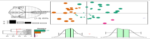
4. rész

Nevezetes eloszlások és generálásuk

Játék a véletlennel

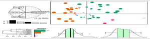
Komputerstatisztika kurzus

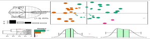
Barczy Mátyás és Ispány Márton 2010
Informatikai Kar
Debreceni Egyetem



A 4. rész témái

1. **Valós eloszlások**
2. **Nevezetes diszkrét eloszlások**
3. **Nevezetes abszolút folytonos eloszlások**
4. **Eloszlások generálása**





Valós eloszlások

A statisztikában a valószínűségszámításban megismert eloszlások alapvető szerepet játszanak. Emellett néhány eloszlás kifejezetten statisztikai problémáknál jelenik meg. Az eloszlásokkal alapvetően két okból találkozunk:

- velük modellezzük, írjuk le a minta viselkedését,
- a mintából származtatott számos fontos statisztikát jellemezhetünk általuk.

Definíció (Kolmogorov–féle valószínűségi mező)

$(\Omega, \mathcal{A}, P)$ hármas, ahol

- Ω egy nemüres halmaz (eseménytér),
- $\mathcal{A}$ Ω részhalmazainak egy σ–algebrája (események halmaza),
- $P : \mathcal{A} \to \mathbb{R}$ halmazfüggvény (valószínűség), melyre
  - $P(A) \geq 0, \forall A \in \mathcal{A}$ (nemnegatív),
  - $P(\Omega) = 1$ (1–re normált),
  - $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$, ha $A_i \in \mathcal{A}$, $A_i \cap A_j = \emptyset$, $i \neq j$ (σ–additív).

Definíció (valószínűségi változó)

Legyen $(\Omega, \mathcal{A}, P)$ egy valószínűségi mező. Az $X : \Omega \to \mathbb{R}$ leképezést valószínűségi változónak hívjuk, ha $\forall\, x \in \mathbb{R}$ esetén

$$\{\omega \in \Omega : X(\omega) < x\} \in \mathcal{A}.$$

Definíció (valószínűségi változó eloszlása)

A X valószínűségi változó eloszlásán a

$$P_X(B) := P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}), \ \ B \in \mathcal{B}(\mathbb{R}),$$

halmazfüggvényt értjük, ahol $\mathcal{B}(\mathbb{R})$ az $\mathbb{R}$ Borel–halmazainak a σ–algebráját jelöli ($\mathbb{R}$ nyílt halmazai által generált σ–algebra).

Állítás

A X valószínűségi változó eloszlása valószínűségi mérték a számegyenes Borel–halmazain, azaz $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ egy valószínűségi mező.

Megj.: P_X csak az $\mathbb{R}$ Borel–halmazain értelmezett, azaz $\mathbb{R}$ összes részhalmazának nem tudjuk a P_X szerinti mértékét definiálni.

Definíció (eloszlásfüggvény)

Az X valószínűségi változó eloszlásfüggvényén az $F_X : \mathbb{R} \to \mathbb{R}$,

$$F_X(x) := P(X < x) = P_X((-\infty, x)), \qquad x \in \mathbb{R},$$

függvényt értjük.

Tétel

Egy $F : \mathbb{R} \to \mathbb{R}$ *függvény akkor és csak akkor eloszlásfüggvénye valamely* X *valószínűségi változónak, ha*

- F *monoton növekvő,*
- F *balról folytonos,*
- $$\lim_{x \to -\infty} F(x) = 0 \qquad \textit{és} \qquad \lim_{x \to \infty} F(x) = 1.$$

Eloszlások típusai

Diszkrét statisztikai változók esetén (pl. nem, iskolai végzettség, processzorok száma) a minta értékei $\mathbb{R}$ egy diszkrét részhalmazára koncentrálódnak. Egy ilyen mintát diszkrét eloszlásokkal tudunk majd jól jellemezni.

Definíció (Diszkrét valószínűségi változó)

Egy X valószínűségi változót diszkrétnek hívunk, ha értékkészlete megszámlálható. Ilyenkor P_X véges vagy megszámlálhatóan végtelen sok pontra koncentrálódik, azaz létezik olyan $x_1, x_2, \ldots$ véges vagy végtelen valós sorozat, hogy $P_X(\{x_1, x_2, \ldots\}) = 1$.

Ha $p_i := P_X(\{x_i\})$, $i \in \mathbb{N}$, akkor fennáll, hogy

$$\text{(i)}\;\; p_i \geq 0,\; i \in \mathbb{N}, \qquad \text{(ii)}\;\; \sum_{i=1}^{\infty} p_i = 1.$$

Sokszor diszkrét eloszlás alatt egy, a fenti két tulajdonsággal rendelkező $p_1, p_2, \ldots$ számsorozatot értenek.

Sok statisztikai változó (pl. fékút, feszültség, hálózati sebesség) folytonos értékeket vesz fel.

Definíció (Abszolút folytonos valószínűségi változó)

Ha létezik egy olyan $f : \mathbb{R} \to \mathbb{R}$ függvény, hogy

$$F_X(x) = P_X((-\infty, x)) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \qquad x \in \mathbb{R},$$

akkor az X valószínűségi változót (abszolút) folytonos eloszlásúnak hívjuk. Az f függvényt pedig X sűrűségfüggvényének nevezzük.

Megj.: Ha X abszolút folytonos eloszlású, akkor eloszlásfüggvénye folytonos függvény.

Tétel

Egy $f : \mathbb{R} \to \mathbb{R}$ *függvény akkor és csak akkor sűrűségfüggvénye valamely* X *valószínűségi változónak, ha*

- f *Borel–mérhető,*
- $f(x) \geq 0$ *Lebesgue-m.m.* $x \in \mathbb{R}$ *esetén,*
- $$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1.$$

Bernoulli eloszlás

Indikátor, bináris (kétértékű) események jellemzésére használhatjuk. Jellegzetes kísérlet, melyre gondolhatunk vele kapcsolatban, az érmedobás. Az X v. v. Bernoulli eloszlást követ, ha

$$P(X = 1) = p, \qquad P(X = 0) = q := 1 - p,$$

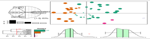
ahol $p \in [0, 1]$ paraméter. Jelölés: $\mathrm{Be}(p)$.

X várható értéke és szórásnégyzete:

$$E(X) = p, \qquad D^2(X) = p(1 - p).$$

Feladat (Szimmetrikus Bernoulli eloszlás)

Gyakran hasznos a Bernoulli eloszlás $Y := 2X - 1$ szimmetrikus változatával dolgozni. Számoljuk ki Y várható értékét és szórásnégyzetét!
Y eloszlását hívjuk Rademacher–eloszlásnak is.



Binomiális eloszlás

Dobjunk fel egy érmét n–szer egymás után. A fejek vagy írások számának eloszlása ún. binomiális eloszlás. Az X v.v. binomiális eloszlást követ, ha

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

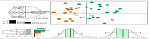
ahol $n \in \mathbb{N}$ és $p \in [0,1]$ paraméterek. Jelölés: $\mathrm{Bi}(n,p)$.

X várható értéke és szórásnégyzete:

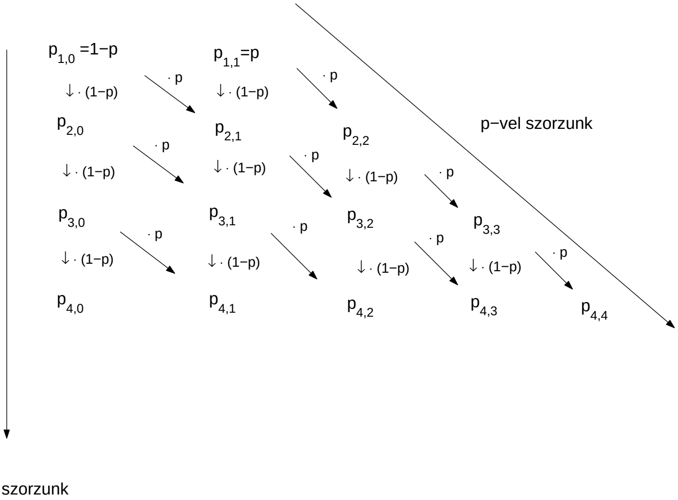
$$E(X) = np, \qquad D^2(X) = np(1-p).$$

Rekurzió a $p_{n,k} := \binom{n}{k} p^k (1-p)^{n-k}$ binomiális valószínűségek számolására:

$$p_{1,0} = 1 - p, \qquad p_{1,1} = p,$$
$$p_{n,0} = (1-p)p_{n-1,0}, \qquad p_{n,n} = p \cdot p_{n-1,n-1}, \ n \geq 2,$$
$$p_{n,k} = p \cdot p_{n-1,k-1} + (1-p)p_{n-1,k}, \ n \geq 2, \ k = 1, \ldots, n-1.$$



A rekurziót az alábbi ábra szemléletesen mutatja:



Binomiális valószínűségekre vonatkozó rekurzió.

A rekurzió helyességének igazolása:

$$p_{1,0} = \binom{1}{0} p^0 (1-p)^{1-0} = 1-p,$$

$$p_{1,1} = \binom{1}{1} p^1 (1-p)^{1-1} = p,$$

$$p_{n,0} = \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n$$

$$\Longrightarrow \quad p_{n,0} = (1-p) p_{n-1,0}, \quad n \geq 2,$$

$$p_{n,n} = \binom{n}{n} p^n (1-p)^{n-n} = p^n$$

$$\Longrightarrow \quad p_{n,n} = p \cdot p_{n-1,n-1}, \quad n \geq 2,$$

továbbá $n \geq 2$ és $k = 1, \ldots, n-1$ esetén

$$
\begin{aligned}
p \cdot p_{n-1,k-1} &+ (1-p)p_{n-1,k} \\
&= p\binom{n-1}{k-1}p^{k-1}(1-p)^{n-1-(k-1)} \\
&\qquad + (1-p)\binom{n-1}{k}p^{k}(1-p)^{n-1-k} \\
&= \left(\binom{n-1}{k-1} + \binom{n-1}{k}\right)p^{k}(1-p)^{n-k} \\
&= \binom{n}{k}p^{k}(1-p)^{n-k} = p_{n,k},
\end{aligned}
$$

ahol a harmadik lépés a Pascal–háromszög egyik nevezetes tulajdonsága alapján következik.

Rekurzió a binomiális eloszlás farok–valószínűségeinek számolására:

Alsó farok–valószínűségek:

$$a_{n,k} := \sum_{i=0}^{k} p_{n,i}, \qquad n \in \mathbb{N},\ k = 0, \dots, n.$$

Felső farok–valószínűségek:

$$f_{n,k} := \sum_{i=n-k}^{n} p_{n,i}, \qquad n \in \mathbb{N},\ k = 0, \dots, n.$$

Alsó farok–valószínűségekre rekurzió:

$$a_{1,0} = 1 - p, \qquad a_{1,1} = 1,$$
$$a_{n,0} = (1-p)a_{n-1,0},\ n \geq 2, \qquad a_{n,n} = 1,\ n \geq 1,$$
$$a_{n,k} = p \cdot a_{n-1,k-1} + (1-p)a_{n-1,k},\ n \geq 2,\ k = 1, \dots, n-1.$$

Poisson eloszlás

Ritka események, pl. sajtóhibák száma egy lapon, meghibásodások száma egy gépnél, mazsolák egy kalácsban, eloszlásának jellemzésére szolgál. Az X v. v. Poisson eloszlást követ, ha

$$P(X = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

ahol $\lambda > 0$ paraméter. Jelölés: $\mathrm{Po}(\lambda)$.
X várható értéke és szórásnégyzete:

$$E(X) = \lambda, \qquad D^2(X) = \lambda.$$

Rekurzió a Poisson eloszlás $p_k := \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$ valószínűségeinek számolására:

$$p_0 = \mathrm{e}^{-\lambda}, \qquad p_n = \frac{\lambda}{n} p_{n-1}, \qquad n \in \mathbb{N}.$$

Feladat

Adjunk rekurziót a farok–valószínűségek számolására!

Egyenletes eloszlás

Az egyik legtermészetesebb eloszlás az (egy halmazon adott) egyenletes eloszlás. Pl.: gyakran mondjuk, hogy dobjunk le egy pontot egyenletesen egy intervallumra. Ez azt jelenti, hogy egy részintervallumba esés valószínűsége arányos a részintervallum hosszával.

Az X egyenletes eloszlást követ az $[a, b]$ intervallumon, ha sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{ha } a < x < b, \\ 0 & \text{egyébként.} \end{cases}$$

Jelölés: $\mathrm{Uni}(a, b)$.

X várható értéke és szórásnégyzete:

$$E(X) = \frac{a+b}{2}, \qquad D^2(X) = \frac{(b-a)^2}{12}.$$

Feladat

Vezessük le a fenti összefüggéseket! Határozzuk meg X eloszlásfüggvényét!

Normális eloszlás

A természetben egyik leggyakrabban előforduló eloszlás, az alakja kapcsán gyakran harang–görbe eloszlásról beszélnek. Ennek elméleti alapja a központi határeloszlás tétel, melyet szokás úgy interpretálni, hogy sok független, azonos eloszlású kis hatás összesítése (megfelelő normálás után) normális eloszlást eredményez.

Az X normális eloszlást követ, ha sűrűségfüggvénye:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \qquad x \in \mathbb{R},$$

ahol $m \in \mathbb{R}$ és $\sigma > 0$ paraméterek. Jelölés: $\mathcal{N}(m, \sigma^2)$.
X várható értéke és szórásnégyzete (azaz a paraméterek jelentése):

$$E(X) = m, \qquad D^2(X) = \sigma^2.$$

Az $\mathcal{N}(0, 1)$ eloszlást standard normális eloszlásnak nevezik.

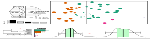
A standard normális eloszlás eloszlásfüggvénye

Ezt a függvényt Φ–vel jelöljük és alapvető szerepet játszik a statisztikában (ld. próbák, konfidencia intervallumok). Φ nem állítható elő elemi függvények segítségével, numerikus módszerekre van szükségünk a kiszámolásához.
Ha $x \geq 0$, akkor

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x \mathrm{e}^{-t^2/2} \mathrm{d}t.$$

Felhasználva az exponenciális függvény jól ismert hatványsorát kapjuk, hogy

$$\mathrm{e}^{-t^2/2} = \sum_{n=0}^{\infty} (-1)^n \frac{t^{2n}}{2^n n!}, \qquad t \in \mathbb{R}.$$



Felcserélve az integrálást és az összegzést, az alábbi alternáló sort kapjuk

$$\begin{aligned}\Phi(x) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{2^n n!} \int_0^x t^{2n}\,\mathrm{d}t \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2^n(2n+1)n!},\end{aligned} \tag{1}$$

amelynek elég nagy a kerekítési pontatlansága.

Numerikusan stabil sorfejtés Φ–re

Egy kis trükkel az alábbi sorfejtést kaphatjuk:

$$\Phi(x) = \frac{1}{2} + \sum_{n=0}^{\infty} a_n(x), \qquad x \geq 0, \tag{2}$$

ahol $a_n(x)$ eleget tesz az alábbi rekurziónak:

$$a_0(x) := \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} x, \qquad a_n(x) := \frac{x^2}{2n+1} a_{n-1}(x).$$

A közelítés (kerekítés) pontossága:

Legyen $x > 0$, $0 < \varepsilon < 1$ és $n_0 \in \mathbb{N}$ olyan, hogy $a_{n_0}(x) < 1$ és $x^2/(2n_0+1) < \varepsilon$. Ekkor $a_n(x) < \varepsilon^{n-n_0}$, $n \geq n_0 + 1$ és így

$$\left| \Phi(x) - \left(\frac{1}{2} + \sum_{n=0}^{n_0} a_n(x) \right) \right| = \sum_{n=n_0+1}^{\infty} a_n(x)$$

$$\leq \sum_{n=n_0+1}^{\infty} \varepsilon^{n-n_0} = \frac{\varepsilon}{1-\varepsilon}.$$

A (2) sorfejtés igazolása:

Bevezetve a $g : \mathbb{R} \to \mathbb{R}$,

$$g(x) := \mathrm{e}^{\frac{x^2}{2}} \int_0^x \mathrm{e}^{-\frac{t^2}{2}} \, \mathrm{d}t, \qquad x \geq 0$$

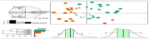
függvényt azt kell belátnunk, hogy

$$\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} g(x) = \sum_{n=0}^{\infty} a_n(x), \qquad x \geq 0,$$

azaz azt, hogy

$$g(x) = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!!} =: h(x), \qquad x \geq 0,$$

ahol $(2n+1)!! := 1 \cdot 3 \cdot 5 \cdots (2n+1)$.
Ellenőrizzük, hogy g és h ugyanazt a differenciálegyenletet elégíti ki ugyanazzal a kezdeti feltétellel, amiből következik, hogy $g = h$.



Mivel

$$g'(x) = x e^{\frac{x^2}{2}} \int_0^x e^{-\frac{t^2}{2}} \, dt + 1 = x g(x) + 1, \qquad x > 0,$$

$$\begin{aligned} h'(x) &= \sum_{n=0}^{\infty} \frac{2n+1}{(2n+1)!!} x^{2n} = 1 + \sum_{n=1}^{\infty} \frac{2n+1}{(2n+1)!!} x^{2n} \\ &= 1 + \sum_{n=0}^{\infty} \frac{2n+3}{(2n+3)!!} x^{2n+2} = 1 + x \sum_{n=0}^{\infty} \frac{1}{(2n+1)!!} x^{2n+1} \\ &= 1 + x h(x), \qquad x > 0, \end{aligned}$$

és $g(0) = h(0) = 0$, kapjuk a (2) sorfejtést.

Gamma eloszlás

Ez az eloszlás számos (exponenciális, χ^2) eloszlás közös általánosítása. Az X v. v. $p > 0$ és $\lambda > 0$ paraméterű gamma eloszlást követ, ha sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{\lambda^p x^{p-1} \mathrm{e}^{-\lambda x}}{\Gamma(p)} & \text{ha } x > 0, \\ 0 & \text{ha } x \leq 0, \end{cases}$$

ahol

$$\Gamma(p) := \int_0^\infty t^{p-1} \mathrm{e}^{-t} \mathrm{d}t$$

az ún. (teljes) gamma függvény. Jelölés: $\mathrm{Ga}(p, \lambda)$.

Ismert, hogy $\Gamma(p+1) = p\Gamma(p)$, $p > 0$.

X várható értéke és szórásnégyzete:

$$E(X) = \frac{p}{\lambda}, \qquad D^2(X) = \frac{p}{\lambda^2}.$$

Az eloszlásfüggvény számolása az ún. nemteljes gamma függvény segítségével történik, amelyre sorfejtés adható.

Nemteljes gamma függvény:

$$\Gamma(p,x) := \int_0^x t^{p-1}\mathrm{e}^{-t}\mathrm{d}t, \qquad x \geq 0.$$

Parciális integrálással ellenőrizhető, hogy

$$\Gamma(p,x) = \frac{x^p \mathrm{e}^{-x}}{p} + \frac{1}{p}\Gamma(p+1,x),$$

ami alapján X eloszlásfüggvénye:

$$\begin{aligned} F_X(x) &= \frac{1}{\Gamma(p)}\Gamma(p,x) = \frac{x^p \mathrm{e}^{-x}}{\Gamma(p+1)} + \frac{\Gamma(p+1,x)}{\Gamma(p+1)} \\ &= \frac{x^p \mathrm{e}^{-x}}{\Gamma(p+1)} + \frac{x^{p+1} \mathrm{e}^{-x}}{\Gamma(p+2)} + \frac{\Gamma(p+2,x)}{\Gamma(p+2)} \\ &= \cdots = x^p \mathrm{e}^{-x} \sum_{n=0}^{\infty} \frac{x^n}{\Gamma(p+n+1)}, \qquad x > 0. \end{aligned}$$

A fenti sorfejtés alkalmazhatósága abban rejlik, hogy a (teljes) gamma függvény jól közelíthető az ún. Stirling–formula segítségével:

$$\lim_{x\to\infty} \frac{\Gamma(x+1)}{\left(\frac{x}{\mathrm{e}}\right)^x \sqrt{2\pi x}} = 1.$$

Speciálisan nagy n–re:

$$n! \approx \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}.$$

Feladat (Speciális gamma eloszlások)

- Mutassuk meg, hogy $\mathrm{Ga}(1,\lambda)$ megegyezik a λ paraméterű exponenciális eloszlással, ahol $\lambda > 0$.
- Igazoljuk, hogy $\mathrm{Ga}(\frac{n}{2},\frac{1}{2})$ megegyezik $\xi_1^2 + \ldots + \xi_n^2$ eloszlásával, ahol $\xi_1,\ldots,\xi_n$ független $\mathcal{N}(0,1)$ eloszlásúak.

χ^2–eloszlás

Definíció

Az X valószínűségi változót n szabadsági fokú χ^2–eloszlásúnak nevezzük, ha eloszlása megegyezik $\xi_1^2 + \cdots + \xi_n^2$ eloszlásával, ahol $\xi_1, \ldots, \xi_n$ független standard normális eloszlású valószínűségi változók. Jelölés: χ_n^2.

Az előző feladat alapján $\chi_n^2 \sim \mathrm{Ga}(\frac{n}{2}, \frac{1}{2})$ és így X sűrűségfüggvénye:

$$f(x) = \begin{cases} \dfrac{2^{-\frac{n}{2}} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}}}{\Gamma(n/2)} & \text{ha } x > 0, \\ 0 & \text{ha } x \leq 0, \end{cases}$$

X várható értéke és szórásnégyzete:

$$\mathbb{E}(X) = n, \qquad \mathbb{D}^2(X) = 2n.$$

χ^2 **addíciós tétel:** ha X n szabadsági fokú χ^2–eloszlású, Y m szabadsági fokú χ^2–eloszlású és függetlenek, akkor $X + Y$ $n + m$ szabadsági fokú χ^2–eloszlású.

t– és F–eloszlás

Definíció (t–eloszlás)

Az X valószínűségi változót n szabadsági fokú t–eloszlásúnak nevezzük, ha eloszlása megegyezik

$$\xi/\sqrt{\eta/n}$$

eloszlásával, ahol ξ standard normális eloszlású, η n szabadsági fokú χ^2–eloszlású, ξ és η függetlenek. Jelölés: t_n.

Definíció (F–eloszlás)

Az X valószínűségi változót n, m szabadsági fokú F–eloszlásúnak nevezzük, ha eloszlása megegyezik

$$(\xi/n)/(\eta/m)$$

eloszlásával, ahol ξ és η független n, illetve m szabadsági fokú χ^2–eloszlásúak. Jelölés: $F_{n,m}$.

Eloszlások generálása

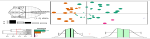
A statisztikában (és máshol is) gyakran szükség van a véletlen mesterséges előállítására. Erre példák az alábbi területek:

- Adott eloszlású minta generálása statisztikák, illetve statisztikai eljárások vizsgálatára.
- Determinisztikus mennyiségek (terület, valószínűség stb.) meghatározása Monte–Carlo típusú módsze–rekkel.
- Titkosítás, kódok generálása.

Ezek mindegyikének hatékonysága azon múlik, hogy rendelkezünk–e elég jó véletlenszám generátorral az adott platformon. (Kérdés: mit jelent az, hogy elég jó?)

Általános módszerek

- direkt (fizikai, pszeudo–véletlen),
- inverz,
- Neumann–féle.



Egyenletes eloszlású számsorozat

Nagyon sok véletlen számsorozatot az egyenletes eloszlásból állítunk elő.

A valószínűségszámításból ismert független egyenletes eloszlású sorozatot nehéz előállítani, a függetlenség garantálása önmagában is nagyon nehéz probléma.

Definíció

Egy $[0,1)$–beli $x_1, x_2, \ldots$ számsorozatot a $[0,1)$ intervallumon egyenletes eloszlásúnak nevezünk, ha minden $x \in [0,1]$ esetén fennáll, hogy

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i < x) \to x, \qquad \text{ha} \qquad n \to \infty,$$

ahol $I(A)$ az A esemény indikátorfüggvényét jelöli.

Állítás

Legyen $x_1, x_2, \ldots$ *egy* $[0,1)$*–beli, a* $[0,1)$ *intervallumon egyenletes eloszlású számsorozat. Ekkor tetszőleges* $x \leq y$*,* $x, y \in [0,1]$ *esetén*

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i \in [x,y)) \to y - x, \qquad \text{ha} \qquad n \to \infty, \tag{3}$$

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i \in [x,y]) \to y - x, \qquad \text{ha} \qquad n \to \infty. \tag{4}$$

Bizonyítás.

Mivel $x \leq y$, $x, y \in [0,1]$ esetén $[x,y) = [0,y) \setminus [0,x)$, kapjuk (3)–at.

Megmutatjuk, hogy $x \in [0,1]$ esetén

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x) \to x, \qquad \text{ha} \qquad n \to \infty.$$

Ez akkor és csak akkor teljesül, ha bármilyen $x \in [0,1]$ esetén

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i = x) \to 0, \qquad \text{ha} \qquad n \to \infty.$$

Indirekt módon tegyük fel, hogy létezik olyan $x \in [0,1]$, hogy az

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i = x), \quad n \in \mathbb{N},$$

sorozat nem tart 0–hoz, amint $n \to \infty$. Mivel ez a sorozat korlátos, a Bolzano–Weierstrass tétel alapján létezik konvergens részsorozata. A továbbiakban az általánosság megszorítása nélkül feltételezhetjük, hogy ez a részsorozat az eredeti sorozat. Így létezik $c_x > 0$, hogy

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i = x) \to c_x, \qquad \text{ha} \qquad n \to \infty.$$

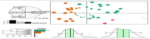
Mivel az $x_1, x_2, \ldots$ sorozat a $[0,1)$ intervallumon egyenletes eloszlású, kapjuk, hogy

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i \in [x, x + c_x/2)) \to \frac{c_x}{2}, \qquad \text{ha} \qquad n \to \infty.$$

Azonban

$$\frac{1}{n}\sum_{i=1}^{n} I(x_i = x) \leq \frac{1}{n}\sum_{i=1}^{n} I(x_i \in [x, x + c_x/2)), \qquad n \in \mathbb{N},$$

így annak kellene teljesülnie, hogy $c_x \leq c_x/2$. Ez azonban ellentmondás, hiszen $c_x > 0$. $\square$



Tétel (Weyl–kritérium)

Legyen $x_1, x_2, \dots$ *egy* $[0,1)$*–beli számsorozat. Ekkor az alábbi állítások ekvivalensek:*

- $x_1, x_2, \dots$ *a* $[0,1)$ *intervallumon egyenletes eloszlású.*
- *tetszőleges* $k \neq 0$, $k \in \mathbb{Z}$ *esetén*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} e^{2\pi i k x_j} = 0.$$

- *tetszőleges* $f : [0,1] \to \mathbb{C}$ *Riemann–integrálható függvény esetén*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(x_j) = \int_0^1 f(x)\,dx.$$

Példa

A korábbi definíció értelmében egyenletes eloszlású számsorozatok:

- $\{\alpha n\}$, $n \in \mathbb{N}$, ahol α egy irracionális szám,
- $\{f(n)\}$, $n \in \mathbb{N}$, ahol f egy olyan polinom, aminek van legalább egy, a konstans tagtól különböző irracionális együtthatója (Ez Weyl–tétele),
- $\{\sqrt{n}\}$, $n \in \mathbb{N}$,
- $\{\alpha p_n\}$, $n \in \mathbb{N}$, ahol α egy irracionális szám, $p_1, p_2, \ldots$ pedig a prímszámok sorozata (Ez Vinogradov–tétele, 1935),

ahol $\{\cdot\}$ a törtrészt jelöli.

Ezek bizonyítása mély számelméleti eredményeken alapszik.

Pszeudo–véletlenszámok

Valójában a legtöbb egyenletes véletlenszám generátor a kongruencia módszeren alapszik.

Lineáris kongruencia generátor (LCG):

$$x_k \equiv a \cdot x_{k-1} + b \mod m, \qquad k \geq 1, \tag{5}$$

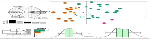
ahol m egy nagy szám (rendszerint prímszám) és $a, b, x_0 \in \{0, 1, \ldots, m-1\}$ adott.

A lineáris kongruencia generátorok nagyon érzékenyek a paraméterek megválasztására. Jól választott paraméterek esetén az eredményül kapott számsorozat a $\{0, 1, \ldots, m-1\}$ halmazon lesz egyenletes.

Ez leképezhető egy $[0, 1)$–beli számsorozattá, melyet pszeudo–véletlenszám sorozatnak is hívunk:

$$\widetilde{x_k} := \frac{x_k}{m}, \qquad k \geq 0.$$

Az $(\widetilde{x_k})$, $k \in \mathbb{Z}_+$, sorozat nem egyenletes eloszlású számsorozat (a korábbi definíció értelmében).



A módszert Lehmer vezette be 1949–ben, ő az alábbi generátort használta:

$$a = 23, \quad b = 0, \quad m = 10^8 + 1, \qquad x_0 = 47594118.$$

A lineáris kongruencia generátor periódusa (ciklusa): a legkisebb olyan $c \in \mathbb{N} \cup \{0\}$, melyre

$$x_{k+c} \equiv x_k \mod m, \quad k \geq 0.$$

Nyilván $c \leq m$.

Tétel (Greenberger, Hull, Dobell)

Az (5) lineáris kongruencia generátor periódusa akkor és csak akkor m (teljes periódusú LCG), ha

- $(b, m) = 1$,
- $p|(a-1)$ *minden olyan* p *prímre, melyre* $p|m$,
- $4|(a-1)$, *ha* $4|m$.

A SAS rendszer RANUNI egyenletes generátora az alábbi paraméterekkel működik:

$$a = 397204094, \qquad b = 0, \qquad m = 2^{31} - 1.$$

Egy másik lineáris kongruencia generátor (Random class, Java API),

$$a = 25214903917, \qquad b = 11, \qquad m = 2^{48}.$$

A lineáris kongruencia generátorok kezdőértékét seed-nek nevezik. Ennek használatával elérhető, hogy a generálás, a folyamat determinisztikussága miatt, újra megismételhető lehessen, és eredményül pontosan ugyanazt a számsorozatot állítsa elő.

Három fontos irányelv az LCG paramétereinek megválasztásához:

- tetszőleges x_0 kezdőérték esetén a kapott pszeudo–véletlen számsorozat „közelítőleg” egyenletes számsorozat legyen,
- tetszőleges x_0 kezdőérték esetén a generátor periódusa elég „nagy” legyen,
- hatékonyan számolható legyen a pszeudo–véletlen számsorozat.

Példa lineáris kongruencia generátorra

- $$x_k \equiv 10 \cdot x_{k-1} + 9 \mod 11, \quad k \geq 1,$$
  $$x_0 = 9.$$

  Ekkor $x_0 = 9$, $x_1 = 0$, $x_2 = 9$, és így a periódus 2. Továbbá,

  $$\widetilde{x}_{2n} = \frac{9}{11}, \quad n \in \mathbb{Z}_+, \qquad \widetilde{x}_{2n+1} = 0, \quad n \in \mathbb{Z}_+.$$

  Így, ha $0 < x \leq 9/11$, akkor

  $$\frac{1}{2n+1} \sum_{i=1}^{2n+1} I(\widetilde{x}_i < x) = \frac{n}{2n+1} \to \frac{1}{2} \qquad \text{ha } n \to \infty,$$

  ami alapján az $(\widetilde{x}_k)$, $k \in \mathbb{Z}_+$, sorozat nem egyenletes eloszlású számsorozat.

-

$$x_k \equiv 7 \cdot x_{k-1} + 9 \mod 11, \quad k \geq 1,$$
$$x_0 = 9.$$

Ekkor

$$x_0 = 9,\ x_1 = 6,\ x_2 = 7,\ x_3 = 3,\ x_4 = 8,\ x_5 = 10,$$
$$x_6 = 2,\ x_7 = 1,\ x_8 = 5,\ x_9 = 0,\ x_{10} = 9,$$

és így a periódus 10.

Nem egyenletes eloszlású számsorozat.

Az inverz módszer

Az inverz módszerrel, mely az alábbi tételen alapszik, az egyenletes eloszlásból állíthatunk elő más eloszlásokat.

Tétel

Legyen az X v. v. F eloszlásfüggvénye szigorúan monoton növekvő és folytonos.

(i) *Ekkor $F(X)$ egyenletes eloszlású v. v. $(0,1)$–en.*

(ii) *Ha U egyenletes eloszlású v. v. $(0,1)$–en, akkor $F^{-1}(U)$ eloszlásfüggvénye F, ahol F^{-1} jelöli F inverzét.*

Bizonyítás. A feltételek alapján F inverze létezik és $F^{-1} : (0,1) \to \mathbb{R}$ szigorúan monoton növekvő és folytonos.
(i): Minden $u \in (0,1)$–re

$$\{x \in \mathbb{R} : F(x) < u\} = \{x \in \mathbb{R} : x < F^{-1}(u)\}.$$

Ezért $F(X)$ eloszlásfüggvénye:

$$P(F(X) < u)$$

$$= \begin{cases} 0 & \text{ha } u \leq 0, \\ P(X < F^{-1}(u)) = F(F^{-1}(u)) = u & \text{ha } u \in (0,1), \\ 1 & \text{ha } u \geq 1. \end{cases}$$

Így $F(X)$ a $(0,1)$–en egyenletes eloszlású v. v..

(ii): Hasonlóan kapjuk, hogy minden $x \in \mathbb{R}$–re

$$\{u \in \mathbb{R} : F^{-1}(u) < x\} = \{u \in \mathbb{R} : u < F(x)\}.$$

Így $F^{-1}(U)$ eloszlásfüggvénye:

$$P(F^{-1}(U) < x) = P(U < F(x)) = F(x), \quad x \in \mathbb{R},$$

hiszen $F(x) \in (0,1)$, $x \in \mathbb{R}$. $\square$

Az előző tétel általánosítható.

Kvantilis–transzformáció

Egy F eloszlásfüggvény kvantilis–transzformáltja (általánosított inverze) alatt a $Q : (0,1) \to \mathbb{R}$,

$$Q(y) := \sup\{x \in \mathbb{R} : F(x) < y\}, \qquad y \in (0,1),$$

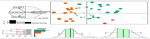
függvényt értjük.

Tétel

Ha U egyenletes eloszlású v. v. $(0,1)$–en, akkor $Q(U)$ eloszlásfüggvénye F.

Feladat

Mutassuk meg, hogy egy szigorúan monoton növekvő, folytonos eloszlásfüggvény kvantilis–transzformáltja éppen a függvény inverze.



Példa (Cauchy eloszlás generálása)

Generáljunk (μ, σ)–paraméterű Cauchy eloszlású véletlen számsorozatot, ahol $\mu \in \mathbb{R}$ és $\sigma > 0$.

A (μ, σ)–paraméterű Cauchy eloszlás eloszlásfüggvénye $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\mu}{\sigma}\right)$, $x \in \mathbb{R}$.

Ekkor F szigorúan monoton növekvő és inverze

$$Q(y) = F^{-1}(y) = \sigma \tan\left(\pi\left(y - \frac{1}{2}\right)\right) + \mu, \quad y \in (0,1).$$

Így, ha $U \sim \mathrm{Uni}(0,1)$, akkor

$$\sigma \tan\left(\pi\left(U - \frac{1}{2}\right)\right) + \mu$$

(μ, σ)–paraméterű Cauchy eloszlású v. v..

Példa (Exponenciális eloszlás generálása)

Legyen X λ–paraméterű exponenciális eloszlású v. v., ahol $\lambda > 0$. Ekkor X eloszlásfüggvénye:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{ha } x \geq 0, \\ 0 & \text{ha } x < 0. \end{cases}$$

Továbbá, $F|_{(0,\infty)}$ inverze

$$Q(y) = \left(F|_{(0,\infty)}\right)^{-1}(y) = -\frac{1}{\lambda}\ln(1-y), \quad y \in (0,1).$$

Az inverz módszer szerint (általánosított változat)

$$Q(U) = -\frac{1}{\lambda}\ln(1-U) \sim \mathrm{Exp}(\lambda),$$

ahol $U \sim \mathrm{Uni}(0,1)$. Egy további egyszerűsítés: mivel $1 - U \sim \mathrm{Uni}(0,1)$, így $-\ln(U)/\lambda \sim \mathrm{Exp}(\lambda)$.

A dominó módszer

Az inverz módszert alkalmazhatjuk diszkrét eloszlások generálására is. Ehhez az inverz függvény fogalmának kiterjesztésére van szükségünk.
Az X diszkrét v. v. lehetséges értékei legyenek $x_1 < \ldots < x_k < \ldots$, jelölje $p_1, \ldots, p_k, \ldots$ ezek valószínűségeit. Ekkor X eloszlásfüggvénye:

$$F(x) = \sum_{\ell=1}^{i} p_\ell \quad \text{ha} \quad x_i < x \leq x_{i+1}, \; i \in \mathbb{Z}_+,$$

ahol $x_0 := -\infty$, az üres szumma pedig definíció szerint 0.
Definiáljuk az alábbi $Q : (0,1) \to \mathbb{R}$ függvényt:

$$Q(u) := x_i, \quad \text{ha} \quad \sum_{\ell=1}^{i-1} p_\ell < u \leq \sum_{\ell=1}^{i} p_\ell. \tag{6}$$

Ekkor $Q(U)$ eloszlása, ahol $U \sim \mathrm{Uni}(0,1)$, éppen X eloszlásával egyezik meg.

Bizonyítás. A Q függvény definíciója alapján

$$P(Q(U) = x_i) = P\left(\sum_{\ell=1}^{i-1} p_\ell < U \leq \sum_{\ell=1}^{i} p_\ell\right) = \sum_{\ell=1}^{i} p_\ell - \sum_{\ell=1}^{i-1} p_\ell = p_i$$

minden $i \in \mathbb{N}$ esetén. $\square$

A módszer elnevezése az alábbiból ered. Képzeljünk el sorban olyan p_i magasságú dominókat, melyekre az x_i számok vannak felírva, $i \in \mathbb{N}$. Legyen U egyenletes eloszlású v. v. a $(0, 1)$–en. Döntsük fel sorban egymás után a dominókat addig, míg a feldöntöttek együttes magassága kisebb, mint az U érték. Válasszuk ezután a legelső nem feldöntött dominóra írt számot!

Feladat

Írjunk programot nevezetes diszkrét eloszlások (Bernoulli, binomiális, Poisson) generálására!

Feladat (Dominó módszer és kvantilis–transzformált)

Mutassuk meg, hogy egy diszkrét eloszlású v. v. eloszlásfüggvényének a kvantilis–transzformáltja éppen a (6) alatt definiált Q függvény.

Normális eloszlás generálása

Mivel Φ–re nincs zárt alak (nem állítható elő elemi függ–vényekből), ezért az inverz módszer nem alkalmazható.

Egy durva közelítés az IBM módszere. Generáljunk egy $(0,1)$–en egyenletes eloszlású számsorozatot, majd bontsuk 12 hosszú blokkokra. Legyenek a blokkok:

$$u_{1+12n}, u_{2+12n}, \ldots, u_{12+12n}, \quad n = 0, 1, \ldots.$$

Ekkor a különböző blokkok segítségével előállított

$$u_{1+12n} + u_{2+12n} + \cdots + u_{12+12n} - 6, \quad n = 0, 1, \ldots$$

számok közelítőleg standard normális eloszlásúak.

Indoklás: használjuk a központi határeloszlás tételt. Eszerint, ha U_n, $n \in \mathbb{N}$ független $\mathrm{Uni}(0,1)$–eloszlásúak és $S_n := U_1 + \ldots + U_n$, $n \in \mathbb{N}$, akkor

$$\frac{S_n - E(S_n)}{D(S_n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

ahol $\xrightarrow{\mathcal{L}}$ az eloszlásbeli konvergenciát jelöli.
Speciálisan, ha $n = 12$, akkor $E(S_{12}) = 6$ és $D(S_{12}) = 1$.
Így minden $n = 0, 1, \ldots$ esetén

$$U_{1+12n} + U_{2+12n} + \cdots + U_{12+12n} - 6 \stackrel{\mathcal{L}}{=} \frac{S_{12} - E(S_{12})}{D(S_{12})},$$

ahol $\stackrel{\mathcal{L}}{=}$ az eloszlásbeli egyenlőséget jelöli. Továbbá, $(S_{12} - E(S_{12}))/D(S_{12})$ eloszlása a központi határeloszlás tétel alapján közelítőleg $\mathcal{N}(0,1)$.

Megjegyzés

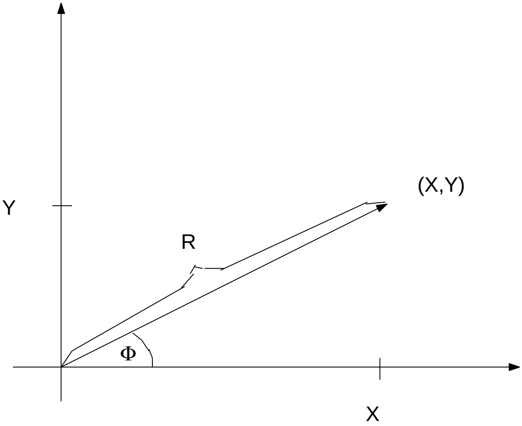
- A fentiekben az $(S_{12} - E(S_{12}))/D(S_{12})$ eloszlás standard normális eloszlással való „jól” közelíthetősége azon múlik, hogy az

$$\frac{S_n - E(S_n)}{D(S_n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

eloszlásbeli konvergencia elég „gyors”. Gondoljunk ugyanis a Berry–Esseen–tételre, ahol is a konvergenciasebesség nagysága az abszolút ferdeségtől (harmadik abszolút centrált momentum osztva a szórás köbével) függ, ami $\mathrm{Uni}(0,1)$–eloszlás esetén relatíve kicsi.
- $U_1 + \cdots + U_n$ eloszlását Irwin–Hall–eloszlásnak is nevezik.

A Box–Muller módszer

A módszer azon az észrevételen alapszik, hogy ha X és Y független standard normális eloszlású v. v.-k, akkor az (X, Y) párt polárkoordináta rendszerbe átírva szintén nevezetes eloszlásokat kapunk. Polárkoordinátákra való áttérésnél $X = R\cos\Phi$ és $Y = R\sin\Phi$:



Polárkoordináták.

Tétel (sűrűségfüggvény transzformációja)

Legyen a $\xi : \Omega \to \mathbb{R}^n$ *v. v. (együttes) sűrűségfüggvénye* f_ξ. *Tegyük fel továbbá, hogy létezik egy olyan* $D \subset \mathbb{R}^n$ *nyílt halmaz, hogy* $P(\xi \in D) = 1$, *és legyen* $g : D \to \mathbb{R}^n$ *egy folytonosan differenciálható, kölcsönösen egyértelmű leképezés, melynek Jacobi–determinánsa seholsem nulla. Jelölje* $h : g(D) \to D$ *a* g *inverzét. Ekkor* $\eta := g(\xi)$ *is abszolút folytonos és sűrűségfüggvénye:*

$$f_\eta(y) = \begin{cases} f_\xi(h(y))|J_h(y)|, & \text{ha } y \in g(D), \\ 0 & \text{ha } y \notin g(D), \end{cases}$$

ahol $J_h(y)$ *a* h *Jacobi-determinánsa az* y *helyen, azaz* h y*–beli elsőrendű parciális deriváltjaiból álló mátrixának a determinánsa.*

Alkalmazzuk az előző tételt az alábbi választásokkal:
$n = 2,\ \ \xi = (X, Y),\ \ D = \mathbb{R}^2,\ \ \ g : D \to \mathbb{R}^2,$

$$g(x, y) := \begin{cases} \left(\sqrt{x^2+y^2}, \arctan\left(\frac{y}{x}\right)\right) & \text{ha } x > 0 \text{ és } y \geq 0, \\ \left(\sqrt{x^2+y^2}, \arctan\left(\frac{y}{x}\right) + 2\pi\right) & \text{ha } x > 0 \text{ és } y < 0, \\ \left(\sqrt{x^2+y^2}, \arctan\left(\frac{y}{x}\right) + \pi\right) & \text{ha } x < 0, \\ \left(\sqrt{x^2+y^2}, \frac{\pi}{2}\right) & \text{ha } x = 0 \text{ és } y > 0, \\ \left(\sqrt{x^2+y^2}, \frac{3\pi}{2}\right) & \text{ha } x = 0 \text{ és } y < 0. \end{cases}$$

Ekkor $g(D) = g(\mathbb{R}^2) = [0, \infty) \times [0, 2\pi)$.
Továbbá, g inverze, $h : g(D) \to D$,

$$h(R, \Phi) = (R\cos\Phi, R\sin\Phi), \qquad (R, \Phi) \in [0, \infty) \times [0, 2\pi).$$

Felhasználva, hogy a $\xi = (X, Y)$ v. v. (együttes) sűrű-ségfüggvénye:

$$f_\xi(x,y) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{y^2}{2}} = \frac{1}{2\pi} \mathrm{e}^{-\frac{x^2+y^2}{2}}, \qquad (x,y) \in \mathbb{R}^2,$$

a sűrűségfüggvény transzformációjára vonatkozó tétel alapján az (R, Φ) pár együttes sűrűségfüggvénye:

$$\begin{aligned} f_{(R,\Phi)}(r,\varphi) &= \frac{1}{2\pi} \mathrm{e}^{-\frac{(R\cos\Phi)^2+(R\sin\Phi)^2}{2}} \left| \det \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} \right| \\ &= \frac{1}{2\pi} r \mathrm{e}^{-r^2/2}, \quad \text{ha } \varphi \in [0, 2\pi) \text{ és } r \geq 0, \end{aligned}$$

egyébként pedig 0.

Ekkor $f_{(R,\Phi)}$ a Φ és az R v. v.-k sűrűségfüggvényeinek a szorzata:

$$f_\Phi(\varphi) = \begin{cases} \frac{1}{2\pi} & \text{ha } \varphi \in [0, 2\pi), \\ 0 & \text{ha } \varphi \notin [0, 2\pi), \end{cases}$$

és

$$f_R(r) = \begin{cases} r\mathrm{e}^{-r^2/2} & \text{ha } r > 0, \\ 0 & \text{ha } r \leq 0 \end{cases}$$

Ezért Φ és R függetlenek.
Továbbá, R^2 sűrűségfüggvénye:

$$f_{R^2}(x) = \begin{cases} f_R(\sqrt{x})\frac{1}{2\sqrt{x}} = \frac{1}{2}\mathrm{e}^{-\frac{x}{2}} & \text{ha } x > 0, \\ 0 & \text{ha } x \leq 0. \end{cases}$$

Így $R^2 \sim \mathrm{Exp}(\frac{1}{2})$.

Ezért, ha $U, V \sim \mathrm{Uni}[0,1)$ függetlenek, akkor, felhasználva, hogy $2\pi U \sim \mathrm{Uni}[0, 2\pi)$, illetve azt, hogy az inverz módszer alkalmazásaként beláttuk, hogy

$$-2 \ln V \sim \mathrm{Exp}(1/2) \sim R^2,$$

kapjuk, hogy

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} R\cos\Phi \\ R\sin\Phi \end{pmatrix} \stackrel{\mathcal{L}}{=} \begin{pmatrix} \sqrt{-2\ln V}\cos(2\pi U) \\ \sqrt{-2\ln V}\sin(2\pi U) \end{pmatrix},$$

ahol $\stackrel{\mathcal{L}}{=}$ az eloszlásbeli egyenlőséget jelöli.
Így

$$\sqrt{-2\ln V}\cos(2\pi U) \quad \text{és} \quad \sqrt{-2\ln V}\sin(2\pi U)$$

független standard normális eloszlásúak.

A Marsaglia módszer

A Box-Muller módszerrel való generálás során szükségünk van a képletekben szereplő π–re, erre azonban csak közelítő értékünk lehet.

A Marsaglia módszer során ettöl is megszabadulhatunk.

Legyenek $U, V \sim \mathrm{Uni}(-1, 1)$ függetlenek.

$\mathrm{Uni}(-1, 1)$ eloszlást az $\mathrm{Uni}(0, 1)$ eloszlásból az $(0, 1) \ni x \mapsto 2x - 1$ transzformációval kaphatunk.

Tartsuk meg ezután azokat az (U, V) párokat, amelyekre $R^2 := U^2 + V^2 < 1$.

Ekkor az origó középpontú, K–val jelölt nyílt egység-körlapon egyenletes eloszlást kapunk, melynek sűrűség-függvénye $\frac{1}{\pi}$ ezen a körlapon, egyébként pedig 0.

Valóban, az (U,V) párnak az $\{U^2+V^2<1\}=\{(U,V)\in K\}$ feltételre vonatkozó feltételes eloszlásfüggvénye:

$$F_{(U,V)|\{U^2+V^2<1\}}(u,v)=\frac{P(U<u,V<v,U^2+V^2<1)}{P(U^2+V^2<1)}$$
$$=\begin{cases}\frac{4}{\pi}F_{(U,V)}(u,v) & \text{ha } (u,v)\in K,\\ 0 & \text{ha } (u,v)\notin K,\end{cases}$$

így az (U,V) párnak az $\{U^2+V^2<1\}$ feltételre vonatkozó feltételes sűrűségfüggvénye:

$$f_{(U,V)|\{U^2+V^2<1\}}(u,v)=\begin{cases}\frac{4}{\pi}\cdot\frac{1}{4}=\frac{1}{\pi} & \text{ha } (u,v)\in K,\\ 0 & \text{ha } (u,v)\notin K.\end{cases}$$

Legyenek

$$X := \sqrt{-2R^{-2}\ln R^2}\,U, \quad Y := \sqrt{-2R^{-2}\ln R^2}\,V.$$

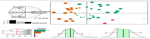
Alkalmazzuk a sűrűségfüggvény transzformációjára vonatkozó tételt az alábbi választásokkal: $n = 2$, $\xi = (U, V)$, $D = K \setminus \{(0,0)\}$, $g : D \to \mathbb{R}^2$,

$$g(u,v) := \begin{pmatrix} \sqrt{-2(u^2+v^2)^{-1}\ln((u^2+v^2))}\,u \\ \sqrt{-2(u^2+v^2)^{-1}\ln((u^2+v^2))}\,v \end{pmatrix}, \qquad (u,v) \in D.$$

Ekkor $P(\xi \in D) = 1$, $g(D) = g(K) = \mathbb{R}^2 \setminus \{(0,0)\}$, továbbá ellenőrizhető, hogy $h : g(D) \to D$,

$$h(x,y) = \begin{pmatrix} h_1(x,y) \\ h_2(x,y) \end{pmatrix} = \frac{e^{-\frac{x^2+y^2}{4}}}{\sqrt{x^2+y^2}} \begin{pmatrix} x \\ y \end{pmatrix},$$

ahol $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$.



Szintén ellenőrizhető, hogy minden $(x, y) \in \mathbb{R}^2 \setminus \{(0,0)\}$ esetén

$$\begin{bmatrix} \frac{\partial h_1}{\partial x}(x,y) & \frac{\partial h_1}{\partial y}(x,y) \\ \frac{\partial h_2}{\partial x}(x,y) & \frac{\partial h_2}{\partial y}(x,y) \end{bmatrix} = \frac{e^{-\frac{x^2+y^2}{4}}}{\sqrt{x^2+y^2}} \begin{bmatrix} -\frac{x^2}{2} + \frac{y^2}{x^2+y^2} & -\frac{xy}{2} - \frac{xy}{x^2+y^2} \\ -\frac{xy}{2} - \frac{xy}{x^2+y^2} & -\frac{y^2}{2} + \frac{x^2}{x^2+y^2} \end{bmatrix}.$$

Így h ((x, y) pontbeli) Jacobi determinánsának az abszolút értéke $\frac{1}{2}e^{-\frac{x^2+y^2}{2}}$, és ezért

$$f_{(X,Y)}(x,y) = f_{(U,V)}(h(x,y))\frac{1}{2}e^{-\frac{x^2+y^2}{2}} = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$$

minden $(x, y) \in \mathbb{R}^2 \setminus \{(0,0)\}$ esetén. Így X és Y függetlenek és $\mathcal{N}(0,1)$–eloszlásúak.

A Neumann–féle elfogadás–elutasítás módszer

Legyen $f = \kappa \widetilde{f}$ egy bonyolult sűrűségfüggvény, ahol $\widetilde{f}$–ot viszonylag egyszerűen tudjuk számolni, κ pedig egy nem feltétlenül explicit módon ismert normalizáló konstans. Nyilván, $\kappa^{-1} = \int_{-\infty}^{\infty} \widetilde{f}(x)\mathrm{d}x$.
Legyen g egy olyan ún. javasló eloszlás sűrűségfüggvénye, amelyet könnyen tudunk generálni. Feltételezzük továbbá, hogy létezik olyan $c > 0$, hogy $\widetilde{f} \leq cg$ (majorálási feltétel).

Minden lépésben két véletlen számot generálunk:

- egy x–et a g alapján,
- egy u–t, amely egyenletes a $[0, cg(x)]$ intervallumon (itt x az előbb generált véletlen szám).

Végül az így generált minta elemeit (pontosabban az x koordinátákat) elfogadjuk, ha az alábbi feltétel teljesül: $u \leq \widetilde{f}(x)$, ellenkező esetben pedig elutasítjuk.

Ezért nevezik a módszert gyakran elfogadás–elutasítás módszernek. Az így kapott véletlen számsorozat elosz–lása pontosan az f sűrűségfüggvényt fogja követni, amelyet az alábbi módon lehet belátni.

Legyen X egy g sűrűségfüggvényű v. v., illetve legyen U egy olyan v. v., melynek X–re vonatkozó feltételes eloszlása $\mathrm{Uni}[0, cg(X)]$, azaz U–nak az X–re vonat-kozó feltételes sűrűségfüggvénye:

$$f_{U|X}(u \mid x) = \begin{cases} \frac{1}{cg(x)} & \text{ha } g(x) \neq 0 \text{ és } u \in [0, g(x)], \\ 0 & \text{egyébként.} \end{cases}$$

Ekkor

$$f_{(X,U)}(x, u) = f_{U|X}(u|x) f_X(x) = \frac{1}{cg(x)} g(x) = \frac{1}{c},$$

ha $g(x) \neq 0$ és $u \in [0, g(x)]$, egyébként pedig $f_{(X,U)}(x, u) = 0$. Így az (X, U) pár egyenletes eloszlású a cg függvény és az $y = 0$ egyenes által határolt síkrészen, ugyanis $\int_{-\infty}^{\infty} cg(x)\, dx = c$.

Jelöljük (X^*, U^*)–gal az elfogadás után kapott véletlen mennyiséget. Ez egyenletes eloszlású lesz az $\widetilde{f}$ függvény és az $y = 0$ egyenes által határolt síkrészen, ugyanis (X, U)–nak az $\{U \leq \widetilde{f}(X)\}$ feltételre vonatkozó feltételes eloszlása egyenletes eloszlás az előbb említett síkrészen. Mivel $\int_{-\infty}^{\infty} \widetilde{f}(x)\mathrm{d}x = \kappa^{-1}$, kapjuk, hogy

$$f_{(X^*,U^*)}(x,u) = \begin{cases} \kappa & \text{ha } 0 \leq u \leq \widetilde{f}(x), \\ 0 & \text{egyébként.} \end{cases}$$

Így

$$f_{X^*}(x) = \int_{-\infty}^{\infty} f_{(X^*,U^*)}(x,u)\mathrm{d}u = \int_0^{\widetilde{f}(x)} \kappa \mathrm{d}u$$
$$= \kappa\widetilde{f}(x) = f(x), \quad x \in \mathbb{R}.$$

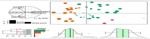
A c konstanst úgy érdemes megválasztani, hogy a (következőkben belátandó)

$$P(\text{elfogadás}) = \frac{1}{c\kappa}$$

valószínűség minél nagyobb legyen. Ez c minimalizálását jelenti az $\widetilde{f} \leq cg$ feltétel mellett.

Felhasználva, hogy $\widetilde{f} \leq cg$ alapján $g(x) = 0$ esetén $\widetilde{f}(x) = 0$ is teljesül, kapjuk, hogy

$$\begin{aligned} P(\text{elfogadás}) &= P(U \leq \widetilde{f}(X)) \\ &= \int_{-\infty}^{\infty} P(U \leq \widetilde{f}(x) \mid X = x) f_X(x)\,\mathrm{d}x \\ &= \int_{\{x \in \mathbb{R} : g(x) > 0\}} \frac{\widetilde{f}(x)}{cg(x)} g(x)\,\mathrm{d}x = \frac{1}{c} \int_{\{x \in \mathbb{R} : g(x) > 0\}} \widetilde{f}(x)\,\mathrm{d}x \\ &= \frac{1}{c} \int_{-\infty}^{\infty} \widetilde{f}(x)\,\mathrm{d}x = \frac{1}{c\kappa}. \end{aligned}$$



A módszer egyaránt alkalmazható diszkrét és folytonos eloszlások generálására.

Feladat

Generáljunk mintát $\mathrm{Ga}(p,\lambda)$ eloszlásra, ahol $p > 0$ és $\lambda > 0$.

1. Megoldás. (Iglói Endre nyomán)
Elég $\lambda = 1$ paraméterűt generálni, ugyanis ha $\xi \sim \mathrm{Ga}(p,1)$, akkor $\frac{1}{\lambda}\xi \sim \mathrm{Ga}(p,\lambda)$.
Másrészt, elég $0 < p < 1$ paraméterűt generálni, ugyanis felhasználva, hogy $x = [x] + \{x\}$, $x > 0$, kapjuk, hogy ha $\xi_1, \dots, \xi_{[p]} \sim \mathrm{Ga}(1,1) \sim \mathrm{Exp}(1)$, $\eta \sim \mathrm{Ga}(\{p\},1)$, és $\xi_1, \dots, \xi_{[p]}, \eta$ függetlenek, akkor a gamma–eloszlásra vonatkozó addíciós tétel miatt
$\xi_1 + \cdots + \xi_{[p]} + \eta \sim \mathrm{Ga}(p,1)$.
A továbbiakban feltételezzük tehát, hogy $0 < p < 1$.

A $\mathrm{Ga}(p, 1)$ eloszlás sűrűségfüggvénye:

$$f(x) = \begin{cases} \frac{1}{\Gamma(p)} x^{p-1} \mathrm{e}^{-x} & \text{ha } x > 0, \\ 0 & \text{ha } x \leq 0. \end{cases}$$

Legyen $\kappa := 1/\Gamma(p)$,

$$\widetilde{f}(x) := \begin{cases} x^{p-1} \mathrm{e}^{-x} & \text{ha } x > 0, \\ 0 & \text{ha } x \leq 0, \end{cases}$$

és

$$\widetilde{g}(x) := \begin{cases} x^{p-1} & \text{ha } 0 < x < 1, \\ \mathrm{e}^{-x} & \text{ha } x \geq 1, \\ 0 & \text{ha } x \leq 0. \end{cases}$$

Ekkor $\widetilde{f}(x) \leq \widetilde{g}(x)$, $x \in \mathbb{R}$, és

$$\int_{-\infty}^{\infty} \widetilde{g}(x)\,\mathrm{d}x = \int_0^1 x^{p-1}\,\mathrm{d}x + \int_1^{\infty} \mathrm{e}^{-x}\,\mathrm{d}x = \frac{1}{p} + \frac{1}{\mathrm{e}}.$$

A $\vartheta := \frac{1}{1/p+1/e}$ jelölést használva legyen

$$g(x) := \vartheta\widetilde{g}(x) = \begin{cases} \vartheta x^{p-1} & \text{ha } 0 < x < 1, \\ \vartheta e^{-x} & \text{ha } x \geq 1, \\ 0 & \text{ha } x \leq 0. \end{cases}$$

Ekkor g sűrűségfüggvény és

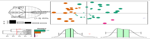
$$\widetilde{f}(x) \leq \frac{1}{\vartheta} g(x), \qquad x \in \mathbb{R},$$

így $c := 1/\vartheta$ választással teljesül az $\widetilde{f} \leq cg$ feltétel.

Már csak az a kérdés, hogy tudunk–e g sűrűség-függvényű eloszlást generálni?

Megmutatjuk, hogy az inverz módszer segítségével ez lehetséges.

Ehhez meg kell határoznunk a g sűrűségfüggvény G eloszlásfüggvényének az inverzét.



G meghatározása:

$$G(x) = \int_{-\infty}^{x} g(y)\,\mathrm{d}y, \qquad x \in \mathbb{R}.$$

Ha $x \leq 0$, akkor $G(x) = 0$.

Ha $0 < x < 1$, akkor

$$G(x) = \vartheta \int_0^x y^{p-1}\,\mathrm{d}y = \frac{\vartheta}{p} x^p.$$

Ha $x \geq 1$, akkor

$$G(x) = \vartheta \int_0^1 y^{p-1}\,\mathrm{d}y + \vartheta \int_1^x \mathrm{e}^{-y}\,\mathrm{d}y = \frac{\vartheta}{p} + \vartheta(\mathrm{e}^{-1} - \mathrm{e}^{-x}).$$

Összefoglalva,

$$G(x) = \begin{cases} \frac{\vartheta}{p} x^p & \text{ha } 0 < x < 1, \\ \frac{\vartheta}{p} + \vartheta(\mathrm{e}^{-1} - \mathrm{e}^{-x}) & \text{ha } x \geq 1, \\ 0 & \text{ha } x \leq 0. \end{cases}$$

$G|_{(0,\infty)}$ **inverzének a meghatározása:**

Ha $0 < y < \vartheta/p$, akkor

$$G(x) = y \quad\Longleftrightarrow\quad \frac{\vartheta}{p} x^p = y \quad\Longleftrightarrow\quad x = \left(\frac{p}{\vartheta} y\right)^{1/p},$$

így $G^{-1}(y) = \left(\frac{p}{\vartheta} y\right)^{1/p}$, ha $0 < y < \vartheta/p$.

Ha $y \geq \vartheta/p$, akkor

$$\begin{aligned} G(x) = y \quad&\Longleftrightarrow\quad \frac{\vartheta}{p} + \vartheta(\mathrm{e}^{-1} - \mathrm{e}^{-x}) = y \\ &\Longleftrightarrow\quad x = -\ln\left(\mathrm{e}^{-1} + p^{-1} - \frac{y}{\vartheta}\right), \end{aligned}$$

így $G^{-1}(y) = -\ln\left(\mathrm{e}^{-1} + p^{-1} - \frac{y}{\vartheta}\right)$, ha $y \geq \vartheta/p$.

Összefoglalva,

$$(G|_{(0,\infty)})^{-1}(y) = \begin{cases} \left(\frac{p}{\vartheta}y\right)^{1/p} & \text{ha } 0 < y < \vartheta/p, \\ -\ln\left(\mathrm{e}^{-1} + p^{-1} - \frac{y}{\vartheta}\right) & \text{ha } y \geq \vartheta/p. \end{cases}$$

Az inverz módszer szerint (általánosított változat) a $Q(U) = (G|_{(0,\infty)})^{-1}(U)$ valószínűségi változó sűrűségfüggvénye g, ahol $U \sim \mathrm{Uni}(0,1)$.

Az elfogadás valószínűsége:

$$\frac{1}{c\kappa} = \frac{1}{\frac{1}{\vartheta}\frac{1}{\Gamma(p)}} = \vartheta\Gamma(p) = \frac{\Gamma(p)}{\frac{1}{p}+\frac{1}{\mathrm{e}}} = \frac{\mathrm{e}p\Gamma(p)}{p+\mathrm{e}},$$

ahol $p \in (0,1)$.

Látjuk, hogy $p \uparrow 1$, illetve $p \downarrow 0$ esetén az elfogadás valószínűsége $\mathrm{e}/(\mathrm{e}+1)$–hez, illetve 1–hez konvergál.

2. Megoldás. (csak a $p > 1$ esetre érvényes)

Az exponenciális eloszlást, mint javasló eloszlást használva generálunk mintát $\mathrm{Ga}(p, 1)$ eloszlásra (ahol $p > 1$), azaz

$$g(x) := \begin{cases} \mu e^{-\mu x}, & \text{ha } x > 0, \\ 0, & \text{ha } x \leq 0, \end{cases}$$

ahol feltételezzük, hogy $0 < \mu < 1$.

Mivel exponenciális eloszlást az inverz módszer alapján tudunk generálni (lásd korábban), az elvetés–elfogadás módszer alkalmazásához azt kell ellenőriznünk, hogy létezik olyan $c > 0$ konstans, hogy $\widetilde{f}(x) \leq cg(x)$, $x \in \mathbb{R}$, ahol $\widetilde{f}$ ugyanaz, mint korábban.

Ehhez elég azt megmutatni, hogy az $\widetilde{f}(x)/g(x)$, $x > 0$ függvény korlátos.

Mivel

$$\frac{\widetilde{f}(x)}{g(x)} = \frac{x^{p-1}\mathrm{e}^{-x}}{\mu\mathrm{e}^{-\mu x}} = \frac{1}{\mu}x^{p-1}\mathrm{e}^{(\mu-1)x}, \quad x > 0,$$

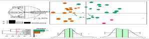
$$\lim_{x\downarrow 0}\frac{\widetilde{f}(x)}{g(x)} = 0, \qquad \lim_{x\to\infty}\frac{\widetilde{f}(x)}{g(x)} = 0,$$

felhasználva azt is, hogy kompakt halmazon folytonos függvény korlátos, kapjuk, hogy az $\widetilde{f}(x)/g(x)$, $x > 0$ függvény korlátos.

A maximális elfogadási arányt akkor kapjuk, ha c minimális. Ehhez az $\widetilde{f}(x)/g(x)$, $x > 0$ függvény maximumát kell meghatároznunk.

Mivel

$$\left(\frac{\widetilde{f}}{g}\right)'(x) = \frac{1}{\mu}x^{p-2}\mathrm{e}^{(\mu-1)x}(p-1+(\mu-1)x), \quad x > 0,$$



kapjuk, hogy az $\widetilde{f}(x)/g(x)$, $x > 0$ függvénynek az

$$x = \frac{p-1}{1-\mu}$$

helyen van a globális maximuma, és a maximum értéke

$$c(\mu) = \frac{1}{\mu}\left(\frac{p-1}{1-\mu}\right)^{p-1} e^{(\mu-1)\frac{p-1}{1-\mu}} = \frac{1}{\mu}\left(\frac{p-1}{1-\mu}\right)^{p-1} e^{1-p}.$$

Szokás ún. optimális elfogadási valószínűségről is beszélni abban az értelemben, hogy melyik $\mu \in (0,1)$ paraméterű exponenciális eloszlás (mint javasló eloszlás) esetén lesz az elfogadási valószínűség maximális.

Ez a $(0,1) \ni \mu \mapsto c(\mu)$ függvény minimumhelyének meghatározását jelenti.
Ehhez a (a fentiek alapján) a

$$(0,1) \ni \mu \mapsto \mu(1-\mu)^{p-1}$$

függvény maximumhelyét kell megkeresnünk.

Mivel

$$
\begin{aligned}
(\mu(1-\mu)^{p-1})' &= (1-\mu)^{p-1} - \mu(p-1)(1-\mu)^{p-2} \\
&= (1-\mu)^{p-2}(1-\mu-(p-1)\mu) \\
&= (1-\mu)^{p-2}(1-p\mu), \qquad 0 < \mu < 1,
\end{aligned}
$$

kapjuk, hogy a $\mu = 1/p$ helyen van a $\mu(1-\mu)^{p-1}$, $\mu \in (0,1)$ függvény globális maximuma.

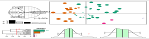
Így

$$
c(1/p) = \frac{1}{1/p}\left(\frac{p-1}{1-1/p}\right)^{p-1} \mathrm{e}^{1-p} = p^p \mathrm{e}^{1-p},
$$

és az optimális elfogadási valószínűség:

$$
\frac{1}{\kappa c(1/p)} = \frac{\Gamma(p)}{c(1/p)} = \Gamma(p) p^{-p} \mathrm{e}^{p-1}.
$$

Vegyük észre, hogy $1/p$ nem más, mint a $\mathrm{Ga}(p,1)$ eloszlás várható értéke.



3. Megoldás. (csak a $p > 1$ esetre érvényes)

A Cauchy eloszlást, mint javasló eloszlást használva generálunk mintát $\mathrm{Ga}(p, 1)$ eloszlásra (ahol $p > 1$), azaz

$$g(x) := \frac{1}{\sigma\pi(1 + (x - \mu)^2/\sigma^2)} = \frac{\sigma}{\pi(\sigma^2 + (x - \mu)^2)}, \quad x \in \mathbb{R},$$

ahol $\mu > 0$ és $\sigma > 0$.

Mivel Cauchy eloszlást az inverz módszer alapján tudunk generálni (lásd korábban), az elvetés–elfogadás módszer alkalmazásához azt kell ellenőriznünk, hogy létezik olyan $c > 0$ konstans, hogy $\widetilde{f}(x) \le cg(x)$, $x \in \mathbb{R}$, ahol $\widetilde{f}$ ugyanaz, mint korábban.

Ehhez elég azt megmutatni, hogy az $\widetilde{f}(x)/g(x)$, $x \geq 0$, függvény korlátos.

Mivel a szóban forgó függvény folytonos, $\widetilde{f}(0)/g(0) = 0$ és

$$\lim_{x\to\infty} \frac{\widetilde{f}(x)}{g(x)} = 0,$$

kapjuk, hogy korlátos is.

Megj. Megjegyezzük, hogy a $0 < p < 1$ esetben sem az exponenciális eloszlás, sem a Cauchy eloszlás, mint javasló eloszlás nem használható, mert nem létezik olyan $c > 0$, melyre $\widetilde{f}(x) \leq cg(x)$, $x \in \mathbb{R}$, hiszen $0 < p < 1$ esetén $\widetilde{f}$ nem korlátos.

$\square$

Irodalomjegyzék

1. Fazekas I.: (szerk.), Bevezetés a matematikai statisztikába. Kossuth Egyetemi Kiadó. Debrecen, 2003.
2. Knuth, D. E.: The Art of Computer Programming, 2nd ed., Volume 2. Addison–Wesley Publishing Company, 1981.
3. Lange, K.: Numerical Analysis for Statisticians. Springer. New York, 1998.
4. Weyl, H.: Uber die Gleichverteilung von Zahlen mod. Eins, Math. Ann. 77 (1916), 313–352.

Összefoglalás

1. Valószínűségi mező, valószínűségi változó, eloszlás, eloszlásfüggvény.
2. Nevezetes diszkrét eloszlások.
3. Nevezetes abszolút folytonos eloszlások.
4. Egyenletes generátorok.
5. Inverz módszer.
6. Dominó módszer.
7. Box–Muller módszer.
8. Marsaglia módszer.
9. Neumann–féle elfogadás–elutasítás módszer.