Vysoké učení technické v Brně, FIT

Dokumentace k projektu do předmětu IMS

Barbora Skřivánková (xskriv01)
Jan Kročil (xkroci02)

6. prosince 2013

Obsah

1 Úvod

1.1 Souhrn problematiky řešené v naší práci

V této práci se zabýváme možností aproximvat každodenní jevy, které se zdají býti náhodnými, pomocí matematických funkcí. Aproximace běžných jevů matematickými funkcemi může být velmi užitečná při zobrazování samotných jevů, nebo i celých systémů pomocí simulačních modelů v počítači. Simulační modely nám mohou přinést nové poznatky o modelovaném systému, které mnohdy z reálného systému pro velkou finanční nebo časovou náročnost získat nelze.

Naše práce se zabývá konkrétně generátory pseudonáhodných čísel, které v simulačních modelech hrají velmi důležitou roli - v modelovém prostředí veškeré události probíhají v nulovém (nebo jiném konstantním) čase a právě generátory pseudonáhodných čísel zajišťují, že se simulační model svými zpožděními, četnostmi výskytů zkoumaných jevů nebo intervaly mezi příchody jednotlivých požadavků na služby přiblíží reálnému systému.

Jevy, které jsme v naší práci řešili, jsou následující:

1. **Intervaly mezi vjezdy jednotlivých automobilů do Královopolského tunelu**
2. **Četnosti příchodů zákazníků na poštu Brno 12 v průběhu pracovního dne**
3. **Četnosti příjezdů nákladních automobilů do skladu během pracovního dne**

1.2 Způsoby získávání použitých dat

Data pro všechny tři jevy jsme získali měřením v terénu.

Měření intervalů mezi vjezdy automobilů do Královopolského tunelu jsme provedli sami u vjezdu do zmíněného tunelu. Měření jsme prováděli poloautomatickou metodou s využitím základních funkcí knihovny time.h jazyka C.

Data o příchodech klientů na poštu Brno 12 nám oficiální cestou poskytl oblastní manažer České Pošty, pan Tomáš Křepela.

Měření hodinových četností příjezdu nákladních aut do skladu ve firmě, která nechtěla být jmenována, bylo provedeno na vrátnici dané firmy, přes kterou musela projet všechna auta, která se do areálu skladu v den měření dostala.

1.3 Metody ověření validity získaného modelu

Výstupem našich simulačních modelů jsou histogramy, znázorňující zkoumané vlastnosti jednotlivých jevů. Pro ověření validity našich simulačních modelů jsme porovnávali histogramy naměřených hodnot s histogramy hodnot generovaných našimi generátory.

2 Koncepce

V této kapitole podrobně popíšeme okolnosti měření jednotlivých jevů a všechna fakta, která by mohla přesnost měření ovlivnit. Zároveň přesně definujeme, pro jaké okolní podmínky je náš model platný, protože změny vnějších podmínek mohou velmi výrazně ovlivnit průběhy jednotlivých jevů. Pokud vezmeme v úvahu například první z našich měřených jevů - Intervaly mezi vjezdy aut do Královopolského tunelu - jeho platnost je omezená ryze na danou dobu dne, kterou je poledne. V odpolední špičce by průběh daného jevu mohl vypadat velmi odlišně.

2.1 Intervaly mezi vjezdy jednotlivých automobilů do Královopolského tunelu

2.1.1 Měřený jev

První měřený jev má následující vlastnosti: jedná se o měření intervalů mezi vjezdy jednotlivých aut do Královopolského tunelu z Žabovřeské ulice Hradecká směrem do Králova pole na ulici Svitavská. Za vjezd do tunelu je považován okamžik, ve kterém se zadní část auta dostane do vnitřního prostoru tunelu.

Měření bylo prováděno v poledních hodinách (11.16 - 13.16) v úterý 19.11.2013, provoz který byl zaznamenán lze tedy považovat za standardní polední provoz ve všední den. Ze všedních dní je však pro přesnost nutné vyloučit pátek, během kterého se situace výrazně mění, jak bylo zmíněno výše.

2.1.2 Způsob měření

Měření bylo prováděno z mostu, který je bezprostředně u portálu Královopolského tunelu, takže je z něj dostačující výhled pro určení přesného okamžiku vjezdu do tunelu. Pro eliminaci lidské chyby jsme vytvořili jednoduchý počítačový program reagující na stisk klávesy a zaznamenávající dobu od minulého stisku klávesy. Jako jednotku jsme použili sekundu s matematickým zaokrouhlením, což je vzhledem k rozptylu hodnot 0-34 s dostačující (maximální chyba je 0.5 s, tedy 1.47% rozsahu hodnot).

2.1.3 Zpracování naměřených dat

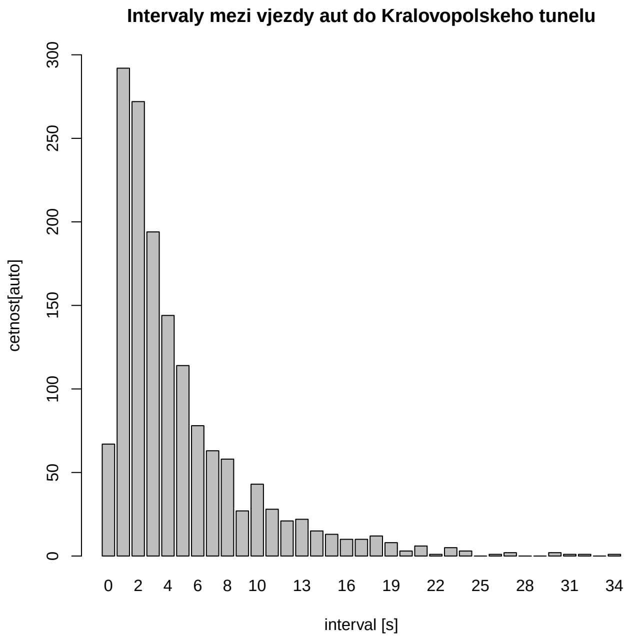
Při měření se data automaticky ukládala do souboru ve formátu csv, který jsme dále zpracovávali pomocí jazyka R v prostředí R Studio, kde jsme vypočítali četnost výskytů jednotlivých hodnot v celém souboru 1517 hodnot a vektor těchto četností jsme následně zobrazili ve formě histogramu, jak lze vidět na obrázku 1 a v tomto histogramu jsme hledali nastudovaná rozložení.

2.2 Četnosti příchodů zákazníků na poštu Brno 12 v průběhu pracovního dne

2.2.1 Měřený jev

Druhý jev, který jsme v rámci projektu zpracovali, jsou četnosti příchodů zákazníků na poštu Brno 12 na Mojmírově náměstí v Králově Poli. Zákazník se do datového souboru zařadí v okamžiku, kdy si vezme pořadový lísteček z vyvolávacího systému.

Pracovní den jsme rozdělili do 21 intervalů pokrývajících celou otvírací dobu pošty (8.00 - 18.30), kdy každý interval má délku 30 minut. Měření bylo provedeno v pondělí 11.11.2013, které je dle slov pana Křepely vždy nejfrekventovanější den pracovního týdne.



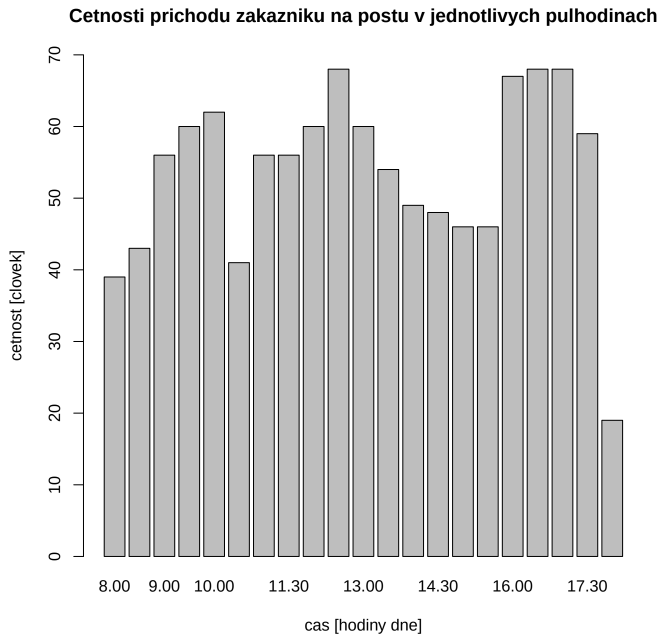
Obrázek 1: Histogram měření

2.2.2 Způsob měření

Měření bylo provedeno zcela automaticky vyvolávacím systémem, který o každém požadavku zaznamenává přesný čas zmáčknutí tlačítka (v sekundách). Mezi každodenní výstupy vyvolávacího systému patří tabulka obsahující 30minutové intervaly a počty zákazníků, kteří v daných intervalech přišli. Tuto tabulku nám Česká Pošta byla ochotna poskytnout.

2.2.3 Zpracování naměřených dat

Tištěnou tabulku získanou od České Pošty jsme převedli do elektronické podoby a v prostředí R Studio jsme ji převedli na histogram zobrazený v obrázku 2 a s ním jsme dále pracovali při hledání rozložení klientů během dne. V histogramu jsou jednotlivé intervaly popsány počátečními časy jednotlivých intervalů.



Obrázek 2: Histogram měření

2.3 Četnosti příjezdů nákladních automobilů do skladu během pracovního dne

2.3.1 Měřený jev

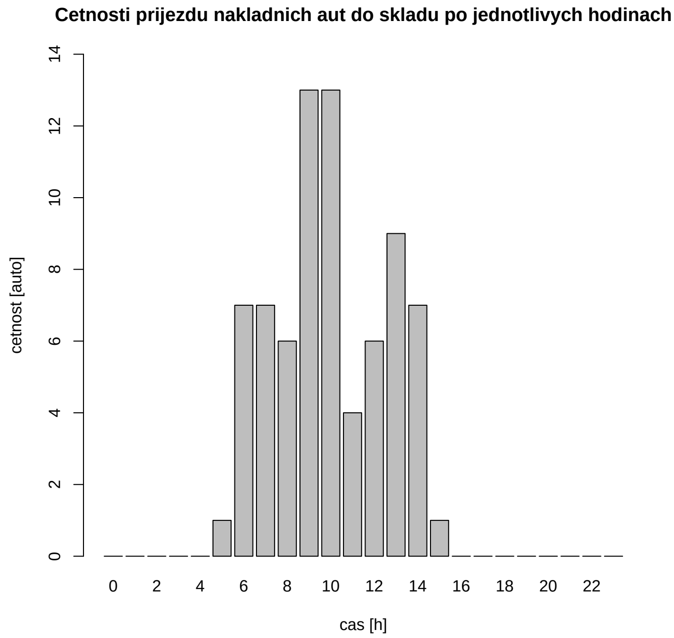
Posledním z našich jevů jsou četnosti příjezdů nákladních automobilů do skladu firmy působící ve Zlínském kraji, která ale bohužel nechtěla být jmenována. Tyto četnosti byly měřeny na vrátnici dané firmy ve středu 27.11.2013 během celého dne (od půlnoci do půlnoci). Tato vrátnice je jedinnou cestou do areálu firmy, takže vypovídající hodnota našeho měření pro střed týdne by měla být 100%.

2.3.2 Způsob měření

U nákladních automobilů se standardně zapisuje čas příjezdu a čas odjezdu. Statistika, kterou nám firma poskytla seskupovala vždy počet přijíždějících nákladních aut během 60minutového intervalu. Výsledkem měření tedy jsou počty přijíždějících automobilů v jednotlivých hodinách dne.

2.3.3 Zpracování naměřených dat

Naměřená data jsme získali v psané podobě, převedli jsme je tedy do podoby elektronické a stejně jako v ostatních případech jsme si v R Studiu vykreslili histogram četností jednotlivých hodnot (Obrázek 3) a provedli jsme teoretické prokládání histogramu funkcí hustoty pravděpodobnosti daného rozložení.



Obrázek 3: Histogram měření

3 Způsob řešení

V této kapitole se budeme věnovat způsobu, jakým jsme dosáhli exaktních aproximací jedntolivých jevů reálného světa pomocí matematických funkcí hustoty rozložení. Pro naše řešení jsme využívali pouze rozložení dostupná v knihovně SIMLIB v. 3.02 a jejich kombinace.

3.1 Aproximace funkcí hustoty pravděpodobnosti

Každý histogram jsme tvarově porovnali s obecnými tvary funkcí hustoty pravděpodobnosti. Pokud byl tvar histogramu pro různé části časové osy rozdílný, rozdělili jsme ji na několik částí a v těchto částech jsme potom prováděli aproximaci odděleně. V implementaci jsme tyto jednotlivé části složili dohromady takovým způsobem, že při každém generování požadavku se použilo právě jedno z odvozených rozložení. Rozhodování mezi použitými rozloženími bylo prováděno s pomocí vestavěné funkce Random() v závislosti na tom, jaká část (v %) naměřených hodnot patřila do danou funkcí aproximovaného intervalu.

4 Testování

4.1 Postup experimentování a okolnosti

Hlavní částí přesné aproximace změřeného jevu bylo testování. Po teoretickém proložení histogramu funkcí hustoty pravděpodobnosti jsme získali pro každý jev cca tři vyhovující varianty aproximace. Během testování jsme program opakovaně spouštěli a porovnávali jeho výsledky vynesené do histogramu s histogramy reálně naměřených jevů. Tímto způsobem jsme vždy vybrali nejvhodnější z teoretických aproximací a tu jsme potom experimentálně upravovali pro dosažení přesnějších výsledků.

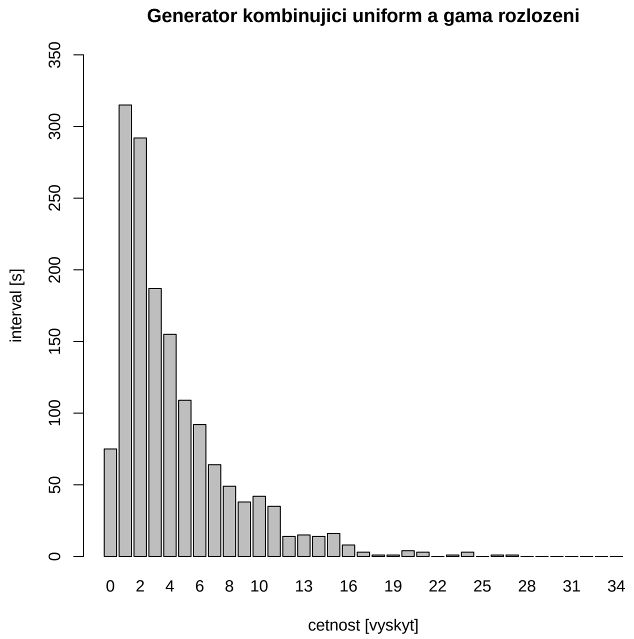
Parametry, které byly nejčastěji během testování v rozloženích měněny:

- **Charakteristiky polohy** - funkce jsme posouvali jak přičítáním konstanty k výsledku, tak pomocí parametrů jednotlivých funkcí (např. střední hodnota u exponenciálního rozložení nebo horní a dolní mez u rovnoměrného rozložení)
- **Charakteristiky variability a špičatosti**

4.2 Závěr experimentů

4.2.1 Intervaly mezi vjezdy jednotlivých automobilů do Královopolského tunelu

Po aproximaci histogramu naměřených hodnot pomocí kombinace rozložení Gama, jehož počátek byl posunut do hodnoty 1 a doplnění 4.6% aut, která vjela do tunelu speciálním způsobem - zároveň s jiným autem - pomocí rovnoměrného rozložení v intervalu ¡0,1¿ jsme získali graf zobrazený na obrázku 4. Můžeme zde pozorovat aproximací s minimální chybou. O přesnosti aproximace mluví také fakt, že pokud jsme simulaci nechali běžet dvě hodiny modelového času, bylo vygenerováno v průměru 1537 aut oproti 1517 autům při dvouhodinovém měření.



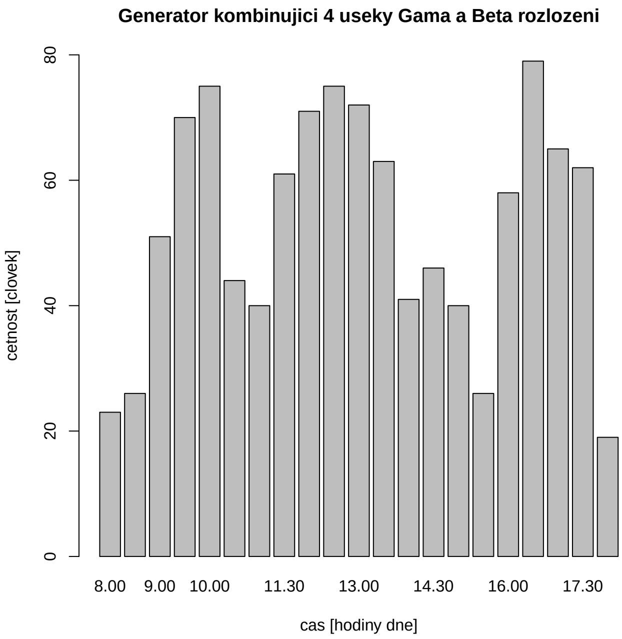
Obrázek 4: Histogram z jednoho běhu generátoru pro tunel

4.2.2 Četnosti příchodů zákazníků na poštu Brno 12 v průběhu pracovního dne

Výsledné rozložení jsme složili ze čtyř různých rozložení, ve výsledku se nám tedy objevuje dvakrát rozložení Beta a dvakrát rozložení Gama. Histogram výsledků jednoho z běhů je zobrazen na obrázku 5. Lze z něj vyčíst o něco menší, stále však pro model dostačující přesnost. Soustředit bychom se tady měli hlavně na tvar výsledného histogramu, který zhruba odpovídá skutečnosti.

4.2.3 Četnosti příjezdů nákladních automobilů do skladu během pracovního dne

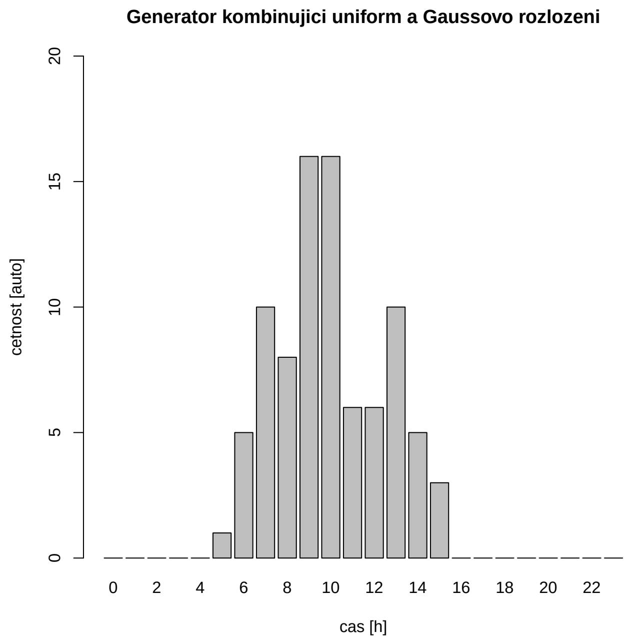
Při hledání vhodné aproximace jsme opět zvolili metodu rozdělit rozložení na několik samostatných rozložení, kdy jsme okrajové hodnoty modelovali pomocí normálního rozložení a maximální hodnoty pomocí rozložení rovnoměrného. Histogram výsledků je zobrazen na obrázku 6. Jeho tvar i hodnoty v jednotlivých hodinách odpovídají hledanému jevu znázorněném v obrázku 3.



Obrázek 5: Histogram z jednoho běhu generátoru pro poštu

5 Závěr

Během řešení této práce se nám povedlo nasimulovat náhodnost jevů probíhajících v reálném světě pomocí generátorů pseudonáhodných čísel. Dokázali jsme tím tedy, že matematický aparát generování pseudonáhodných čísel je dostatečně flexibilní na to, aby jeho prostřednictvím bylo možné aproximovat i jevy, které na první pohled nejsou učebnicovým příkladem nám známých rozložení. Matematika definuje mnohem více různých rozložení než bylo uvažováno a využíváno v rámci řešeného projektu, proto tedy na závěr této práce můžeme prohlásit, že jakákoliv struktura reálných hodnot se dá do simulačních modelů nahradit méně či více složitým generátorem pseudonáhodných čísel s aproximovaným rozložením.



Obrázek 6: Histogram z jednoho běhu generátoru pro sklad