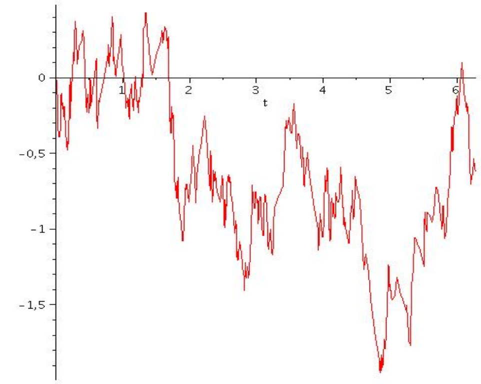
GP modellek II.

Fegyverneki Sándor
Miskolci Egyetem
Alkalmazott Matematikai Intézeti Tanszék
matfs@uni-miskolc.hu

2021. február 15.

1 Bevezetés

Figure 1: Mi ez?



Full Screen

Search

Close

(1) $$W(t) = \xi_0 \frac{t}{\sqrt{\pi}} + \frac{2}{\sqrt{\pi}} \sum_{j=1}^{\infty} \frac{\sin(jt)}{j} \xi_j,$$

ahol $t \in [0, \pi]$, $j \in \mathbf{N}$, $\xi_j \sim N(0,1)$, azaz standard Gauss-eloszlású.

Tétel:

(nagy számok gyenge törvénye) Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók sorozata. Létezik a szórásnégyzet. Ekkor tetszőleges $\varepsilon > 0$ esetén

$$\lim_{n \to +\infty} P\left(\left| \frac{X_1 + \cdots + X_n}{n} - E(X_1) \right| \geq \varepsilon \right) = 0.$$

Megjegyzés:

Legyen A esemény, $P(A) = p$, és S_n az A esemény gyakorisága az első n kísérletből egy Bernoulli kísérletsorozatnál.

Legyen

$$X_i = \begin{cases} 1, & \text{ha } A \text{ bekövetkezik az } i\text{-edik kísérletnél,} \\ 0, & \text{ha } A \text{ nem bekövetkezik az } i\text{-edik kísérletnél.} \end{cases}$$

Tehát

$$S_n = X_1 + \cdots + X_n.$$

Ekkor tetszőleges $\varepsilon > 0$ esetén

$$\lim_{n\to+\infty} P\left(|\frac{S_n}{n} - p| \geq \varepsilon\right) = 0. \tag{2}$$

$S_n \sim B(n, p)$, így

$$P\left(|\frac{S_n}{n} - P(A)| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}. \tag{3}$$

Tétel:

(centrális határeloszlás-tétel) Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók sorozata és létezik az $E(X_i) = \mu$ és $D^2(X_i) = \sigma^2 > 0$.

Ha $S_n = \sum_{k=1}^{n} X_k$, akkor

$$\lim_{n\to+\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < x\right) = \Phi(x), \qquad x \in \mathbf{R}, \tag{4}$$

ahol Φ a standard normális eloszlásfüggvény.

Centrális határeloszlás-tétel

Tétel:

Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók sorozata és létezik az $E(X_i) = \mu$ és $D^2(X_i) = \sigma^2 > 0$. Ha $S_n = \sum_{k=1}^{n} X_k$, akkor

$$\lim_{n \to +\infty} P\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} < x \right) = \Phi(x), \qquad x \in \mathbf{R}, \tag{5}$$

ahol Φ a standard normális eloszlásfüggvény.

Tétel:
(Moivre-Laplace) Legyen a ξ valószínűségi változó binomiális eloszlású n és p paraméterrel és $0 \leq a < b \leq n$ egész, akkor

$$P(a \leq \xi \leq b) = \sum_{k=a}^{b} \binom{n}{k} p^k q^{n-k} \approx \tag{6}$$

$$\approx \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right). \tag{7}$$

2 Monte Carlo módszerek

Definíció:

*Monte Carlo módszer*eknek nevezzük matematikai feladatok megoldásának véletlen mennyiségek modellezését felhasználó numerikus módszereit.

Véletlen számok

Kísérletezés – a szimuláció olcsóbb.

Kezdetek – véletlenszám táblázatok készítése.

Kísérletek – érme, tű, kocka (nehézségek, idő).

Full Screen

Search

Close

A szimuláció alapvető problémái:

egy determinisztikus számítógépen közelítjük a véletlent.

Diszkréttel közelítünk folytonosat vagy fordítva.

Végtelen feladat korlátos modell, véges szimuláció.

Ne felejtsük el, hogy a valószínűség-számítás fogalmai, tételei feltételezik, hogy az elemzés tömegjelenségre vonatkozik.

3 Pszeudovéletlen számok

Azokat az $x_1, x_2, \ldots, x_n$ számokat, amelyeket egy adott algoritmus alapján számítottunk ki, és a véletlen számok helyett használhatók, pszeudovéletlen számoknak nevezzük. A generálásuknak és ellenőrzésüknek (egyenletesség, véletlenszerűség) külön elmélete alakult ki. Ezzel itt nem foglalkozunk. A legtöbb magasszintű számítógépi programozási nyelv elég jó generátort tartalmaz beépített eljárásként. Azért ajánlatos az ellenőrzés. Itt most két egyszerű pszeudovéletlenszám generátort adunk meg.

Példa:

$$(8) \qquad x_1 = 1, \quad x_{n+1} \equiv 125 x_n \quad (mod \quad 8192).$$

Példa:

$$(9) \quad x_1 = 1, \quad x_{n+1} \equiv 16807 x_n \quad (mod \quad 2147483647).$$

Napjainkban majdnem minden számítógép (programozási nyelv, programcsomag) az előző példákhoz hasonló beépített kongruenciális generátort használ. Az $x_1, x_2, \ldots, x_k$ számok generálására ilyen a lineáris kongruencia vagy hatványmaradék módszer, ekkor a következő rekurzív kapcsolat adott:

$$x_{i+1} \equiv \alpha x_i + c \qquad (mod \quad m), \tag{10}$$

ahol α konstans szorzó, c a növekmény és m a modulus. Az x_0 kezdő érték az ún. "seed". Ha megoldjuk a 10 egyenletet, akkor azt kapjuk, hogy

$$x_n \equiv \left[\alpha^n x_0 + c\frac{\alpha^n - 1}{\alpha - 1}\right] \qquad (mod \quad m). \tag{11}$$

Nyilván a paraméterek határozzák meg a generátor "jóságát". A szokásos követelmények egy véletlenszám-generátorral szemben:

1. Jó statisztikai tulajdonságok. Tehát legyenek függetlenek (korrelálatlanok) és aznos eloszlásúak.
2. Az ismétlődési periódus legyen hosszú, hogy sok és változatos problémánál legyen alkalmazható.
3. Ismételhető legyen. Tehát ugyanazokra a paraméterekre ugyanazt a sorozatot adja.
4. A szimulációk többsége sok véletlen számot igényel, ezért legyen gyors és könnyen számolható.
5. Legyen könnyű a szeparált sorozatok készítése.

A paraméterek választására javasoljuk a [1] irodalmat.

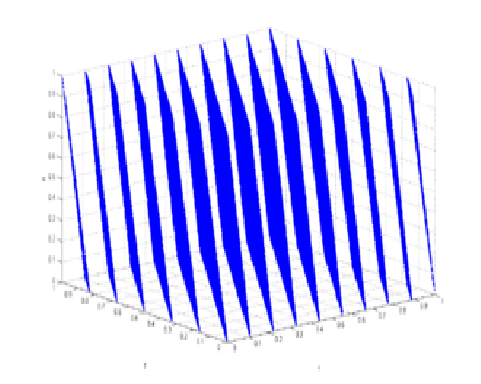
Megjegyzés:
Diszkrét egyenletes (klasszikus valószínűségi mező) eloszlást közelítenek a megadott rekurzív algoritmusok. Az 8 példa még számítógép nélkül is jól használható.

Példa: RANDU

$$x_1 = 1, \quad x_{n+1} \equiv 65539 x_n \quad (mod \quad 2^{31}). \tag{12}$$

A számítógépi algoritmusok legtöbbször (valójában mindig diszkrétet, hiszen véges a számábrázolásuk) a $[0,1]$ intervallumon egyenletes eloszlást próbálják közelíteni, mert ebből különböző módszerek segítségével – a tanult eloszlások tulajdonságainak felhasználásával – más eloszlású véletlen számokat tudunk előállítani.

Figure 2: RANDU



4 Eloszlások generálása

Inverzfüggvény módszer

Ha F szigorúan monoton növő eloszlásfüggvény és X F eloszlású, akkor $Y = F(X)$ egyenletes eloszlású a $[0,1]$ intervallumon. Fordítva, ha $X \sim U(0,1)$, akkor $Y = F^{-1}(X)$ éppen F eloszlású.

$$F_Y(x) = P(Y < x) =$$

$$= P(F^{-1}(X) < x) = P(X < F(x)) = F(x).$$

Full Screen

Search

Close

Következmény:

1. Ha $X \sim U(0,1)$, akkor $Y = (b-a)X + a \sim U(a,b)$.

2. Ha $X \sim U(0,1)$, akkor $Y = -\frac{1}{\lambda}\ln(X) \sim Exp(\lambda)$.

3. Ha $X \sim U(0,1)$, akkor $Y = tg(\pi(X-0.5))$ standard Cauchy eloszlású.

4. Ha $X \sim U(0,1)$, akkor $Y = \Phi^{-1}(X)$ standard normális eloszlású.

Az elfogadás-elvetés módszere

Legyen az X valószínűségi változó sűrűségfüggvénye f, amelyhez létezik egy olyan g sűrűségfüggvény, hogy $f(x) \leq cg(x)$ (minden x-re és c egy véges konstans) és a g könnyen generálható eloszlású. Legyen az Y valószínűségi változó g sűrűségfüggvényű és $U \sim U(0,1)$, amely független Y-tól, ekkor

$$(Y| \text{ ha } cUg(Y) < f(Y)) \sim X, \tag{13}$$

azaz a feltételes valószínűségi változó éppen megfelel az X eloszlásának.

Bizonyítás: Valójában

$$c = \int cg(y)dy \geq \int f(y)dy > 0$$

és

$$f(y) = 0, \quad \text{ha} \quad g(y) = 0.$$

Full Screen

Search

Close

Tehát

$$P(Y \in [x, x+dx] | cUg(Y) < f(Y)) =$$

$$= \frac{g(x)dxP\left(U < \frac{f(x)}{cg(x)}\right)}{P\left(U < \frac{f(Y)}{cg(Y)}\right)} =$$

$$= \frac{g(x)\frac{f(x)}{cg(x)}dx}{\int \frac{f(y)}{cg(y)} g(y)dy} =$$

$$= f(x)dx = P(X \in [x, x+dx]). \tag{14}$$

Megjegyzés:

Ez a módszer akkor praktikus, ha Y könnyen generálható és c nem nagyon nagy (tehát az elutasítás nem gyakori). Ha lehetséges, akkor az optimális választás a c konstansra

$$c = \sup_x \frac{f(x)}{g(x)}.$$

Példa: Ha $Y \sim \Gamma(1,a)$, akkor $X = \dfrac{Y}{\lambda} \sim \Gamma(\lambda,a)$. Továbbá, ha $X \sim \Gamma(\lambda,a)$ és $Z \sim \Gamma(\lambda,b)$, akkor $X+Z \sim \Gamma(\lambda,a+b)$. Tehát elegendő csak olyan $Y \sim \Gamma(1,a)$ eloszlású véletlen számokat generálni, ahol $a \in (0,1]$, ekkor a sűrűségfüggvény

$$f_Y \leq \frac{a+e}{ae\Gamma(a)} g,$$

ahol

$$g(x) = \frac{eg_1(x) + ag_2(x)}{a+e},$$

amikor is

$$g_1(x) = ax^{a-1}, \text{ ha } 0 < x < 1,$$

míg

$$g_2(x) = e^{-x+1}, \text{ ha } 1 < x < +\infty.$$

Ekkor g_1 és g_2 is sűrűségfüggvény és mind a kettő szimulálható az inverzfüggvény módszerrel. g pedig a kettő keveréke, ahol a súlyok

$$\frac{e}{a+e} \text{ és } \frac{a}{a+e}.$$

Generálunk egy egyenletest a $(0,1)$ intervallumon ez eldönti, hogy melyik függvénnyel folytatjuk felhasználva az inverzfüggvény módszert és utána az elfogadás-elvetés módszerével kapjuk az Y értékét. Tehát három $U(0,1)$ típusú véletlen számot használunk fel.

Normális eloszlás generálása

A normális eloszlás eloszlásfüggvénye nehezen kezelhető, ezért számos generátort találtak ki a tulajdonságai alapján. Néhány példa.

Példa: Ha $X_i \sim U(0,1)(i = 1, \ldots, 12)$, akkor

$$Y = \sum_{i=1}^{12} X_i - 6$$

közelítőleg standard normális eloszlású.

Ez a centrális határeloszlás-tétel egy véges alkalmazása.

Nem hatékony, mert sok véletlen számot használ.

A 3 ábrán látható, ha csak három összegét tekintjük.

Full Screen

Search

Close

Figure 3:

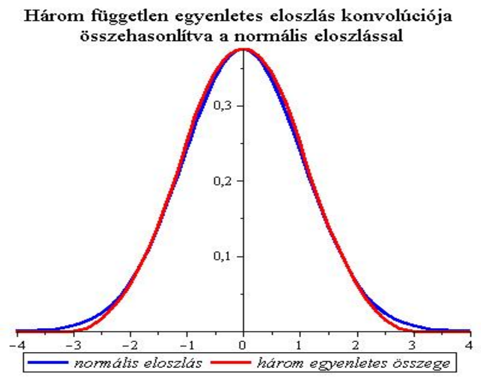

Három független egyenletes eloszlás konvolúciója
összehasonlítva a normális eloszlással
0,3
0,2
0,1
−4
−3
−2
−1
0
1
2
3
4
normális eloszlás
három egyenletes összege

Figure 4:

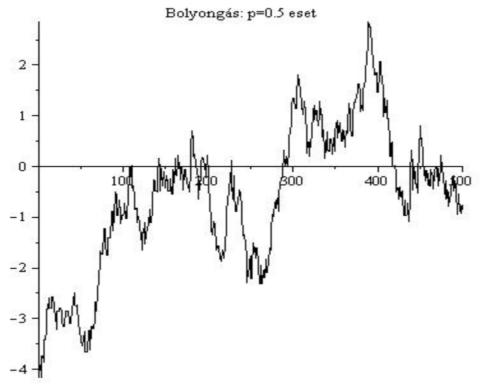
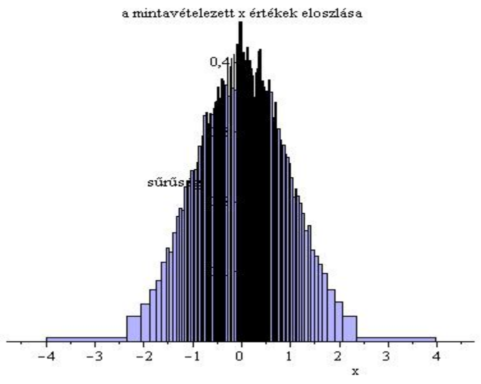


Figure 5:


a mintavételezett x értékek eloszlása
0,4
sűrűsé
-4
-3
-2
-1
0
1
2
3
4
x

Példa: A legtöbb statisztikai programcsomag a következő ún. Box-Müller módszert használja. Legyen

$$U_i \sim U(0,1)(i=1,2),$$

ekkor

$$X_1 = \sqrt{-2\ln U_1}\cos(2\pi U_2),$$

$$X_2 = \sqrt{-2\ln U_1}\sin(2\pi U_2)$$

közelítőleg standard normális eloszlásúak.

Full Screen

Search

Close

5 A közelítő integrálás hibája

Az egyszerű Monte-Carlo módszer esetén a hibabecslés jellemzésére általában a szórást használjuk.

Legyen h egy tetszőleges valós függvény, amely esetén az

$$\int_{-\infty}^{\infty} h^2(x)dF(x) \tag{15}$$

létezik. Ez szükséges és elégséges feltétele, hogy az $Y = h(X)$ valószínűségi változó, ahol X F eloszlásfüggvényű, szórásnégyzete létezzen. Továbbá legyen

$$E(h(X)) = \mu, \quad \text{és} \quad D^2(h(X)) = \sigma^2, \tag{16}$$

akkor az $X_1, X_2, \ldots, X_n$ minta esetén (X_i F eloszlású) a hibabecslés szórásnégyzete

(17)

$$D^2(\varepsilon) = \frac{1}{n^2} D^2(h(X_1) + h(X_2) + \cdots + h(X_n)) = \frac{\sigma^2}{n}.$$

Ebből leolvashatjuk a Monte-Carlo módszer egy igen lényeges tulajdonságát: ha a mintaelemek számát növeljük a hiba illetve a jellemzését adó szórás csak $\sqrt{n}$ arányában csökken. Látszólag ez azt jelenti, hogy azok a jó becslések, amelyeknek kicsi a szórása. De azzal, hogy a robusztus tulajdonságok nem változnak meg egy konstans tényező hatására az következik, hogy más szempontból kell összehasonlítani az integrálási tulajdonságokat, illetve érzékenységeket. Ezeket a további vizsgálatokat célszerű úgy elvégezni, hogy a szórások legyenek egyenlőek a becsléseknél. Legyen ez a közös érték 1, s az ilyen egyenletet nevezzük *kanonikus egyenletnek*.

Full Screen

Search

Close

Példa: Hány darab véletlen számot kell generálni ahhoz, hogy az

$$I = \int_0^{\frac{\pi}{2}} \sin x dx \tag{18}$$

integrált megbecsüljük úgy, hogy a becslés abszolút hibája legfeljebb I 0.1% legyen legalább 0.99 valószínűséggel?

Bizonyítás: Tudjuk, hogy

$$\frac{2}{\pi}I = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \sin x dx = \int_0^{\frac{\pi}{2}} \sin x \frac{2}{\pi} dx = E(\sin X), \tag{19}$$

ahol $X \sim U\left(0, \frac{\pi}{2}\right)$. Tehát I egy közelítő értéke

$$I_n = \frac{\pi}{2n}\sum_{i=1}^{n} \sin X_n, \tag{20}$$

ahol X_n pszeudovéletlenszám a $\left(0, \frac{\pi}{2}\right)$ intervallumból. Felhasználva, hogy

$$\frac{I_n - I}{D(I_n)} \sim N(0, 1), \tag{21}$$

ahol

$$D^2(I_n) = \frac{\pi^2}{4n} D^2(\sin X) = \frac{\pi^2 - 8}{8n} \tag{22}$$

kapjuk, hogy $n \approx 1550579$.

Köszönöm
a figyelmet!

References

[1] Deák I.: *Véletlenszámgenerátorok és alkalmazásaik,* Akadémiai Kiadó, Budapest, 1986. 13

[FR11] Fegyverneki Sándor, Raisz Péter: *Sztochasztikus modellezés,*elektronikus jegyzet, 2011, TÁMOP 4.1.2-08/1/A-2009-0001 project, https://www.uni-miskolc.hu/~ matfs/

[FS11] Fegyverneki Sándor: *Valószínűség-számítás és matematikai statisztika,* elektronikus jegyzet, Kempelen Farkas elktronikus könyvtár, 2011, TÁMOP 4.1.2-08/1/A-2009-0001 project, https://www.uni-miskolc.hu/~ matfs/

[FE78] W. Feller: *Bevezetés a valószínűségszámításba és alkalmazásaiba,* Műszaki Könyvkiadó, Budapest, 1978.

Full Screen

Search

Close

[2] I.M. Szobol: *A Monte-Carlo módszerek alapjai*, Műszaki Könyvkiadó, Budapest, 1981.