Generisanje slučajnih brojeva

- Ključni element, neophodan za sprovodjenje simulacije jesu slučajni brojevi. Mada je generisanje brojeva koji su zaista slučajni u principu moguće, zahvaljujući fizičkim fenomenima u makro (bacanje kocke, rulet,...) i mikro svetu (beli šum, kvantni fenomeni,...), za praktično korišćenje u simulaciji i primenu metode Monte Karlo koriste se pseudo slučajni brojevi.
- Pseudo slučajni brojevi po svojoj suštini, odnosno načinu nastanka, nisu slučajni, već su posledica primene odgovarajućeg algoritma, ali poseduju osobine koje u potrebnoj meri odgovaraju nizu slučajnih brojeva. Te osobine podrazumevaju pre svega nezavisnost, saglasnost sa odgovarajućim raspodelama verovatnoća koja se dokazuje statističkim testovima, kao i zadovoljenje testa autokorelativnosti.

- Prvi od pristupa generisanju pseudoslučajnih brojeva predložio je Džon fon Nojman 1946, za potrebe korišćenja metoda Monte Karlo na računaru ENIAC (Electronic Numerical Integrator And Computer). To je bio jednostavan algoritam zasnovan na metodu “sredine kvadrata”.
- Ovaj metod podrazumeva da se za generisanje niza pseudoslučajnih izabere jedan početni (semeni), te da se nakon kvadriranja tog broja, sledeći dobija kao niz središnjih cifara dobijenog kvadrata, nakon čega se postupak ponavlja.

- Ako je izabrani semeni broj 6457, njegov kvadrat iznosi 41692849, te je prvi “slučajni broj” predstavljen sa četiri središnje cifre 6928. Kvadriranjem ovog broja i ponavljanjem procesa dobija se niz čija su prva četiri člana: 6928, 9971, 4208 i 7072.
- Medjutim, kod ovog algoritma, period do ponavljanja za n-to cifrene brojeve ne može biti duži od 10^n, te je primena ovog pristupa ograničena.
- Pored ovoga razvijen je i niz drugih algoritama kao što su: *linearni kongruentni generator*, ili *inverzni kongruentni generator*, odnosno savremeniji, kao što su *BlumBlumShub-ov generator* (Blum L., et al. 1986) ili *Mersenne twister* (Matsumoto M., Nishimura T., 1998).
- ***Linearni kongruentni generator*** predstavlja jedan od najstarijih i najpoznatijih algoritama

Ovaj algoritam, predložen pre više od pola veka (Lehmer, 1951) baziran je na primeni sledećeg rekurzivnog izraza:

$$X_{j+1}=(aX_j+b)\ \mathbf{mod}\ m$$

gde su:

X_j – pseudo slučajni brojevi ($\mathbf{X_0}$, semeni broj čija se vrednost zadaje)
a,b,m – celobrojne konstante, **a>0, b>0,m>0**
mod – operator modulnog deljenja (označava ostatak pri deljenju dva broja, što znači da **u mod v** označava ostatak pri deljenju **u** brojem **v**)

- Period linearnog kongruentnog generatora iznosi najviše **m**, ali je često manji, te primena ovog generatora u mnogome zavisi od izbora vrednosti za **a,b i m**. S obzirom da se u simulaciji koriste slučajni brojevi u intervalu (0,1), najčešće se koristi trensformacija $X_j^{(0,1)} = X_j / m$.
- U tabeli je dat primer sekvence pseudo slučajnih brojeva dobijenih primenom linearnog kongruentnog generatora, za m=11123, a=1341, c=2977 i X_0=8543.

Primer korišćenja linearnog kongruentnog generatora

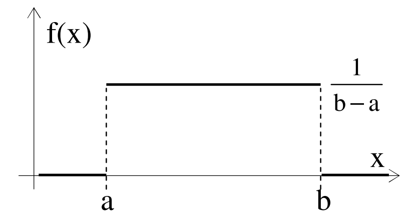
REDNI BROJ	$aX_j + m$	X_j	$X_j^{(0,1)} = X_j / m$
1	11459140	2450	0.22026
2	3288427	7142	0.64209
3	9580399	3496	0.31430
4	4691113	8330	0.74890
5	11173507	6015	0.54077
6	8069092	4917	0.44206
7	6596674	735	0.06608

Statističke metode i teorija verovatnoće u simulaciji

- Simulaciono modeliranje povezano je i sa primenom metoda matematičke statistike i teorije verovatnoće te podrazumeva znanja iz ovih oblasti. To se se pre svega odnosi na poznavanje osnovnih postavki teorije verovatnoće, pojma slučajne promenljive, osobina gustine i funkcije raspodele slučajnih promenljivih, parametara koji karakterišu slučajne promenljive – najpre matematičkog očekivanja i disperzije, kao i statističkih metoda na bazi kojih se ove veličine ocenjuju na uzorku, odnosno na empirijskim ili simulacijom dobijenim podacima.
- Metode matematičke statistike i teorije verovatnoće podjednako su važne za:
  - ✓ generisanje slučajnih vrednosti promenljive sa poznatim teorijskim ili empirijskim zakonom raspodele
  - ✓ analizu ulaznih podataka koji se koriste u simulacionom modelu
  - ✓ analizu rezultata simulacije

Generisanje slučajnih promenljivih iz kontinualnih raspodela

- Kada je reč o kontinualnim vrednostima, najčešće se koriste ravnomerna, eksponencijalna, normalna, odnosno neka empirijska raspodela verovatnoća
- **RAVNOMERNA RASPODELA** neprekidne slučajne promenljive X, na intervalu (a,b), kako je poznato, definiše se gustinom raspodele verovatnoća



$$f(x) = \begin{cases} 0, \text{ za } x < a \\ \dfrac{1}{b-a}, \text{ za } a \le x \le b \\ 0, \text{ za } x > b \end{cases}$$

- Funkcija raspodele definiše se izrazom

$$F(x) = \begin{cases} 0, \text{ za } x < a \\ \displaystyle\int_0^x \frac{1}{b-a}dx = \frac{x-a}{b-a}, \text{ za } a \le x \le b \\ 0, \text{ za } x > b \end{cases}$$

- Matematičko očekivanje i disperzija

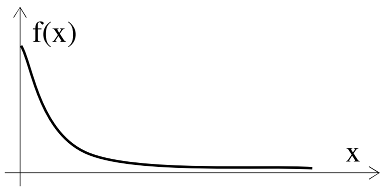
$$M(X) = \frac{a+b}{2}, \quad D(X) = \frac{(b-a)^2}{12}$$

- Za ***generisanje vrednosti ravnomerno rasporedjene slučajne promenljive*** na intervalu (a,b), koristiti se direktna transformacija funkcije raspodele slučajne promenljive

- Za datu vrednost slučajnog broja $\gamma \in (0,1)$, rešavanjem $x = F^{-1}(\gamma)$, lako se dobija vrednost slučajne promenljive $x \in (a, b)$.

$$\gamma = F(x) \Rightarrow \gamma = \frac{x-a}{b-a} \Rightarrow x = a + \gamma(b-a)$$

- **EKSPONENCIJALNA RASPODELA** neprekidne slučajne promenljive X, sa parametrom λ, kako je poznato, definiše se gustinom raspodele verovatnoća datom izrazom



$$f(x) = \lambda e^{-\lambda x}, 0 \le x < \infty$$

- Funkcija raspodele eksponencijalno rasporedjenje slućajne promenljive, M(X), D(X) definišu se izrazima

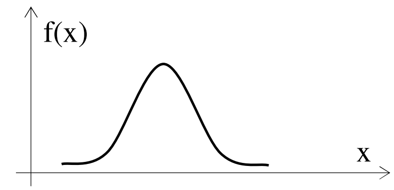
$$F(x) = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}, \quad M(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}$$

- Za generisanje vrednosti eksponencijalno rasporedjene slučajne promenljive sa parametrom λ, analogno ravnomernoj raspodeli, može se koristiti direktna transformacija funkcije raspodele slučajne promenljive.

- Za datu vrednost slučajnog broja $\gamma \in (0,1)$, rešavanjem $x = F^{-1}(\gamma)$, lako se dobija vrednost slučajne promenljive x.

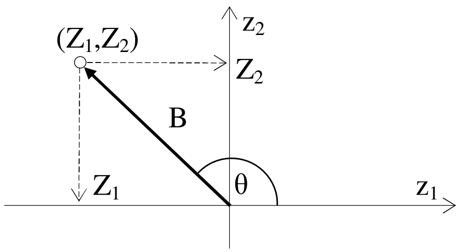
$$\gamma = F(x) \Rightarrow \gamma = 1 - e^{-\lambda x} \Rightarrow x = -\frac{1}{\lambda}\ln(1-\gamma) \Leftrightarrow x = -\frac{1}{\lambda}\ln\gamma$$

- Ekvivalencija 1-γ ⇔ γ, korišćena u prethodnom izrazu, posledica je činjenice da ukoliko je $\gamma \in (0,1)$ slučajni broj, razlika 1-γ, je takodje slučajni broj u istom intervalu, te se za utvrđivanje vrednosti slučajne promenljive x, mogu koristiti oba oblika izraza.
- **NORMALNA RASPODELA** neprekidne slučajne promenljive X, sa parametrima μ i σ, kako je poznato, definiše se gustinom raspodele verovatnoća datom narednim izrazom, gde su date i vrednosti matematičkog očekivanja i disperzije.



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty \quad M(X) = \mu, D(X) = \sigma^2$$

- Funkciju raspodele normalno rasporedjenje slučajne promenljive nije moguće dobiti u praktično primenjivoj formi, s obzirom da nije moguće naći primitivnu funkciju funkcije f(x). Otuda i nije moguće generisanje vrednosti normalno rasporedjene slučajne promenljive koje bi bilo bazirano na transformaciji $x = F^{-1}(\gamma)$. Kao posledica toga, za generisanje normalno rasporedjenih slučajnih promenljivih koriste se različiti algoritmi.



- ***Metod direktne transformacije*** predložili su (Box G.E.P, Muller M.F., 1958), a ideja je bazirana je na položaju tačke u pravouglom koordinatnom sistemu sa polarnim koordinatama, za slučaj da su koordinate tačke normalno rasporedjene slučajne veličine Z_1 i Z_2.
- U tom slučaju $B^2 = Z_1^2 + Z_2^2$ ima Hi-kvadrat raspodelu sa dva stepena slobode, što odgovara eksponencijalnoj raspodeli sa parametrom λ=0.5.
- S obzirom na to, radijus B moguće je generisati kao kvadratni koren eksponencijalno rasporedjene slučajne veličine, tj. $B = \sqrt{(-2\ln\gamma)}$.
- Kako je, takođe, zbog simetričnosti gustine normalne raspodele ugao θ ravnomerno raspoređen na intervalu [0,2π], pri čemu su B i θ nezavisni, imajući u vidu relacije $Z_1 = B\cos\theta$ i $Z_2 = B\sin\theta$, metod za generisanje dve nezavisne, standardizovane, normalno rasporedjene slučajne promenljive (μ=0, σ=1), za dva data, nezavisna, ravnomerno rasporedjena slučajna broja γ_1 i γ_2 moguće je predstaviti izrazom

$$Z_1 = \sqrt{(-2\ln\gamma_1)} \cdot \cos(2\pi\gamma_2)$$

$$Z_2 = \sqrt{(-2\ln\gamma_1)} \cdot \sin(2\pi\gamma_2)$$

- Nakon što se generišu vrednosti standardizovane normalno rasporedjenje slučajne promenljive Z_1, odnosno Z_2, vrednost slučajne promenljive X, proizvoljne normalne raspodele sa parametrima $N(\mu, \sigma)$ utvrdjuje se na osnovu izraza

$$X = \mu + \sigma Z_1, \text{ odnosno } X = \mu + \sigma Z_2$$

- ***Metod zbira slučajnih promenljivih*** baziran je na primeni zakona velikih brojeva i centralne granične teoreme.
- Naime, jedna od posledica zakona velikih brojeva je da ukoliko postoji n nezavisnih slučajnih promenljivih X_1, X_2,...X_n sa proizvoljnom, istom, raspodelom verovatnoća, pri čemu je $M(X_i)= \mu$ i $D(X_i)= \sigma^2$, onda raspodela verovatnoća slučajne promenljive

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

teži normalnpoj raspodeli s parametrima μ i $\frac{\sigma}{\sqrt{n}}$, tj.

$$\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

- Prema tome, zbir "dovoljno velikog broja" ravnomerno rasporedjenih slučajnih brojeva zapravo ima normalnu raspodelu. Pokazuje se da je za praktičnu primenu dovoljno koristiti 6 sabiraka, mada je mnogo jednostavnije, a uz to i tačnije, koristiti zbir od 12 ravnomerno raspoređenih slučajnih brojeva.

- Primena ovog algoritma podrazumeva formiranje zbira 12 ravnomerno raspoređenih slučajnih brojeva $\sum_{i=1}^{12} \gamma_i$, čiji zbir ima normalnu raspodelu sa parametrima $\mu_{\sum \gamma_i} = 12 \cdot \frac{1}{2} = 6$ i $\sigma_{\sum \gamma_i} = \sqrt{12 \cdot \frac{1}{12}} = 1$.
- Na bazi ovoga, vrednost standardizovane normalno rasporedjenje slučajne promenljive Z, može se dobiti kao

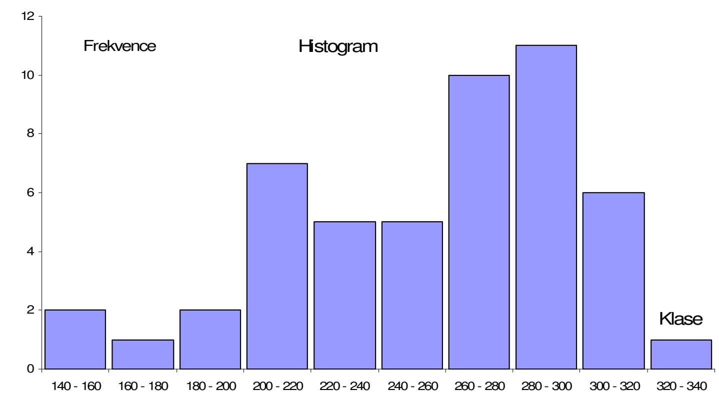
$$Z = \frac{\sum_{i=1}^{12} \gamma_i - \mu_{\sum \gamma_i}}{\sigma_{\sum \gamma_i}}$$

- Da bi se dobila vrednost X proizvoljne normalne raspodele sa parametrima $N(\mu, \sigma)$, koristi se izraz

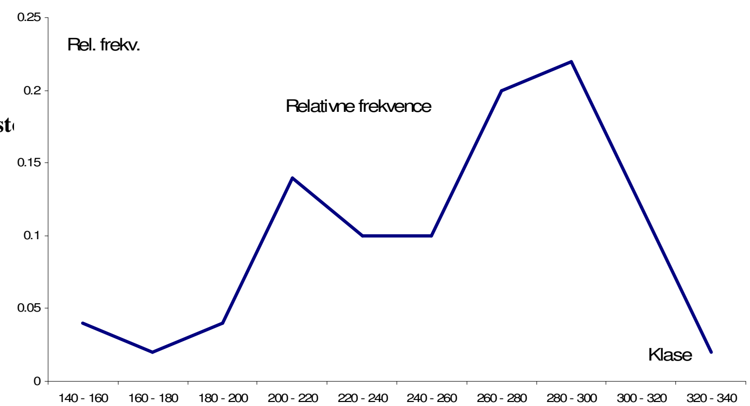
$$X = \mu + \sigma Z \text{, tj. za 12 sabiraka} \quad X = \mu + \sigma\left(\sum_{i=1}^{12} \gamma_i - 6\right)$$

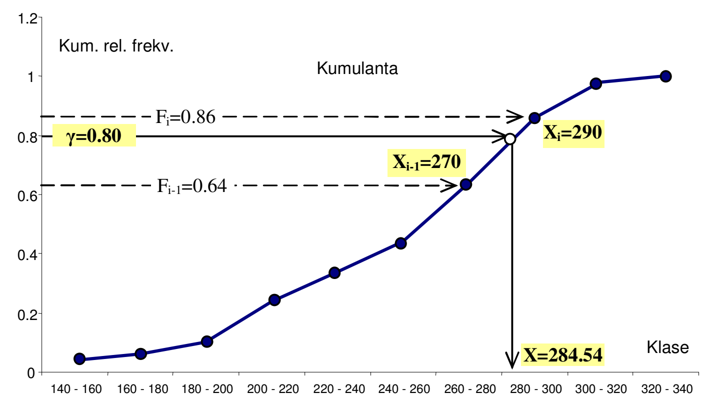
- **KONTINUALNA EMPIRIJSKA RASPODELA** predstavljena je statističkom tabelom koja sadrži klase, frekvence, relativne frekvence i kumulantu relativnih frekvenci kako je to prikazano na primeru u tabeli i grafički histogramom raspodele frekvencija, odnosno raspodelom relativnih frekvenci i kumulantom.

R.B. (i)	KLASE (X_i)	FREKVENCE (f_i)	REL. FREKV. (f_{Ri})	KUMULANTA ($F_i=\sum f_{Ri}$)
1	140 - 160	2	0.04	0.04
2	160 - 180	1	0.02	0.06
3	180 - 200	2	0.04	0.10
4	200 - 220	7	0.14	0.24
5	220 - 240	5	0.10	0.34
6	240 - 260	5	0.10	0.44
7	260 - 280	10	0.20	0.64
8	280 - 300	11	0.22	0.86
9	300 - 320	6	0.12	0.98
10	320 - 340	1	0.02	1.00
	∑	**50**	**1**	



Hist





- Simulacija kontinualne empirijske raspodele, nakon generisanja slučajnog broja γ, svodi se na određivanje klase kojoj pripada vrednost simulirane slučajne promenljive i na linearnu ekstrapolaciju vrednosti X koja odgovara generisanom slučajnom broju γ.
- Odredjivanje klase podrazumeva primenu algoritma kojim se definišu X_{i-1} i X_i, takvi da $(X_{i-1}, X_i \mid F_{i-1} < \gamma < F_i)$.
- Za poznate vrednosti sredina klasa X_{i-1} i X_i, ekstrapolirana vrednost slučajne promenljive X odredjuje se na bazi izraza.

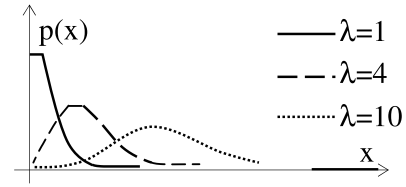
$$X = X_{i-1} + (\gamma - F_{i-1})\frac{X_i - X_{i-1}}{F_i - F_{i-1}}$$

- Za vrednost slučajnog broja γ=0.80, za podatke iz tabele, dobija se

$$X = 270 + (0.80 - 0.64)\frac{290 - 270}{0.86 - 0.64} = 284.54$$

Generisanje slučajnih promenljivih iz diskretnih raspodela

- Kada je reč o diskretnim vrednostima slučajnih promenljivih, najčešće je u primeni Puasonova, ili neka empirijska raspodela verovatnoća.
- **PUASONOVA RASPODELA** diskretne slučajne promenljive X, kako je poznato, definiše se raspodelom verovatnoća datom izrazom



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

- Funkcija raspodele definiše se kao suma prethodnih verovatnoća, a matematičko očekivanje i disperzija dati su u nastavku.

$$F(x) = P(X < x) = \sum_{i=0}^{x} \frac{e^{-\lambda}\lambda^i}{i!}, \quad i = 0,1,2,\ldots,x \,, \quad M(X) = D(X) = \lambda$$

- Za generisanje vrednosti slučajne promenljive sa Puasonovom raspodelom sa parametrom λ, koristiti se interpretacija ove raspodele kao broja pojavljivanja dogadjaja u Puasonovom procesu, pri čemu je poznato da intervali izmedju pojavljivanja dva uzastopna dogadjaja imaju eksponencijalnu raspodelu.
- Otuda, broj pojavljivanja dogadjaja u nekom intervalu jednak je x, ako i samo ako se x-ti dogadjaj pojavi pre kraja intervala dužine 1, a naredni x+1 nakon tog vremena, to jest gde su t_i, intervali pojave dogadjaja

$$X = x \quad \text{ako i samo ako} \quad t_1 + t_2 + \ldots + t_x \le 1 < t_1 + t_2 + \ldots + t_x + t_{x+1}$$

- Ukoliko se iskoristi relacija za generisanje eksponencijalno rasporedjene slučajne promenljive, tada se prethodni izraz može napisati na drugi način:

$$\sum_{i=1}^{x} -\frac{1}{\lambda}\ln\gamma_i \le 1 < \sum_{i=1}^{x+1} -\frac{1}{\lambda}\ln\gamma_i$$

- Koristeći osobinu da se zbir logaritama može predstaviti logaritmom proizvoda, sredjivanjem prethodnog izraza dobija se

$$\ln\prod_{i=1}^{x}\gamma_i \ge -\lambda > \ln\prod_{i=1}^{x+1}\gamma_i \quad \Rightarrow \quad \prod_{i=1}^{x}\gamma_i \ge e^{-\lambda} > \prod_{i=1}^{x+1}\gamma_i$$

- Na bazi prethodnog izraza moguće je konstruisati jednostavan algoritam (Banks et al. 2005) za generisanje vrednosti slučajne promenljive sa Puasonovom raspodelom

Korak 1. *Postaviti $x=0$ i $P=1$*

Korak 2. *Generisati slučajan broj γ_{x+1} i zameniti $P=P\cdot \gamma_{x+1}$*

Korak 3. *Ako je $P<e^{-\lambda}$ onda je $X=x$. U suprotnom, uvećati $x=x+1$ i ponoviti* **Korak 2**

- Primena algoritma prikazana je u tabeli, za Puasonovu raspodelu sa parametrom 0.1, korišćenjem pet slučajnih brojeva.

X	γ	P	REZULTAT KORAKA 3	SIMULIRANA VREDNOST
0	0.6780	0.6780	0.6780<0.9048 (DA)	X=0
0	0.9099	0.9099	0.9099<0.9048 (NE)	-
1	0.3763	0.3424	0.3424<0.9048 (DA)	X=1
0	0.9117	0.9117	0.9117<0.9048 (NE)	-
0	0.3123	0.2847	0.2847<0.9048 (DA)	X=0

- **DISKRETNA EMPIRIJSKA RASPODELA**, analogno kontinualnoj, predstavljena je statističkom tabelom koja sadrži vrednosti slučajne promenljive, frekvence, relativne frekvence i kumulantu relativnih frekvenci.
- Simulacija diskretne empirijske raspodele, nakon generisanja slučajnog broja γ, svodi se na određivanje vrednosti slučajne promenljive X koja odgovara generisanom slučajnom broju γ, što podrazumeva primenu algoritma kojim se definiše X_i, takvo da je $(X_i \mid F_{i-1} < \gamma \leq F_i)$, kako je to prikazano na slici.