

# MC meetodid, praktikum 12

## Teema 1: Permutatsioonitestid

Permutatsioonitestid leiavad teststatistiku täpse jaotuse, vaadeldes argumenttunnuste kõikvõimalikke ümberpaigutusi kas kogu andmestikus või sobivalt valitud rühmade kaupa. Tähtis on aru saada, et permutatsioonitestide korral on nullhüpoteesiks uuritava tunnuse sõltumatus (või grupisisene tinglik sõltumatus, kui ümber paigutatakse mingite rühmade siseselt) vaadeldavatest argumenttunnustest, mistõttu iga ümberpaigutus vastab võimalikule katsetulemuste komplektile. Permutatsioonitestide heaks omaduseks on see, et need on täpsed testid (st nullhüpoteesi kehtimisel võetakse alternatiivseid hüpoteese vastu õige tõenäosusega).

Vaatleme permutatsioonitesti rakendamist ühe konkreetse näite korral. Kasutame selleks `coin` paketti (mis tuleb vajadusel paigaldada).

### Harjutus 1

Vaatleme parema ja vasaku käe pikkuste näiteandmestikku leheküljelt [https://rcompanion.org/handbook/K\\_01.html](https://rcompanion.org/handbook/K_01.html) (Introduction to Permutation Tests). Oletame, et tahame teha kindlaks, kas parema ja vasaku käe pikkuses on erinevusi. Vaatame kõigepealt ideed, et kasutame statistikuna kummagi rühma vaatluste keskmiste erinevust. Kui me ignoreerime, et andmestikus on kordusmõõtmised, siis eeldusel, et vasaku ja parema käe pikkused pärinevad samast jaotusest ja vaatlused on sõltumatud, võime leida nullhüpoteesile vastava statistiku jaotuse kohta informatsiooni, vaadeldes kõikvõimalikke tunnuse “Hand” väärtuste permutatsioone ning leides tekkinud rühmade erinevused nende põhjal.

Kasutame testi rakendamiseks valmi vahendeid, nimelt funktsiooni `oneway_test` paketist `coin`. Leidke testi tulemust. Mis järelduse  $\alpha = 0.05$  korral teeksite? Kas testi eeldused on rahudatud?

### Harjutus 2

Vaatleme eelnevat testi, eeldades, et me võime vahetada tunnuse “Hand” väärtuseid ainult iga inimese korral. Selleks tuleb kasutada käsus `oneway_test` valemit kujul `Length ~ Hand | Individual`. Mis järelduse teed testi tulemuse põhjal nüüd? Kas selle testi eeldused võiksid olla katse jaoks täidetud?

## Teema 2: Keskvärtuse leidmine MC meetodil

### Keskvärtuse hindamine MC meetodil

- Vaatleme juhuslikku suurust  $Y = g(X)$
- Genereerime  $n$ -elemendilise valimi  $X$  jaotusest, saame  $x_1, \dots, x_n$
- Rakendame funktsiooni  $g$ , saame valimi  $Y$  jaotusest:  $g(x_1), \dots, g(x_n)$
- Hindame  $Y$  keskvärtust valimi keskmise abil

$$EY \approx \bar{y} = \frac{1}{n} \sum_{i=1}^n g(x_i)$$

- valimi keskmise jaoks erinevad hinnangud (NB! Kehtivad, kui juhuslik suurus  $Y = g(X)$  on lõpliku dispersiooniga!)
  - ligikaudne usaldusintervall ( $z_\alpha$  tähistab standardse normaaljaotuse  $\alpha$ -kvantiili)

$$(\bar{y} + z_{\frac{\alpha}{2}} \frac{\sigma_Y}{\sqrt{n}}, \bar{y} - z_{\frac{\alpha}{2}} \frac{\sigma_Y}{\sqrt{n}})$$

- Tõenäosusega  $1 - \alpha$  kehtiv (ligikaudne) vahinnang:

$$-z_{\frac{\alpha}{2}} \frac{\sigma_Y}{\sqrt{n}}$$

– Tõenäoline viga - tõenäosusega 1/2 kehtiv veahinnang

### Harjutus 3

Leia MC meetodil  $E\sqrt{|X|^5 + 1}$ ,  $X \sim N(1, 0.5)$ , kasutades valimit suurusega  $n = 10000$ . Leia samuti hinnangu tõenäoline viga.

### Harjutus 4

Leia MC meetodil  $E\sqrt{|X|^5 + 1}$ ,  $X \sim N(1, 0.5)$ , kasutades sellise suurusega valimit, mille korral tõenäoline viga on ligikaudu 0.0001.

### Harjutus 5

Leia Monte-Carlo meetodil ristkülikust  $[0, 2] \times [0, 3]$  juhuslikult valitud punkti keskmine kaugus koordinaatide alguspunktist veaga, mis on tõenäosusega 0.95 väiksem kui 0.001

### Harjutus 6

MC meetodi rakendamise etapid ( $n$  väärtuse genereerimine, keskmise ja veahinnangu arvutamine) on korduvalt vajaminevad tegevused) ja käsuridade kopeerimise asemel on mõistlik defineerida vastav funktsioon. Hea on, kui see funktsioon lubaks lahendada palju erinevaid ülesandeid. Defineeri funktsioon MC, millele saab argumentidena anda ette generaatori nime, funktsiooni  $g$  nime, kasutatavate väärtuste arvu  $n$  ja veahinnangu leidmisel kasutatava  $\alpha$  väärtuse ning mis tagastaks arvupaari, kus esimene on keskväärtuse hinnang ja teine on veahinnang. Kontrolli oma funktsiooni tööd eelmise kahe ülesande juhul.

### Harjutus 7

Kontrolli funktsiooni MC töö korrektsust, leides  $E(e^{2X})$  ligikaudset väärtust  $n = 10000$  genereeritud väärtuse abil juhul  $X \sim \text{Exp}(3)$ . Korda arvutusi  $m = 200$  korda ning tee kindlaks, kui sageli on tegelik viga suurem kui tõenäolise vea hinnang. Kas tulemus on kooskõlas tõenäolise vea definitsiooniga?

## Teema 3: Integraalide leidmine MC meetodil

Iga integraali saab esitada keskväärtusena: kui on vaja arvutada

$$I = \int_D f(x) dx, \quad D \subset \mathbb{R}^m,$$

siis valime ühe tihedusfunktsiooni  $f_X$ , mille korral kehtib  $f_X(x) > 0$ ,  $x \in D$ , defineerime

$$g(x) = \begin{cases} \frac{f(x)}{f_X(x)}, & x \in D, \\ 0, & x \notin D \end{cases}$$

ning leiame MC meetodiga  $I = E g(X)$ , kus  $X$  on tihedusfunktsioonile  $f_X$  vastava jaotusega juhuslik suurus. Saadava meetodi headus sõltub aga oluliselt sellest, millised on funktsiooni  $g(x)$  omadused. Hea on, kui  $g(x)$  on tõkestatud mingi konstandiga. Kõige parem on, kui see on vähemuutuv (lähedane konstandile), mis juhtub siis, kui  $f_X$  on lähedane  $f$ -le. Halvasti valitud  $f_X$  korral aga võib juhtuda, et suuruse  $g(X)$  dispersioon on lõpmatu ning sel juhul on koondumine väga aeglane (ja tüüpilised MC meetodi veahinnangud ei kehti).

### Harjutus 8

Leia MC meetodiga ligikaudne väärtus koos tõenäolise veaga integraalile

$$\int_0^2 x^4 dx,$$

kasutades  $n = 10000$  genereeritud väärtust kolmel erineva  $X$  jaotuse korral: a)  $U(0,2)$ , b)  $N(1.5,1)$  c) tinglik jaotus  $X \sim N(4, 2) \mid 0 \leq X \leq 2$ . Millisel juhul on veahinnang kõige väiksem?

**Harjutus 9**

Leia MC meetodiga ligikaudne väärtus ja tõenäosusega 0.9 kehtiv veahinnang integraalile

$$\iiint_{x^2+y^2+z^2 \leq 1} \sqrt{x^2 + 2y^2 + 3z^2} \, dx dy dz,$$

kasutades selleks 100000 genereeritud väärtust.