

MC meetodid, praktikum 4

Teema 1: Vahemike testi rakendamine sõltumatuse kontrollimiseks

Vahemike test:

- Fikseerime mingi sündmuse A , mille toimumise tõenäosus ühel katsel on p
- Vaatleme, kui palju katseid kulub, et pärast eelmist A toimumist see toimuks jälle, moodustame jada nendest arvudest
- Ooteaegade jaotuseks on $Geom(p)$. Kirjanduses on sellest jaotusest kaks versiooni, mis erinevad vähima võimaliku väärtuse osas (kas väärtused hakkavad nullist või ühest). Meie vaatame versiooni, mis hakkab ühest, st $P(\{X = k\}) = p \cdot (1 - p)^{k-1}$, $k = 1, 2, \dots$ ja $P(\{X \geq \ell\}) = (1 - p)^{\ell-1}$
- Valime ℓ , mis on oluliselt väiksem kui n , võrdleme pikkustega $1, 2, \dots, \ell - 1$ ja pikkusega vähemalt ℓ vahemike arvu oodatava arvuga k -s katses (jälle χ^2 -test)

Harjutus 1

Genereeri 10000 pseudoranduslikku arvu lineaarse kongruentse geneaatoriga juhul $a = 5$, $b = 11$ ja $m = 2^{30}$, kasutades seemet 11. Vaatleme sündmust A , mis vastab saadud katsetulemuse kuulumisele lõiku $[0, 0.1]$. Kontrolli vahemike testiga tulemuste vastavust katsete sõltumatuse tingimusele juhul $\ell = 11$. Mis on otsus olulisuse nivool $\alpha = 0.1$?

Vihjed ülesande lahendamiseks:

- 1) Tekita TRUE/FALSE väärtustega vektor, mis annab iga pseudorandusliku arvu korral teada, kas vaadeldav sündmus toimus või mitte (seda saab teha ühe käsuga, ilma tsükliteta, kasutades vektori võrdlemist arvuga). 2) Seejärel saab `which` käsuga leida toimumiste ajad (nt kui on vektor `kas_toimus` väärtustega FALSE,TRUE,FALSE,FALSE,TRUE, siis `millal_toimus <- which(kas_toimus)` annab tulemuseks 2,5)
- 2) Kui ooteaegadele ette panna `c()` käsuga arv 0, siis `diff()` käsuga saab leida vektori ooteaegadest (eelmisses näites saame `which(millal_toimus)` abil ooteaegadeks 2,3). Ülesande lahendamisel leitud ooteaegadest esimesed 5 peaks olema 1, 1, 1, 1, 1).
- 3) Tabeli tekitamiseks muutke seejärel kõik 11-st suuremad ajad võrseks arvuga 11, mida saab jällegi teha ühe käsuga kogu vektor[tingimus] <- uus_väärtus. NB! Selleks, et sagedustabeli tekitamisel oleks esindatud kõik vaadeldavad väärtused isegi juhul, kui vastava pikkusega ooteaegu reaalselt ei esinenud, tuleks enne tabeli tekitamist leitud ooteaegade vektori faktortunnuseks käsuga kujul `ajad <- factor(ajad, levels=1:ell)`
- 4) Lõpuks tuleb Hii-ruut testi rakendamiseks leida ooteaegade sagedustabel ning geomeetrilise jaotuse põhjal vastavate tõenäosuste vektor. Testi rakendamisel peaks tulemuseks olema

```
##
## Chi-squared test for given probabilities
##
## data:  tabel
## X-squared = 149.23, df = 10, p-value < 0.00000000000000022
```

Teema 2: Seeriade test sõltumatuse kontrollimiseks

Seeriade test

- Fikseerime m vastastikku välistavat võrdvõimalikku sündmust $A_1, \dots, A_m$

- vaatleme, kui mitme katse järel toimub eelnevatest erinev sündmust (st vaatame, kui pikk sama sündmuse jada toimus enne muutumist)
- Võrdleme erinevate seeriade pikkuste esinemissagedusi teoreetiliste tõenäosustega, milleks on $Geom(\frac{m-1}{m})$

Harjutus 2

Tekita käsuga `rnorm` 10000 pseudojuhuslikku arvu standardsest normaaljaotusest. Vaatleme kümmet sündmust, mis vastab reaaltelje jagamisele 0.1, 0.2, ..., 0.9 kvantiilidega osadeks. Kontrolli vahemike testiga tulemuste vastavust sõltumatutele katsetele, vaadeldes seeriade pikkusi 1,2,3,4, ja neljast pikemaid seeriaid. Kas χ^2 -jaotuse kasutamine p -väärtuste arvutamiseks on õigustatud? Mis on tekitatud arvude korral otsus olulisuse nivool $\alpha = 0.01$?

Vihjed lahendamiseks. Kvantiilid saab leida käsuga `qnorm`. Leitud kvantiile ja `cut` käsku kasutades saab teha kindlaks, milline sündmus igale arvule vastab. Kui muutuda `cut` kasu tulemus käsuga `as.numeric()` täisarvude vektoriks (st saame sündmuste numbrid), siis edasi saame `diff` käsu tulemuse ja `which` käsu abil leida kohad, kus üks sündmus läheb teiseks üle. Edasi on tegevused juba tuttavad. Kui selgub, et χ^2 -jaotuse kasutamine ei ole teststatistiku väärtuse arvutamisel õigustatud, saab R käsus `chisq.test()` anda ette suvandi `simulate.p.value=TRUE`, mille tulemusena leitakse teststatistiku p -väärtus arvutisimulatsioonide teel.

Kontrollimiseks: kui otse enne käsu `rnorm` kasutamist anda käsk `set.seed(2019)`, siis esimesele kolmele x väärtusele vastavad sündmuste numbrid on 8, 4, 1 ja esimesed 16 ooteaega on 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1

Teema 3: Etteantud jaotusega pseudojuhuslike arvude genereerimine jaotusfunktsiooni pööramise meetodil.

Teame: Kui F on pidev ja piirkonnas $\{x : 0 < F(x) < 1\}$ rangelt kasvav jaotusfunktsioon ja U on ühtlase jaotusega $U(0, 1)$ juhuslik suurus, siis juhusliku suuruse $X = F^{-1}(U)$ jaotusfunktsiooniks on F

Harjutus 3

Tekita `runif` käsuga pseudojuhuslikku arvu jaotusest $U(0, 1)$ ning nendest jaotusfunktsiooni pöördfunktsiooni kasutades pseudojuhuslikud arvud jaotusest, mille jaotusfunktsiooniks on $F(x) = \frac{e^x}{e^x + 1}$, $-\infty < x < \infty$. Kontrolli Kolmogorov-Smirnovi testiga saadud arvude vastavust soovitud jaotusele. Samuti kontrolli vahemike testiga juhul $A = \{X \geq 1\}$ vastavust katsete sõltumatuse eeldusele.

Vihjed lahendamiseks Kindlasti tasub alati kontrollida leitud pöördfunktsiooni korrektsust. Sobivaks kontrolliks on, et $F(F^{-1}(x)) = x$ iga $x \in (0, 1)$ korral. Seda võib teha näiteks graafiku joonistamisega käsuga `curve(F(Fpoord(x)), 0.01, 0.99)`, arvutada y vastavalt eelnevale valemile ning seejärel `plot` käsu abil veenduda, et tulemus on sirglõik.

Teema 4: Pokker-testi kasutamine katsete sõltumatuse kontrolliks.

Pokker-test

- Fikseerime m vastastikku välistavat võrdvõimalikku sündmust $A_1, \dots, A_m$
- Igale katsetulemusele seame vastavusse sündmuse numbri, mis sellel katsel toimus
- vaatleme viisikuid $(k_1, k_2, k_3, k_4, k_5), \dots$
- Vaatleme erinevaid mustreid: kõik erinevad, üks paar, kaks paari, kolmik, paar ja kolmik, nelik, viisik
- Kasutame χ^2 -testi paaride esinemissageduste võrdlemisel teoreetiliste sagedustega Teoreetilised tõenäosused on

muster	tõenäosus
erinevad	$\frac{(m-1) \cdot (m-2) \cdot (m-3) \cdot (m-4)}{10 \cdot (m-1) \cdot \frac{m^4}{(m-2)} \cdot (m-3)}$
paar	$\frac{15 \cdot (m-1) \cdot \frac{m^4}{(m-2)}}{10 \cdot (m-1) \cdot \frac{m^4}{(m-2)}}$
kaks paari	$\frac{10 \cdot (m-1) \cdot \frac{m^4}{(m-2)}}{10 \cdot (m-1) \cdot \frac{m^4}{(m-2)}}$
kolmik	$\frac{5 \cdot (m-1)}{\frac{m^4}{(m-2)}}$
kolmik ja paar	$\frac{1}{m^4}$
nelik	
viisik	

Harjutus 4

Tekita RANDU generaatoriga 50000 pseudojuhuslikku arvu alates seemnest 11. Vaatleme sündmuseid, mis vastavad hulkadesse $(0, 0.2]$, $(0.2, 0.4]$ jne sattumisele. Testi pokker-testiga, kas saadud pseudojuhuslikud arvud võivad vastata sõltumatutele katsetele.

Vihjed. viiest täisarvust vastava vektori mustri tüübi kindlakstegemiseks võib kasutada järgnevat funktsiooni:

```
muster <- function(viisik){
  abi <- c(1,2,3,5,7)
  number <- c(1,2,4,3,6,5,7)
  return(number[prod(abi[table(viisik)])])
}
```

Katsu aru saada, mida see funktsioon teeb! Tekita jada sündmuse numbritest, mis vastavad pseudojuhuslike suuruste kuulumisele valitud hulkadesse. Seda saab teha kas `cut` ja `as.numeric` käske kasutades või siis arve viiega korrutades ja `floor` funktsiooni rakendades. Seejärel tee maatriks, mille ridades on viisikud ning rakenda kõikidele ridadele funktsiooni muster `apply` käsku kasutades. Edasine peaks juba lihtne olema.

Kodune töö (tähtaeg 4. oktoober 2020) (1 punkt) Tekita vaikeparameetritega runif generaatoriga seemnest 197 lähtuvalt 100000 pseudojuhuslikku arvu ning kasuta neid selleks, et saada jaotusfunktsioonile

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{x}{2}, & 0 \leq x < 1, \\ 1 - \frac{e^{1-x}}{2}, & x \geq 1 \end{cases}$$

vastavad pseudojuhuslikud arvud. Kontrolli saadud arvude vastavust toodud jaotusfunktsioonile Cramer-von Mises testiga ning sõltumatust seeriade testiga (vaadeldes seeriaid pikkusega 1,2,3,4 ja neljast suuremaid), defineerides kvantiilide abil 9 võrdse tõenäosusega sündmuse täissüsteemi. NB! Kindlasti veendu, et leitud pöördfunktsiooni korral kehtib võrdus $F(F^{-1}(u)) = u \ \forall u \in (0, 1)$ (seda võib teha graafiku joonistamise teel).