Sandsynlighedsbaserede metoder

– Et førstehåndsindtryk med pseudotilfældige tal

Daniel Kjær

For nogle uger siden pålagde jeg mig selv den opgave at aflevere to artikler til FAMØS om Monte Carlo-metoden. Af hensyn til bekvemmelighed forbindes disse artikler sekventielt, idet første artikel indeholder en introduktion til metoderne til frembringelsen af stokastiske variable (s.k. *pseudotilfældige tal*), og den anden artikel omhandler den egentlige Monte Carlo-metode.

De problemer, vi i det følgende vil betragte, er forholdsvist enkle, og vores fremstilling af emnet vil kun kræve et indledende kendskab til målteori.

Ved udfærdigelsen af den foreliggende artikel er der på ingen måde gjort et forsøg på at give en fuldstændig fremstilling af emnet.

Pseudotilfældige talgeneratorer

Vi skal nu træde et skridt tilbage og finde en kilde til produktionen af s.k. *pseudotilfældige tal* og konverteringen af disse tal til udfald af ligefordelte stokastiske variable. De fleste algoritmer til frembringelsen af pseudotilfældige tal er på formen

$$\omega_k = \Gamma(\omega_{k-1}), \qquad k \in \mathbb{N}.$$

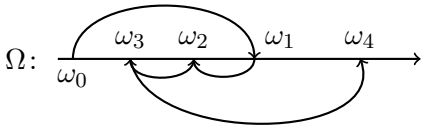
Elementet ω_0 kaldes et *seed* og siges at initialisere følgen af pseudotilfældige, tal og transformationen Γ kaldes en *pseudotilfældig talgenerator*. Det er klart, at det ikke er enhver afbildning, der kan spille rollen som generator.

Definition 1 Lad $(\Omega, \mathfrak{F})$ være et målbart rum og lad T være en målelig afbildning på Ω ind i Ω. Lad $n \in \mathbb{N}$ og definér den n-foldige

sammensætning af T med sig selv, betegnet ved T^n, ved

$$\begin{cases} T^0 = \mathrm{id}_\Omega, \\ T^n = T \circ T^{n-1}, \quad n \in \mathbb{N} \end{cases}$$

hvor id_Ω er den identiske afbildning på Ω. Ved den af ω frembragte omløbsbane under T for et $\omega \in \Omega$, betegnet ved T_ω, forstås følgen $(T^n(\omega))_{n\in\mathbb{Z}_+}$.



Figur 1 En omløbsbane under transformationen T.

Lad $(\Omega, \mathfrak{F})$ være et målbart rum og lad $\omega_0 \in \Omega$ være et element i udfaldsrummet. Lad T være en målelig afbildning på Ω ind i Ω og lad T_{ω_0} betegne den af ω_0 frembragte omløbsbane under T. Omløbsbanen under T kan visualiseres som i ovenstående skema. De idéelle egenskaber hørende til en god pseudotilfældig talgenerator til generelle formål er lette at blive enige om. Tilstedeværelsen af uafhængighed og ensartethed er de gældende kriterier for en generators tilstrækkelighed.

Vi vil gerne matematisk afspejle idéen om, at ligefordelingen bevares af generatoren under iterationer, der bærer et pseudotilfældigt tal over i et andet. Det leder os til studiet af målbevarende afbildninger, *ergodeteorien*.

Definition 2 Lad $(\Omega, \mathfrak{F}, P)$ være et sandsynlighedsfelt. En målelig afbildning, T, på Ω ind i Ω siges at være P-målbevarende, hvis $T(P) = P$.

Definition 3 Lad $(\Omega, \mathfrak{F}, P)$ være et sandsynlighedsfelt og lad T være en P-målbevarende afbildning på Ω ind i Ω. En mængde $A \in \mathfrak{F}$ siges at være T-invariant, hvis $T^{-1}(A) = A$. Ved den T-invariante σ-algebra, betegnet ved $\mathfrak{I}$, forstås samlingen $\{A \in \mathfrak{F} \colon T^{-1}(A) = A\}$ af T-invariante mængder.

Øvelse: Lad $(\Omega, \mathfrak{F}, P)$ være et sandsynlighedsfelt og lad T være en P-målbevarende afbildning på Ω ind i Ω. Bekræft, at den T-invariante σ-algebra $\mathfrak{I}$ faktisk er en σ-algebra.

Øvelse: Lad $(\Omega, \mathfrak{F}, P)$ være et sandsynlighedsfelt. Lad T være en P-målbevarende afbildning på Ω ind i Ω og lad $A \in \mathfrak{I}$ være en mængde i den T-invariante σ-algebra. Antag, at $A \notin \{\emptyset, \Omega\}$ og at $P(A) \in (0, 1)$. Vis, at $T(A) \subset A$ og forklar hvorfor det kan være problematisk at bruge T som generator.

I øvelsen er strukturen af den frembragte omløbsbane antaget så simpel, at den ikke lader sig fordele jævnt over hele udfaldsrummet.

Spørgsmål: Lad $(\Omega, \mathfrak{F}, P)$ være et sandsynlighedsfelt. Lad T være en målelig afbildning på Ω ind i Ω, og lad T_ω være den af ω frembragte omløbsbane under T. Hvilke yderligere betingelser skal T opfylde, for at følgende betingelse gælder?

1. Der findes en nulmængde Λ i $(\Omega, \mathfrak{F}, P)$, så T_ω i en vis forstand er en replika af Ω for alle $\omega \in \Lambda^c$.

En idé er, at man for en generel målbevarende afbildning T kræver, at enhver T-invariant mængde er triviel som f.eks. Ω og $\emptyset$.

Definition 4 (Ergodisk System) Lad $(\Omega, \mathfrak{F}, P)$ være et sandsynlighedsfelt. En målelig afbildning, T, på Ω ind i Ω siges at være ergodisk, hvis følgende betingelser er opfyldt:

1. T er P-målbevarende,
2. For enhver mængde A i den T-invariante σ-algebra $\mathfrak{I}$, så er $P(A) = 0$ eller 1.

Kvadruplet $(\Omega, \mathfrak{F}, P, T)$ kaldes et ergodisk system.

Definitionen af et ergodisk system giver, overraskende nok, løsningen på spørgsmålet. Følgende sætning er endda også ergodeteoriens ubestridte højdepunkt.

Sætning 5 (Birkhoffs ergodesætning) *Lad $(\Omega, \mathfrak{F}, P)$ være et sandsynlighedsfelt. Hvis T er en P-målbevarende afbildning på Ω ind i Ω, og X er en integrabel reel stokastisk variabel på Ω, da findes en nulmængde Λ i $(\Omega, \mathfrak{F}, P)$, således at*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X \circ T^k(\omega) = \mathbf{E}(X \mid \mathfrak{I})(\omega), \quad \textit{dersom } \omega \in \Lambda^c.$$

Hvis yderligere T er ergodisk, gælder at der findes en nulmængde Λ i $(\Omega, \mathfrak{F}, P)$, således at

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X \circ T^k(\omega) = \mathbf{E}(X), \quad \textit{dersom } \omega \in \Lambda^c.$$

Bevis. Se [1]. □

Definition 6 (Pseudotilfældig talgenerator) Lad os betragte sandsynlighedsfeltet $([0,1), \mathfrak{B}_{[0,1)}, m_L)$, hvor m_L er Lebesgue målet på Borel σ-algebraen $\mathfrak{B}_{[0,1)}$ af delmængder af $[0,1)$. Ved en pseudotilfældig talgenerator på $([0,1), \mathfrak{B}_{[0,1)}, m_L)$ mener vi en målelig afbildning Γ på $[0,1)$ ind i $[0,1)$, således at følgende betingelse er opfyldt:

1. Kvadruplet $([0,1), \mathfrak{B}_{[0,1)}, m_L, \Gamma)$ er et ergodisk system.

At bruge Birkhoff ergodesætningen til estimation af integraler er *idéen* bag Monte Carlo-metoder, og fejlen kan estimeres ved hjælp af statistiske standardmetoder.

Korollar 7 *Lad Γ være en pseudotilfældig talgenerator på sandsynlighedsfeltet $([0,1), \mathfrak{B}_{[0,1)}, m_L)$. For enhver mængde $A \in \mathfrak{B}_{[0,1)}$ findes en nulmængde Λ i $([0,1), \mathfrak{B}_{[0,1)}, m_L)$, således at*

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbf{1}_A \circ T^k(\omega) = m_L(A), \quad \textit{dersom } \omega \in \Lambda^c.$$

Bevis. Følger af Birkhoff's ergodesætning. □

Det er i denne forstand, at omløbsbanerne for en ergodisk afbildning skulle forestille at replikere udfaldsrummet, og det er denne ideale egenskab, vi vil kræve fra en pseudotilfældig talgenerator: Omløbsbanerne for en pseudotilfældig talgenerator rammer A uendeligt ofte med asymptotisk relativ frekvens $m_L(A)$.

Den letteste og mest populære algoritme til frembringelsen af pseudotilfældige tal, den s.k. *lineære kongruens-generator*, blev udviklet af D. H. Lehmer i 1949, og algoritmen for den lineære kongruens-generator er modelleret efter en ergodisk transformation.

Definition 8 Lad $a, c, m \in \mathbb{Z}_+$ være ikke-negative heltal, således at $0 < a < m$ og $0 \leqslant c < m$. Den lineære kongruens-generator er en afbildning L på $\{0, \ldots, m-1\}$ ind i $\{0, \ldots, m-1\}$ givet ved

$$L(\gamma) = a\gamma + c \mod m, \qquad \gamma \in \{0, \ldots, m-1\}.$$

Den egentlige produktion af en følge $(\omega_k)_{k\in\mathbb{Z}_+}$ af pseudotilfældige tal i $[0, 1)$ under den lineære kongruens-generator fås ved at lade $\gamma_0 \in \{0, \ldots, m-1\}$ være givet og sætte

$$\begin{aligned} \gamma_k &= L^k(\gamma_0), \\ \omega_k &= \gamma_k/m, \qquad k \in \mathbb{Z}_+. \end{aligned}$$

Bemærkning 9 Lad Γ være en afbildning på $\{\frac{0}{m}, \ldots, \frac{m-1}{m}\}$ ind i $\{\frac{0}{m}, \ldots, \frac{m-1}{m}\}$ givet ved

$$\Gamma(\omega) = L(m\omega)/m, \qquad \omega \in \{\tfrac{0}{m}, \ldots, \tfrac{m-1}{m}\}.$$

Fra et matematisk synspunkt er følgen $(\omega_k)_{k\in\mathbb{Z}_+}$ af pseudotilfældige tal frembragt af den lineære kongruens-generator ækvivalent med den af $\omega_0 \equiv \gamma_0/m$ frembragte omløbsbane under Γ, altså

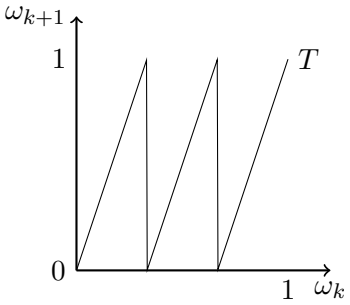
$$\Gamma_{\omega_0} = (\omega_k)_{k\in\mathbb{Z}_+}. \tag{1}$$

Det ses endvidere ved udregning, at

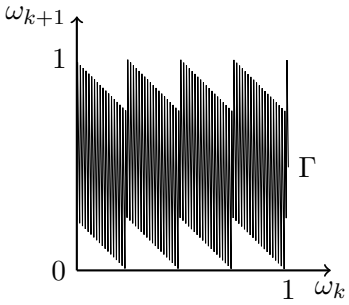
$$\Gamma(\omega) = a\omega + \frac{c}{m} \mod 1, \qquad \omega \in \{\tfrac{0}{m}, \ldots, \tfrac{m-1}{m}\}.$$

Lad os tage en time-out til en sludder for en sladder: Den lineære kongruens-generator er altså ikke en pseudotilfældig generator i den ideale forstand, som vi har lagt op til indtil videre, for eksempel er Γ ikke en afbildning på hele $[0, 1)$, men med en vederstyggelig brug af notation har vi at $\{\frac{0}{m}, \ldots, \frac{m-1}{m}\} \to [0, 1)$ dersom

$m \to \infty$. I øvrigt viser det sig, at L er periodisk i den forstand, at der findes et $n \in \mathbb{N}$, således at den n-foldige sammensætning af L med sig selv er den identiske afbildning. At L er periodisk, er selvfølgelig acceptabelt, *kun hvis* perioden er meget stor.



Figur 2 Den ergodiske transformation $T(\omega = a\omega \mod 1)$ for et $a = 3$.



Figur 3 Den lineære kongruensgenerator for parametre $a = 16838$ og $c/m = 0.236$.

Den lineære kongruens-generator er altså modelleret efter følgende ergodiske transformation. Se figuren for sammenligning.

Sætning 10 *Betragt sandsynlighedsfeltet* $([0,1), \mathfrak{B}_{[0,1)}, m_L)$. *Lad* $a \in \mathbb{N}, a > 1$ *være et naturligt tal. Den målelige afbildning* T *på* $[0,1)$ *ind i* $[0,1)$ *givet ved multiplikation med* a modulo 1, *altså*

$$T(\omega) = a\omega \mod 1, \qquad \omega \in [0,1)$$

er ergodisk.

Note 11 (Bevisskitse) Vi skal bevise, at $([0,1), \mathfrak{B}_{[0,1)}, m_L, T)$ er et ergodisk system. Vi opdeler vores bevis i to trin: Først beviser

vi, at T er m_L-målbevarende. Dernæst beviser vi, at enhver T-invariant mængde $A \in \mathfrak{I}$ er triviel, i den forstand at A har m_L-mål 0 eller 1.

Bevis. Genkald, at $[\lambda] = \max\{k \in \mathbb{Z}_+ \colon k \leqslant \lambda\}$ for $\lambda \in \mathbb{R}_+$ og omskriv T på formen

$$T(\omega) = a\omega - [a\omega], \qquad \omega \in [0,1). \tag{2}$$

Det er klart, at T er målelig, og i øvrigt har vi fra (2), at

$$T(\omega) = a\omega - (k-1), \qquad \omega \in \left[\tfrac{k-1}{a}, \tfrac{k}{a}\right), k = 1, \ldots, a. \tag{3}$$

For at bevise at T er m_L-målbevarende, lad $\alpha, \beta \in [0,1)$ med $\alpha < \beta$ være givet. Bemærk først, at fra (3) har vi at

$$T^{-1}([\alpha,\beta)) = \bigcup_{k=1}^{a} [\tfrac{\alpha+k-1}{a}, \tfrac{\beta+k-1}{a}).$$

Herfor finder vi ved udregning, at

$$\begin{aligned} T(m_L)([\alpha,\beta)) &= m_L(T^{-1}([\alpha,\beta))) \\ &= \sum_{k=1}^{a} m_L([\tfrac{\alpha+k-1}{a}, \tfrac{\beta+k-1}{a})) \\ &= \beta - \alpha \\ &= m_L([\alpha,\beta)). \end{aligned}$$

Bemærk dernæst at $\{[\alpha,\beta) \subset [0,1) \colon \alpha, \beta \in [0,1), \alpha < \beta\}$ er et fællesmængdestabilt frembringersystem for Borel σ-algebraen $\mathfrak{B}_{[0,1)}$ af delmængder af $[0,1)$. Dette beviser at $T(m_L) = m_L$, altså at T er m_L-målbevarende.

Det genstår at bevise den anden halvdel. Lad $A \in \mathfrak{I}$ være en T-invariant mængde, $A = T^{-1}(A)$. Vi skal vise, at $m_L(A) = 0$ eller 1. I tilfældet $m_L(A) = 1$ er der intet at vise. Lad os antage at $m_L(A) < 1$ og vise at $m_L(A) = 0$ ved at vise, at $m_L(A) < \varepsilon$ for alle $\varepsilon > 0$.

Lad $I_{n,m} = [(m-1)a^{-n}, ma^{-n})$ for $n \in \mathbb{N}$ og $m = 1, \ldots, a^n$. Observér først, at

$$T^k(I_{n,m}) = [\frac{m-1 \mod a^{n-k}}{a^{n-k}}, \frac{m-1 \mod a^{n-k}+1}{a^{n-k}}), \quad (4)$$

for $k \in \{0, \ldots, n\}$. Fra (4) har vi, at

$$m_L(T^k(I_{n,m})) = a^k m_L(I_{n,m}). \quad (5)$$

Bemærk dernæst, at

$$T^n(I_{n,m} \cap A^c) \subset T^n(I_{n,m}) \cap T^n(A^c) \subset T^n(I_{n,m}) \cap A^c = A^c, \quad (6)$$

hvor anden inklusion gælder, fordi også A^c er T-invariant og $T(T^{-1}(A^c)) \subset T^{-1}(A^c) = A^c$. Lighedstegnet følger af (4) idet $T^n(I_{n,m}) = [0,1)$. Givet $\varepsilon > 0$ findes et $n \in \mathbb{N}$ og et $m \in \{1, \ldots, a^n\}$, således at

$$\varepsilon > \frac{m_L(I_{n,m} \cap A)}{m_L(I_{n,m})}. \quad (7)$$

Dette er ækvivalent med

$$m_L(I_{n,m} \cap A^c) > (1-\varepsilon) m_L(I_{n,m}). \quad (8)$$

Ved udregning har vi altså, at

$$\begin{aligned} m_L(A^c) &\geqslant m_L(T^n(I_{n,m} \cap A^c)) && (9)\\ &= a^n m_L(I_{n,m} \cap A^c) && (10)\\ &\geqslant a^n(1-\varepsilon) m_L(I_{n,m}) && (11)\\ &= 1 - \varepsilon && (12) \end{aligned}$$

hvor (9) følger af (6), (10) følger af (5), og (11) følger af (8). Da $m_L(A^c) \geqslant 1 - \varepsilon$ konkluderer vi at $m_L(A) < \varepsilon$. Da dette gælder for alle $\varepsilon > 0$, har vi bevist at $m_L(A) = 0$. □

Programmel

En klar-til-brug implementering af den s.k. *Wichmann & Hill*-generator følger. Algoritmen er anstændig nok til personlig brug.

Listing 1 `R` kode

```
## Generate pseudo random numbers uniformly between 0 and 1
uniform <- local({
    # A sequence of initial values
    x = 5
    y = 11
    z = 17

    # Make x, y and z local static variables.
    f <- function(){
        x <<- 171 * (x %% 177) - 2 * (x %/% 177)
        y <<- 172 * (y %% 176) - 35 * (y %/% 176)
        z <<- 170 * (z %% 178) - 63 * (z %/% 178)

        # The part where we deal with negative x, y and z
        if(x < 0)
            x <<- x + 30269
        if(y < 0)
            y <<- y + 30307
        if(z < 0)
            z <<- z + 30323

        return((x / 30269. + y / 30307. + z / 30323.) %% 1)
    }
})

# Print 5 random numbers
for(i in 1:5){
    print(uniform())
}
```

Listing 2 C++ kode

```
//: MC_intro:Uniform.cpp
// Generate pseudo random numbers uniformly between 0 and 1
#include <iostream>
#include <math.h> // For using "fmod()"
using namespace std;

float uniform(){
    // A sequence of initial values
    static int x = 5;
    static int y = 11;
    static int z = 17;

    // Some integer arithmetic required
    x = 171 * (x % 177) - 2 * (x / 177);
    y = 172 * (y % 176) - 35 * (y / 176);
    z = 170 * (z % 178) - 63 * (z / 178);

    /* If both operands are nonnegative then the
    remainder is nonnegative; if not, the sign of
    the remainder is implementation-defined. */
    if(x < 0)
        x = x + 30269;
    if(y < 0)
        y = y + 30307;
    if(z < 0)
        z = z + 30323;

    return fmod(x / 30269. + y / 30307. + z / 30323., 1.);
}

int main(){
    // Print 5 random numbers
    for(int i = 0; i < 5; ++i){
        cout << uniform() << ", ";
    }
}///:~
```

I næste artikel diskuteres, hvorledes pseudotilfældige tal bekvemt og effektivt kan bruges til frembringe af stokastiske variable til brug i en Monte Carlo-model.

Litteratur

[1] Jacobsen, M., *Videregående Sandsynlighedsregning*, Institut for Matematiske Fag, Københavns Universitet, 3. udgave, 2003

[2] Wichmann, B. A. & Hill, I. D., *Algorithm AS 183: An Efficient and Portable Pseudo-Random Number Generator*, Applied Statistics, Vol 31, No. 2 (1982), 188–190