

NÁHODNÁ ČÍSLA

TYPY GENERÁTORŮ, LINEÁRNÍ KONGRUENČNÍ GENERÁTORY, TESTY NÁHODNOSTI, VYUŽITÍ HODNOT NÁHODNÝCH VELIČIN V SIMULACI

CO JE TO NÁHODNÉ ČÍSLO?

Náhodné číslo definujeme jako nezávislé hodnoty z rovnoměrného rozdělení na otevřeném intervalu $(0,1)$. Toto rozdělení označujeme jako $R(0,1)$ a používá se ke generování hodnot náhodných veličin. Náhodná veličina může mít na rozdíl od náhodného čísla různé pravděpodobnostní rozdělení.

Náhodná čísla mají následující vlastnosti:

Hustota pravděpodobnosti: $f(x) = 1$ pro $x \in (0,1)$,

$f(x) = 0$ jinak

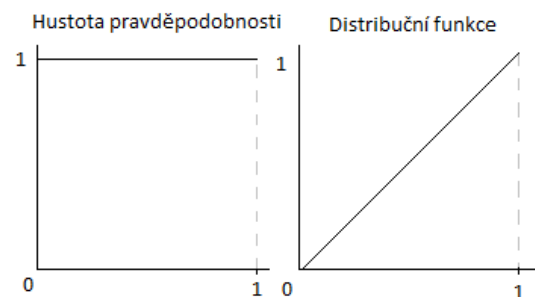
Distribuční funkce: $F(x) = 0$ pro $x \leq 0$,

$F(x) = x$ pro $x \in (0,1)$,

$F(x) = 1$ pro $x \geq 1$

Střední hodnota: $E(x) = 0,5$

Rozptyl: $D(x) = 1/12$



JAK SE GENERUJÍ NÁHODNÁ ČÍSLA?

Náhodná čísla lze generovat některým z následujících **generátorů náhodných čísel**:

- (1) Tabulka náhodných čísel: můžeme ji získat například s využitím telefonního seznamu a vypadá třeba takto:

68 33 95 60 74 94 50 63 90 96 37 32 29 64 79 98
 66 14 80 97 36 56 41 42 21 85 72 43 54 70 24 30
 53 64 83 20 18 28 01 03 14 72 35 83 92 25 41 06
 51 30 27 07 52 58 14 78 08 91 57 16 68 77 12 73

44 48 87 57 84 79 95 32 47 26 35 19 50 14 75 73
 87 30 96 19 86 59 32 00 81 73 70 91 92 67 34 75
 58 76 73 83 31 55 63 53 66 64 55 20 53 67 76 03
 19 46 61 57 86 93 35 27 28 29 68 00 47 47 98 96

ZDROJ: PŘEDNÁŠKY 4EK412

Náhodná čísla by pak mohla být např.: 0,683; 0,395... (záleží na požadovaném počtu desetinných míst).

- (2) Mechanické generátory jsou generátory typu hrací kostka, mince, nebo speciální přístroje, které využívají při losování sázkové kanceláře – tudíž pro počítačové simulace těžko použitelné.
- (3) Fyzikální generátory jsou generátory založené na fyzikálních procesech. Mohou využívat například poločas rozpadu, špičky v síti či pohyb myši (počítačových, ale se skutečnými by to asi taky šlo), dříve se používala třeba i takováto lávová lampa. Tady jde o skutečně náhodná čísla, avšak nevýhoda je, že je k tomu kromě počítače potřeba nějaké další zařízení.



LÁVOVÁ LAMPA

ZDROJ: WIKIPEDIA.ORG

- (4) Aritmetické generátory: fungují tak, že další „náhodné“ číslo počítají z předchozího pomocí určité aritmetické operace, proto taková čísla označujeme jako pseudonáhodná. Patří sem:
- **von Neumannova metoda středu kvadrátu:** vezme nějaké číslo, například 472 654 123, a z něj vybere určitý počet prostředních, řekněme pět, když to nejde, může předsunout nulu. Toto číslo umocní, v našem případě bychom tedy dostali číslo 704 424 681. Opět vybereme prostředních pět číslic atd. Náhodná čísla by v tomto případě byla 0,26541; 0,44246.
 - **Aditivní lineární kongruenční generátor (Fibonacciův):** Zvolíme si číslo x_0 a parametr m . Následující číslo vypočítáme z předchozího pomocí vztahu: $x_{n+1} = (x_n + x_{n-1})(\text{mod } m)$, kde mod značí zbytek po celočíselném dělení. Slovem „kongruence“ označujeme stav, kdy dvě čísla dávají při celočíselném dělení číslem m stejný zbytek, odtud nejspíš název generátoru. Nevýhodou tohoto stejně jako dále uvedeného smíšeného generátoru je, že se maximálně po m krocích začne posloupnost opakovat, proto je třeba zvolit m dostatečně dlouhé.
 - **Multiplikativní kongruenční generátor (Lehmerův):** počítá následující číslo z předchozího vztahem $x_{n+1} = ax_n(\text{mod } m)$. Nevýhoda je, že jakmile dostaneme nulu, budou nedále vycházet už jen samé nuly. Vygenerovaná hodnota se na rozdíl od předchozího generátoru může opakovat, takže perioda může být delší než m .
 - **Smíšený lineární kongruenční generátor:** počítá následující číslo z předchozího vztahem $x_{n+1} = (ax_n + c)(\text{mod } m)$.

Aditivní generátor		Multiplikativní generátor		Smíšený generátor	
$m = 7, x_0 = 100, x_1 = 120$	$r = x/m$	$m = 7, x_0 = 100, a = 31$	$r = x/m$	$m = 7, x_0 = 100, a = 31, c = 4$	$r = x/m$
$x_2 = 100 + 120 (\text{mod } 7) = 3$	0,42857	$x_1 = 31 \cdot 100 (\text{mod } 7) = 6$	0,85714	$x_1 = 31 \cdot 100 + 4 (\text{mod } 7) = 3$	0,42857
$x_3 = 120 + 3 (\text{mod } 7) = 4$	0,57142	$x_2 = 31 \cdot 6 (\text{mod } 7) = 4$	0,57142	$x_2 = 31 \cdot 3 + 4 (\text{mod } 7) = 6$	0,85714

JAK OVĚŘIT, JESTLI JSOU NÁHODNÁ ČÍSLA SKUTEČNĚ NÁHODNÁ?

To, zda je posloupnost čísel vygenerovaná některým z uvedených generátorů skutečně náhodná (resp. pseudonáhodná, ale s vlastnostmi náhodných čísel), je třeba testovat. Použít můžeme například některý z následujících **testů náhodnosti**. Kromě empirických testů náhodnosti, které pracují přímo s vygenerovanou posloupností čísel, existují i testy teoretické, pomocí kterých zkoumáme parametry samotného generátoru. Ty vycházejí z teorie čísel.

Mezi běžné testy náhodnosti patří následující:

- **Frekvenční test** zjišťuje, jestli jsou čísla rovnoměrně rozdělena. Rozdělíme interval (0,1) na k intervalů (nemusí být stejné délky). Pravděpodobnost, že náhodná veličina nabude hodnoty z intervalu (a, b) je $b-a$. Nulovou hypotézou je shoda očekávaných a skutečných četností v jednotlivých intervalech. Nejpoužívanější testem je chí-kvadrát test. Statistika χ^2 se vypočítá podle vzorce:

$$\chi^2 = \sum_{j=1}^k (o_j - e_j)^2 / o_j$$

kde o_j je očekávaná četnost v intervalu j , e_j je empirická četnost v intervalu j . Vypočtená hodnota se porovná s kritickou hodnotou Chí kvadrát rozdělení s $k-1$ stupni volnosti pro zvolenou hladinu statistické významnosti $\alpha-1$, přičemž $k-1$ je tam proto, že známe celkový počet čísel, tudíž vlastně známe i počet čísel v jednom z intervalů: jsou to všechna čísla, která nejsou v ostatních intervalech, takže ztratíme jeden stupeň volnosti. Při překročení kritické hodnoty zamítneme nulovou hypotézu o rovnoměrném rozdělení náhodných čísel.

Například máme 20 čísel, jejichž rozdělení do intervalů je v následující tabulce. V každém intervalu by měla být 4 čísla. Chí-kvadrát statistika se rovná 8,5. Kritická hodnota pro 4 stupně volnosti a hladinu významnosti 0,05 je přibližně 9,5. V Excelu použijeme funkci `chisq.inv`. Nezamítáme nulovou hypotézu, že nejde o náhodná čísla.

četnosti			
interval	empirické	očekávané	statistika
0 až 0,2	2	4	4
0,2 až 0,4	8	4	16
0,4 až 0,6	6	4	4
0,6 až 0,8	1	4	9
0,8 až 1	3	4	1
			8,5

- Kromě testu Chí-kvadrát existuje ještě Kolmogorovův-Smirnovův test, který také umožňuje ověřit, jestli pocházejí hodnoty z určitého rozdělení. Například v předchozím případě se Chí-kvadrát test moc nehodí, protože máme výběr malého rozsahu, a u Chí-kvadrát testu by mělo být v každém intervalu alespoň 5 hodnot. Ukážeme si na předchozím příkladu i KS test:

Nulová hypotéza je, že výběr pochází z rovnoměrného rozdělení (0,1). Celkový počet hodnot označíme n , pořadí hodnot označíme i . Testovou statistiku spočítáme podle následujícího vzorce:

$$D_i^* = \max \left\{ \left| \frac{i}{n} - F_0(x) \right|, \left| \frac{i-1}{n} - F_0(x) \right| \right\}, i = 1 \dots n,$$

kde D_i^* je maximální odchylka mezi teoretickou a pozorovanou distribuční funkcí v bodech nespojitosti pozorované distribuční funkce. Testové kritérium D_n je maximum z dílčích odchylek D_i^* . Takže jak na to:

- (1) Nejprve seřadíme hodnoty a přiřadíme jim pořadí i (druhý sloupec).
- (2) Pak zjistíme teoretickou hodnotu distribuční funkce pro danou hodnotu (pátý sloupec).
- (3) Spočítáme absolutní hodnotu rozdílu i/n a této teoretické distribuční funkce a hodnotu rozdílu $(i - 1)/n$ a této teoretické distribuční funkce (šestý a sedmý sloupec).
- (4) Maximum z těchto dvou absolutních hodnot označíme D_i^* (osmý sloupec).

Testová statistika D_n je maximem z hodnot D_i^* . V našem případě je testová statistika rovna 0,21. V tabulkách pro KS test zjistíme, že kritická hodnota na 5% hladině významnosti je necelých 0,3, takže nulovou hypotézu nezamítáme.

seřazené hodnoty	pořadí (i)	$(i - 1)/n$	i/n	$F_0(x_i)$	$ i/n - F_0(x_i) $	$ (i - 1)/n - F_0(x_i) $	D_i^*
0,15	1	0	0,05	0,15	0,15	0,1	0,15
0,18	2	0,05	0,1	0,18	0,13	0,08	0,13
0,21	3	0,1	0,15	0,21	0,11	0,06	0,11
0,24	4	0,15	0,2	0,24	0,09	0,04	0,09
0,27	5	0,2	0,25	0,27	0,07	0,02	0,07
0,28	6	0,25	0,3	0,28	0,03	0,02	0,03
0,3	7	0,3	0,35	0,3	0	0,05	0,05
0,31	8	0,35	0,4	0,31	0,04	0,09	0,09
0,38	9	0,4	0,45	0,38	0,02	0,07	0,07
0,39	10	0,45	0,5	0,39	0,06	0,11	0,11
0,41	11	0,5	0,55	0,41	0,09	0,14	0,14
0,44	12	0,55	0,6	0,44	0,11	0,16	0,16
0,45	13	0,6	0,65	0,45	0,15	0,2	0,2
0,55	14	0,65	0,7	0,55	0,1	0,15	0,15
0,56	15	0,7	0,75	0,56	0,14	0,19	0,19
0,59	16	0,75	0,8	0,59	0,16	0,21	0,21
0,7	17	0,8	0,85	0,7	0,1	0,15	0,15
0,89	18	0,85	0,9	0,89	0,04	0,01	0,04
0,91	19	0,9	0,95	0,91	0,01	0,04	0,04
0,99	20	0,95	1	0,99	0,04	0,01	0,04

→ **Poker test** testuje četnost výskytu různých číslic ve vygenerovaných náhodných číslech. Například když máme číslo s pěti desetinnými místy, v některých by se měla určitá číslice opakovat jednou, v jiných dvakrát, v dalších by mělo být pět různých číslic, a jednotlivým situacím lze přiřadit určitou pravděpodobnost, kterou někdy někdo spočítal a kterou pro pět číslic uvádí tabulka níže. Tento test tedy považuje číslo 0,22145 za stejné jako 0,93123 (v obou je jedna číslice dvakrát). Pro testování shody empirických a očekávaných četností se pak opět může použít chí-kvadrát test dobré shody.

Varianta	Název	Pravděpodobnost	Příklad
a b c d e	Všechny různé	0,3024	0,12345
a a b c d	Jedna dvojka	0,5040	0,11456
a a b b c	Dvě dvojky	0,1080	0,44515
a a a b c	Trojka	0,0720	0,43383
a a a b b	Dvojka a trojka	0,0090	0,98998
a a a a b	Čtyřka (poker)	0,0045	0,77773
a a a a a	Pětka	0,0001	0,55555

ZDROJ: DLOUHÝ, M. A KOL: SIMULACE PODNIKOVÝCH PROCESŮ, 2007.

- **Test autokorelace** ověřuje existenci či neexistenci vazeb mezi prvky posloupnosti náhodných čísel $r_1 \dots r_N$. Testovou statistiku počítá podle vztahu: $C_k = \frac{1}{N-k} \sum_{i=1}^{N-k} (r_i \cdot r_{i-k})$, kde například pro zkoumání závislosti po sobě jdoucích čísel bude $k = 1$.
- **Hammingův test** se snaží zjistit, jestli se některé hodnoty nevyskytují s větší četností. Testovou statistiku spočítáme jako $\sum_{i=1}^{n-1} (r_i - 0,5)(r_{i+1} - 0,5)$ a měla by se blížit 0.
- **Test mezer** pracuje s relacemi tří sousedních čísel. Uvažujme tři sousední čísla v posloupnosti a, b, c: pak existuje 6 možných relací: $a > b > c$, $a > c > b$, $b > a > c$, $b > c > a$, $c > b > a$, $c > a > b$. Správně by v posloupnosti čísel měla být každá relace zastoupena v jedné šestině případů. Jestli tomu tak je, ověříme opět Chí-kvadrát testem dobré shody.

JAK SE VYUŽÍVAJÍ NÁHODNÉ VELIČINY V SIMULACI?

Simulační modely můžeme rozdělit na deterministické a stochastické. Ve stochastických procesech nemůžeme některé jevy či procesy popsat přesně, protože jsou náhodné, můžeme ale specifikovat jejich pravděpodobnostní rozdělení. Události v simulaci, jako třeba příchod zákazníka, výskyt vadného výrobku apod., budou tedy náhodnou veličinou ze specifikovaného rozdělení. Interval mezi příchody se nejčastěji popisuje exponenciálním rozdělením, počet příchodů za jednotku času diskrétním Poissonovým rozdělením. Zahrnutí náhodných veličin hraje roli při interpretaci výsledků: výsledkem simulace není přesně řešení, nýbrž jen statistický odhad.

ZDROJE:

Přednášky 4EK421 Simulační modely ekonomických procesů, VŠE Praha, 2013.

Dlouhý, M. a kol.: Simulace podnikových procesů. Computer Press, Brno 2007.