Universitatea din București
Facultatea de Matematică și Informatică

CURSURILE nr. 4 – 5
METODE MODERNE DE CALCUL ȘI SIMULARE

Validarea generatorilor

Metoda amestecării – compunerii

Lect. dr. Bianca Mogoș

Conținut

Validarea generatorilor/algoritmilor de simulare a unor variabile aleatoare

1. Histograma – o validare empirică a algoritmului de simulare
2. Test bazat pe momentele de selecție
3. Testul X^2

Metoda amestecării – compunerii

1. Definiții: amestecare discretă și amestecare continuă
2. Descrierea metodei
3. Algoritm general de amestecare – compunere
4. Fundamentarea matematică a metodei
5. Exemple: Simularea v.a. Laplace și Lomax

Validarea generatorilor/algoritmilor de simulare a unor variabile aleatoare (1)

- *Algoritm de simulare*: definirea unei v.a. X având o funcție de repartiție dată $F(x) = P(X \leq x), \forall x \in \mathbb{R}$ și observarea variabilei X
- *Validarea algoritmilor de simulare* înseamnă:
  - verificarea corectitudinii formale a algoritmilor: se arată că v.a. X construită în algoritm are funcția de repartiție $F(x)$
  - analiza valorilor de selecție asupra v.a. X returnate de algoritm: pe baza unei mulțimi de selecție $X_1, X_2, \ldots, X_n$, se verifică ipoteza statistică $H_0 : X \hookrightarrow F(x)$.

Validarea generatorilor (2): Histograma – o validare empirică a algoritmului de simulare (1)

- Se verifică intuitiv dacă *repartiția empirică (de selecție)* este asemănătoare cu cea *teoretică*
- *Histograma* asociată mulțimii de valori de selecție $x_1, \ldots, x_n$ asupra variabilei aleatoare X având funcția de repartiție $F(x)$ și densitatea $f(x)$:
  - se determină $m = \min\{x_1, x_2, \ldots, x_n\}$ și $M = \max\{x_1, x_2, \ldots, x_n\}$
  - se alege k numărul de intervale/dreptunghiuri ale histogramei
  - se împarte intervalul $[m, M]$ în k intervale egale $I_i = (a_{i-1}, a_i]\,, 2 \leq i \leq k, I_1 = [a_0, a_1]$, $a_0 = m, a_k = M$
  - se determină frecvențele relative $r_i = \dfrac{f_i}{n}$, unde f_i : numărul de valori de selecție ce cad în intervalul I_i, $1 \leq i \leq k$
  - se reprezintă grafic: se iau pe abscisă intervalele I_i și se construiesc dreptunghiurile având ca bază aceste intervale și ca înălțimi $h_i = r_i$.
- *Înălțimile h_i ale dreptunghiurilor se scalează a.î. $h_i \approx f(x), x \in I_i$. Definim $h_i = \dfrac{f_i}{nl} \approx f(x), l = a_i - a_{i-1}$.*

Validarea generatorilor (3): Histograma – o validare empirică a algoritmului de simulare (2)

- Din *Teorema lui Bernoulli – forma slabă* rezultă că pentru orice $\epsilon > 0$ avem

$$\lim_{n\to\infty} P\left(\left|\frac{f_i}{n} - p_i\right| \leq \epsilon\right) = 1, \quad (1)$$

unde $p_i = P(X \in I_i)$, $I_i = (a_{i-1}, a_i]$

- Numărul k de dreptunghiuri ale histogramei se consideră a.î. să se minimizeze *"media pătratelor erorilor (MSE)"* estimatorului $\hat{f}(x)$ (definit în orice punct x) al densității de repartiție $f(x)$, definită prin

$$MSE(\hat{f}(x)) = E\left[\left(\hat{f}(x) - f(x)\right)^2\right]. \quad (2)$$

Astfel, avem *Regula Sturges*

$$k = [1 + \log_2 n]. \quad (3)$$

Validarea generatorilor (4): Test bazat pe momentele de selecție

- Se determină *momentele teoretice* ale v.a. X:

$$\mu = E[X] \text{ și } \sigma^2 = Var(X) \tag{4}$$

- Pe baza mulțimii de valori de selecție $\{x_1, x_2, \ldots, x_n\}$ se calculează *momentele de selecție*:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ și } s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{X}^2 \tag{5}$$

- Ca o consecință a *Legii numerelor mari*, putem considera că generatorul este bun dacă pentru n suficient de mare ($n > 1000$)

$$\overline{X} \approx \mu \text{ și } s^2 \approx \sigma^2 \tag{6}$$

Validarea generatorilor (5): Testul X^2 (1)

- Considerăm *testul de concordanță* X^2 pentru verificarea ipotezei $H_0 : X \hookrightarrow F(x)$.
- Definim *variabila aleatoare* X^2:

$$X^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \tag{7}$$

unde $p_1 = F(a_1), p_i = F(a_i) - F(a_{i-1}), 2 \leq i \leq k-1, p_k = 1 - F(a_{k-1})$.

- *Observații*:
  - $f_i : \Omega \to \{0, 1, \dots, n\}$ este o v.a. $Bin\,(n, p_i)$
  - Mulțimea tuturor valorilor posibile ale lui X^2 se obține făcând ca $f_1, f_2, \dots, f_k$ să parcurgă toți întregii nenegativi a.î. $\sum_{i=1}^{k} f_i = n$
  - Pentru $n \to \infty$, X^2 este repartizată χ^2_{k-1} (hi pătrat cu $k-1$ grade de libertate).

Validarea generatorilor (6): Testul X^2 (2)

- Fie α *eroarea de tip I*: probabilitatea de a respinge ipoteza nulă H_0 când este adevărată. Valorile clasice pentru α sunt 0.01, 0.05, 0.1.
- Se determină α *– cuantila superioară*, notată $\chi^2_{k-1,\alpha}$ astfel încât

$$P\left(X^2 \leq \chi^2_{k-1,\alpha}\right) = 1 - \alpha \tag{8}$$

- Ipoteza *H_0 se acceptă* dacă în urma experimentului aleator s-a obținut evenimentul $\omega \in \Omega$ a.î.:

$$X^2(\omega) \leq \chi^2_{k-1,\alpha}, \tag{9}$$

  în caz contrar se respinge.

Metoda compunerii (1): Definiții amestecare discretă și amestecare continuă

- Spunem că funcția de repartiție $F(x)$ este o *amestecare (compunere sau mixtură) discretă* a mulțimii de funcții de repartiție $\{F_j(x)\}_{1\leq j\leq m}$ cu repartiția discretă J dacă

$$F(x) = \sum_{j=1}^{m} p_j F_j(x) \text{ și } J : \begin{pmatrix} 1 & 2 & \dots & m \\ p_1 & p_2 & \dots & p_m \end{pmatrix}, \sum_{j=1}^{m} p_j = 1 \quad (10)$$

- Spunem că funcția de repartiție $F(x)$ este o *amestecare continuă* a familiei de funcții de repartiție $\{G(x, Y)\}_{Y\in\mathbb{R}}$, cu funcția de repartiție continuă $H(y)$ a lui Y dacă este de forma

$$F(x) = \int_{\mathbb{R}} G(x, y) dH(y). \quad (11)$$

Metoda compunerii (2): Descrierea metodei

- Fie X o v.a. cu funcția de repartiție $F(x)$ definită în (10) și $X_j, 1 \leq j \leq m$ v.a. având funcțiile de repartiție $F_j(x)$. Atunci

$$X = X_j \text{ cu probabilitatea } p_j \tag{12}$$

- Fie X o v.a. cu funcția de repartiție $F(x)$ definită în (11) și $Z_Y, Y \in \mathbb{R}$ v.a. având funcțiile de repartiție $G(x, Y)$, unde Y are funcția de repartiție $H(y)$. Atunci

$$X = Z_y \text{ unde } y \text{ este generat cu funcția de repartiție } h(y) \tag{13}$$

- În definiția amestecării se pot considera în loc de funcții de repartiție, densități de repartiție și formulele devin

$$f(x) = \sum_{i=1}^{m} p_i f_i(x), \text{ respectiv } f(x) = \int_{\mathbb{R}} g(x, y) h(y) dy \tag{14}$$

Metoda compunerii (3): Algoritm pentru simularea unei variabile aleatoare bazat pe noțiunea de amestecare discretă

Intrare	Se cunosc algoritmi pentru simularea v.a. X_j având funcțiile de repartiție $F_j(x)$ Repartiția v.a. discrete J: $P(J=j)=p_j, \sum_{j=1}^{m} p_j = 1$
Pas 1	Se generează un indice j având repartiția J
Pas 2	Se generează x_j cu funcția de repartiție $F_j(x)$
Pas 3	Se definește $x = x_j$
Ieșire	Valoarea de selecție, x, a v.a. X având funcția de repartiție $F(x)$

Metoda compunerii (4): Algoritm pentru simularea unei variabile aleatoare bazat pe noțiunea de amestecare continuă

Intrare	Se cunosc algoritmi pentru simularea v.a Z_Y având funcțiile de repartiție în familia $\{G(x, Y)\}_{Y \in \mathbb{R}}$ Se cunoaște un algoritm pentru simularea v.a. Y cu funcția de repartiție $H(y)$
Pas 1	Se generează y cu funcția de repartiție $H(y)$
Pas 2	Se generează z_y cu funcția de repartiție $G(x, y)$
Pas 3	Se definește $x = z_y$
Ieșire	Valoarea de selecție, x, a v.a. X având funcția de repartiție $F(x)$

Metoda compunerii (5): Exemplu amestecare discretă

- Fie X variabila având repartiția Laplace a cărei densitate de repartiție este

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}, x \in \mathbb{R}, \lambda > 0 \tag{15}$$

- Observăm că

$$f(x) = p_1 f_1(x) + p_2 f_2(x), p_1 = p_2 = \frac{1}{2}$$

unde

$$f_1(x) = \begin{cases} 0 & \text{dacă } x \leq 0 \\ \lambda e^{-\lambda x} & \text{dacă } x > 0 \end{cases}, f_2(x) = \begin{cases} \lambda e^{\lambda x} & \text{dacă } x \leq 0 \\ 0 & \text{dacă } x > 0 \end{cases}$$

Metoda compunerii (6): Algoritmul pentru simularea v.a. X având repartiția Laplace

Intrare	V.a. $X_1 \sim Exp(\lambda)$ cu densitatea de repartiție $f_1(x)$ și V.a. $X_2 = -X_1$ cu densitatea de repartiție $f_2(x)$
Pas 1	Se generează val. de selecție u a v.a. $U \sim \mathcal{U}(0, 1)$
Pas 2	Dacă $u \leq 0.5$ atunci mergi la Pas 3; altfel mergi la Pas 4
Pas 3	s $=$ 1; mergi la Pas 5
Pas 4	s $=$ -1
Pas 5	Se generează val. de selecție x_1 a v.a. $X_1 \sim Exp(\lambda)$
Pas 6	Se definește $x = sx_1$
Ieșire	Valoarea de selecție, x, a v.a. X.

Metoda compunerii (7): Exemplu amestecare continuă

- Fie variabila $X, X > 0$, durata de funcționare a unui aparat, repartizată Exponențial$(\eta\lambda)$, unde
  - λ este un parametru determinat de producător, iar
  - η este un parametru aleator care indică influența mediului în care este studiat aparatul.
- Repartiția de probabilitate a lui η este de tip $Gamma(0, b, a)$:

$$h(\eta) = \begin{cases} 0, & \text{dacă } x < 0 \\ \frac{b^a}{\Gamma(a)}\eta^{a-1}e^{-b\eta}, & \text{altfel} \end{cases} \tag{16}$$

- Densitatea de repartiție a lui X, pentru $x \geq 0$ va fi de forma

$$\begin{aligned} f(x) &= \int_0^\infty \eta\lambda e^{-\lambda\eta x}\frac{b^a}{\Gamma(a)}\eta^{a-1}e^{-b\eta}d\eta = \\ &= \frac{a\theta}{(\theta x+1)^{a+1}}, \text{ unde } \theta = \frac{\lambda}{b} \end{aligned} \tag{17}$$

- Repartiția având densitatea (17) se numește *repartiție Lomax*.

Bibliografie I

- M. Craiu (1998), *Statistică matematică: teorie și probleme*, Editura Matrix Rom, București
- W. L. Martinez, A. R. Martinez (2002), *Computational Statistics Handbook with MATLAB*, Chapman & Hall/CRC, Boca Raton London New York Washington, D.C.
- I. Văduva (2004), *Modele de simulare: note de curs*, Editura Universității din București, București