ZÁPADOČESKÁ
UNIVERZITA
V PLZNI

KIV/VSP - Průběžná práce

Generování náhodných čísel – okruh 1

Příklad: 9

Jan Šobíšek (A08N0044P)
Narozen 30.7.1984
jsobisek@students.zcu.cz
Datum odevzdání 13.3.2009

Obsah

1. Zadání

Vytvořte generátor rozdělení jako funkci v jazyce Java, C či Pascal/Delphi (parametry rozdělení jsou zároveň parametry funkce generátoru) s využitím vhodné metody (inverzní transformace, kompoziční, vylučovací, atd.).

Napište hlavní program, který bude možné spustit s následujícím formátem parametrů (argumentů programu):
`program (počet generovaných čísel) (parametr rozdělení1) [parametr rozdělení2] [...]`

Počet parametrů rozdělení závisí samozřejmě na zadaném pravděpodobnostním rozdělení. Po spuštění vygeneruje program zadaný počet náhodných veličin dle zadaného rozdělení a bude (pokud možno průběžně) počítat dále uvedené statistiky, které po doběhnutí vypíše na standardní výstup. Tímto se tedy bude testovat správnost vytvořené funkce generátoru. Respektujte přesně zadané parametry programu (hlavně počet generovaných čísel je jako první parametr), nepřidávejte žádné svoje (povinné) další či nešetřete tam, kde si myslíte, že je některý zbytečný! Chcete-li přidat nějakou svou funkcionalitu "nad" rámec zadání, je to jen vítáno, ale musíte zachovat základní zadané rozhraní programu!

Při spuštění bez parametrů bude program tuto činnost provádět pro nejméně dvě množiny vhodně zvolených parametrů (tedy se bude chovat jako by byl spuštěn dvakrát s různými parametry na příkazové řádce).

Sledované statistiky jsou: střední hodnota (E), rozptyl (D) a charakter rozdělení - histogram. Ve výstupu programu budou tedy uvedeny jednak teoretické (vypočtené ze zadaných parametrů rozdělení) hodnoty E_teorie a D_teorie, poté vypočítané z vygenerovaných čísel E_vypocet a D_vypocet a nakonec bude uveden histogram (textový, pokud možno třeba hvězdičkový, viz. příklad).

Vygenerované zadání:
k = (3**den_narození* + 3**den_odevzdání*) % 10 = (3*30 + 3*13) % 10 = 129 % 10 = 9

9 - Trojúhelníkové rozdělení

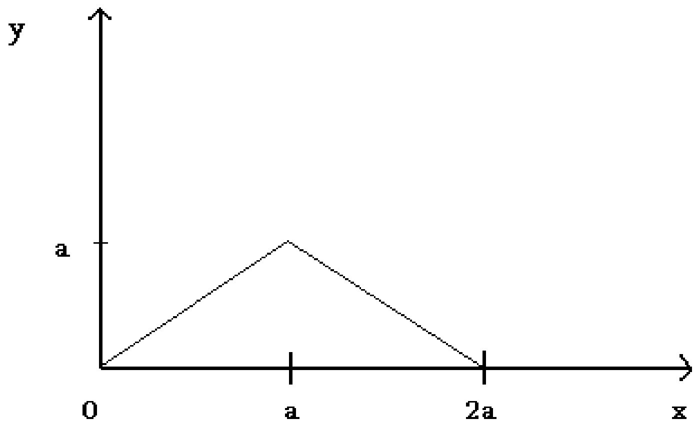
- $f(x) = 0$ pro $x<0$
- $f(x)$ lin. roste pro $a > x > 0$
- $f(x)$ lin. klesá pro $2a > x > a$
- $f(x) = 0$ pro $x>2a$

Poznámka: U tohoto příkladu není pro dosažení plného počtu bodů nutné počítat teoretické D. Pokud jej spočítáte, dostanete bod navíc.

2. Teoretická část

2.1 Trojúhelníkové (Simpsonovo) rozdělení

Následující graf demonstruje trojúhelníkové rozdělení. Vstupní parametr *a* je na ose *x* pod vrcholem tohoto rovnoramenného trojúhelníka.



Pro usnadněné práce při výpočtech předpokládejme vstupní parametr *a =1*.
Na intervalu *(0, a)* bude funkce definována $y = x$.
Na intervalu *(a, 2a)* bude funkce definována $y = |\, -x + 2a \,|$

2.2 Střední hodnota

Nejlepším odhadem střední hodnoty náhodné veličiny je aritmetický průměr z množiny realizací, tedy hodnot získaných v jednotlivých pokusech. Vzorec pro výpočet střední hodnoty:

$$E\{x\} \doteq \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Teoretickou střední hodnotu náhodné veličiny bychom získali (spojitá náhodná veličina) jako:

$$E = \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

Z grafu trojúhelníkového rozdělení je ale možné vypozorovat, že právě střední hodnota se rovná parametru *a*.

Střední hodnota je průběžně vypočítávána ze sledovaných veličin. Výpočet sumy je realizován jako průběžný součet výsledků jednotlivých pokusů, který na závěr celého výpočtu dělíme ještě počtem pokusů N.

Pokud předpokládáme $a = 1$, podle našeho grafu by měla střední hodnota vyjít rovno *1*, dosadíme tedy do předchozího vzorce:

$$E = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_0^1 x\, x\, dx + \int_1^2 x(2-x)\, dx = \left[\frac{x^3}{3}\right]_0^1 + \left[x^2 - \frac{x^3}{3}\right]_1^2$$

$$= \frac{1}{3} + \left(\left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right)\right) = 1$$

2.3 Rozptyl

Rozptyl je definován jako střední hodnota čtverce odchylky realizací náhodné veličiny x od střední hodnoty. Odhad našeho rozptylu vypočítáme jako:

$$D\{x\} \doteq \frac{1}{N-1}\sum_{i=1}^{N}(\bar{x} - x_i)^2$$

Pokud uvažujeme velké hodnoty, můžeme nahradit jmenovatel N-1 pouze číslem N, dostaneme tedy vzorec pro výpočet:

$$\begin{aligned} D\{x\} &\doteq \overline{(x_i - \bar{x})^2} = \\ &= \frac{1}{N}\sum_{i=1}^{N}(\bar{x} - x_i)^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2\bar{x}\frac{1}{N}\sum_{i=1}^{N} x_i + \frac{1}{N}N\bar{x}^2 = \\ &= \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right) - \bar{x}^2 \end{aligned}$$

Pro výpočet zavádíme dvě proměnné pro sumy, kterým průběžně v cyklu přičítáme získávané hodnoty:

$$S_x = S_x + x_i$$

$$S_{x^2} = S_{x^2} + x_{i^2}$$

Na závěr vypočítáme:

$$\bar{x} = \frac{S_x}{N}; E\{x\} = \bar{x}; D\{x\} = \frac{S_{x^2}}{N} - \bar{x}^2$$

Podle materiálu [KACK] budeme počítat v programu teoretický rozptyl jako:

$$D = \frac{a^2}{6}$$

K tomu samému výsledku bychom s největší pravděpodobností došli vypočítáním:

$$D = \int_{-\infty}^{+\infty} x^2 f(x)\, dx = \int_0^1 x^2\, x\, dx + \int_1^2 x^2(-x + 2a)\, dx$$

Nakonec bychom odečetli rozptyl a střední hodnotu:

$$D(x)_t = \int_{-\infty}^{+\infty} x^2 f(x)\, dx - \left(\int_{-\infty}^{+\infty} x\, f(x)\, dx \right)^2$$

3. Praktická část

3.1 Implementace

3.1.1 Volba programovacího jazyka

Jako programovací jazyk byl zvolen objektově orientovaný jazyk JAVA z důvodu snadné přenositelnosti na různých OS (*Operation Systém.* Programovací jazyk Java byl vyvinut firmou *Sun Microsystems* která ho představila v roce 1995. V roce 2007 byly uvolněny zdrojové kódy (přibližně 2,5 miliónů řádků kódu) a Java bude dále vyvíjena pouze jako open source (*otevřené zdrojové kódy*).

Celkově vzniklo mnoho platforem tohoto jazyka, platforma *JavaCard* je určena pro čipové karty, platforma *Java ME* je určena pro mobilní a vestavěné (*embedded*) zařízení, pro aplikace běžící na desktopu je určena platforma *Java SE*, pro rozsáhlejší distribuované systémy je určena *Java EE.* Všechny tyto technologie jsou souhrnně označovány jako platforma Java.

Mezi základní vlastnosti patří jednoduchost (zjednodušená syntaxe jazyka C a C++), robustnost (správa paměti pomocí *Garbage collectoru*) a přenositelnost (vytváření *bajtkódu*, který je nezávislý na architektuře, spuštění probíhá pomocí interpretu Javy - JVM (*Java*

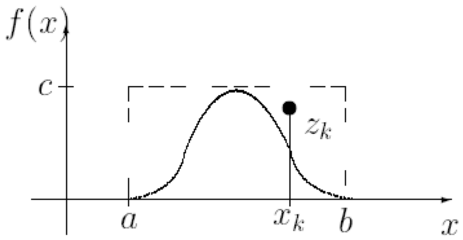
Virtual Machine)). Dále můžeme jmenovat výkonnost, bezpečnost a snadnou čitelnost zdrojového kódu. V současné době (jaro 2009) se pro desktopové aplikace využívá Java SE 1.6, která přináší další výrazná výkonové a bezpečnostní vylepšení.

Semestrální práce byla vyvíjena pod JDK verze 1.6.

```
C:\Users\sobi>java -version
java version "1.6.0_16"
Java(TM) SE Runtime Environment (build 1.6.0_16-b01)
Java HotSpot(TM) Client VM (build 14.2-b01, mixed mode, sharing)
```

3.1.2 Generování náhodných čísel

V programu je využívána třída *Random*, která generuje hodnoty s rovnoměrným rozložením pravděpodobnosti. Pro splnění trojúhelníkového rozložení pravděpodobnosti je používána vylučovací metoda:



- Generujeme náhodné číslo x_k z výběru s rovnoměrným rozdělením na intervalu <a, b), tedy $x_k = a + (b - a) * y$, kde y je číslo poskytnuté základním generátorem
- Generujeme náhodné číslo z_k z výběru s rovnoměrným rozdělením na intervalu <0, c), tedy $z_k = c * y$, kde y je další číslo poskytnuté základním generátorem
- Čísla x_k a z_k interpretujeme jako souřadnice bodu v rovině
- Pokud je náhodně generovaný bod pod křivkou funkce *f(x)*, postup končí a výstupem generátoru je hodnota x_k, v opačném případě postup opakujeme od prvního bodu.

3.1.3 Histogram

Odhad hustoty pravděpodobnosti v určitém intervalu je počítán jako:

$$f_k = \frac{c_k}{N * (velikost\_intervalu)}$$

Kde c_k je počet hodnot vyskytujících se v každém intervalu, N je počet celkových pokusů, a velikost_intervalu je velikost jednoho intervalu (intervaly by měly být stejně velké).

3.2 Uživatelská příručka

Spustíme skript s příslušnými parametry *run.cmd [počet_gen_ čísel] [a] [počet_intervalů]*.

Pokud spustíme script bez parametrů, jsou použity dvě množiny defaultních hodnot. Program se tedy chová jako kdyby byl spuštěn dvakrát.

4. Závěr

Semestrální práce odpovídá danému zadání. Během řešení se vyskytlo pár drobných problémů hlavně s teoretickými výpočty a převody čísel v programovacím jazyce JAVA. Je možné pozorovat mírné rozdíly ve vypočítaném a teoretickém rozptylu.

Literatura a zdroje

[HART]	HARTMAN, David. *Vyhodnocení výsledků stochastické metody Monte Carlo* . 2004 [cit. 2009-10-10]. Dostupný z WWW: <http://portal.zcu.cz>.
[HERO]	HEROUT, Pavel. *Učebnice jazyka Java* . 3. rozš. vyd. České Budějovice : KOPP, 2007. 381 s. ISBN 978-80-7232-323-4.
[KACK]	KACKER, Raghu, LAWRENCE, James. *Trapezoidal, triangular, and rectangular distributions for Type B evaluations* . 2007 [cit. 2009-10-10]. Dostupný z WWW: <http://docs.google.com/gview?a=v&q=cache:ByFLzuZLy8UJ:a00096.berlin.ptb.de/pls/portal/url/ITEM/350EB471A0561870E0401EAC35C87A24+Trapezoidal+triangular+and+rectangular+distributions+for+Type+B+evaluations&hl=cs&gl=cz&sig=AFQjCNF8Q2zqnMY7QJdZZ4yl_rG-k7nyDw>.
[RACE]	RACEK, Stanislav. *Pravděpodobnostní modely počítačů* . 2002 [cit. 2009-10-10]. Dostupný z WWW: <http://portal.zcu.cz>.
[URL-FRIE]	*Pravděpodobnost a statistika HYPERTEXTOVĚ* [online]. [2002-2004] [cit. 2009-10-10]. Dostupný z WWW: <http://home.zcu.cz/~friesl/hpsb/tit.html>.
[URL-WIKI]	*Wikipedia : the free encyclopedia that anyone can edit* [online]. [2001-] [cit. 2009-10-10]. Dostupný z WWW: <http://en.wikipedia.org>.

Příloha A – Ukázkový výstup programu

Pro počet generovaných čísel = 100000, parametr a=20, pro počet intervalů = 10

```
E_teorie=20.0
D_teorie=66.66666666666667
E_vypocet=19.717447221823512
D_vypocet=72.5142669344969

00,0000:****
04,0000:*************
08,0000:********************
12,0000:*************************
16,0000:****************************
20,0000:****************************
24,0000:*************************
28,0000:*******************
32,0000:**********
36,0000:***
```

Pro počet generovaných čísel = 100000 * 100000, parametr a=20 * 20, pro počet intervalů = 10 * 10

```
E_teorie=40.0
D_teorie=266.6666666666667
E_vypocet=39.43560949632585
D_vypocet=290.2908454213152

00,0000:****
04,0000:************
08,0000:*********************
12,0000:*******************************
16,0000:***************************************
20,0000:********************************************
24,0000:************************************************
28,0000:*************************************************
32,0000:*****************************************************
36,0000:********************************************************
40,0000:***********************************************************
44,0000:********************************************************
48,0000:*****************************************************
52,0000:**************************************************
56,0000:*******************************************
60,0000:**********************************
64,0000:*************************
68,0000:*****************
72,0000:**********
76,0000:***
```