Skilaverkefni
Inngangur að Líkindafræði

Monte Carlo hermun, heildun, teninga- og krónuköst

Skilið uppkasti í hólf Guðmundar (númer 72) í anddyri VR-II
fyrir föstudaginn 21.mars

Ykkur hefur verið skipt upp í **tveggja** manna hópa og skal hver hópur leysa það verkefni sem honum er úthlutað ásamt þeim lesnu dæmum sem Hermann setur fyrir. Þið haldið síðan kynningu á ykkar lausnum **föstudagana 21. og 28.mars**. Góð lausn og kynning á verkefninu gildir **1 heilan** til upphækkunar á prófinu. Í öllum dæmum þar sem forritun er þörf eigið þið að nota R. Ef þið lendið í vandræðum getið þið sent póst á Guðmund Einarsson, netfang: `gue20@hi.is`.

Verkefni 1: Mat á π

Í þessu dæmi ætlum við að meta π með svokallaðri Monte Carlo hermun.

Við getum ímyndað okkur að við séum að henda pílum á hringskífu með radíus 1 og að pílurnar lendi alltaf innan kassans $[-1,1]\times[-1,1]$. Ef við gerum ráð fyrir að pílurnar lendi hvar sem er á kassanum með jöfnum líkum, þá stefnir hlutfall þeirra pílna sem lenda innan hringskífunnar á hlutfall flatarmáls hringskífunnar af flatarmáli kassans, sem er einmitt $\frac{\pi}{4}$.

Nú skulum við skoða hvað við erum að gera fræðilega. Gerum ráð fyrir að við höfum tvær óháðar jafndreifðar stærðir U og V á $[-1,1]$. Þá viljum við meta

$$\mathbb{E}[\mathbf{1}_{\{U^2+V^2\le 1\}}] = \frac{\pi}{4}$$

þar sem $\mathbf{1}_{\{U^2+V^2\le 1\}}$ er kennifall mengisins, (kennifallið tekur gildið 1 ef skilyrðið er uppfyllt, annars er það 0). Þetta væntigildi er auðvitað heildi sem einfalt er að reikna, en við viljum sjá hvernig hægt er að meta það með slembiaðferðum.

Athugum nú að skv. lögmáli mikils fjölda (*e. law of large numbers*) gildir að ef við slembum úrtak úr samþéttleika (U,V), köllum það $((U_1,V_1),(U_2,V_2),...,(U_n,V_n))$, fáum við mat á væntigildið að ofan með:

$$\frac{\hat{\pi}}{4} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{U_i^2+V_i^2\le 1\}}$$

sem er nálægt $\frac{\pi}{4}$ þegar n er stór tala.

Verkefnið felst í því að herma svona úrtak í R og meta hversu margar tvenndir lenda innan hringskífunnar. Reikniritið er í grunninn:

1 Ákveða stærð úrtaks n og setja $C = 0$, þar sem C er teljarinn okkar.

2 Herma tvær uniform stærðir U, V á $[-1,1]\times[-1,1]$, hægt að gera með `runif(2,min=-1,max=1)`.

3 Athuga hvort $U^2 + V^2 \leq 1$, ef svo er hækka C um einn, annars breyta teljara breytunni ekkert, fara svo aftur í skref 2.

4 Þegar n ítrunum er lokið er matið okkar á π einfaldlega $4 \cdot \frac{C}{n}$.

Þið eigið að vinna eftirfarandi verkefni eins vel og þið getið: (d) og (e) eru aukaverkefni, þið þurfið ekki að svara þeim, en gaman er að reyna við þær.

(a) Er þetta gott mat á π? Prófið fyrir $n = 10^k$ þar sem $k = 3, 4, ..., 9$.

(b$_i$) Hvaða töluleg vandamál geta komið upp?

(b$_{ii}$) Eru tölurnar sem við erum að herma í R raunverulega slembnar?

(c) Teiknið mynd af matinu á móti fjölda ítrana, þar sem fjöldi ítrana er á lograskala, gerið þetta fyrir mest 10^6 ítranir. Virðist matið vera að stefna á π?

(d) Gefið mat á stærð úrtaks sem þarf til að fá 4 aukastafi rétta með 95% líkum. Hérna er skynsamlegt að nota höfuðmarkgildissetninguna.

(e) Búið til myndband af ferlinu, þar sem við sjáum punktana bætast við á hringskífuna og matið á π breytast jafnóðum.

Verkefni 2: Monte Carlo heildun

Í seinasta dæmi vorum við að beita almennri aðferð sem er kölluð Monte Carlo heildun. Hún gengur út á það að fyrir gefna slembistærð X og fall h viljum við meta

$$\mathbb{E}[h(X)] = \int_{\mathcal{X}} h(x)f(x)dx$$

þar sem f er þéttleiki X og $\mathcal{X}$ er úrtaksrúm X. Skv. lögmáli mikils fjölda gildir að

$$\overline{h}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

kemst nálægt væntigildinu $\mathbb{E}[h(X)]$ þegar n er stór tala, ef $(X_1, X_2, ..., X_n)$ er slembið úrtak úr dreifingu X. Þið eigið að framkvæma eftirfarandi útreikninga:

(a) Látum $X \sim \mathcal{N}(\mu, \sigma^2)$, hvað er gildi $\mathbb{E}[X^2]$?

(b) Notið aðferðina sem lýst er að ofan til að meta $\mathbb{E}[X^2]$ fyrir $\mu = 5$ og $\sigma = 10$, þar sem $n = 10^6$. Hvernig ber niðurstöðunni saman við gildið í (a)-lið?

(c) Plottið mynd af matinu $\overline{h}_n$ á móti fjölda ítrana. Látið ítranirnar vera á lograskala og gerið myndina fyrir $n = 1, 2, ..., 10^5$.

(d) Veljið fall eitthvert h og einhverja slembistærð X og metið $\mathbb{E}[h(X)]$ með aðferðinni.

Verkefni 3: Krónuköst

Gerum ráð fyrir að við köstum krónupeningi í röð. Táknum bergrisann með H og þorskinn með T. Dæmi um runu sem gæti komið í 5 köstum er þá $HTHHT$. Við höfum áhuga á að meta væntan fjölda kasta þar til seinustu þrjú táknin í rununni eru HTH, við viljum gera eins fyrir HTT.

Hægt er að herma þetta í R með því að geyma alltaf seinustu þrjú stökin í rununni og telja fjölda kasta þar til staðan HTH kemur upp og gera svo eins fyrir HTT. Ef þið framkvæmið 10^5 slíkar tilraunir og takið meðaltalið af fjölda kasta sem þarf þar til staðan komi upp ættuð þið að fá gott mat á væntan fjölda kasta.

Verkefnið er því

(a) Hermið tilraunina í R. Er sami væntur fjöldi kasta fyrir báðar stöðurnar HTT og HTH?

(b) Plottið mynd af matinu á móti fjölda ítrana, þar sem fjöldi ítrana er á lograskala.

(c) Reiknið í höndunum hver væntur fjöldi kasta er. Það er ekki skylda að gera þennan lið, en reynið samt við hann.

Hægt er að lýsa þessu ferli sem *Markovkeðju*, en ítarlega er farið í þær í námskeiðinu slembiferli.

Verkefni 4: Summa jafndreifðra stærða og normaldreifingin

Látum $U_1, U_2, ..., U_{12}$ vera óháðar jafndreifðar slembistærðir á bilinu $[0, 1]$. Skilgreinum þá nýja slembistærð

$$Z_U = \left(\sum_{i=1}^{12} U_i \right) - 6$$

Þessi slembistærð líkist stöðluðu normalstærðinni á margan hátt. Leysið nú eftirfarandi dæmi:

(a) Reiknið væntigildi og dreifni Z_U. Hvernig ber því saman við stöðluðu normaldreifinguna? (þ.e.a.s. $\mathcal{N}(0, 1)$)

(b) Hermið úr dreifingunni í R og gerið *stöplarit* af úrtakinu, þar sem flatarmál stöplanna er 1. Þetta gerið þið í R með því að kalla á `hist(x,freq=FALSE)`, þar sem `x` eru hermdu gögnin.

($\mathbf{c}_i$) Þessi dreifing var notuð fyrr á tímum í stað þess að herma úr staðlaðri normaldreifingu. Af hverju ætti það að vera í lagi?

($\mathbf{c}_{ii}$) Undir hvaða kringumstæðum myndum við ekki vilja nota svona nálgun?

Verkefni 5: Skrýtnir teningar

Gefnir eru fjórir sex hliða teningar, köllum þá α, β, γ og δ, þeir hafa eftirfarandi tölur á sínum hliðum:

$$\alpha : 5, 5, 5, 5, 1, 1$$
$$\beta : 4, 4, 4, 4, 4, 4$$
$$\gamma : 7, 7, 3, 3, 3, 3$$
$$\delta : 6, 6, 6, 2, 2, 2$$

Látum nú A, B, C og D tákna slembistærðir sem eru útkomur úr teningaköstum teninganna α, β, γ og δ. Leysið nú eftirfarandi dæmi:

(a) Hvað er væntigildið á A, B, C og D?

(b) Beitið hermun í R til að meta $P(A > B)$, $P(B > C)$, $P(C > D)$ og $P(D > A)$. Notið úrtak með a.m.k. 10^6 teningaköstum. Eftirfarandi kóði hermir í R eitt teningakast úr α:

```
teningurA <- c(5,5,5,5,1,1)
kast <- sample(teningurA,1)
```

(c) Þið spilið eftirfarandi leik við andstæðing:

(i) Annar spilarinn velur einn af teningunum

(ii) Hinn velur einn tening af restinni

(iii) Báðir kasta teningunum og sá sem fær hærra vinnur.

Byggt á niðurstöðunum í (b)-lið, viljið þið vera á undan eða á eftir að velja tening?

(d) Nú er leiknum breytt þannig að þið verðið að velja fyrst og síðan velur andstæðingurinn tening af handahófi, hvaða tening er best að velja?

Verkefni 6: Fleiri skrýtnir teningar

Í þessu dæmi ætlum við að skoða venjulega sex hliða teninga ásamt sex hliða teningunum

$$\eta : 1, 2, 2, 3, 3, 4$$
$$\theta : 1, 3, 4, 5, 6, 8$$

Köllum slembistærðina sem er útkoma úr teningakasti venjulegs sex hliða tenings D og köllum slembistærðirnar sem eru útkomur úr η og θ einfaldlega D_η og D_θ. Við höfum sérstakan áhuga á **summu útkoma** úr tveimur köstum **venjulegra** sex hliða teninga, köllum þær stærðir D_1 og D_2 og summan er þá $X = D_1 + D_2$. Látum að sama skapi $Y = D_\eta + D_\theta$. Okkur grunar að slembistærðirnar X og Y séu keimlíkar! Leysið fyrstu tvo liðina og reynið við (c)-lið.

(a) Reiknið væntigildi slembistærðanna X og Y. Hvert er útkomurúm X? Hefur Y sama útkomurúm?

(b) Sýnið að dreifing Y sé sú sama og dreifing X, þið megið nota hermun til þess. Eftirfarandi kóði hermir í R eitt teningakast úr η:

```
teningurEta <- c(1,2,2,3,3,4)
kast <- sample(teningurEta,1)
```

Verið viss um að herma nógu mörg köst til að fá gott mat líkunum á sérhverri útkomu Y.

(c) **Aukaspurning!** Er til annað par af sex hliða teningum (eins og η og θ), þannig að summa útkomanna hafi sömu dreifingu og X? Hér verða hliðar teninganna að vera jákvæðar heilar tölur.

Verkefni 7: Bjagaður peningur

Eftirfarandi spurningar eru úr atvinnuviðtölum hjá stórum hugbúnaðarfyrirtækjum!

(a) Gerum ráð fyrir að ykkur sé gefinn bjagaður krónu peningur. Þegar þessum peningi er kastað kemur upp bergrisi með líkunum p, en þorskur með líkunum $1-p$, þar sem $p \neq \frac{1}{2}$. Þið lendið í vandræðum, þar sem þörf er á óbjöguðum pening! Eina sem þið hafið í höndunum er bjagaði krónupeningurinn. Hvernig getið þið búið til tvo atburði með þessum pening, óháð líkunum p, þannig að helmings líkur eru á að hvor atburðurinn gerist? Hermið þessa atburði í R og sýnið niðurstöðurnar.

(b) Fyrir framan ykkur eru tveir lokaðir kassar. Í öðrum kassanum eru 50 bláir pennar og hinum eru 50 rauðir. Eftir ákveðinn tíma kemur maður sem velur annan kassann handahófskennt og dregur úr honum penna af handahófi. Hvernig getið þið hámarkað líkurnar á því að maðurinn dragi rauðann penna? Allir pennarnir verða að vera í kössunum þegar maðurinn kemur inn, annars megið þið raða þeim eins og þið viljið.