Atsitiktiniu vektorių generavimas

Lukas Klusis

2014 birželio 20 d.

1 Įvadas

Yra žinoma daugybė variantų, kaip generuoti 1-dimensijos (pseudo) atsitiktinius skaičius. Turbūt populiariausi yra Tiesinis kongruentinis bei Tiesinis rekurentinis metodai, kurie leidžia mums gauti norimo ilgio tolygiai pasiskirsčiusių atsitiktinių skaičių seką greitai ir efektyviai. Tačiau dažnai neužtenka 1-dimensijos skaičių. Šie metodai taip pat taikomi norint gauti n-dimensijų vektorius, tačiau reikia papildomų veiksmų ir sąlygų. Generuojant atsitiktinį vektorių svarbu atsižvelgti į šio vektoriaus individualių kintamųjų koreliaciją. Yra žinoma, jeigu $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ yra atsitiktinis skaičius, su pasiskirstymo funkcija $G(x_1, x_2, \ldots, x_n) = P\{\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n\}$, tai ji gali būti išskaidyta į dvi pasiskirstymo funkcijas:

$$G(x_1, x_2, \ldots, x_n) = G_1(x_1) G_{2\ldots n}(x_2, \ldots, x_n | x_1) \tag{1}$$

kur G_1 yra pirmosios komponentės besąlygiška pasiskirstymo funkcija, o $G_{2\ldots n}$ yra sąlygiškai pasiskirsčiusi funkcija komponenčių $x_2, \ldots, x_n$ su sąlyga $x_1 = \xi_1$. Pritaikius šią idėją funkcijai $G_{2\ldots n}$ ir paėmus (1) lygybės dešinę pusę, po $(n-1)$ karto mes išskaidytume šią funkciją tokiu būdu:

$$\begin{aligned} G(x_1, x_2, \ldots, x_n) = G_1(x_1) G_2(x_2|x_1) G_3(x_3|x_1, x_2) \\ \ldots G_n(x_n|x_1, x_2, \ldots x_{n-1}) \end{aligned} \tag{2}$$

kur $G_i(x_i|x_1, x_2, \ldots x_{i-1})$ yra vektoriaus ξ nario ξ_i sąlyginio pasiskirstymo funkcija. Pasinaudoję (2) lygtimi galime sukonstruoti bendrą procedūra, kuri generuoti atsitiktinius vektorius su pasiskirstymo funkcija $G(x_1, x_2, \ldots, x_n)$:

Vektoriaus generavimas atitinkantis pasiskirstymo funkciją:

1. Generuojame reikšmę z_1, tokią kad atitiktų pasiskirstymo funkciją $G_1(x_1)$ ir nustatome kintamąjį $i \longleftarrow 2$

2. Generuojame skaičių z_i pagal sąlyginę pasiskirstymo funkciją $G_i(x_i|x_1 = z_1, \ldots x_{i-1} = z_{i-1})$.

3. Jeigu $n > i$, tai padidiname i reikšmę vienetu $i \longleftarrow i + 1$ ir grįžtame į Žingsnį 2.

4. Gauname vektorių $(z_1, z_2, \ldots, z_n)$, kuris yra pasiskirstęs pagal pasiskirstymo funkciją $G(x_1, x_2, \ldots, x_n)$.

Naudojantis panašiu principu, šio darbo tikslas bus suprasti ir pademonstruoti kaip galima standartizuoti atsitiktinių vektorių generavimą. Arba tiksliau tariant, nagrinėsime kaip reikia generuoti vektorius pasiskirsčiusius pagal Normalųjį skirstinį.

2 Reikšmės pasiskirsčiusios pagal normalųjį dydį

Iš pradžių mums reikia išnagrinėti kaip galima generuoti 1-dimensijos atsitiktinius skaičius, kurie būtų pasiskirstę pagal normalųjį dydį, t.y. generuojamos sekos pasiskirstymo funkcija $G(x)$ atitiktų funkciją $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t$ Naudokimės tokiu algoritmu:

Normaliai pasiskirsčiusio atsitiktinio dydžio gavimas

1. Generuojame u_1 ir u_2, kur u_1 ir u_2 yra nepriklausomi ir tolygiai pasiskirstę intervale $[0, 1)$ atsitiktiniai dydžiai.

2. Apskaičiuojame dydžius $v_1 = 2u_1 - 1$ ir $v_2 = 2u_2 - 1$, t.y. gauname nepriklausomus atsitiktinius dydžius pasiskirsčiusius intervale $[-1, 1)$

3. Apskaičiuojame reikšmę:

$$s = v_1^2 + v_2^2. \tag{3}$$

4. Jeigu $S \geq 1$ grįžtame į Žingsnį 1. Priešingu atveju jeigu $0 \leq S < 1$ tai tęsiame toliau.

5. Apskaičiuojame dydžius:

$$x_1 = v_1 \sqrt{\frac{-2 \ln s}{s}}, x_2 = v_2 \sqrt{\frac{-2 \ln s}{s}} \tag{4}$$

gauti dydžiai x_1 ir x_2 yra nepriklausomi ir normaliai pasiskirstę atsitiktiniai dydžiai, t.y. $x_1, x_2 \in \mathcal{N}(0,1)$

Įrodymas: [1] Pereikime prie polinių koordinačių

$$\begin{aligned} v_1 = r\cos\theta, v2 = r\sin\theta \\ \Rightarrow s = r^2 \\ \Rightarrow x_1 = \sqrt{-2\ln s}\cos\theta, x_2 = \sqrt{-2\ln s}\sin\theta \end{aligned}$$

Pažymėkime

$$r' = \sqrt{-2lns}$$

Dydžiai r ir θ yra nepriklausomi atsitiktiniai dydžiai, kintantys vienetinio skritulio viduje, nes v_1 ir v_2 nepriklausomi. Tuomet r' ir θ taip pat nepriklausomi dydžiai. Tikimybė patekti taškui į vienodo kampo skritulio išpjovas vienoda, todėl θ yra tolygiai pasiskirstęs intervale $[0, 2\pi)$.

Tikimybė

$$\begin{aligned} P(r' < R) = P(\sqrt{-2lns} < R) = P(-2lnS < R^2) \\ = P(lnS > -\frac{R^2}{2}) = P(S > e^{-\frac{R^2}{2}}) = P(r^2 > e^{-\frac{R^2}{2}}) \\ = 1 - P(r^2 \leq e^{\frac{R^2}{2}}) = 1 - e^{-\frac{R^2}{2}} \end{aligned}$$

nes dydis r^2 yra tolygiai pasiskirstęs intervale [0,1). Tai galima įsitikinti, paskaičiavus geometrinę tikimybę:

$$P(r^2 < t) = P(r < \sqrt{t}) = \frac{\text{skritulio, su spinduliu } \sqrt{t}\text{, plotas}}{\text{skritulio, su spinduliu 1, plotas}} = \frac{\pi t}{\pi} = t$$

Taigi atsitiktinio dydžio r' pasiskirstymo funkcija

$$F_{r'}(R) := P(r' < R) = 1 - e^{-\frac{R^2}{2}}, r \geq 0.$$

Jau minėjome, kad dydis θ tolygiai pasiskirstęs intervale $[0, 2\pi)$, todėl jo pasiskirstymo funkcija

$$F_\theta(\varphi) := (\theta < \varphi) = \frac{\varphi}{2\pi}, 0 \leq \varphi < 2\pi.$$

[1] Pagal G. Stepanausko pateiktą mokomąją medžiagą „Monte Karlo Metodas", 2008 [1, p. 70]

Dydžiai x_1 ir x_2 bus nepriklausomi ir normaliai pasiskirstę pagal $\mathcal{N}(0,1)$, jei

$$P(x_1 < x_1', x_2 < x_2') = P(x_1 < x_1')P(x_2 < x_2') = \Phi(x_1')\Phi(x_2').$$

Įsitikinsime, kad taip ir yra:

$$\begin{aligned}
P(x_1 < x_1', x_2 < x_2') &= P(r'\cos\theta < x_1', r'\sin\theta < x_2') \\
&= \iint\limits_{\{(R,\varphi)|R\cos\varphi < x_1', R\sin\varphi < x_2'\}} \mathrm{d}F_{r',\theta}(R,\varphi) \\
&\overset{\text{r' ir } \theta \text{ nepriklausomi}}{=} \iint\limits_{\{(R,\varphi)|R\cos\varphi < x_1', R\sin\varphi < x_2'\}} \mathrm{d}F_{r'}(R)\mathrm{d}F_{\theta}(\varphi) \\
&= \iint\limits_{\{(R,\varphi)|R\cos\varphi < x_1', R\sin\varphi < x_2'\}} Re^{-\frac{R^2}{2}}\frac{1}{2\pi}\mathrm{d}R\mathrm{d}\varphi \\
&= \left[\begin{smallmatrix} x=R\cos\varphi \\ y=R\sin\varphi \end{smallmatrix} \begin{vmatrix} \frac{\partial x}{\partial R} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial R} & \frac{\partial y}{\partial \varphi} \end{vmatrix} = R \right] = \frac{1}{2\pi} \iint\limits_{\{(x,y)|x<x_1', y<x_2'\}} e^{-\frac{x^2+y^2}{2}}\mathrm{d}x\mathrm{d}y \\
&= \left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_1'} e^{-\frac{x^2}{2}}\mathrm{d}x\right)\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_2'} e^{-\frac{y^2}{2}}\mathrm{d}y\right) = \Phi(x_1')\Phi(x_2')
\end{aligned}$$

3 Vektoriai pasiskirstę pagal normalųjį dydį

Tarkime n-dimensijų normaliai pasiskirstęs dydis **O** ir šis dydis turi koreliacijos matricą **R**, tai šio dydžio pasiskirstymo funkcija yra

$$\varphi(\mathrm{x}) = (2\pi)^{-\frac{n}{2}}|R|^{-\frac{1}{2}}e^{-\frac{1}{2}\mathrm{x}'R^{-1}\mathrm{x}} \tag{5}$$

čia matrica R yra teigiamai apibrėžta ir simetriška. Tarkime matrica **T** yra tokia, kuri tenkina lygybę $\mathbf{R} = \mathbf{T}\mathbf{T}'$. Toliau mūsų apibrėžta procedūra generuoja vektorius, kurie yra pasiskirstę pagal pasiskirstymo funkciją φ.

Normaliai pasiskirsčiusio atsitiktinio dydžio gavimas

1. Generuojame skaičių $x_i, i = 1, \ldots, n$ naudojant 1-dimensijos standartinių normaliųjų dydžių generavimo metodą aptartą skyriuje 2

2. Skaičiuojame vektorių y $\longleftarrow$ **T**x. Gautas rezultatas yra atsitiktinis dydis pasiskirstęs pagal funkciją φ

Norint įrodyti, kad šis metodas yra korektiškas reikia parodyti du dalykus:

1. Parodyti, kaip yra gaunama matrica **T**, kai turime matricą **R** taip, kad:
$$\mathbf{R} = \mathbf{T}\mathbf{T}' \tag{6}$$
2. parodyti, kad sugeneruoto vektoriaus y pasiskirstymo funkcija yra φ

1) Pradėkime nuo parodymo, kaip gauti matricą T, atitinkančią lygtį (6). Ši matrica yra gaunama, pagal žinomas lygtis, vadinamas Cholesky transformacijomis:

$$\begin{aligned} t_{ij} &= 0, \text{ jeigu i<j}, \\ t_{i1} &= \frac{r_{i1}}{\sqrt{r_{11}}}, \text{ jeigu } 1 \le i \le n, \\ t_{ii} &= \sqrt{r_{ii} - \sum_{k=1}^{i-1} t_{ik}^2}, \text{ jeigu } 1 \le i \le n, \\ t_{ij} &= \frac{r_{ij} - \sum_{k=1}^{j-1} t_{ik}t_{jk}}{t_{jj}}, \text{ jeigu } 1 < j < i. \end{aligned} \tag{7}$$

2) Įrodymas [2] Tarkime, kad atsitiktinis kintamasis dydis x_i yra pasiskirstęs pagal funkciją $f_i(z_i) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}(z_i^2)}, i = 1, \ldots, n$. Tada vektoriaus x $= (x_1, x_2, \ldots, x_n)$ tankio funkcija yra šių funkcijų sandauga:

$$g(\mathrm{z}) = \prod_{i=1}^{n} f_i(z_i) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(z_1^2 + \ldots + z_n^2)}.$$

Vektorius y yra gautas iš vektoriaus x naudojant transformaciją y $=$ **T**x (arba kitaip galima parašyti x $= \mathbf{T^{-1}}$y). Atsitiktinio vektoriaus y $=$ **T**x tankio funkciją galima išreikšti tokiu būdu:

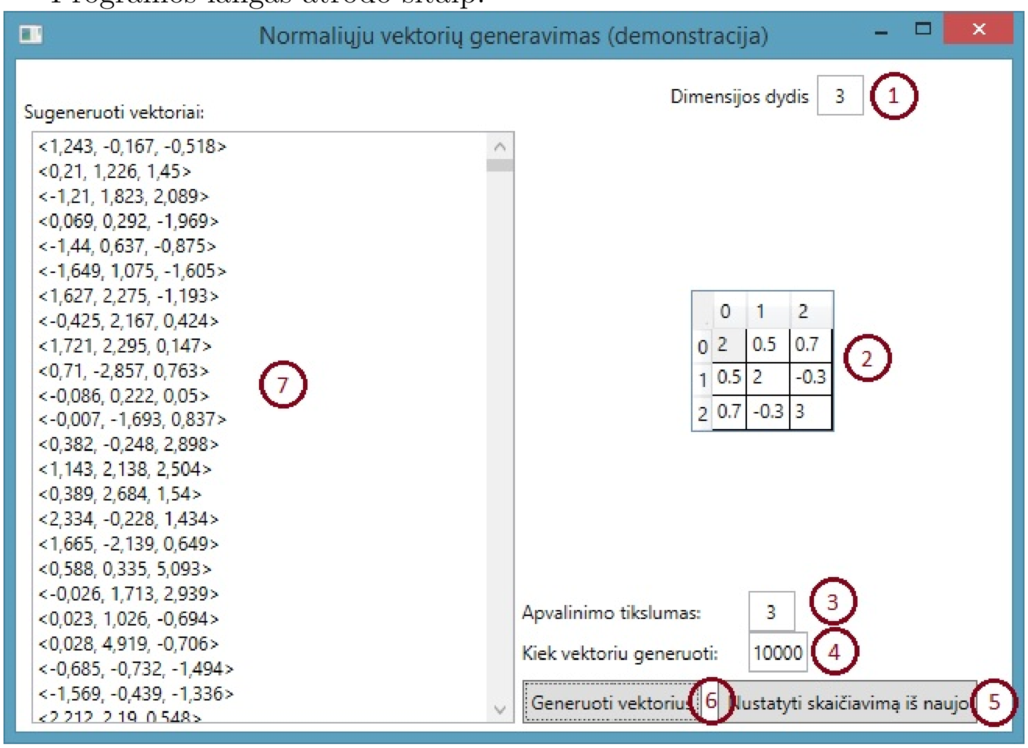
$$\begin{aligned} h(s) = g(T^{-1}s) \left| \frac{\partial T^{-1}s}{\partial s} \right| &= (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(T^{-1}s)'(T^{-1}s)} \left| T^{-1} \right| \\ &= (2\pi)^{-\frac{n}{2}} \left| R^{-\frac{1}{2}} \right| e^{-\frac{1}{2}s'R^{-1}s} = \varphi(s). \end{aligned}$$

[2] Pagal I. Deak pateiktą mokomąją medžiagą „Random number generators and simulation", 1990 [2, p. 187]

4 Demonstracinis atsitiktinių vektorių generavimas

Dabar jau galime sukonstruoti programą, kuri realizuoti aukščiau aprašytus metodus bei generuotų atsitiktinius vektorius pagal mūsų pateiktus parametrus. Programos realizaciją ir išeities kodus galima pasiekti adresu goo.gl/F14smU.

Programos langas atrodo šitaip:



Trumpai galime pakomentuoti šios programos funkcionalumą. Generuojame vektorių y $= (y_1, \ldots, y_n)$:

1. Galime numatyti kelių dimensijų vektorius programa generuos,

2. Pateikiame matricą $\mathbf{R}$, verta pastebėti, kad matricos įstrižainės reikšmė r_{ii} nurodo, kokia bus generuojamo vektoriaus $i - tosios$ dydžio standartinis nuokrypis σ_i, o matricos reikšmės $r_{ij}, i \neq j$ nurodo koreliaciją ρ_{ij} tarp dydžių y_i ir y_j, taip pat $r_{ij} = r_{ji}$,

3. nurodome, kiek skaitmenų po kablelio turės vektoriaus y kintamieji dydžiai $(y_1, \ldots, y_n)$,

4. nurodome, kiek vektorių bus sugeneruota,

5. siekiant pakartoti (pseudo) atsitiktinę vektorių seką galime nustatyti sekos skaičiavimą iš naujo. Taip galima sugeneruoti visiškai identišką seką sugeneruotai anksčiau,

6. paspaudus šį mygtuką pradedami generuoti vektoriai,

7. laukas į kurį yra išvedami pagal mūsų pateiktus parametrus sugeneruoti vektoriai.

Literatūra

[1] prof. G. Stepanauskas, *Monte Karlo Metodas*, http://uosis.mif.vu.lt/~stepanauskas/MK/Monte%20Karlo%20metodas.pdf

[2] I. Deak, *Random number generators and simulation*, Akademiai Kiado, Budapest, 1990, p.166-196.