Informační Bulletin
České Statistické Společnosti
č. 1, květen 1998, ročník 9.

JAK POMOCÍ SIMULACÍ DOKÁZAT NEMOŽNÉ

Jaromír ANTOCH

MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8 – Karlín
tel. (+02) 2191 3275, fax. (+02) 23 23 316
e-mail: jaromir.antoch@karlin.mff.cuni.cz

ABSTRAKT: Tento příspěvek ukazuje na nebezpečí, jež může nastat v případě některých simulací, použijeme-li jako zdroj náhody generátor pseudonáhodných čísel proceduru *ran1* (nebo *ran2*) z prvního vydání knihy *Numerical Recipes*. Několik příkladů opětovně ukazuje potřebu důkladného testování generátorů pseudonáhodných čísel a zároveň varuje před přílišnou důvěrou v ne vždy platné přísloví „*Co je psáno je i dáno*".

KLÍČOVÁ SLOVA: Generátory pseudonáhodných čísel kongruenčního typu, centrální limitní věta, Kolmogorovovův-Smirnovovův test dobré shody, chí-kvadrát test dobré shody, Monte-Carlo integrace, Monte-Carlo simulace, DIEHARD.

ÚVOD

Spolu s velmi rychle rostoucí výpočetní silou (superpočítač poloviny osmdesátých let má dnes na stole pomalu každá sekretářka) můžeme pozorovat, jak stále více a více lidí „cosi simuluje". Velmi často se jedná o problémy, jež dotyční neumí řešit jinak. Představíme-li si, že sem patří mimo jiné simulace v automobilovém a leteckém průmyslu, o jaderných či armádních výzkumných projektech nemluvě, není divu, že se objevují obavy z toho, že naše bytí může být ohroženo. Zvláště silné jsou tyto obavy tehdy, vyskytne-li se podezření, že jsou založeny na špatných „základních kamenech". Je přitom zřejmé, že mezi klíčová místa většiny simulací patří generátory produkující náhodná čísla, která do jednotlivých simulací vstupují.

Tento příspěvek vznikl šťastnou náhodou. Když jsem se vrátil z konference ISI'97 v Istanbulu, nalezl jsem na svém stole několik nových čísel CSDA. V jednom z nich byl uveřejněn článek nazvaný *A note on the proper use of Numerical Recipes RAN1 pseudorandom number generator*, viz Baker (1997). Vzhledem k tomu, že jsem ve svém životě uskutečnil mnoho simulací všeho druhu, přečetl jsem si článek s velkým zájmem. Přispělo k tomu i to, že monografie Numerical Recipes je více než dobře známa a rozšířena všude po světě. Ostatně i já sám jsem v ní nejednou hledal inspiraci, podobně jako tisíce studentů, inženýrů, statistiků a matematiků (a to nejen těch aplikovaných). Připomeňme, že jednotlivá vydání zmíněné monografie vedle fundovaného popisu řady klíčových numerických postupů obsahují též zdrojové programy v jazycích *C, FORTRAN* a *Pascal*, čímž čtenářům odpadá (mnohdy pracné) programování. O její popularitě svědčí, že první vydání bylo dokonce přeloženo do češtiny. Pokud je mi však známo, nebylo nakonec (především z finančních důvodů) vydáno.

Poznamenejme, že Baker v druhé části svého příspěvku líčí ne vždy dobrou zkušenost s generátorem *ran1*, popisuje chybu v jeho (Fortranské) implementaci a ukazuje, jak ji odstranit. V okamžiku, kdy jsem článek četl podruhé, řekl jsem si, že je to skutečně zvláštní, a slíbil si, že bych se měl k tomuto problému někdy vrátit. Jelikož jsem (alespoň ne vědomě) tento generátor nikdy nepoužil a vzhledem k tomu, že trpím chronickému nedostatku času, odložil jsem celou záležitost dočasně *ad acta* a věnoval se záležitostem zdánlivě urgentnějším. Nikterak jsem přitom nečekal, že se tento problém vrátí rychle zpět na můj stůl.

Po dvou týdnech poté jsem odjel na Univerzitu do Bordeaux, kde jsem se měl zúčastnit obhajoby disertační práce pana Durrieu. Na navštíveném pracovišti jsem se mimo jiné setkal s panem Purnabou, disertantem tamní University, který ve své práci testoval řadu generátorů pseudonáhodných čísel[1], jež chtěl užít pro své simulace. Mezi testované generátory zahrnul:

- tři generátory z prvního vydání zmíněné monografie označené *ran1, ran2* a *ran3*;
- tři generátory z druhého vydání zmíněné monografie označené opět *ran1, ran2* a *ran3*;
- generátory *rand* a *drand48* dodávané standardně s operačním systémem UNIX.

Jedním ze závěrů práce Purnaba (1997) bylo, že generátor *ran1* (a částečně i generátor *ran2*) z prvního vydání *Numerical Recipes in C* nelze pro

[1]Toto je činnost více než chvályhodná a měla by být stejnou samozřejmostí, jako je ověřování předpokladů vět a postupů, které pro analýzu svých dat používáme.

rutinní použití doporučit, neboť výsledky některých jeho testů „nejsou dvakrát uspokojující". Když jsem si tyto výsledky prohlížel a snažil se pochopit, co by z nich mohlo vyplývat pro simulace běžně užívané ve statistice, uvědomil jsem si, že použití těchto generátorů může být extrémně nebezpečné pro některé typy simulací, jež statistici rutinně provádějí. To znamená, že nejenom může být, ale občas i je mnohem horší, než by se zdálo ze stručné Bakerovy poznámky „ ... *Clearly, something was amiss*". V následujícím textu se to na třech jednoduchých příkladech pokusím ukázat.

JAK VYPADAJÍ SLEDOVANÉ GENERÁTORY?

Generátor *ran1* z prvního vydání Numerical Recipes, o němž bude především řeč, je založen na třech lineárních kongruenčních generátorech tvaru

$$x_{n+1} = (7\,141\,x_n + 54\,773)\ mod\ 259\,200,$$
$$y_{n+1} = (8\,121\,y_n + 28\,411)\ mod\ 134\,456,$$
$$z_{n+1} = (4\,561\,z_n + 51\,349)\ mod\ 243\,000.$$

Náhodná čísla, jež procedura vrací, jsou tvaru

$$rnd1 = \left(x_{n+1} + \frac{y_{n+1}}{134\,456}\right) \Big/ 259\,200.$$

Přitom první generátor slouží pro generování nejpodstatnější části generovaných čísel, tj. horních bitů, druhý generátor pro generování méně významných dolních bitů, a třetí generátor k míchání (shuffling). V praxi to znamená, že procedura nevrátí přímo číslo *rnd1*, nýbrž jej uloží do pomocné tabulky délky 97 na „náhodné místo". Volbu tohoto místa řídí třetí generátor a *rnd1* se v pomocném vektoru ukládá na pozici $1 + floor(97 \times z_{n+1}/243\,000)$, zatímco číslo, jež bylo na této pozici, je vlastním výstupem z procedury *ran1*.

Generátor *ran2* z prvního vydání Numerical Recipes je lineární kongruenční generátor tvaru

$$w_{n+1} = (1\,366\,w_n + 150\,889)\ mod\ 714\,025,$$

který opět používá míchání pomocí tabulky délky 98. Pro míchání je použito generovaných hodnot w_{n+1}, dolní bity se neznáhodňují.

Poznamenejme, že přestože generátory v prvním i v druhém vydání citované monografie mají tatáž jména, pouze třetí generátor (*ran3* založený

na Knuthově návrhu pro přenosný generátor) je v obou monografiích tentýž. Co se týče generátorů *ran1* a *ran2*, jejich názvy v obou vydáních jsou sice stejné, nicméně použité algoritmy jsou zásadně různé, takže se liší i odpovídající zdrojové kódy. Snažil jsem se nalézt důvody, které vedly k této změně, bohužel bezvýsledně. Pouze na straně xi_5 druhého vydání lze nalézt malou poznámku informující o tom, že *„Rutiny pro generování náhodných čísel byly vylepšeny"*.

JAK DOPADLY TESTY DOBRÉ SHODY?

Příklad 1.

Mezi nejběžnější testy dobré shodu patří testy Kolmogorovova-Smirnovova (KS) typu. Jejich provedení je snadné, máme-li k disposici potřebné kritické hodnoty. Nemáme-li je, celá věc se značně komplikuje. Na druhé straně si můžeme vypomoci počítačem, máme-li k němu přístup. Je to ostatně běžná praxe řady spíše teoreticky zaměřených statistiků v případě, chtějí-li si vyzkoušet chování nových procedur, spočítat kritické hodnoty pro konečné výběry, apod.

Abychom nasimulovali kritické hodnoty výše zmíněného KS testu dobré shody, stačí připravit (v našem oblíbeném programovacím jazyce) program podle následujícího jednoduchého algoritmu.

```
SET n    ... Stanovení počtu pozorování.
SET opak ... Stanovení počtu opakování simulace.
SET seed ... Stanovení počáteční hodnoty pro generátor
             pseudonáhodných čísel.


Inicializuj generátor (s použitím hodnoty seed).
FOR i=1:opak
    Vygeneruj U_1,...,U_n, tj. náhodný výběr z žádaného
              rozdělení (v našem případě rovnoměrného).
    Vypočti hodnotu Kolmogorovovy-Smirnovovy statistiky
              (s využitím U_1,...,U_n) a ulož ji do KS(i).
ENDFOR
Vytvoř empirickou distribuční funkci z hodnot KS(i),
i=1,...,opak, a použij empirické kvantily místo kvantilů
teoretických.
```

Lze očekávat, že jestliže hodnoty n a *opak* budou dosti velké, potom empirická distribuční funkce odpovídající realizacím KS testových statistik

by se měla prakticky shodovat s teoretickou (asymptotickou) distribuční funkcí. Většinou tomu tak skutečně je.

Podívejme se však na situaci, kdy pro generování $U_1, \ldots, U_n$ použijeme generátor *ran1* z prvního vydání citované monografie. Místo dlouhých tabulek budou výsledky prezentovány graficky. Na obrázku 1 můžeme vidět teoretickou distribuční funkci (označenou $F(x)$) spolu s empirickými distribučními funkcemi KS testových statistik (označenými $\widehat{F}_n(x)$) pro různé hodnoty n. *Co je velice překvapující, je fakt, že s rostoucím n se odpovídající empirické distribuční funkce sledovaných KS statistik počítané pro různé rozsahy výběru* **vzdalují od teoretické distribuční funkce**.

Mimo jiné to znamená, že při pevném α se s rostoucím rozsahem výběru odpovídající kritické hodnoty vzdalují od hodnot teoretických, takže je nelze prakticky použít!!! Pro ukázkovou simulaci jsem zvolil *opak* = 10 000 a rozsah výběru n postupně 100, 300, 500, 1 000, 2 000, 3 000, 5 000, 7 000 a 10 000. Samozřejmě bylo použito i jiných hodnot *opak*, *seed* a n, ale výsledky byly vždy téhož typu. Poznamenejme, že již 1 000 opakování stačí, aby bylo možno ukázat základní problém.

Situace je analogická, i když ne tak špatná, použijeme-li generátor *ran2* z prvního vydání Numerical Recipes, viz. obrázek 3. Výsledky jsou zde jasně lepší, ale stále zdaleka ne optimální, takže ani tento generátor nelze příliš doporučit.

Příklad 2.

Abychom nepodlehli klamu, že se něco podezřelého děje pouze s Kolmogorovovým-Smirnovovým testem, podíval jsem se též na některé jiné velmi často používané statistiky, mezi nimi na statistiku odpovídající klasickému chí-kvadrát testu dobré shody. Provedení simulace bylo analogické simulaci popsané v příkladu 1. Rozdíl byl pouze v tom, že pro jednotlivé náhodné výběry byla konstruována statistika odpovídající chí-kvadrát testu dobré shody. Ve výsledcích shrnutých na obrázcích 4 a 5 můžeme vidět jev podrobně popsaný výše, tj. empirické distribuční funkce se při zvětšujícím se rozsahu vzdalují od rozdělení teoretického.

PLATÍ VŮBEC CENTRÁLNÍ LIMITNÍ VĚTA?

Příklad 3.

V každém základním kurzu matematické statistiky se probírá centrální limitní věta. Mezi její typické aplikace patří příklad ukazující, že je-li $X_1, \ldots, X_n$ náhodný výběr z rozdělení se střední hodnotou μ, rozptylem σ^2 a konečným třetím momentem, potom rozdělení výběrového průměru je přibližně normální. Přesněji, ukazuje se, že

$$\mathcal{L}(\overline{X}_n) \approx \mathcal{N}(\mu, \sigma^2/n).$$

Této vlastnosti se ostatně často využívá pro konstrukci rychlého generátoru z normálního rozdělení.

Jaké překvapení nám však přináší použití generátorů *ran*1 a *ran*2 z prvního vydání *Numerical Recipes*? Na obrázku 6 vidíme pomalu již standardní situaci. Empirické distribuční funkce normovaných výběrových průměrů $\overline{U}_k = n^{-1} \sum_{i=1}^{n} U_{k,i}$, $k = 1, \ldots, opak$, kde $\{U_{k,1}, \ldots, U_{k,n}\}$ jsou výběry z R(0,1) generované pomocí generátoru *ran*1, se s rostoucím rozsahem výběru n vzdalují od teoretické distribuční funkce. *Nic nového pod sluncem*!

Výsledky pro generátor *ran*2 jsou lepší, ale zdaleka ne dobré.

CO KDYBYCHOM NÁŠ GENERÁTOR TROCHU MODIFIKOVALI?

Příklad 4.

První vydání Numerical Recipes (viz Press a kol., 1988) uvádí na straně 211 rozsáhlou tabulku konstant, jež je možné užít pro modifikaci generátoru *ran1*. Tyto konstanty byly vybrány tak, aby odpovídající generátory měly nejdelší možnou periodu. Jak tvrdí autoři na straně 210, „ ... *volba těchto konstant je prakticky libovolná, podobně jako délka vektoru pro míchání (shuffling)*". Abych si toto tvrzení ověřil, použil jsem pro generování nejpodstatnější části generovaných čísel, tj. horních bitů, vztah

$$x_{n+1} = (84\,589\, x_n + 45\,989) \; mod \; 217\,728.$$

Zatímco obrázek 7 shrnuje výsledky pro Kolmogorovovy-Smirnovovy testové statistiky, obrázek 8 pro situaci popsanou v příkladu 3. Vlastní obrázky nejsou příliš překvapující, neboť empirické distribuční funkce se opět vzdalují od teoretické distribuční funkce. Překvapující však je fakt, že **tentokrát se vzdalují na druhou stranu než v příkladech 1 a 3**. Jinými slovy to znamená, že podle potřeby je možné (minimálně v našich speciálních příkladech) získat pomocí simulací kritické hodnoty jak malé tak velké – stačí pouze vhodně zvolit konstanty definující generátor a dostatečně velký rozsah výběru.

Když jsem o celé věci přemýšlel hlouběji, celá tato metoda mi přišla téměř tak účinná jako metoda *cílené stratifikace* kolegy Klaschky, a přitom mnohem hůře napadnutelná. Vždyť přeci konstanty pro mé generátory splňují všechny standardní teoretické požadavky, viz například Knuth (1981).

A CO INTEGRACE MONTE CARLO?

Nechť f je funkce s ohraničenou totální variací $V(f)$ definovaná na intervalu $[0,1)$, a $\boldsymbol{u} = (u_1, \ldots, u_n)$ je posloupnost reálných čísel z $[0,1)$. Neznámá hodnota integrálu

$$I = \int_0^1 f(x)\,dx$$

se často aproximuje pomocí

$$S_n(\boldsymbol{u}) = \frac{1}{n} \sum_{i=1}^{n} f(u_i).$$

Je známo, že rozdíl mezi I a $S_n(\boldsymbol{u})$ může být omezen pomocí totální variace $V(f)$ integrované funkce diskrepance $D_n(\boldsymbol{u})$ posloupnosti $\boldsymbol{u}$. Přesněji, z Věty 5.1 v Kuipers a Niederreiter (1974) vyplývá, že

$$\left| I - S_n(\boldsymbol{u}) \right| \leq V(f)\, D_n(\boldsymbol{u}).$$

Připomeňme, že *diskrepance* je definována vztahem

$$D_n(\boldsymbol{u}) = sup_{x \in [0,1)} \left| nx - Card\{k \leq n \,|\, u_k \leq x\} \right|,$$

takže pro výběr $\boldsymbol{X} = (X_1, \ldots, X_n)$ z rovnoměrného rozdělení $R(0,1)$ přechází D_n v

$$D_n(\boldsymbol{X}) = sup_{x \in [0,1)}\, n \left| \widehat{F}_n(x) - x \right|, \qquad x \in [0,1).$$

Tedy, čím menší je diskrepance posloupnosti $\boldsymbol{u}$, tím lepší může být výsledek integrace Monte-Carlo. Vzhledem k této vlastnosti jsou proto často pro numerickou integraci preferovány tzv. *quasi-náhodné posloupnosti* s malou diskrepancí. Podrobnou diskusi uvádí například Niederreiter (1992).

Tou lepší stranou generátoru *ran1* (částečně též generátoru *ran2*) z prvního vydání *Numerical Recipes* je, že generované posloupnosti pseudonáhodných čísel mají malou diskrepanci, takže použití těchto generátorů pro integraci metodou Monte-Carlo může být výhodnější než „aplikace mnohem náhodnějších generátorů". Jak jsme si však ověřili s Purnabou na případě některých standardních testových příkladů, chování generátorů *ran1* a *ran2* je pro numerickou integraci sice vcelku vyhovující, nicméně výsledky jsou vesměs horší než například chování generátoru *drand48* standardně dodávaného s operačním systémem UNIX. Z tohoto důvodu nelze použití generátorů *ran1* a *ran2* příliš doporučit ani pro tento typ aplikací. Navíc, jak

ukazuje příklad 4, „nešťastná volba“ konstant definujících generátor může vést také k velké diskrepanci, a tudíž ke špatné kvalitě integračního schématu.

KDE JE PROBLÉM?

V první chvíli mne napadlo, *zda problém není* (podobně jako ve verzi pro FORTRAN) *pouze v implementaci*, jak by se čtenář mohl domnívat na základě Bakerova článku. Po pečlivém prostudování kódu v jazyce C si to nemyslím ani já, ani větší experti na programování v *Céčku*. Oba kódy, tj. ve FORTRANu i v C jsou identické s tou výjimkou, že v kódu v jazyce C jsou veškeré klíčové proměnné deklarovány jako statické (*static*), čemuž odpovídá použití *COMMON* bloku navrhované v Baker (1997). ‘

Jako další krok jsem proto reimplementoval kód generátoru *ran1* (a poté i generátoru *ran2*) z prvního vydání *Numerical Recipes* v *Matlabu v. 4.2* pro MS Windows a využil přitom pouze možností tohoto makrojazyka. Obrázek 2 ukazuje, že výsledky spočtené v MATLABu jsou obdobné výsledkům vypočteným pomocí originální implementace v jazyce C. *To posílilo mé podezření, že vada není v implementaci, nýbrž v generátorech samotných.*

Hlavním zdrojem problémů je samozřejmě „krátká perioda“ obou generátorů. Ve skutečnosti je perioda generátoru *ran1* díky své druhé složce, jež znáhodňuje dolní (nejméně významné) bity, prakticky nekonečná. Nicméně praktický vliv této složky ve výše uvedených příkladech je zanedbatelný. Abych si ověřil tuto hypotézu, modifikoval jsem generátor *ran1* tak, že jsem složku y_i znáhodňující dolní bity vypustil. Výsledky byly dle očekávání naprosto analogické těm jež jsou popsány výše. Proto jsem se pustil do dalšího testování.

V literatuře byla navržena řada testů náhodnosti, a některé z nich byly dokonce implementovány. Mezi nejznámější baterii testů patří *DIEHARD – a battery of tests of randomness* pana G. Marsaglii, která je volně k dispozici na Internetu. Naneštěstí není možné tuto baterii testů přímo použít na *ran1*, neboť jeho výstupem nejsou celá čísla. Nicméně, vypustíme-li z generátoru *ran1* jeho druhou složku znáhodňující dolní bity, je již možné některé z Marsagliových testů použít. Výsledek byl více než tristní. Podobně špatně jsem pochodil při použití generátoru *ran2*.

NĚKOLIK POZNÁMEK DO DISKUSE

K přípravě tohoto příspěvku mne vedly především tyto důvody:

a) Mnoho lidí je přesvědčeno, že v současné době je již většina běžně užívaných generátorů pseudonáhodných čísel prosta hrubých chyb (ať již

algoritmických či implementačních). Většina autorů, výrobců či distributorů, a to nejenom statistických systémů, ale i překladačů apod., se obvykle mimo jiné zaštiťuje nejnovějšími výsledky z literatury a důkladnými testy. Zpravidla však nejsou uveřejněny výsledky těchto testů, takže nezbývá než autorům věřit. Nejsou však výjimkou situace, typické pro řadu překladačů i některé statistické programy, kdy tvar generátoru je buď *taktně zamlčen* či jej jenom s maximálním úsilím objevíme schovaný kdesi v appendixu *xxx* nebo ve speciální technické zprávě výrobce. Těžko pochopitelná bývá mnohdy také implementace. Jaká úskalí na nás v nedávné minulosti číhala u kalkulátorů firem Texas Instruments nebo Hewlett Packard, o překladačích jazyka *Pascal* nemluvě, se čtenář může dozvědět například z velmi zajímavé práce Lehn (1994).

b) Bakerova poznámka o chybě *ran1* mi zdaleka nesignalizovala, že by chování tohoto generátoru mohlo být tak špatné. Teprve výsledky výše uvedených pokusů mi způsobily jisté mrazení. z

c) Vezmu-li v úvahu, kolik lidí dnes a denně rutině používá zmíněné generátory, rád bych je varoval. Pokud někteří z čtenářů pomocí těchto generátorů uskutečnili své simulace, měli by se k nim vrátit a alespoň pro klid duše si je (s využitím některého jiného generátoru) přepočítat. Konečně ti, kteří tyto generátory implementovali do svých systémů, by měli co nejdříve připravit novou (opravenou, oprášenou a podstatně vylepšenou) verzi a implementovat některý z lepších generátorů.

d) V současné době se objevují (po kolikáté již?) návrhy používat iracionální čísla jako „zdroje náhody", blíže viz například Dodge (1996). Když jsem se ptal na názor kolegu Břetislava Nováka, známého specialisty v teorii čísel, pouze se na mne smutně podíval a pravil cosi jako „*Ani jsem netušil, jaká překvapení nás na tomto poli ještě čekají*". A poté mi začal vysvětlovat, že takové e je žhavým kandidátem na potenciálně velmi špatný zdroj náhody.

Jak jsem se zmínil již dříve, generátory *ran1* a *ran2 mají to samé jméno v obou vydáních Numerical Recipes, používají však úplně jiné algoritmy. Naneštěstí tento fakt není nikde v druhém vydání dostatečně zdůrazněn*! Pátý řádek zdola na straně *xi* mi nepřipadá dostatečný, neboť jeho obsah není podporován žádnými argumenty. Navíc na straně 3–4 ve druhém vydání Numerical Recipes autoři explicitně uvádějí: „ ... *If you have written software of any appreciable complexity that is dependent on the First Edition routines, we do no recommend blindly replacing them by the corresponding routines in this book*" ... Navíc stále existuje řada lidí (všude na světě), kteří druhé vydání nemají k dispozici, ať již je důvodem cena, neznalost

faktu, že druhé vydání existuje, či nedostatek místa v příruční knihovně.[2]

Zdůrazněme na druhé straně, že s podobným typem problémů (popsaných v příkladech 1 – 4) jsem se zatím nesetkal ani u algoritmu *ran3*, který je stejný v obou vydáních, ani u algoritmů *ran1* a *ran2* z druhého vydání *Numerical Recipes*. Podobně též generátory *rand* a *drand48* distribuované spolu s UNIXem či generátory v systémech *Matlab v. 4.2c.1* a *S+ v. 3.3* prošly tímto typem testů bez problémů.

TECHNICKÉ DETAILY

Kódy generátorů a testů dobré shody v jazyce C byly převzaty z prvního a druhého vydání monografie *Numerical Recipes*. Původní testování generátorů připravil Purnaba v jazyce C a výpočty byly provedeny na počítači SPARC Station 10 za použití kompilátoru GNU Project GCC verze 2.7.2. na Université v Bordeaux. Kontrolní simulace byly provedeny s použitím *Matlabu v. 4.2c.1* a *S+ v. 3.3* na osobních počítačích s procesorem Pentium 120 a 166 MHz a na počítači *Digital alpha* na MFF UK. Další kontrolní výpočty nezávisle provedl G. Sawitzki na Université v Heidelbergu.

ZÁVĚR

Když jsem se ptal řady kolegů, co si o daném problému myslí, většina vyjádřila více méně zdvořilý údiv a bezprostředně dodala, že oni používají generátor jiný. Někteří z dotázaných ihned opatrně dodali *otázečku*, zda jsem náhodou též netestoval zrovna ten jejich generátor. Nicméně jsem již nalezl dva kolegy, kteří používání „inkriminovaného" generátoru ve větším měřítku připustili. Mám však osobní podezření, že stejnou cestou jako tito dva se ubíral dlouhý zástup mlčenlivých.

Je bohužel s podivem, že přes velký pokrok v oblasti generování náhodných čísel uživatel stále může narazit, a to i v klasické referenční knize, jakou Numerical Recipes bezesporu jsou, na tak závažné nedostatky. Jak uvádějí Press et al. (1992) na straně 275, „ ... *a reliable source of random uniform deviates is an essential building block for any sort of stochastic modelling or Monte Carlo computer work*". Doufejme, že nové algoritmy i programy

[2]Koncem března mne G. Sawitzki upozornil na článek dvou autorů Numerical Recipes, viz Press a Teukolski (1992b). Po jistém úsilí, neboť se jedná o časopis možná známý fyzikům, ale prakticky neznámý mezi statistiky a matematiky, se mi jej podařilo sehnat. Silně negativní kritika generátorů z prvního vydání od jeho vlastních autorů je pro mne určitou satisfakcí, ovšem pouze částečnou. Domnívám se totiž, že se měla objevit (tučně vytištěna) v druhém vydání Numerical Recipes, aby byl čtenář dostatečně varován a přinucen se zamyslet nad tím, co vlastně používá a rozhodně přejít na jiné generátory, třeba ty nově publikované v druhém vydání.

uvedené v druhém vydání *Numerical Recipes* jsou bez jakýchkoliv algoritmických problémů a implementačních chyb. Naneštěstí algoritmy uvedené v prvním vydání jdou spíše cestou světoznámé rutiny RANDU, kterou po dlouhá léta *úspěšně šířila* společnost IBM.

PODĚKOVÁNÍ

Základ tohoto příspěvku byl připraven během autorovy návštěvy na Université Victora Segalena Bordeaux 2, Francie, jež se uskutečnila díky podpoře University Bordeaux 2 CNRS (UMR 9936) a grantu GAČR 1163/201/97. Podklady pro přípravu obrázků 1 a 3–5 spočetl G. P. Purnaba. Vřelé díky patří kolegovi Janu Klaschkovi za pečlivé přečtení, podnětné diskuse a vyostření nejenom názvu, ale i některých závěrů. První verze (rozsahem i obsahem mnohem stručnější, jak ostatně odpovídalo výsledkům, jež měl autor k dispozici počátkem loňského podzimu) byla připravena pro sborník konference *Analýza dat'97*. Podnětná byla též e-mailová diskuse s G. Sawitzkim, který iniciativně přepočetl část výsledků a upozornil mne na dvě podstatné citace.

LITERATURA

Antoch J., *Jak empiricky ukázat nedokazatelné*. Sborník Analýza dat'97 (Kupka K., ed.), 10–20, Pardubice, 1997.

Antoch J., *Několik poznámek o generování pseudonáhodných čísel*. Sborník ROBUST'90 (Antoch J. a Havránek T., eds.), JČSMF 1990, 14–36.

Baker F. B., *A note on the proper use of the Numerical Recipes RAN1 random number generator*. SSNinCSDA, July 1997, 495–497.

Dodge Y., *A natural random number generator*. International Statistical Review 64, 1996, 329–344.

Knuth D. E., *The Art of Computer Programming. Vol. 2. Seminumerical Algorithms*. Addison-Wesley, New York, 1981 (2. vyd.).

Klaschka J., O jedné dosti obecené metodě statistického důkazy čehokoliv. Sborník ROBUST'86 (Antoch J. a Havránek T., eds.), JČSMF 1986, 55–59.

Kuipers L. and Niederreiter H., *Uniform Distribution of Sequences*. J. Wiley, New York, 1974.

Lehn J. and Rettig S., *On the choice and implementation of pseudorandom number generators*. Computational Statistics (Dirschedl P. and Ostermann R., eds.), Physica Verlag, Heidelberg, 1994.

Marsaglia G., *DIEHARD: a battery of tests of randomness. Version: DOS, January 7, 1997*. (viz geo@stat.fsu.edu nebo http://stat.fsu.edu/ geo/diehard.html).

Niederreiter H., *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, NSF-CBMS Regional Conference Series in Applied Mathematics, Philadelphia, 1992.

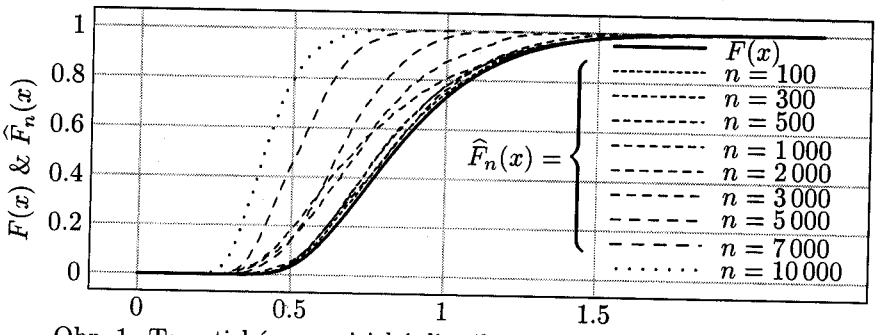
Press W. H., Flannery B. P., Teukolsky S. A. and Wetterling W. T., *Numerical Recipes: The Art of Scientific Computing, First Ed., Fourth Reprint*. Cambridge University Press, 1987.

Press W. H., Flannery B. P., Teukolsky S. A. and Wetterling W. T., *Numerical Recipes in C: The Art of Scientific Computing, First Edition*. Cambridge University Press, 1988.

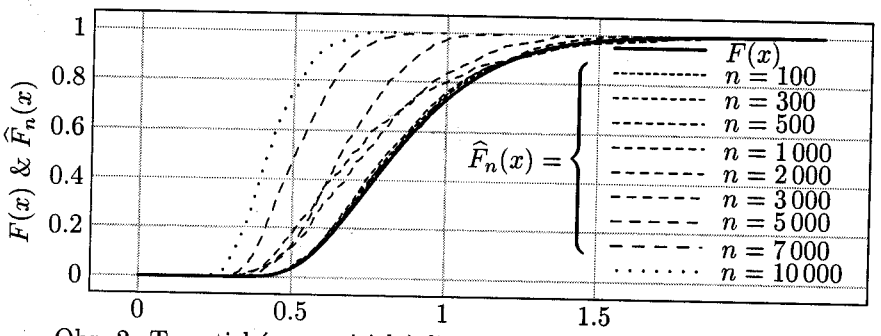
Press W. H., Flannery B. P., Teukolsky S. A. and Wetterling W. T., *Numerical Recipes in C: The Art of Scientific Computing, Second Edition*. Cambridge University Press, 1992a, software version 2.02.

Press W. H. and Teukolsky S. A., *Portable random number generators*. Computers in Physics **6**, 522–524, 1992b.

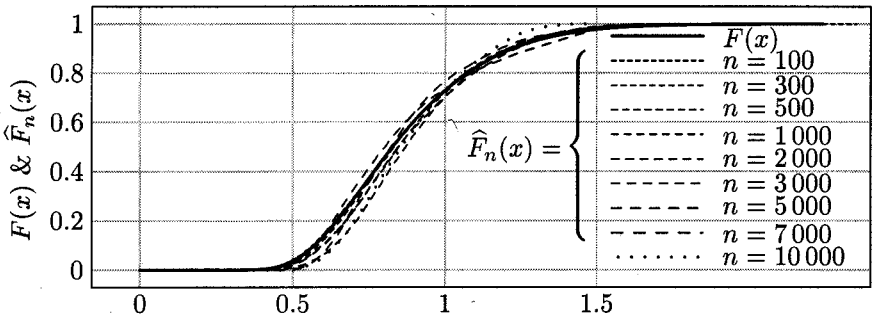
Purnaba G. P., *L'étude de divers problèmes statistiques liés aux valeurs extrêmes: Modélisation et simulations*. Thèse, Université Bordeaux 1, 1997.



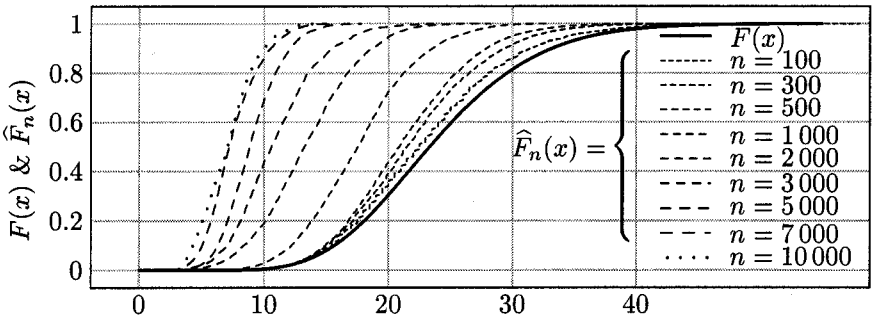
Obr. 1. Teoretická a empirické distribuční funkce Kolmogorovových-Smirnovových testových statistik při použití generátoru *ran1* implementovaného v C.



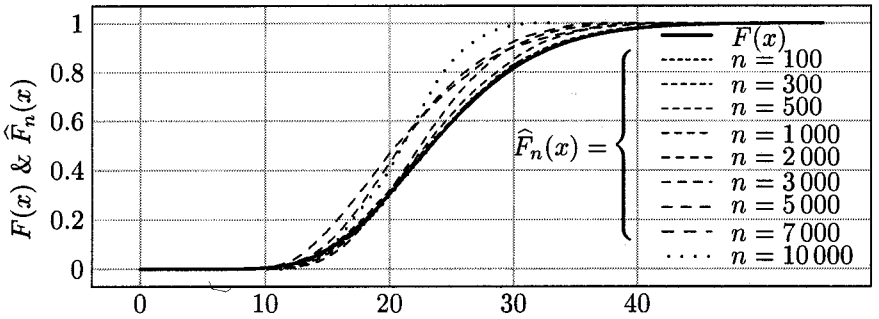
Obr. 2. Teoretická a empirické distribuční funkce Kolmogorovových-Smirnovových testových statistik při použití generátoru *ran1* implementovaného v MATLABu.



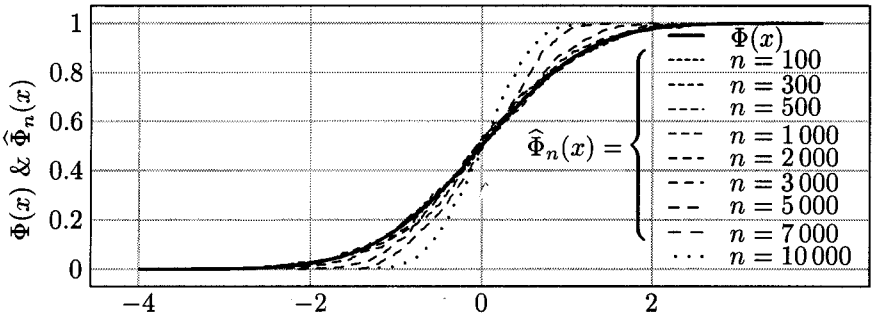
Obr. 3. Teoretická a empirické distribuční funkce Kolmogorovových-Smirnovových testových statistik při použití generátoru *ran2* implementovaného v C.



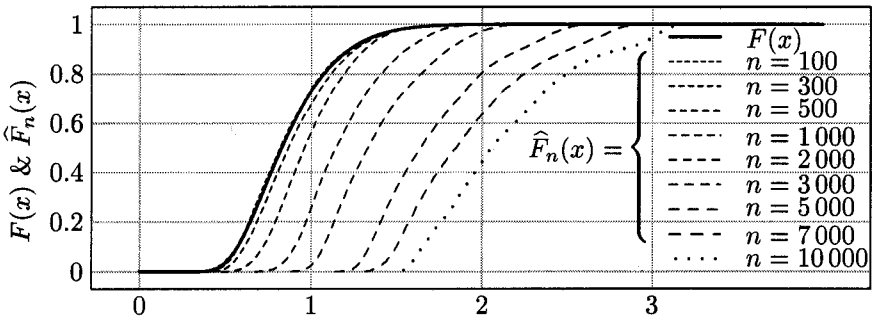
Obr. 4. Teoretická a empirické distribuční funkce χ^2 testu dobré shody při použití generátoru *ran1* implementovaného v C.



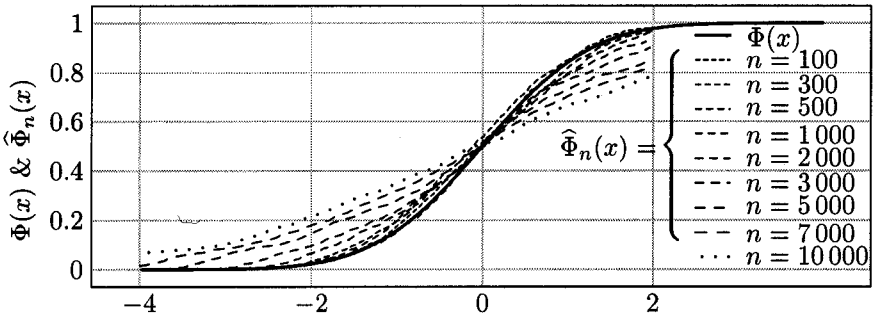
Obr. 5. Teoretická a empirické distribuční funkce χ^2 testu dobré shody při použití generátoru *ran2* implementovaného v C.



Obr. 6. Teoretická a empirické distribuční funkce výběrového průměru (normovaného) při použití generátoru *ran1* implementovaného v MATLABu.



Obr. 7. Teoretická a empirické distribuční funkce Kolmogorovových-Smirnovových testových statistik při použití generátoru *ran1* modifikovaného dle příkladu 4 a implementovaného v Matlabu.



Obr. 8. Teoretická a empirické distribuční funkce výběrového průměru (normovaného) při použití generátoru *ran1* modifikovaného dle příkladu 4 a implementovaného v MATLABu.

Výběrové metody ve zdravotnické statistice

Mgr. Jiří Holub, ÚZIS ČR

Ve svém krátkém sdělení bych vás rád seznámil s tím, co dělá náš Ústav zdravotnických informací a statistiky České republiky (ÚZIS ČR) v oblasti zdravotnické statistiky a jak využívá výběrová šetření ve své práci. Předpokládám, že mnozí z vás mají větší či menší informace o naší činnosti. Pro ty, kteří se s názvem ústavu a jeho činností nesetkávají tak často, bych rád úvodem řekl několik vět k obecné náplni ÚZIS ČR.

Ústav zdravotnických informací a statistiky České republiky

Ústav zdravotnických informací a statistiky České republiky je samostatnou rozpočtovou organizací se sídlem v Praze a detašovanými pracovišti v okresech. Ústav je podřízen přímo ministerstvu zdravotnictví. Jeho působnost se vztahuje na území celé České republiky.

Základní úkoly ústavu

Hlavním posláním ústavu je všestranné zajištění funkce Národního zdravotnického informačního systému (dále jen „NZIS") v souladu s jeho koncepcí, tj. především metodické, organizační a technické zajištění sběru dat ze zdravotnických zařízení, jejich zpracování a analýza, publikování a poskytování informací uživatelům.
Jsme v podstatě malý zdravotnický statistický úřad jehož úkolem je monitorovat prakticky většinu činností, zdrojů a aktivit v rezortu zdravotnictví, monitorovat stav a vývoj zdravotního stavu obyvatel ČR a mít informace i z ostatních oblastí ať již se zdravotnictvím souvisí přímo i nepřímo.

Ústav plní tyto základní úkoly:

- Vytváří koncepci rozvoje a vývoje NZIS, stanovuje a reviduje obsah NZIS a každoročně připravuje program statistických zjišťování.
- Definuje standardní vstupy a výstupy, zavádí jednotné klasifikace a číselníky, vydává směrnice a pokyny k vyplňování sběru a zpracování dat, zajišťuje vazby mezi NZIS a zdravotnickými zařízeními (státními i nestátními).

- Zajišťuje sběr, kontrolu a zpracování dat celorepublikových informačních subsystémů, vede databáze zdravotnických informací, analyticky a projekčně řeší systémy a subsystémy NZIS, kontroluje dodržování jednotných postupů při realizaci NZIS, kvalitu a včasnost předávaných dat.
- Analyzuje zpracované informace, vytváří systém publikování a poskytování informací a usměrňuje jejich používání, ochranu a archivaci, poskytuje informace uživatelům.
- Zajišťuje spolupráci s provozovateli ostatních IS, zajišťuje vazby na tyto systémy, zajišťuje mezinárodní spolupráci a výměnu informací, spolupracuje se státními orgány, zejména se zdravotními referáty okresních úřadů.
- Zajišťuje mezinárodní spolupráci v oblasti zdravotnické informatiky a statistiky, zejména se Světovou zdravotnickou organizací a OECD.

Z toho co jsem řekl vyplývá, že:

(1) Charakter naší práce směřuje spíše k plošným šetřením než k výběrovým, respektive plošná šetření budou mít vždy převahu nad výběrovými.

(2) Vedeme celostátní registry poskytovatelů zdravotní péče (RZZ), registr lékařů, registr zhoubných novotvarů, registr TBC a dalších povinně hlášených nemocí (VV, přenosné nemoci) a to buď plně v naší režii nebo ve spolupráci s dalšími zdravotnickými organizacemi. Nezbytná je pochopitelně dobrá spolupráce se zpravodajskými jednotkami.

(3) Protože data jsou sbírána především pro statistické účely, výsledky jsou publikovány v agregované podobě a musí být zabezpečena ochrana individuálních údajů. Přesto jsou často požadovány konkrétní údaje o jednotlivých subjektech, ty lze pochopitelně poskytnout pouze výjimečně a jen v souladu s našimi zákony.

(4) V náplni ústavu není naopak vyhodnocování klasických lékařských experimentů jako testování různých metod léčby, hodnocení účinnosti léků apod. K hlavním činnostem také nepatří v dnešní době již zcela běžné průzkumy veřejného mínění, v této oblasti působí řada komerčních agentur.

(5) V podstatě jedinou oblastí kde se zaměřujeme na výběrová šetření jsou zjišťování zdravotního stavu obyvatel, kde ani jiný způsob organizace není myslitelný. Pochopitelně výběrová šetření používáme všude tam, kde hlavním cílem je snížit časové, finanční i kapacitní nároky na sběr a zpracování dat při zachování vypovídací schopnosti dané statistiky. To je ostatně jeden z obecných důvodů pro realizaci výběrových šetření.

Nyní bych přešel již k vlastním výběrovým šetřením. Účelem tohoto sdělení není detailní popis metodiky, zpracování a výsledků jednotlivých šetření zadání, ale spíše informativní přehled jaká výběrová šetření jsme dělali, děláme nebo připravujeme. Svůj přehled začnu zhruba v polovině 80–tých let. Není třeba se bát, nebude toho moc, neboť jak jsem již naznačil nejsou výběrová šetření naší hlavní doménou.

Historický přehled výběrových šetření

1) Ukončené případy pracovní neschopnosti - 19,7 % výběr (narození 5., 6., 15., 16., 25. a 26 dne každého měsíce, rozsah cca 1 milión PN, do roku 1992 to bylo jedno z nejstarších rutinně prováděných výběrových šetření); nyní tuto statistiku zpracovává ČSSZ a sleduje všechny případy pracovní neschopnosti; ÚZIS ČR data přebírá a zpracovává i nadále statistiku ukončených případů PN.
2) Hospitalizace, různé způsoby výběru, každých 5 let plošně a v mezidobí určitá oddělení.

Tato dvě výběrová šetření jsou typickým příkladem, kdy jsou k dispozici data za celý soubor (potvrzení pracovní neschopnosti dostává každý, záznam hospitalizace se vyplňuje za každého hospitalizovaného pacienta), ale výběr se provádí pouze z důvodu omezených kapacit jak pracovních tak ekonomických.

3) CINDI (Countrywide Integrated Noncommunicable Diseases Intervention – Integrovaný program prevence nejdůležitějších neinfekčích nemocí – IPPNNN); monitoring působení preventivních opatření, životního stylu a zdravotního stavu respondentů; součástí šetření byla i lékařská prohlídka a laboratorní testy. Rozsah výběru cca 4000 osob ve věku 15–64 let (0,45%), výběry z CRO, prostý náhodný výběr osob; 1983–4 pilotní studie, první rok realizace v ČR 1985, 12 okresů (BE, KH, KT, LI, PM, P10, KV, PE, ST, ŠU, VY, PA); rozesláno 4262 dotazníků – návratnost 90%, plně použitelných 2887 – 68%; opakování po pěti letech v 1990 (podobná metodika, nyní v působnosti SZÚ).
4) Šetření o ošetřované nemocnosti 1986 (ošetřovaná (léčená) nemocnost = celková nemocnost–skrytá nemocnost), plošné šetření z hlediska lékařů; výběr osob z kartoték; rozsah 1,64 % = 131 097 osob z celé populace; (výběr podle data narození – 7. den a každý lichý měsíc); výpis údajů o zdravotním stavu vybrané osoby ze zdravotnické dokumentace. Naše dosud nejrozsáhlejší klasické výběrové

šetření, ale pokud jsme chtěli pochytit i méně častá onemocnění a stavy, byl tento rozsah nutný. U každé osoby se sledovaly hlavní socioekonomické charakteristiky (věk, pohlaví, bydliště, délka pobytu v okrese, ekonomická aktivita, zaměstnání, charakter práce, pracovní schopnost, soběstačnost , zajištění péče a dále chronická onemocnění a registrovaná u praktického lékaře, akutní onemocnění a úrazy v posledním roce a další informace vztahující se k diagnóze (závažnost, PN, hosp., léky).Mělo se opakovat každých pět let, ale ....

5) Anketa názorů občanů na zdravotní péči – listopad 1988, poštovní anketa, rozesláno 3100 dopisů do zpracování zařazeno 1550, tj. 50%; 18 let a více; výběry z CRO. Možnost porovnávat předlistopadové a polistopadové názory.

6) HIS ČR 93 (Health Interview Survey); rozsah 1600 (753 mužů, 847 žen). Šetření bylo prováděné metodou přímých rozhovorů prostřednictvím vyškoleného tazatele (vytvořily jsme vlastní tazatelskou síť). Odpovědi se zaznamenávaly do standardizovaného dotazníku podle mezinárodně doporučené metodiky. Šetření patří do skupiny průřezových epidemiologických studií, jejichž hlavním úkolem je stanovení četnosti (frekvence), s jakou se nemoc, příznak, jev nebo zdravotní potíž vyskytuje v populaci a v jejích podskupinách. Jednotlivými respondenty byly osoby starší 15 let, náhodně vybrané stratifikovaným výběrem na základě dané demografické struktury. Šetření proběhlo současně ve všech okresech České republiky. Náhodný výběr respondentů šetření HIS ČR 93 probíhal v jednotlivých okresech navzájem nezávisle, a to ve třech stupních. Přitom struktura vybraných respondentů musela odpovídat věkovému složení obyvatelstva daného okresu (každému okresu byl předepsán počet respondentů podle pohlaví a 15letých věkových skupin). Prvním stupněm byl výběr reprezentantů měst a obcí podle pětistupňové velikostní škály. Druhým stupněm byl výběr takzvaných záchytných adres v městech a obcích vybraných v prvním stupni. Tento krok byl prováděn podle telefonních seznamů bytových stanic, přičemž náhodně vybrané obce do 1000 obyvatel byly považovány za záchytnou adresu samy o sobě. Posledním třetím stupněm náhodného výběru byl výběr takzvané modelové osoby, tedy osoby kontaktované na záchytné adrese nebo v jejím okolí, starší 15 let, zapadající do předepsané demografické struktury a především ochotné odpovídat na dotazy tazatele. Dotazník má celkem 26 otázek, které lze orientačně rozdělit do 5 okruhů

I. Identifikační a socioekonomická charakteristika (otázky 1–9)
II. Zdravotní stav (otázky 10–16)
III. Sociální zdraví (otázky 17–19)
IV. Rizika chování (otázky 20–23)
V. Názory na transformační změny ve zdravotnictví po roce 1989 (otázky 24–26)

7) Anketa k povinnému členství v profesních komorách 13. – 16. 2. 1995; telefonická anketa; 1 537 dotázaných lékařů; realizace podobných šetření závisí na poptávce od našeho zřizovatele.
8) Těchto požadavků obecně není mnoho, souvisí to s uměním používat a využívat informace při řízení a rozhodování. Teprve nyní se u nás dochází k poznání, že informace pro řízení a rozhodování jsou potřebné.
9) Výběrové šetření ekonomických údajů od nestátních zařízeních s jednoduchým účetnictvím 1995 a 1997
10) HIS ČR 96, opakování po třech letech (viz HIS ČR 93); rozsah 3396 (1624 mužů, 1772 žen), mírně modifikovaný způsob výběru, slabinou zůstává poslední stupeň výběru, kdy výběr konkrétního respondenta závisí na subjektivní volbě tazatele a na ochotě spolupráce ze strany respondenta. Výsledky jsou zpracovány, připravuje se publikace. Další šetření by se mělo konat v roce 1999. Velkou výhodou je možnost sledování vývoje v čase u šetření probíhajících za srovnatelných podmínek (dotazník, metodika, apod.).
11) CIDI – 98 Composite International Diagnostc Interview. Jde o zjišťování výskytu psychosomatických poruch, sociální patologie, duševních a neurologických poruch. Jedná se o mezinárodní projekt, hlavní řešitel je Psychiatrické centrum, ÚZIS ČR je spolupracující organizace. Rozsah výběru cca 3000, metodika obdobná jako u HIS snad s pomocí CRO. Počítačový dotazník, všichni tazatelé vybaveni notebooky (30 tazatelů). Zatím zahájena přípravná fáze, vlastní šetření začne v dubnu 1998.
12) Spokojenost hospitalizovaných pacientů s pobytem v nemocnici. Poštovní anketa, výběr nemocnice oddělení a prostý náhodný výběr pacientů ze souboru hospitalizovaných; připravuje se projekt šetření, realizace cca duben – květen 1998.

Problematika výběrů

Opora výběru při populačních šetřeních – prakticky všechny zdroje nejsou 100% a navíc jsou obtížně sehnatelné. V posledních letech se vše točí kolem

ochrany individuálních dat a je komplikované získat nějaký seznam osob i pro zcela dobrovolné šetření. Chybí jasná legislativa. Tam kde oporu výběru představují naše databáze a registry lze výběr udělat snáze. Objevuje se nám zde druhý problém, patrně spojený s českou povahou, tj. „proč jsem byl vybrán já = byl jsem poškozen, a ne ten druhý". Neboli, lékaři raději snesou když se šetření týká všech a ne jen náhodného vzorku, zejména pokud se dostali do výběru.

Závěr

Pro výběry se snažíme použít většinou prostý náhodný výběr, někde kombinovaný kvótními výběry. Výběrová šetření používáme jednak tehdy, když chceme pouze omezit rozsah zpracování z důvodu úspornosti, a za druhé při realizaci šetření o zdravotního stavu obyvatel pomocí dotazníkových akcí u vybraných respondentů ať již pomocí vlastní tazatelské sítě nebo poštovní ankety.

ROBUST'98

Ve dnech 26.–30.1.1998 se ve školicím středisku Českého statistického úřadu v Radešíně konala jubilejní desátá zimní škola Jednoty českých matematiků a fyziků ROBUST'98, jež byla zorganizována skupinou pro výpočetní statistiku MVS JČMF za podpory České statistické společnosti, KPMS MFF UK a KTM FS ČVUT. Zimní školy se spolu s hosty zúčastnilo okolo šedesáti účastníků.

Tak jako předchozí letní a zimní školy, i ROBUST'98 byl věnován současným moderním trendům matematické statistiky, teorie pravděpodobnosti a analýzy dat. K přednesení bloků hlavních přednášek byli pozváni:

- *Beneš V., J. Rataj a I. Saxl*, Stochastická geometrie;
- *Fischer J. a K. Kudláka*, O problémech „oficiální statistiky";
- *Hušková M., D. Jarušková a J. Antoch*, Detekce změn ve statistických modelech;
- *Jurečková J. a J. Picek*, Užití pořadových testů pro detekci závislosti v časových řadách;
- *Volf P. a A. Linka*, Metody Monte Carlo Markov Chains.

Celkem bylo předneseno 31 přednášek. K naší velké radosti bylo mezi řečníky i tentokrát tolik doktorandů, že jsme mohli z jejich vystoupení vytvořit samostatný půldenní blok.

Poměrně mnoho času bylo věnováno též diskusím, nejenom těm neformálním v krásné okolní přírodě. Pondělní večer byl zasvěcen problematice grantového systému agentury GAČR. Na úterní večer přijali pozvání organizátorů zástupci firem ELKAN a TRILOBYTE, kteří předvedli nejnovější verze programů MATHEMATICA a S+.

Zkrátka nepřišel ani kulturní a poznávací program. Vedle středečního výletu za historií lyží se mohli účastníci potěšit krásou přírody na Samotíně a okolí rybníku Devět skal.

Příští ROBUST bude bude svého druhu též jubilejní. Všichni totiž doufáme, že se sejdeme i po dvaceti letech. Zatím není určeno kde se akce bude konat, takže uvidíme. Podaří-li se nám sehnat hezké a nepříliš drahé místo, rádi s Vámi vstoupíme do nového tisíciletí.

Jaromír ANTOCH

Ze společnosti

Jak ten čas letí ...

... a je tu zase čas složenek. Pro ty z Vás, kteří ještě v tomto roce nezaplatili členský příspěvek (a samozřejmě i pro ty, kteří jej nezaplatili v letech minulých - bohužel, jsou i tací!) přikládáme složenky, které Vám tuto radostnou povinnost usnadní. Připomínáme, že valná hromada odhlasovala zvýšení minimálního členského příspěvku na 100,- Kč, pro studenty a doktorandy 60,- Kč, přičemž horní hranice zůstává neomezena (doplatky za minulá léta očekáváme v původní výši 60,- Kč za rok).
Dále přikládáme výpis z členské databáze. V případě, že zde uvedené údaje nesouhlasí se skutečností, sdělte prosím správné údaje na adresu:
Marek Malý, SZÚ MSP, Šrobárova 48, 100 42 Praha 10, nebo prostřednictvím e-mailu na adresu: *marek.maly@szu.cz*

Valná hromada České statistické společnosti

Ve čtvrtek 12. února 1998 se v budově MFF UK v Praze–Karlíně konala mimořádná valná hromada České statistické společnosti. Zasedání řídil předseda Společnosti ing. Zdeněk Roth, CSc., který v úvodu přednesl zprávu o činnosti Společnosti v minulém roce. Ze zprávy vyjímáme následující údaje. Členskou základnu společnosti tvoří asi 260 členů, z nichž však někteří již delší dobu neplatí členské příspěvky. V současnosti probíhá akce ke zjištění přesného počtu řádných členů. Výbor společnosti se scházel pravidelně v intervalech 4–6 týdnů, od minulé valné hromady (30. 1. 1997) se konalo 9 schůzek; nikdo neabdikoval ani neměl přechodný výpadek ve výkonu své funkce. O každém jednání výboru byl pořízen zápis, který je všem zájemcům k dispozici.

Z odborných aktivit Společnosti v uplynulém období je třeba zmínit vydávání Informačního bulletinu (4 čísla), přílohu 11. čísla časopisu Statistika věnovanou Společnosti, zřízení www stránky společnosti a ustavení odborné skupiny pro problematiku norem. Společnost se podílela na uspořádání několika odborných akcí:

a) Seminář ke stému výročí veřejné statistické služby v českých zemích (17. 4. 1997, ve spolupráci s Českým statistickým úřadem)
b) Statistický den a školení „Statistický software a výpočetní statistika“ v Liberci (28. – 29. 5. 1997, ve spolupráci s Technickou univerzitou v Liberci a pobočkou JČMF v Liberci)
c) Seminář o počítačově intenzivních metodách, který se konal v první den konference „Analýza dat“ v Lázních Bohdaneč (4. 11. 1997, hlavní pořadatel TriloByte, Pardubice)
d) Společnost byla jedním ze spolupořadatelů zimní školy „Robust'98“, který se konal ve školícím středisku Českého statistického úřadu v Radešíně (26. – 30. 1. 1998, hlavní pořadatel Skupina pro výpočetní statistiku MVS JČMF).

Po zprávě o činnosti následovala zpráva o hospodaření, kterou prezentovala doc. ing. H. Řezanková, CSc. Stav finančních prostředků Společnosti se velmi mírně zvýšil, zvýšily se však také náklady na pořádání seminářů, určitý obnos je nutno vyčlenit na finanční podporu mladým statistikům při jejich účasti odborných akcích. Pozitivem je, že díky pochopení ČSÚ vydáváme náš bulletin s minimálními materiálními náklady.

Na programu byla dále úprava stanov Společnosti. Za dobu existence Společnosti totiž praxe ukázala, že některé body stanov (týkající se zejména činnosti orgánů Společnosti a regionálního zastoupení v nich) nelze splnit

přesně podle původního znění stanov. Účastníci valné hromady proto schválili příslušné změny, které rozhodně neznamenají žádný zásadní obrat v činnosti a orientaci Společnosti. S novou podobou stanov se můžete seznámit na internetové adrese Společnosti.

Valná hromada dále schválila zvýšení minimálního členského příspěvku na 100 Kč s tím, že pro studenty a doktorandy zůstává zachována původní výše (60 Kč). Tak jako v minulosti lze členský příspěvek poukázat složenkou, ale hospodářka doc. ing. D. Blatná, CSc. (VŠE) dává přednost přímém platbám z důvodu jejich snadnější identifikace a úspory manipulačního poplatku. Společnost samozřejmě velmi uvítá, pokud bude někdo chtít přispět částkou větší než je stanovené minimum.

Valná hromada byla uzavřena bohatou diskusí, ve které se mj. diskutovalo o budoucích akcích Společnosti, o otázce zvýšení členské základny a o snaze o zvýšení zahraniční činnosti.

Po valné hromadě pokračovalo setkání členů Společnosti odborným programem. Nejprve přednesla doc. RNDr. D. Jarušková, CSc. ze Stavební fakulty ČVUT informace o 51. symposiu ISI v Istanbulu doplující článek dr. Z. Fabiána, který jste si mohli přečíst v samostatném čísle bulletinu. Dále proběhl seminář věnované výběrovým metodám ve zdravotnické statistice. Ing. Z. Roth, CSc. ze Státního zdravotního ústavu promluvil o problematice chybějících pozorování a Mgr. J. Holub z Ústavu zdravotnických informací a statistiky seznámil posluchače s metodikou výběrových šetření při zjišťování zdravotního stavu.

Věříme, že účastníci zasedání Společnosti uvítali i přes poměrně malou účast možnost setkat se se svými kolegy z jiných institucí a z jiných oblastí statistiky a že si ze semináře a valné hromady odnesli zajímavé informace.

Připomenutí

Připomínáme, že Česká statistická společnost bude pořádat příští statistický den v Českých Budějovicích ve dnech 16.-17. září 1998. Jeho hlavní náplní budou „Aplikace statistiky v přírodních vědách". Účast na akci přislíbil Prof. Dr. Michal Schemper z Vídně, předseda rakousko-švýcarské sekce *Mezinárodní biometrické společnosti*, který prosloví zvanou přednášku na téma *Explained variation in Cox and logistic regression*. Vítáme zájem členů ČSS a žádáme všechny zájemce jak o vystoupení, tak o účast, aby kontaktovali *Dr. Marii Kletečkovou, CSc., Karla Tomana 9, 370 06 České Budějovice*, e-mail: kletecko@home.zf.jcu.cz.

Česká statistická společnost ve spolupráci s Katedrou pravděpodobnosti a matematické statistiky MFF UK

pořádají dne 27. května 1998 seminář k problematice cenzorovaných a useknutých pozorování. Seminář se koná od 13,00 hodin v posluchárně K9 v budově Matematicko-fyzikální fakulty UK v Praze 8 – Karlíně, Sokolovská 83.

Program semináře:

1. Mgr. M. Kulich, PhD. (MFF UK Praha): *Modely aditivního rizika*

RNDr. L. Tomášek, CSc. (SÚRO Praha): *Odhady parametrů normálního rozdělení z cenzorovaných dat*

Mgr. K. Hrach (UJEP Ústí n/L): *Modely proporcionálního rizika pro cenzorovaná data přežití*

Ing. J. Machek, CSc. (MFF UK Praha): *Problematika statistických odhadů při neúplných informacích z cenzorovaných a useknutých dat*

Srdečně zveme všechny zájemce jak z řad členů České statistické společnosti, tak ze všech matematických a statistických pracovišť.

↩ * * * ↪

Informační Bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání a jednou v roce v anglické verzi. Předseda společnosti: Ing. Zdeněk Roth, CSc., SZÚ Praha, MSP, Šrobárova 48, 100 42 Praha 10, E-mail: szumsp@earn.cvut.cz. ISSN 1210 – 8022
Redakce: Dr. Gejza Dohnal, Jeronýmova 7, 130 00 Praha 3, E-mail: dohnal@fsik.cvut.cz.