

MC meetodid, praktikum 2

Ülesanne 1: generaatori korrektsuse uurimine

Pseudojuhuslike arvude generaatorite koodi kirjutamisel peab teadma ka arvuti hingeelust!

Kõigepealt kontrolli, et Sul on korrektselt töötav generaator LKG eelmisest praktikumist. Selleks võrdle käsu `LKG(6,seeme=15,a=17,b=3,m=256)` tulemust allpool toodud näitetulemusega:

```
## [1] 0.0078125 0.1445312 0.4687500 0.9804688 0.6796875 0.5664062
```

Kui väärtused on samad, siis on töötab generaator vähemalt toodud parameetrite korral õigesti. Vaatleme aga nüüd SAS tarkvaras RANUNI käsuga kasutatavat generaatorit, mis vastab juhule $a = 397204094, b = 0, m = 2^{31} - 1$. SAS tarkvara korral on seemne 10 korral selle generaatori tekitatud esimesed 8 väärtust

0.8496257, 0.7008872, 0.9982431, 0.5939865, 0.2160258, 0.6927735, 0.4297917, 0.3169172

Vaata, kas Sinu generaator annab täpselt samad arvud? Kui ei anna, siis katsu aru saada, milles on probleem. Vihje: uuri järgmiste käskude tulemusi:

```
print(397204094*1000000001,digits = 20)
print(2^53,digits=20)
print(2^53+1,digits=20)
```

Ülesanne 2: generaatori parandamine

Installeerige pakett `gmp` (*Gnu Multiple Precision arithmetics*) ning laadi see käsuga `library(gmp)`. Uuri järgnevaid käsked ning muuda nende abil oma generaatorit nii, et see annaks õiged tulemused parameetrite a, b, m korral, mis on väiksemad kui 2^{53} . Generaatori väljundiks peaks ikka olema vektor n reaalarvust 0 ja 1 vahel. Kontrolli, kas parandatud generaator annab juhul $a = 397204094, b = 0, m = 2^{31} - 1$ õiged tulemused.

```
as.bigz(397204094)*1000000001
as.bigz(10001)%%276
as.numeric(as.bigz(10001)%%276)
```

Ülesanne 3: Spektraaltestist

Vaatleme generaatorit parameetritega $a = 2^{11} + 1, b = 3, m = 2^{16}$. Mis on selle generaatori perioodiks?

Lihtne on veenduda, et selle generaatori korral $(x_{k+1} - x_k) \bmod m = (2^{11} x_k + 3) \bmod m$. Siit saab näidata, et vahel $x_{k+1} - x_k$ saab olla küllaltki vähe erinevaid väärtuseid (miks?) ja seega on ka vahedel $u_{i+1} - u_i$ küllaltki vähe erinevaid väärtuseid, mistõttu paarid (u_{2i-1}, u_{2i}) on kaetud suhteliselt väikese arvu sirgetega kujul $y = x + c$.

- Genereeri perioodile vastav arv punkte lähtudes seemnest 1 ning tekita neist kaheveeruline maatriks M , täites selle ridade haaval (st ridades on üksteisele vahetult järnevad arvud). Tekita graafik, kus on näha punktide paiknemine ühikruudus.
- Leia, kui palju erinevaid väärtuseid omavad vaadeldud juhul vahed $y - x$, kus y on maatriksi teises veerus olev arv ja x on samas reas esimeses veerus olev arv. Selleks saab lihtsalt lahutada maatriksi teisest veerust esimese veeru ja rakendada tulemusele funktsiooni `table` (et näha erinevaid väärtuseid ja kui mitu korda need esinevad).
- Sirgete $y = ax + c_1$ ja $y = ax + c_2$ vaheline kaugus on $\frac{|c_1 - c_2|}{\sqrt{1+a^2}}$. Leida kaugus punkte katvate naabersirgete vahel.

- See tähendab, et selle generaatori abil võime mingi sündmuse tõenäosuse hindamisel eksida vähemalt leitud kauguse võrra. Näiteks ühikruudus juhuslikult valitud punkti piirkonda $D = \{0 \leq x, y \leq 1; 0,01 < y-x < 0,06\}$ sattumise tõenäosuse hinnaguks saame vaadeldava generaatoriga 0 sõltumata genereeritud punktide arvust. Veendu selles väites ning leia hinnang mainitud piirkonda sattumise tõenäosusele, kasutades selleks generaatoriga `runif` saadud 200000 arvu. Vihje: eelnevad võrratused on samaväärsed tingimusega $|y - x - 0,035| < 0,025$, seega saab kasutada eelmise praktikumi funktsiooni `loendaja`.

Ülesanne 4: Testimine (algus)

Testimiseks on vaja teada, kuidas peaksid käituma juhuslikud arvud ehk sõltumatute katsete tulemused. Siis saame võrrelda pseudojuhuslike arvude käitumist juhuslike arvude käitumisega. Kui liiga erinevad, siis loeme generaatori halvaks.

Alustame lihtsa ideega, kus vaatleme mingi fikseeritud sündmuse toimumist juhuslikel katsetel. Fikseerime näiteks sündmuse $A = \{0,2 \leq X \leq 0,5\}$. Kui me sooritame sõltumatuid katseid n korda ja loeme kokku A toimumiste arvu N_A , siis on teada, kuidas N_A peaks käituma. Mis on N_A jaotus?

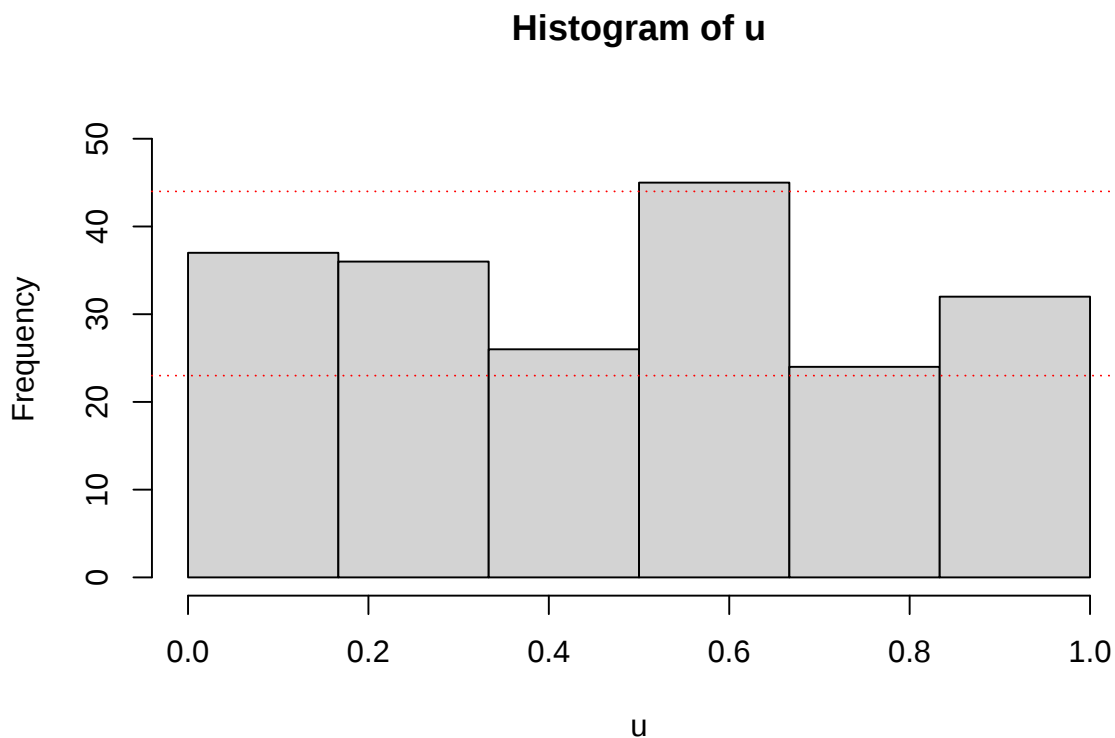
- Kasutame RANDU generaatorit ($a = 65539, b = 0, m = 2^{31}$). Leia seemnest 1000 lähtudes genereeritud 100 pseudojuhusliku arvu põhjal sündmuse A toimumiste arv. Kas see jäi binoomjaotuse 0,025 ja 0,975 kvantiilide vahele? (Vihje: kasuks tuleb käsk `qbinom`)
- Päris juhuslike katsete korral peaks leitud arv jääma tõenäosusega 0,05 ka piiridest välja. Korda katset 200 korda (võtteks esimeseks seemneks 100 ja järgmistel kordadel viimasest generaatori väljastatud arvule u_n vastav x_n) ning leia, mitu korda jäi leitud tulemus piiridest välja. Kui suur on tõenäosus, et see kordade arv erineks oodatavast väljajäämistest arvust vähemalt niipalju, kui praguse katse tulemus näitas?

Ülesanne 5: Jaotuse kontroll histogrammi abil

Selle asemel, et vaadelda juhusliku suuruse sattumist ühte lõiku, võib vaadelda näiteks omavahel mittelõikuvate intervallide süsteemi, mis katavad kõik võimalused. Kui nüüd lugeda igasse intervalli sattuvate punktide arvud ühe katse põhjal, siis saame histogrammi.

- Jätkame RANDU generaatori kasutamist. R-s on üheks võimaluseks histogrammi joonistamiseks käsk `hist` ning tekitavate lõikude süsteemi saab ise ette anda suvandiga `breaks=vektor_jaotuspunktidest`. Genereeri seemnest 71293529 lähtuvalt 200 punkti ning tekita histogramm, kus on kujutatud intervallidesse $[0, \frac{1}{6}), [\frac{1}{6}, \frac{2}{6}), \dots$ sattumiste arve. Kuna intervallid on sama pikkusega ja randu peaks väljastama lõigus $[0,1]$ ühtlase jaotusega pseudojuhuslikke arve, on veapiirid igasse lõiku sattuvate punktide arvule samad. Lisa graafikule neile veapiiridele vastavad horisontaalsed sirged, mis annavad iga lõiku korral teoreetilise jaotuse 0,025 ja 0,975 kvantiilid.

Tulemus peaks välja nägema selline:



Kodutöö 2

Esitamine Moodle kaudu, tähtaeg 18.09.2022. Esitada tuleb ka .R või .Rmd fail ning lisaks kas pdf või HTML file, kus on näha kood koos tuletuskäikude, tulemuste ja kommentaaridega.

Ülesanne 2.1 (0,75p). Defineeri funktsioon, mis etteantud kahekomponedilisest seemnest (x_0, x_1) ja positiivsest täisarvust n lähtuvalt genereeriks n pseudojuhuslikku arvu vastavalt valemile $u_{i-1} = \frac{x_i}{2^{30}}$, $i = 2, \dots, n+1$, kus arvud x_i leitakse vastavalt valemile

$$x_i = (5x_{i-1}^2 + 171x_{i-2} + 3) \mod 2^{30}.$$

Kontrolli $n=10000$ ja seemne (1974, 42) korral, kas sündmuse $0.2 \leq U \leq 0.3$ toimumine on kooskõlas ühtlase jaotuse eeldusega.

Ülesanne 2.2 (0,75p) Kirjuta Wichmann-Hilli generaator (kirjeldus näiteks 2. loengu 11. slaidil või loengukonspektis peatükis 2.1.1) R funktsioonina, mis võtaks argumentideks arvu, mitu juhuslikku suurus on vaja teha, ja seemne (3 arvu). Genereeri selle generaatoriga 1000 pseudojuhuslikku arvu, lähtudes seemnest $(x_0, y_0, z_0) = (166, 9, 32)$ ning kirjeldada joonisel nende arvude sattumist vahemikesse $[0, 0.3)$, $[0.3, 0.8)$, $[0.8, 1)$. Samal joonisel kujutada teoreetilised 0,05 ja 0,95 kvantiilid (mis on iga vahemiku jaoks erinevad). Kommenteeri tulemust - kas see katse tekitab mingeid kahtluseid generaatori korrektsuse osas või mitte.

Ülesanne 2.3 (Boonusülesanne, 1 punkt) Leidke täpsed valemid (koos põhjendustega) ülesandes 3 vaadeldud generaatori poolt väljastatavate vahede $u_{i+1} - u_i$ võimalikele väärtustele. Samuti leidke ülesandes 3 vaadeldud piirkonna D täpne pindala (koos selgitustega, kuidas tulemuseni on jõutud).