, \dots, A_m$ se stejnou Lebesgueovou mírou, je testová statistika tvaru

$$\sum_{j=1}^m \frac{(\mathbb{X}(A_j) - n/m)^2}{n/m}.$$

Testová statistika má za hypotézy, že funkce intenzity je konstantní, opět asymptoticky χ^2 -rozdělení, tentokrát s $m - 1$ stupni volnosti. Hypotézu zamítáme pro vysoké hodnoty testové statistiky.

Pro ukázkou použití testu se podíváme na data ze švédského jehličnatého lesa (data jsou volně dostupná v balíčku `spatstat` pro R). Jde o polohu semenáčků borovic (obrázek 4, vlevo). Jejich výskyt v pozorovacím okně se zdá být homogenní. Pozorovací okno proto rozdělíme na $m = 9$ disjunktních částí se stejným tvarem a velikostí (obrázek 4, vpravo). Počty pozorovaných bodů v jednotlivých množinách také ukazuje obrázek 4. Hodnota testové statistiky je přibližně 4,68, počet stupňů volnosti je 8 a p -hodnota testu je přibližně 0,42. Proto nezamítáme (řekněme na hladině 5 %) hypotézu o homogenitě uvažovaného procesu. Upozorňujeme, že případné interakce mezi body jsme v tomto ilustračním příkladu zanedbali.

3.3. Závislost Poissonova procesu na dané kovariátě

Homogenní Poissonův bodový proces není vhodným modelem v situacích, kdy je intenzita výskytu bodů různá v různých částech pozorovacího okna. V takovém případě je funkce intenzity nekonstantní. Analogii předchozího postupu můžeme použít při testování, zda funkce intenzity Poissonova procesu závisí na dané kovariátě Z . Kovariátou rozumíme funkci, jejíž hodnoty pozorujeme v okně W spolu s body procesu $\mathbb{X}$. Kovariáta může například udávat nadmořskou výšku či pH půdy v daném místě nebo vzdálenost od vodních zdrojů či zdroje znečištění.

VZPOMÍNKY NA PEPU TVRDIKA MEMORIES OF PEPA TVRDIK

Jan Klaschka

E-mail: klaschka@cs.cas.cz

Koncem listopadu 2020 přišla z Ostravy smutná zpráva, že nás ve věku 77 let opustil Pepa Tvrdík. Tedy správně doc. Ing. Josef Tvrdík, CSc., ale dovolte mi zůstat u toho Pepy.

Moje nejstarší vzpomínka na Pepu je vtip: *Parta havířů udělala rekord a radili se, co s odměnou. „Vyrazíme si s novými babama,“ navrhl jeden. (Překlad: S mladými děvčaty.) „Ožereme se jako vždycky,“ oponoval druhý. „Stará pořád vyrývá, že jí nikam nevezmu. Půjdemo s manželkami,“ zněl další návrh. A nakonec jeden uvědomělý soudruh přišel s tím, že by nejlepší bylo peníze dát na Fond solidarity. (Pro mladší: Za předlistopadového režimu jeden ze způsobů rituálního vyznání lásky „státostraně“.) Rozhodnout musel předák. „Uděláme to, chlapi, spravedlivě, losem. Hodím touhle korunou. Když padne panna, půjdemo s novými babama. Když lev, ožereme se jako vždycky. Když padne na hranu, půjdemo s manželkami. A když zůstane ve vzduchu, dáme to na Solidaritu.“* Baval tak o přestávce před přednáškovým blokem na Robustu 1986 hlouček přátel. Tuším, že mezi nimi byl Tomáš Havránek, Dan Pokorný a šéfprogramátor z IKEMu Erhard Tschernoster. Pro mě vesměs „šlechta“ – chci tím říct, že Pepovi už tehdy patřilo v domácí statistické komunitě dobře viditelné místo.

Měl v té době už za sebou v životě i práci pěkný kus cesty. Poznal jsem ho jako statistika a Ostraváka, ale nebyl původně jedno ani druhé. Narodil se a vyrůstal ve Zruči nad Sázavou a do Ostravy v roce 1972 následoval manželku. S představou, že pro chemického inženýra po aspirantuře nebude těžké se tam uchytit. Když proti očekávání nemohl sehnat místo ve svém oboru, přeškolil se na jiné povolání – dávno ještě ne na statistika, ale na programátora. Dnes už je těžko k uvěření, jak dobrodružná práce to tehdy a tady byla. Ale také jak obrovské neprobádané území k objevování a seberealizaci pro tak zvědavého a činorodého člověka, jakým Pepa byl. (Škoda, že zde není místo alespoň na část jeho zápisků o tom, jaké netriviální problémy během své programátorské kariéry řešil.)

V profesi programátora působil postupně v několika institucích, začínal např. v Báňských projektech a v době, kdy jsem ho na Robustu 1986 poznal, pracoval na Krajské hygienické stanici, kde se přirozeně setkával s biostatistickou problematikou. Největší vliv na jeho pozdější profesní směřování



*Dnes, jenom pro mne, zvoní z věže!
ač jen tím malým nejtišším...
pro mne a za mne, smutné je, že...
já jediný to neslyším*
J. H. Krchovský

VZPOMÍNÁME NA JOSEFA TVRDÍKA THE LAST GOODBYE TO JOSEF TVRDÍK

Výbor společnosti

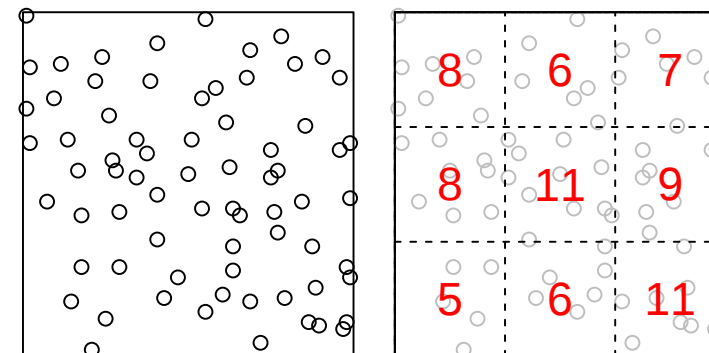
Výbor ČStS s velkou lítostí oznamuje, že v pondělí 23. listopadu nás ve věku 77 let opustil

doc. Ing. Josef Tvrdík, CSc.,

uznávaný statistik a vynikající vysokoškolský pedagog, který stál u vzniku České statistické společnosti, byl dlouholetým členem výboru společnosti a v letech 2005–2007 byl jejím místopředsedou. Vzpomínáme na jeho osobnost s hlubokou úctou a pocitem ztráty.

Připomeňme jeho osobnost výběrem jeho článků v Informačním bulletinu ČStS:

- Medici, lékaři a statistika (3/2012, str. 74–78).
- Je větší rozdíl méně významný? (3/2009, str. 22–25).
- Grafy a tabulky ve statistice aneb Na co ve výuce obvykle není čas (4/2007, str. 6–20).
- O intuici pravděpodobnosti (3/2005, str. 8–11).
- Objektivní příčiny nespokojenosti v divadelních souborech (2/2002, str. 26).



Obrázek 4: Švédský dataset. Levý panel: poloha 71 semenáčků borovic v obdélníkovém okně o rozměrech $9,6 \times 10$ metrů. Právý panel: rozdělení pozorovacího okna na 9 disjunktních podmnožin se stejným tvarem a velikostí, čísla udávají pozorovaný počet bodů v jednotlivých množinách.

Test závislosti bodového procesu na kovariátě Z formulujeme následovně. Předpokládáme, že pozorovaný bodový proces je Poissonův, a testujeme hypotézu, že jeho funkce intenzity není ovlivněna hodnotami Z , proti alternativě, že ovlivněna je. Pro přehlednost test popíšeme ve speciálním případě, kdy za nulové hypotézy je funkce intenzity konstantní a proces je tedy homogenní. Test je možné přímočaře upravit pro situaci, kdy je za nulové hypotézy funkce intenzity (známou) nekonstantní funkcí.

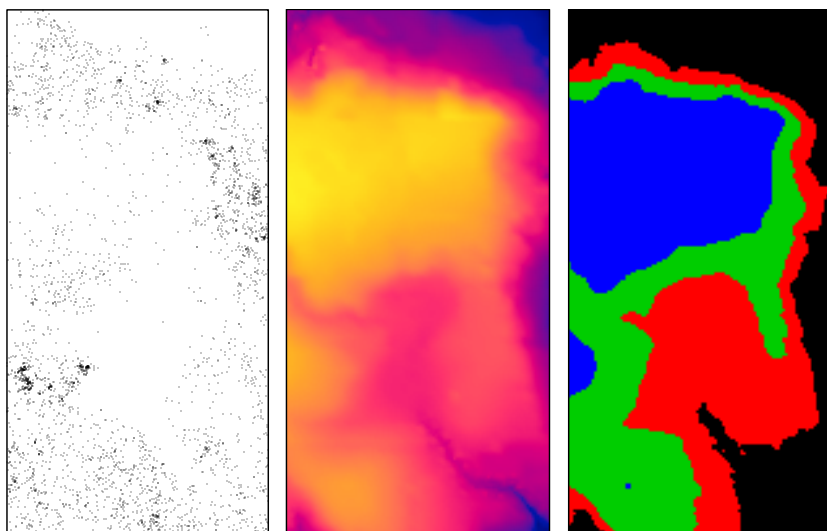
Testová statistika i její asymptotické rozdělení budou stejné jako v části 3.2. Využijeme ale toho, že pozorovací okno W můžeme na disjunktní množiny $A_1, \dots, A_m$ rozdělit zcela libovolně. Tyto množiny tentokrát zvolíme tak, aby reflektovaly hodnoty dané kovariáty Z . Ve snaze dosáhnout vysoké síly testu proti uvažované alternativě tedy například v oboru hodnot kovariáty zvolíme hodnoty $a_1, \dots, a_{m-1}$ a položíme $A_1 = \{u \in W : Z(u) \leq a_1\}$, $A_2 = \{u \in W : a_1 < Z(u) \leq a_2\}$, ..., $A_m = \{u \in W : Z(u) > a_{m-1}\}$. Testová statistika je pak ve tvaru

$$\sum_{j=1}^m \frac{(\mathbb{X}(A_j) - np_j)^2}{np_j}, \quad \text{kde } p_j = \frac{|A_j|}{|W|}.$$

Tato testová statistika má za hypotézy opět asymptoticky χ^2 -rozdělení s $m-1$ stupni volnosti, hypotézu zamítáme pro vysoké hodnoty testové statistiky.

Zde popsáný test (s konkrétnější alternativní hypotézou) a test z předchozí sekce (s velmi obecnou alternativou) spolu úzce souvisí, ale nejsou totožné. V případě s konkrétnější alternativou (závislost na kovariátě) jsme chytrou volbou dělení okna schopni dosáhnout vyšší síly testu, naopak v případě velmi obecné alternativy musíme akceptovat nižší sílu testu.

Pro ukázkou použití testu se podíváme na data z tropického deštného pralesa, konkrétně z panamského ostrova Barro Colorado (data jsou volně dostupná v balíčku `spatstat` pro R). Jde o polohu stromů druhu *Beilschmiedia pendula* (obrázek 5, vlevo). Výskyt těchto stromů v pozorovacím okně je zjevně nehomogenní a zdá se být ovlivněn nadmořskou výškou (obrázek 5, uprostřed). Pozorovací okno proto rozdělíme na $m = 4$ disjunktní části podle hodnot této kovariáty tak, aby výsledné množiny měly stejnou plochu (obrázek 5, vpravo). Z celkových $n = 3\,604$ stromů jich pak 714 leží



Obrázek 5: Barro Colorado Island dataset. Levý panel: poloha všech 3604 stromů druhu *Beilschmiedia pendula* v obdélníkovém okně o rozměrech 500×1000 metrů. Prostřední panel: nadmořská výška, světlé odstíny značí vysoké hodnoty, tmavé odstíny značí nízké hodnoty. Pravý panel: rozdělení pozorovacího okna na čtyři podmnožiny podle hodnot kovariáty (nadmořské výšky).

jméno doc. Chajdiaka. 23. listopadu 2020 nás opustil doc. Tvrdík. Všichni nám moc chybí.

Zřejmě každý alespoň tuší, jak náročná je práce všech, kteří se podílejí na přípravě časopisu. Od přijímání rukopisů, zajištění recenzních posudků, přes komunikaci s autory při úpravě článků, technickou přípravu jednotlivých článků a sestavení celého čísla, až po zajištění tisku a distribuci. Kromě členů redakční rady je do celého procesu zapojena řada recenzentů. Součástí přípravy časopisu je i zveřejnění na webových stránkách, neboť on-line verze je vlastně samostatný časopis. První webové stránky ČStS vytvořil doc. Jiří Žváček. Bylo to v době, kdy po úraze trávil svůj život ve zdravotnických zařízeních a počítač ovládal tyčinkou, kterou držel v ústech. Na webové stránky umisťoval informace o časopise, včetně jednotlivých čísel. Bohužel, na 25. listopadu 2020 připadlo smutné páté výročí, kdy nás Jiří opustil. Život však jde dál, vyvíjí se i technické možnosti, včetně nástrojů pro správu webových stránek.

<https://www.statapol.cz> → Informační bulletin

Od roku 2019 existuje již třetí verze webových stránek České statistické společnosti. Správcovství (tj. funkci webmastera) od druhé verze zajišťuje dr. Martina Litschmannová. Také díky ní můžeme listovat jednotlivými čísly Informačního bulletinu ČStS on-line. Můžeme však také listovat tištěnými čísly časopisu. V poslední době byla většina čísel vytištěna s podporou Českého statistického úřadu, samozřejmě díky jeho předchozím i současnému vedení. Velmi si toho vážíme.

V předchozích řádcích se objevila řada jmen. Zejména prvního redaktora Gejzy Dohnala, současného šéfredaktora časopisu, a současného odpovědného redaktora Ondřeje Vencálka, který je rovněž předsedou České statistické společnosti. Na něm nyní leží většina činností, zejména sběr článků, komunikace s autory a zajištění recenzí.

S časopisem jsou však spojena i další jména. Jsou to jména autorů článků a informací, jména recenzentů a vývojářů webových stránek. Na kontrole jednotlivých čísel jsem se dříve částečně podílela, řadu let se podílí dr. Marek Malý a dr. Jan Klaschka, o tisk na Českém statistickém úřadu se stará Ing. Ondřej Vozár. Do výroby se od počátku zapojuje Gejza Dohnal a spolu s Martinou Litschmannovou a Markem Malým zajišťují i distribuci.

Cílem tohoto příspěvku není podat úplný výčet jmen. Pouze naznačit, že časopis *Informační bulletin ČStS* již prošel poměrně dlouhým vývojem díky obětavosti a pili řady členů České statistické společnosti. Děkuji všem těmto kolegům za péči o časopis jménem svým i jménem všech ostatních, kteří si práce spojené s přípravou jednotlivých čísel časopisu nesmírně cení.

stránkách zveřejňováno každé číslo tak, jako by ho čtenář četl v papírové podobě. To znamená přední část obálky, články časopisu a zadní část obálky s obsahem a dalšími informacemi. Od č. 3/2020 bylo změněno pořadí tak, že je nejprve přední strana obálky, pak zadní strana obálky následovaná články. Pro čtenáře je tento způsob výhodnější, protože je obsah k dispozici před vlastními články a není třeba se přesouvat na konec dokumentu. Zároveň tak stránkově sedí blok tištěné a elektronické verze.

3. Lidé kolem Informačního bulletinu

Přípravě časopisu se od počátku věnuje Gejza Dohnal, který se ujal funkce redaktora. Jména dalších kolegů, kteří se do přípravy časopisu zapojovali, zůstávala zpočátku skryta a objevovala se spíše náhodně. Například č. 1/2001 začíná G. Dohnal větami: „Je úterý, 10. dubna 2001. Sedíme v pracovně docenta Antocha a připravujeme Informační Bulletin.“ I když jméno Jaromíra Antocha nebylo oficiálně v časopise uváděno, jeho aktivita při přípravě jednotlivých čísel byla všeobecně známá. A právě jeho zásluhou je časopis sázen v $\text{\TeX}$ u. Právě on vytvořil první verzi $\text{\TeX}$ ové šablony.

V č. 1/2005 jsou za redakci už uvedena dvě jména, druhé jméno je Pavel Stríž. V době, kdy se předsedou ČStS stal G. Dohnal, převzal funkci redaktora časopisu P. Stríž (tj. začátkem roku 2007). V souvislosti s přechodem na nový formát časopisu s obálkou začíná být publikována sedmičlenná redakční rada v čele s předsedou redakční rady prof. Ing. Václavem Čermákem, DrSc. Funkce redaktora byla zpřesněna na „technický redaktor“. Později se v redakční radě znovu objevuje jméno G. Dohnala, který samozřejmě i ve funkci předsedy ČStS o časopis pečoval, jen nějakým nedopatřením nebyl uveden (na obálce byl uváděn pouze jako předseda ČStS). V č. 1/2011 se začíná také objevovat poděkování Českému statistickému úřadu ve formulaci „Toto číslo bylo vytištěno s laskavou podporou Českého statistického úřadu“.

V roce 2013 byla redakční rada rozšířena o zahraničního člena doc. Jozefa Chajdiaka a o dr. Ondřeje Vencálka, který se stal odpovědným redaktorem časopisu (výsledné technické sestavení jednotlivých čísel i nadále až do současné doby zajišťuje P. Stríž). Z důvodu potřeby reprezentace časopisu při různých aktivitách (např. žádosti o zařazení do databáze Scopus), byl G. Dohnal ustanoven šéfredaktorem časopisu (změněno v č. 4/2014). V č. 4/2015 se v redakční radě poprvé objevuje jméno ženy. Je to doc. Iveta Stankovičová, předsedkyně SŠDS – dosud stále jediná žena v redakční radě.

Bohužel nejsme na světě věčně a s některými statistickými osobnostmi jsme se postupem let museli rozloučit. Začal se zužovat i seznam jmen v redakční radě časopisu. Od č. 2/2018 chybí jméno prof. Čermáka, od č. 3/2019

v množině s nejmenší nadmořskou výškou, v druhé množině jich leží 883, ve třetí množině 1344, a v množině s nejvyšší nadmořskou výškou jich leží 663. Hodnota testové statistiky je v tomto případě přibližně 320, počet stupňů volnosti je 3 a p -hodnota testu je tedy velmi blízká 0. Nulovou hypotézu zamítáme (na libovolné rozumně zdůvodnitelné hladině testu) a činíme závěr, že funkce intenzity uvažovaného procesu závisí na dané kovariátě.

Stojí za povšimnutí, že jsme tuto závislost odhalili, přestože funkce intenzity není monotónní funkcí kovariáty – hodnoty nadmořské výšky nejvíce preferované tímto druhem stromů ležely mezi mediánem a třetím kvantilem hodnot naměřených v daném pozorovacím okně. Upozorňujeme, že případné interakce mezi body jsme v tomto ilustračním příkladu zanedbali, vzhledem k velikosti pozorovacího okna (stovky metrů) a realistickému odhadu dosahu interakcí mezi dvěma stromy (vyšší jednotky metrů) je však vliv tohoto zanedbání na výsledky testu jen minimální.

3.4. Test dobré shody Poissonova procesu

V předchozí části jsme se zabývali závislostí funkce intenzity bodového procesu na kovariátě Z , blíže jsme ale nespecifikovali tvar funkce intenzity, tedy jak konkrétně tato kovariáta ovlivňuje rozdělení bodového procesu. V následující části budeme chtít sestavit formální test hypotézy, že bodový proces je Poissonův bodový proces s funkcí intenzity pevného tvaru závislého pouze na konečně mnoha neznámých parametrech. Tato část vychází z bakalářské práce [4], kde byl tento test podrobně odvozen pro funkci intenzity tvaru $\lambda(x, y) = \exp(\alpha + \beta_1 Z_1(x) + \beta_2 Z_2(y))$, $x, y \in \mathbb{R}$, kde α, β_1, β_2 jsou neznámé reálné parametry a $Z_1(x), Z_2(y)$ jsou známé spojitě kovariáty.

Chceme tedy testovat hypotézu, že naše pozorování je realizací Poissonova bodového procesu $\mathbb{X}$ v rovině ($d = 2$), jehož funkce intenzity je tvaru $\lambda(x, y) = \exp(\alpha + \beta_1 Z_1(x) + \beta_2 Z_2(y))$, $x, y \in \mathbb{R}$. Pozorovací okno $W \subset \mathbb{R}^2$ budeme tentokrát předpokládat ve tvaru kartézského součinu $W = W_1 \times W_2$. Opět budeme vycházet z věty 1, která říká, že podmíněně při n pozorovaných bodech v okně W se na pozorovanou množinu bodů můžeme dívat jako na realizaci binomického bodového procesu, tedy jako na realizaci náhodného výběru $\{X_1, \dots, X_n\}$ z rozdělení, jehož hustota je tvaru

$$\begin{aligned} f(x, y) &= \frac{\lambda(x, y)}{\int_W \lambda(z, w) \, d(z, w)} \cdot \mathbb{1}_W(x, y) \\ &= \frac{e^{\beta_1 Z_1(x)} \cdot e^{\beta_2 Z_2(y)}}{\int_{W_1} e^{\beta_1 Z_1(z)} \, dz \cdot \int_{W_2} e^{\beta_2 Z_2(w)} \, dw} \cdot \mathbb{1}_W(x, y), \quad x, y \in \mathbb{R}. \end{aligned}$$

Uvažujme nyní rozklad množiny W_1 na k disjunktních množin $B_1, \dots, B_k$ a rozklad množiny W_2 na l disjunktních množin $C_1, \dots, C_l$. Pak přirozeně dostáváme rozklad množiny W na disjunktní množiny $A_{1,1}, \dots, A_{k,l}$, kde $A_{i,j} = B_i \times C_j$. Podmíněné rozdělení (při daném celkovém počtu n pozorovaných bodů ve W) počtu bodů procesu $\mathbb{X}$ v množině $A_{i,j}$ je binomické s parametry n a $p_{i,j}(\beta_1, \beta_2)$, neboť můžeme psát

$$\mathbb{X}(A_{i,j}) = \sum_{\alpha=1}^n \mathbb{1}\{X_\alpha \in A_{i,j}\}.$$

Pravděpodobnosti $p_{i,j}(\beta_1, \beta_2)$ jsou přitom tvaru

$$p_{i,j}(\beta_1, \beta_2) = \int_{B_i \times C_j} \frac{e^{\beta_1 Z_1(x)} \cdot e^{\beta_2 Z_2(y)}}{\int_{W_1} e^{\beta_1 Z_1(u)} du \cdot \int_{W_2} e^{\beta_2 Z_2(v)} dv} d(x, y).$$

Vezmeme-li pak náhodný vektor $(\mathbb{X}(A_{1,1}), \dots, \mathbb{X}(A_{k,l}))$, bude mít za hypotézy, podmíněně při daném n , multinomické rozdělení závislé na konečně mnoha parametrech $M(n; p_{1,1}(\beta_1, \beta_2), \dots, p_{k,l}(\beta_1, \beta_2))$.

Nyní využijeme věty z knihy [1, str. 273], která říká, že za určitých předpokladů (pro tento model ověřených v práci [4]) a za podmínky, že existuje právě jeden odhad neznámých parametrů β_1 a β_2 modifikovanou metodou minimálního χ^2 [1, str. 272], platí, že testová statistika

$$\sum_{i=1}^k \sum_{j=1}^l \frac{(\mathbb{X}(A_{i,j}) - np_{i,j}(\hat{\beta}_1, \hat{\beta}_2))^2}{np_{i,j}(\hat{\beta}_1, \hat{\beta}_2)}$$

má asymptoticky χ^2 -rozdělení s $kl-3$ stupni volnosti ($\hat{\beta}_1, \hat{\beta}_2$ jsou dané odhady neznámých parametrů). Také v tomto případě zamítáme nulovou hypotézu pro vysoké hodnoty testové statistiky.

Analogický test lze samozřejmě odvodit i pro modely Poissonova bodového procesu s jinou funkcí intenzity závislé na neznámých parametrech, pokud tato splňuje potřebné předpoklady. Příbuzný postup lze nalézt v literatuře [2, sekce 10.4.1].

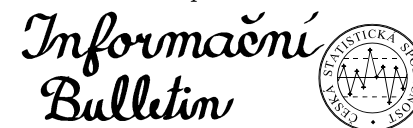
3.5. Separabilita funkce intenzity Poissonova procesu

Nyní se zaměříme na časoprostorové bodové procesy, tedy procesy v $\mathbb{R}^d \times \mathbb{R}$, kde d složek vyjadřuje polohu bodu v prostoru a poslední složka vyjadřuje časovou souřadnici. Časoprostorové procesy se liší od procesů v $\mathbb{R}^{d+1}$ například v tom, že pro určování vzdálenosti mezi body (či spíše událostmi) není

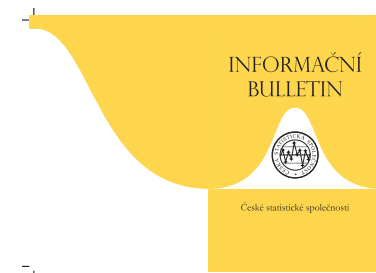
byly díky péči prof. Ing. J. Jílka, CSc. a doc. RNDr. G. Dohnala, CSc. publikovány již tradičně v příloze časopisu Statistika (č. 2, únor 2001), čímž pokračuje spolupráce, jíž si Česká statistická společnost velmi váží.“

2. TeXnická stránka věci

Od č. 4/1991 je časopis sázen v TeXu. Od č. 1/1995 je u časopisu uváděno ISSN 1210-8022. Od č. 1/1999 byl symbol počítače s histogramem a normální křivkou na monitoru (původně byl pouze monitor bez počítače) nahrazen oficiálním logem České statistické společnosti.



V roce 2009 byl „novinový“ formát (s názvem časopisu, číslem, ročníkem a datem na první straně nahoře) nahrazen formátem s obálkou, obvyklým pro časopisy. Obálku ve „Springerovské žluti“ s logem a grafickými symboly pro označení ročníku a čísla časopisu na hřbetu navrhl Gejza Dohnal. Její podoba byla projednávána výborem České statistické společnosti, který ji s nadšením přijal.



V roce 2011 přibýlo DOI: 10.5300/IB. Protože časopis byl (a je dosud) zveřejňován i na webové stránce ČStS, přibýlo v č. 2/2011 také ISSN pro on-line verzi časopisu (1804-8617). Od č. 1/2013 je uváděno též evidenční číslo registrace vedené Ministerstvem kultury ČR, které je E 21214. Kromě těchto formálních záležitostí stojí za zmínku, že od č. 1/2020 jsou hypertextové odkazy webové verze klikatelné.

Od roku 1993 do roku 2008 byla poslední strana vyhrazena pro *Obsah* a pro detailnější informace o časopise, které zahrnují vždy též jméno předsedy České statistické společnosti. S přechodem na nový formát s obálkou byly tyto informace přesunuty na zadní část obálky. Od roku 2010 bylo na webových

1. Listování aneb o čem se psalo v Informačním bulletinu

Ve druhém čísle časopisu převládaly články zaměřené na odborný život statistiků, např. o konferenci COMPSTAT 1990, o pracovišti „biometrie lékařské fakulty Univerzity Palackého v Olomouci“, o novém obsazení řídicích funkcí ve Federálním statistickém úřadě. Byly zavedeny rubriky *Recenze*, *Software*, *Nenechte si ujít* (oznámení o konferencích). A především byl publikován první odborný článek s názvem *Co je nového ve statistické analýze časových řad* od pana profesora Jiřího Anděla. Třetí číslo již obsahovalo dva odborné články a opět různé informace týkající se např. statistického softwaru. V roce 1990 vyšlo také číslo mimořádné, věnované konferenci COMPSTAT 1990 a statistickému softwaru.

Postupně byly zařazovány rubriky *Ze světa* (např. o zkušenostech z Indie či o statistice v Nizozemí), *Ze společnosti*, *Jubilea* a jiné (oznámení o konferencích lze nalézt také v rubrice *Hurá do Evropy!*). Zájemci naleznou řadu zajímavých článků, úvah nad některými problematikami, dozví se mnohé z historie statistiky a pravděpodobnosti, také o ekonomické statistice, či o statistice ve výzkumech veřejného mínění. V brzké době nás čeká Sčítání lidu, domů a bytů 2021. Kdo by se chtěl poohlédnout do historie, v č. 4/1991 vyšel článek *Sčítání lidu, domů a bytů 1991 – problémy a jejich řešení* (autoři L. Balík, P. Čtrnáct, Š. Morávková a H. Vorlová). Kdo by chtěl porovnat význam vybraných statistických časopisů podle IF (impact factor) s rokem 1988, doporučuji článek J. Anděla *Publikace a citace* v č. 1/1993. Zajímavé jsou samozřejmě všechny články, některé zaujmou svým zvláštním tématem, jako např. článek S. Komendy *Korelace, která nás šatí a obouvá* v č. 2/1993.

Jednotlivé rubriky se postupně z časopisu vytratily, i když různé informace byly publikovány stále. Od roku 2013 byla rozlišena rubrika *Vědecké a odborné statě* od ostatních rubrik, označovaných např. *Jiné statě*, *Zprávy a informace* nebo *Pozvánky na akce*. V č. 4/2015 byla základní rubrika přejmenována na *Vědecké, odborné a přehledové články*, od č. 1/2016 je její název *Vědecké a odborné články*. Vše ostatní je obvykle v rubrice *Zprávy a informace*.

Z textů publikovaných v časopise si lze připomenout některá důležitá fakta, jako např. dlouholetou spolupráci s Českým statistickým úřadem (ČSÚ). V č. 1/2001 se píše: „11. ledna 2001 byli členové výboru České statistické společnosti doc. RNDr. J. Antoch, CSc., ing. Z. Roth, CSc. a doc. RNDr. J. Á. Víšek přijati předsedkyní Českého statistického úřadu doc. Ing. Marií Bohatou, CSc., která se vyslovila pro další pokračování spolupráce ČSÚ s naší společností.“ Konkrétní spolupráci lze zaznamenat s časopisem *Statistika*, vydávaným ČSÚ. Ve stejném čísle se píše: „Dva články z loňského bulletinu

přirozené používat eukleidovskou vzdálenost. Pro přehlednost se zde opět zaměříme na nejběžnější případ $d = 2$. Přirozenou a důležitou otázkou je, zda funkce intenzity $\lambda(x, y, t)$, $x, y, t \in \mathbb{R}$, je *separabilní*, tedy zda existují funkce $\lambda_1(x, y)$ a $\lambda_2(t)$ tak, že

$$\lambda(x, y, t) = \lambda_1(x, y) \cdot \lambda_2(t). \quad (1)$$

Na následujících řádcích odvodíme formální test hypotézy, že časoprostorový Poissonův bodový proces má separabilní funkci intenzity. Tento test je založen na vlastnostech Poissonova procesu a jeho zobecnění na procesy s interakcemi mezi body není možné provést při současném požadavku na dodržení nominální hladiny testu [3].

Předpokládejme tedy, že pozorujeme realizaci Poissonova bodového procesu $\mathbb{X} = \{X_1, \dots, X_n\}$ v okně $W = W_1 \times W_2$, kde $W_1 \subset \mathbb{R}^2$ je pozorovací okno v prostoru a $W_2 \subset \mathbb{R}$ je pozorovací okno, či lépe pozorovací interval, v čase. Uvažujme nulovou hypotézu, která říká, že funkce intenzity tohoto procesu je tvaru (1) (tedy je separabilní), proti alternativě, že separabilní není. Využijeme opět věty 1, která říká, že se na pozorované body můžeme dívat jako na náhodný výběr z rozdělení s hustotou

$$f(x, y, t) = \frac{\lambda(x, y, t)}{\int_W \lambda(x, y, t) d(x, y, t)} \cdot \mathbb{1}_W(x, y, t), \quad x, y, t \in \mathbb{R},$$

což můžeme díky výrazu (1) zapsat v součinném tvaru jako

$$f(x, y, t) = \frac{\lambda_1(x, y) \cdot \mathbb{1}_{W_1}(x, y)}{\int_{W_1} \lambda_1(x, y) d(x, y)} \cdot \frac{\lambda_2(t) \cdot \mathbb{1}_{W_2}(t)}{\int_{W_2} \lambda_2(t) dt}, \quad x, y, t \in \mathbb{R}. \quad (2)$$

Speciálně tedy vidíme, že prostorová a časová složka náhodného vektoru jsou na sobě nezávislé (podmíněně při daném počtu bodů v pozorovacím okně). Označme dále činitele ve výrazu (2) postupně jako $f_1(x, y)$ a $f_2(t)$.

Rozdělme nyní, stejně jako v části 3.4., pozorovací okno W na disjunktní podmnožiny $A_{i,j} = B_i \times C_j$, kde $B_i \subset W_1$, $i \in \{1, \dots, k\}$, a $C_j \subset W_2$, $j \in \{1, \dots, l\}$. Opět dostáváme, že náhodný vektor $(\mathbb{X}(A_{1,1}), \dots, \mathbb{X}(A_{k,l}))$ má multinomické rozdělení s parametry n a $(p_{1,1}, \dots, p_{k,l})$, kde $p_{i,j}$ je (podmíněná) pravděpodobnost, že bod procesu padne do oblasti $A_{i,j}$, tedy

$$p_{i,j} = \int_{B_i \times C_j} f(x, y, t) d(x, y, t) = \int_{B_i} f_1(x, y) d(x, y) \cdot \int_{C_j} f_2(t) dt.$$

Tedy pravděpodobnosti $p_{i,j}$ jsou za hypotézy v součinném tvaru a můžeme použít test nezávislosti v kontingenční tabulce [1, str. 279]. Pokud ozna-

číme $\mathbb{X}(B_i)$, resp. $\mathbb{X}(C_j)$, počet bodů procesu $\mathbb{X}$ s prostorovými, resp. časovými, souřadnicemi v množině B_i , resp. C_j , pak testová statistika je tvaru

$$\sum_{i=1}^k \sum_{j=1}^l \frac{(\mathbb{X}(A_{i,j}) - \mathbb{X}(B_i)\mathbb{X}(C_j)/n)^2}{\mathbb{X}(B_i)\mathbb{X}(C_j)/n}$$

a platí, že má asymptoticky χ^2 -rozdělení s $(k-1)(l-1)$ stupni volnosti, přičemž hypotézu separability funkce intenzity opět zamítáme pro vysoké hodnoty testové statistiky. S ohledem na obtížnou vizualizaci časoprostorových dat zde neuvádíme konkrétní příklad a zájemce odkazujeme na článek [3].

4. Závěr

V tomto článku jsme ukázali několik možností využití χ^2 testu k testování různých hypotéz o bodových procesech. Předpoklad, že data pocházejí z Poissonova procesu, byl při odvozování těchto testů stěžejní. Díky tvaru podmíněného rozdělení Poissonova procesu při daném počtu pozorovaných bodů bylo možné využít metody testování pro náhodné výběry. Pro jiné druhy bodových procesů by stejný postup využít nešlo.

Poděkování

Tento příspěvek vznikl s podporou projektu GAČR 19-04412S.

Literatura

- [1] Anděl, J. (2007) : *Základy matematické statistiky*. Matfyzpress, Praha. cit. 8, 14 a 15
- [2] Baddeley, A., Rubak, E., Turner, R. (2015) : *Spatial Point Patterns: Methodology and Applications with R*. CRC Press. cit. 8, 9 a 14
- [3] Ghorbani, M., Vafaei, N., Dvořák, J., Myllymäki, M. (2021+) : Testing the first-order separability hypothesis for spatio-temporal point patterns. Předáno. Preprint dostupný online: <https://arxiv.org/abs/2009.04747>. cit. 15, 16
- [4] Petráková, M. (2020): *Separabilita funkce intenzity Poissonova bodového procesu*. Bakalářská práce, MFF UK, Praha. cit. 7, 13 a 14

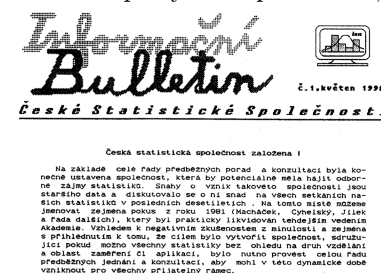
INFORMAČNÍ BULLETIN ČESKÉ STATISTICKÉ SPOLEČNOSTI – 30. VÝROČÍ VYDÁVÁNÍ ČASOPISU

THE INFORMATION BULLETIN OF THE CZECH STATISTICAL SOCIETY – 30TH ANNIVERSARY OF THE JOURNAL PUBLISHING

Hana Řezanková

E-mail: hana.rezankova@vse.cz

I když rok 2020 byl ve znamení koronaviru, on-line výuky, práce v některých profesích ve stylu home office a ve znamení řady jiných omezení, byl spojen také s 30. výročím založení České statistické společnosti, stejně jako s 30. výročím vydávání časopisu *Informační bulletin České statistické společnosti*. Jeho první číslo vyšlo 1. května 1990 s úvodním článkem „Česká statistická společnost založena!“ Toto číslo bylo věnováno nově založené společnosti, jejímu ustavujícímu zasedání, výsledkům voleb předsedy společnosti, hlavního výboru a revizorů a první schůzce hlavního výboru. Dále byla do časopisu zařazena řada sdělení, např. oznámení o přednáškách, seminářích a konferencích. Uveřejněn byl adresář členů ke dni 1.5.1990 a na poslední straně bylo „trochu statistiky“, konkrétně počty členů podle měst, organizací a titulů.



Tím byl odstartován vývoj časopisu, který se v průběhu let vylepšoval po různých stránkách. Časopis vycházel zpočátku pouze v tištěné podobě. S vytvořením webových stránek České statistické společnosti se naskytla možnost uveřejňovat jednotlivá čísla též on-line, což bylo postupně realizováno, včetně doplnění všech starších čísel časopisu. Dnešní doba je úspěšná, ale možná si někteří z čtenářů tohoto informačního článku najdou chvíli času a kromě aktuálního čísla si prolistují i čísla z dnešního pohledu již historická. Předpokládám, že snad každého zaujme nějaký článek či informace, byť staršího data. Sama jsem listovala a vzpomínala.