Dr Slavko Tomović,
dipl. inž.
Tehnički opitni centar KoV,
Beograd

PRIMENA METODE STATISTIČKOG MODELOVANJA EKSPERIMENTA

UDC: 519.24:681.324

Rezime:

U radu je opisan postupak statističkog modelovanja eksperimenta po uzoru na metodu Monte-Karlo i prikazane su razne mogućnosti primene, posebno u matematičkoj statistici i pouzdanosti. U metodi se koriste slučajni i pseudoslučajni brojevi. Na primerima funkcija raspodele vremena rada do pojave otkaza dati su matematički modeli za izračunavanje vremena rada, kao i primer modela za generisanje pseudoslučajnih brojeva sa povoljnim osobinama za primenu na mikroračunarima.

Ključne reči: metoda statističkog modelovanja, slučajni i pseudoslučajni brojevi, matematička statistika, pouzdanost, funkcija raspodele.

APPLICATION OF THE EXPERIMENT STATISTIC SIMULATION METHOD

Summary:

This paper presents models for the simulation of experiments by the Monte-Carlo method. The models are often used in many distributions, especially for the generation of the time to failure in the reliability analysis. The method supposes that the real distribution function is known and has uniform distributed values. The time to failure is calculated as the inverse function. The application of this model is illustrated by some examples.

Key words: statistic simulation method, random and pseudorandom numbers, mathematics statistics, reliability, distribution function.

Uvod

Pri proučavanju ili projektovanju nekog fizičkog sistema vrše se razne analize njegovih funkcija sa ciljem da se optimizira funkcija efektivnosti. Da bi se zadaci lakše rešavali pristupa se posmatranju modela umesto realnog sistema. Na primer, u fazi projektovanja, fizički sistem još i ne postoji, pa je model jedina alternativa.

U oblasti modelovanja sistema koriste se tri vrste modela: fizički, analogni i matematički.

Fizički model sličan je realnom sistemu. On se, međutim, može često uprostiti tako da zahvata samo važne karakteristike realnog sistema, odnosno samo detalj koji se posmatra.

Analogan model ne liči na stvaran sistem. Na primer, električni sistem prikazuje se hidraulički, pa se proticanje elektriciteta prikazuje proticanjem hidrauličkog fluida. U stvari, koriste se analogne matematičke relacije.

Matematički model izražava posmatrana svojstva realnog sistema matematičkim algoritmom.

Fizički modeli su najrealniji, najbliži su realnom sistemu, ali su skupi i komplikovani za izradu i analizu. Analogni su apstraktniji, ali često pogodniji za realizaciju. Sa tog stanovišta, matematički modeli su najpogodniji, mada su najopštiji i najapstraktniji.

Poseban vid matematičkog modelovanja je statističko modelovanje. Metod statističkog modelovanja (simulacije) u stvari je metod rešavanja numeričkih zadataka. Naziva se još i metod Monte-Karlo [2] a dobio je najširu primenu sa razvojem elektronskih računskih sistema. Statističkim modelovanjem obuhvataju se one pojave, procesi i veličine koji imaju slučajni karakter. Pri analizama svih važnijih statističkih karakteristika važni su tačnost i verodostojnost rezultata, što zavisi od broja simulacionih eksperimenata, a greška je reda kvadratnog korena iz $(1/n)$.

Efikasnost metode dolazi do najvećeg izražaja upravo kod najsloženijih matematičkih sistema.

U radu se razmatraju zadaci koji se odnose na modelovanje statističkih eksperimenata iz domena pouzdanosti tehničkih sistema.

Opis metode Monte-Karlo

Metod Monte-Karlo ima jednostavan algoritam za izračunavanje. Druga osobina metode jeste da je greška izračunavanja, po pravilu, proporcionalna kvadratnom korenu iz D/N , gde je D konstanta, a N broj ispitivanja (eksperimenata). Ako želimo da smanjimo grešku $1 \cdot 10^2$ puta broj eksperimenata treba povećati $1 \cdot 10^4$ puta.

Na primer, treba naći približnu vrednost veličine A korišćenjem metode Monte-Karlo. Odredi se raspodela neke slučajne promenljive X , tako da je matematičko očekivanje ove veličine $m(X)=A$. Na osnovu realizovanog uzorka $(x_1, \dots, x_n)$ velikog obima n ,

matematičko očekivanje A ocenjuje se sredinom uzorka:

$$a = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1)$$

Tačnost ove ocene može se proceniti sa datom verovatnoćom. Prema centralnoj graničnoj teoremi može se postaviti izraz:

$$P(|\bar{x} - a| < \varepsilon) = 2\Phi\left(\varepsilon \frac{\sqrt{n}}{s_n}\right),$$

$$s_n^2 = \frac{1}{n} \sum (x_k - \bar{x}_n)^2 \quad (2)$$

Za verovatnoću 0,95 tačnost je reda kvadratnog korena iz $(1/n)$, što se za manje vrednosti n ne smatra za veliku efikasnost. Metoda se primenjuje tamo gde se ne traži znatno veća tačnost od ove.

Za modelovanje neke slučajne veličine (x) potrebno je poznavati ili odrediti zakon raspodele te slučajne veličine, a onda za funkciju raspodele $F(x)$ pretpostavljamo da je uniformno raspodeljena u opsegu $(0,1)$. Zatim se $F(x)$ zamenjuje sa slučajnim brojem iz skupa uniformno raspodeljenih slučajnih brojeva i za svaku vrednost $F(x_i)$ izračuna se vrednost modelovane slučajne veličine iz izraza kojim je definisana funkcija raspodele slučajne veličine.

Modelovanje funkcije raspodele slučajne veličine

Osnovni problem svodi se na određivanje slučajnog uzorka (skupa vrednosti) uniformno raspodeljene funkcije raspodele slučajne promenljive veličine, što se može rešiti korišćenjem tablica slučajnih brojeva ili najčešće putem generisanja slučajnih brojeva u toku procesa modelovanja, koji se dobijaju preko definisanog matematičkog algoritma. Potrebno je da perioda ponavljanja istog niza pseudoslučajnih

brojeva bude što veća. Pored jednostavnosti za generisanje, pseudoslučajni brojevi imaju prednost jer se pomoću njih može realizovati ponovljivost, što je značajno pri upoređenju raznih analitičkih varijanti, radi ocene efikasnosti pojedinih matematičkih modela u toku konkretnog teoretskog istraživanja.

Rekurentna formula za izračunavanje pseudoslučajnih brojeva ima oblik:

$$z_{n+1} = a \cdot z_n \pmod{m} \quad (3)$$

Relacija (3) definiše da je z_{n+1} ostatak pri deljenju z_n sa m . Ovako dobijen pseudoslučajni broj je ceo broj. Da bi dobili slučajan broj $0 < y < 1$ koristi se izraz:

$$y_n = z_n / m \quad (4)$$

U izrazima (3) i (4) pogodno su odabrani veliki brojevi, tako da se postigne što veća perioda ponavljanja. Na primer, mogu se naći primeri gde je $a=997$ i $m=1000$. Znatno jednostavniji za primenu je algoritam (predložila ga je firma Hewlet Packard) u obliku:

$$\begin{aligned} y_{n+1} &= y_n \cdot 997 - \lfloor (y_n \cdot 997) \rfloor, \\ y_0 &= 0,5284163 \end{aligned} \quad (5)$$

Izračunavanje funkcija metodom Monte-Karlo

U matematičkoj statistici, kao i u oblasti pouzdanosti i u drugim oblastima, potrebno je poznavanje određenih funkcija, kao što su gama, beta i Laplasova funkcija. Vrednosti ovih funkcija daju se tabelarno, ali je njihova primena u tom obliku nepodesna.

Metoda Monte-Karlo omogućava približno izračunavanje funkcija, što se u praksi pokazalo kao povoljno zbog jednostavnosti i velike tačnosti.

Približno izračunavanje određenog integrala

Neka je zadata funkcija $f(x)$ na intervalu $a \leq x \leq b$. Zadatak je da se približno izračuna integral:

$$I = \int_a^b f(x) \quad (6)$$

Paralelno sa funkcijom $f(x)$ može se pretpostaviti neka slučajna veličina ξ koja na intervalu a, b ima gustinu raspodele $p_\xi(x)$ i koja zadovoljava uslove [1]:

$$p_\xi(x) > 0 \quad (7)$$

$$\int_a^b p_\xi(x) dx = 1 \quad (8)$$

Pored slučajne veličine ξ formira se na istom intervalu a, b i izvedena slučajna veličina

$$\eta = \frac{f(\xi)}{p_\xi(\xi)}$$

Matematičko očekivanje ove slučajne veličine je

$$M_\eta = \int_a^b [f(x) / p_\xi(x)] \cdot p_\xi(x) dx = I \quad (9)$$

odnosno, to je vrednost integrala datog izrazom (6).

Posmatrajmo niz od N jednorodnih nezavisnih slučajnih veličina $\eta_1, \eta_2, \eta_3, \dots, \eta_N$ i primenimo na njega centralnu graničnu teoremu matematičke statistike. Po njoj za dovoljno veliko N važi verovatnoća:

$$P \left\{ \left| \frac{1}{N} \sum_{i=1}^N \eta_i - I \right| < 3 \sqrt{\frac{D_\eta}{N}} \right\} \approx 0,997 \quad (10)$$

U izrazu (10) N je broj članova niza, D_η je disperzija slučajne veličine η i I vrednost integrala (6).

Izraz (10) znači da se, ako se izabere dovoljno veliko N , može sa velikom tačnošću tvrditi da je:

$$\frac{1}{n} \sum_{i=1}^N \frac{f(\xi_i)}{p_{\xi}(\xi_i)} \sim I \quad (11)$$

Metoda se u praksi svodi na to da se odabere pogodan matematički model za funkciju gustine raspodele $p_{\xi}(x)$. Kada imamo tu funkciju, njoj odgovara za bilo koju vrednost X verovatnoća

$$P(X) = \int_a^x p_{\xi}(x) dx \quad (12)$$

Ako verovatnoća $P(X)$ uzima vrednosti iz uniformno raspodeljenog slučajnog skupa (preko slučajnih ili pseudoslučajnih brojeva), za poznate vrednosti $p(x)$ možemo modelovati vrednosti za X i uneti ih u izraz (10).

Za konkretan određeni integral bira se pogodna funkcija gustine raspodele $p_{\xi}(x)$.

Primena ove metode ilustrovaće se na primerima određivanja vrednosti gama, beta i Laplasove funkcije.

Izračunavanje vrednosti gama funkcije

Gama funkcija ima oblik:

$$\Gamma(p) = \int_0^{\infty} t^{p-1} \cdot e^{-t} dt, \quad p > 0 \quad (13)$$

Ova funkcija ima osobinu da važi izraz [5]:

$$\Gamma(1+p) = p\Gamma(p), \quad (14)$$

koji se koristi u praksi za brže izračunavanje određenih vrednosti.

Za približno izračunavanje vrednosti funkcije (13), po metodi izloženoj u ovoj tački pokazalo se kao izuzetno povoljno da se kao funkcija $p_{\xi}(x)$ upotrebi funkcija gustine raspodele za Vejbulov zakon raspodele, odnosno:

$$p_{\xi}(x) = \gamma \cdot \beta \cdot X^{\beta-1} \exp(-\gamma \cdot X^{\beta}) \quad (15)$$

preko koje se izračunava verovatnoća:

$$P(x) = \int_0^x p_{\xi}(x) dx = \exp(-\gamma \cdot X^{\beta}) \quad (16)$$

gde je:

γ — parametar razmere,

β — parametar oblika.

Između raznih vrednosti ovih parametara izabraćemo one koji su ispitani kao najpovoljniji u praksi, a to su $\beta=0,3$ i $\gamma=0,5$.

Ako za razne vrednosti verovatnoće $P(x_i)$, koja može imati determinisane vrednosti od 0 do 1, izračunamo iz izraza (6) vrednosti x_i , dobijamo:

$$X_i = \left[\frac{\ln \left(\frac{1}{P(x_i)} \right)}{\gamma} \right]^{\frac{1}{\beta}} \quad (17)$$

Relacija (11) sada ima oblik:

$$\Gamma(p) = \frac{1}{N} \sum_{i=1}^N \frac{X_i^{p-1} \cdot e^{-X_i}}{\gamma \cdot \beta \cdot X_i^{\beta-1} \exp(-\gamma \cdot X_i^{\beta})} \quad (18)$$

gde je:

x_i — izračunato iz izraza (17).

Za vrednost verovatnoće $P(X_i)$ koristimo izraz:

$$P(X_i) = P_0 - \sum_{i=1}^N \frac{1}{N} \quad (19)$$

gde je:

$$P_0 = 1 - \frac{1}{2N}$$

Ovako definisano $P(X_i)$ omogućava znatno bržu konvergenciju niza (11) nego kada se za tu verovatnoću koriste slučajni brojevi.

Praktično, već za $N=20$ dobijaju se sa tačnošću 10^{-5} vrednosti gama funkcije $\Gamma(p)$, u opsegu p (od 1 do 2), a za

Tabela 1

Redni broj	1	2	3	4	5	6	7	8	9	10
p	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
$\Gamma(p)$	0,95135	0,91817	0,89747	0,88726	0,88623	0,89352	0,90864	0,93138	0,96177	1,00000

vrednosti p izvan ovog intervala može se koristiti izraz (14).

U tabeli 1 date su ovim putem izračunate vrednosti gama funkcije za 10 vrednosti parametra p (od 1,1 do 2,0).

Izrazi (17), (18) i (19) pogodni su za unošenje u aplikacione softvere. Za vrednosti $n > 3/4 N$ članovi niza (18) veoma brzo teže nuli, pa se za izračunavanje može prekinuti kod $i=3/4 N$, ostali članovi niza zamene se sa nulom.

Izračunavanje vrednosti beta funkcije

Beta funkcija definiše se integralom:

$$\beta(p,q) = \int_0^x t^{p-1}(1-t)^{q-1} dt \quad (20)$$

Za parametre p i q beta funkcija ima vrednost [5]:

$$\beta(p,q) = \frac{\Gamma(p) \cdot \Gamma(q)}{\Gamma(p+q)} \quad (21)$$

pa se njeno izračunavanje svodi na izračunavanje vrednosti gama funkcije $\Gamma(p)$, $\Gamma(q)$ i $\Gamma(p+q)$ koje se uvrste u izraz (21).

Izračunavanje Laplasove funkcije

Laplasova funkcija definiše se integralom:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \quad (22)$$

Za primenu izložene metodologije može se, kao model za gustinu raspodele $p_{\xi}(x)$, koristiti gustina uniformne raspodele definisane u intervalu 0 do X, pa je

$$p_{\xi}(x) = \frac{1}{x} \quad (23)$$

Verovatnoća P(x) je:

$$p(t_i) = \int_0^{t_i} \frac{1}{x} dt = \frac{t}{x} \quad (24)$$

odakle je $t_i = P(t_i) \cdot X_i$, a izraz (6) ima oblik:

$$\Phi(x) = \frac{1}{N\sqrt{2\pi}} \sum_{i=1}^N \frac{e^{-\frac{t_i^2}{2}}}{\frac{1}{x}} \quad (25)$$

Tabela 2

Redni broj	1	2	3	4	5	6	7	8	9	10
t	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
$\Phi(t)$	0,19146	0,34194	0,43319	0,47725	0,49379	0,49865	0,49977	0,49996	0,499997	0,499997

Za $N=50$ dobijaju se, sa tačnošću 10^{-5} , vrednosti Laplasove funkcije za t od nule do 5. Za $N=20$ dobijaju se vrednosti sa tačnošću 10^{-4} . Za 10 raznih vrednosti t (od 0,5 do 5,0) prikazane su ovako izračunate Laplasove funkcije (tabela 2).

Za modelovanje vrednosti t po izrazu (24) koriste se vrednosti za $P(t_i)$ dobijene prema izrazu (19).

Modelovanje eksperimenta primenom metode Monte-Karlo

Pri eksperimentalnim ispitivanjima dobijaju se rezultati slučajnog karaktera, koji se mogu razmatrati kao vrednosti posmatrane slučajne veličine koje podležu određenim statističkim zakonitostima. Zbog toga je pri statističkim analizama ključna funkcija raspodele slučajne veličine, kao i parametri koji iz nje proizilaze.

U praksi se zakon raspodele unapred poznaje, ili se on pretpostavi, a zatim se ta hipoteza testira putem statističkih testova, kao što je, na primer, test Kolmogorova-Smirnova.

Određivanjem parametara ove funkcije dobija se statistička slika o osobinama tehničkog proizvoda, što je cilj izvođenja eksperimenata (ispitivanja).

Modelovanjem se simuliraju eksperimentalni rezultati bez skupih stvarnih ispitivanja, a generisani slučajni uzorci imaju bitne prednosti, kao što su brzina rada i ponovljivost.

Primeri modelovanja vremena rada do pojave otkaza za razne zakone raspodele

Pri modelovanju vremena rada do pojave otkaza, koji se pokorava poznatom zakonu raspodele, posmatramo funkciju raspodele vremena rada do otkaza $F(t)$ koja je definisana izrazom:

$$F(t) = \int_0^t f(t) dt \quad (26)$$

gde je:

$f(t)$ — funkcija gustine raspodele vremena rada do pojave otkaza.

Pretpostavlja se, zatim, da ova funkcija dobija razne slučajne vrednosti iz skupa uniformno raspodeljenih slučajnih ili pseudoslučajnih brojeva $N(0,1)$, a zatim se za svaku ovu vrednost izračunava vreme rada do pojave otkaza kao inverzna funkcija:

$$t = F(t)^{-1} \quad (27)$$

Za neprekidne funkcije, svakoj vrednosti jednoznačno odgovara neko vreme t po izrazu (27).

Normalna raspodela

Funkcija raspodele ima oblik:

$$F(t) = 1 - \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^t e^{-1/2 \cdot \frac{(t-m)^2}{\sigma^2}} \cdot dt \quad (28)$$

Integral u izrazu (28) ne može se izraziti preko elementarnih funkcija, pa se predstavlja preko Laplasove funkcije:

$$\Phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \int_0^x e^{-t^2/2} dt \quad (29)$$

Ako se za vrednost funkcije raspodele uzme slučajan broj iz skupa $N(0,1)$ onda se t može izračunati sukcesivnim približavanjem, tako što se izračuna $1 - F(t)$, a integral se računa kao zbir:

$$\frac{1}{\sqrt{2 \cdot \pi}} \cdot \sum_0^n e^{-\frac{(\sum t_i)^2}{2}} \cdot \Delta t \quad (30)$$

gde je:

$$t_i = \sum_0^n \Delta t$$

Za dovoljno malo Δt i dovoljno veliko n može se, sa bilo kojom tačnošću, odrediti t tako da se zadovolji relacija (28).

Pomoću ovako modelovanog vremena rada do pojave otkaza, koje odgovara normalnoj raspodeli kod koje je standardna devijacija jednaka 1, a srednje vreme $m=0$, izračunava se modelovano vreme rada do pojave otkaza za bilo koje srednje vreme, i bilo koju standardnu devijaciju, po sledećem izrazu:

$$t_i = m + \sigma t \quad (31)$$

Weibulova raspodela

Funkcija raspodele za ovu raspodelu glasi:

$$F(t) = 1 - e^{-\gamma t^\beta} \quad (32)$$

Odatle se izračunava vreme rada do pojave otkaza:

$$t_i = \left[\frac{-\ln(1 - F(t))}{\gamma} \right]^{1/\beta} \quad (33)$$

Ako se ne raspolaže elektronskim računarom koji generiše pseudoslučajne brojeve, ovi brojevi se mogu generisati pomoću sledeće rekurentne formule:

$$Y_{i+1} = (Y_i \cdot 997) - \lfloor (Y_i \cdot 997) \rfloor \quad (34)$$

U izrazu (34) u obzir se uzima samo broj manji od 1, tako što se od proizvoda u zagradi oduzme njegov celobrojni deo.

Literatura:

- [1] Tomović, S.: Izračunavanje i korišćenje nekih važnih funkcija u matematičkoj statistici i pouzdanosti, »Naučno-tehnički pregled«, br. 9, vol. 46, Beograd, 1996.
- [2] Sobol, I. M.: Metod Monte-Karlo, »Nauka« Moskva, 1978.

Zaključak

Statističko modelovanje je savremeni prilaz analitičkim analizama slučajnih veličina. Pojave ili procesi, koji imaju slučajni karakter, mogu se izračunati preko slučajnih brojeva. Najvažniju ulogu imaju uniformno raspodeljeni slučajni i pseudoslučajni brojevi iz skupa (0, 1), kojima se predstavlja verovatnoća nastupanja slučajnih događaja koji se posmatraju. Zakonitosti raspodela posmatranih slučajnih veličina mogu se prema potrebi modelovati. Pri teorijskim istraživanjima modelovanje može da zameni stvaran eksperiment, čime se postižu uštede, u vremenu potrebnom da se prikupe uzorci neophodni za teorijska istraživanja.

Na primer, da bi se pri razvoju nove metode za ocenu parametara neke raspodele, proverila efikasnost date metode i uporedila sa drugim poznatim metodama, mogu se modelovati eksperimentalni skupovi i na osnovu njih odrediti parametar po novoj metodi, a zatim ga statistički analizirati, putem određivanja statističkih pokazatelja, kao što su srednja vrednost, standardna devijacija i dr.

Metoda Monte-Karlo, koja se najčešće primenjuje kod statističkog modelovanja, omogućava efikasne približne proračune pri određivanju vrednosti raznih funkcija.

- [3] Ivković, Z.: Matematička statistika, »Naučna knjiga«, Beograd, 1976.
- [4] Zečević, T.: Matematičko i statističko modeliranje, »Stručna knjiga«, Beograd, 1977.
- [5] Vukadinović S.: Elementi teorije pouzdanosti i teorije obnavljanja tehničkih sistema, »Privredni pregled«, Beograd, 1979.