$$4\int_0^r \sqrt{r^2-x^2}\,dx$$

JOHDANTOA SUURTEN LUKUJEN LAKIIN

Tekijät: Eeli Tamminen, Christian Webb, Mikko Jaskari, Eeva Pyrhönen, Heli Virtanen, Juha-Matti Huusko, Sakari Salonen

Laadittu keväällä 2022.

Kohderyhmä: Lukion kolmas vuosikurssi, pitkä matematiikka (oppitunti tai kerhotunti)

Esitiedot:

- Todennäköisyyslaskenta ja tilastotiede (lukion opintojakso)
- Excelin peruskäyttö

Oppitunnin kesto: 75 min

Oppimistavoitteet: Oppitunnin jälkeen opiskelija

- tietää suurten lukujen lain merkityksen tilastojen käsittelyssä ja
- osaa tehdä simulointia sekä kerätä aineistoa ja käsitellä sitä.

Muut tavoitteet: Oppitunnin tavoitteena on

- inspiroida opiskelijoita todennäköisyyslaskennan aihepiireihin liittyen ja
- konkretisoida ja tuoda matematiikkaa lähemmäs opiskelijan arkielämää.

Sisältö:

- Nopanheiton silmälukujen (tasajakauma) odotusarvon tutkiminen (25 min)
- Suurten lukujen lain todistus (20 min)
- Erilaisten karkkien lukumäärän (hypergeometrinen jakauma) odotusarvojen tutkiminen (20 min)
- Hypergeometrisen jakauman tutkiminen sekä visualisointi Geogebralla (10 min)

Huom. Noppaa heitettäessä tiedetään mikä odotusarvo on, kun vastaavasti karkkipussin sisältämien karkkien määrissä sitä ei tiedetä.

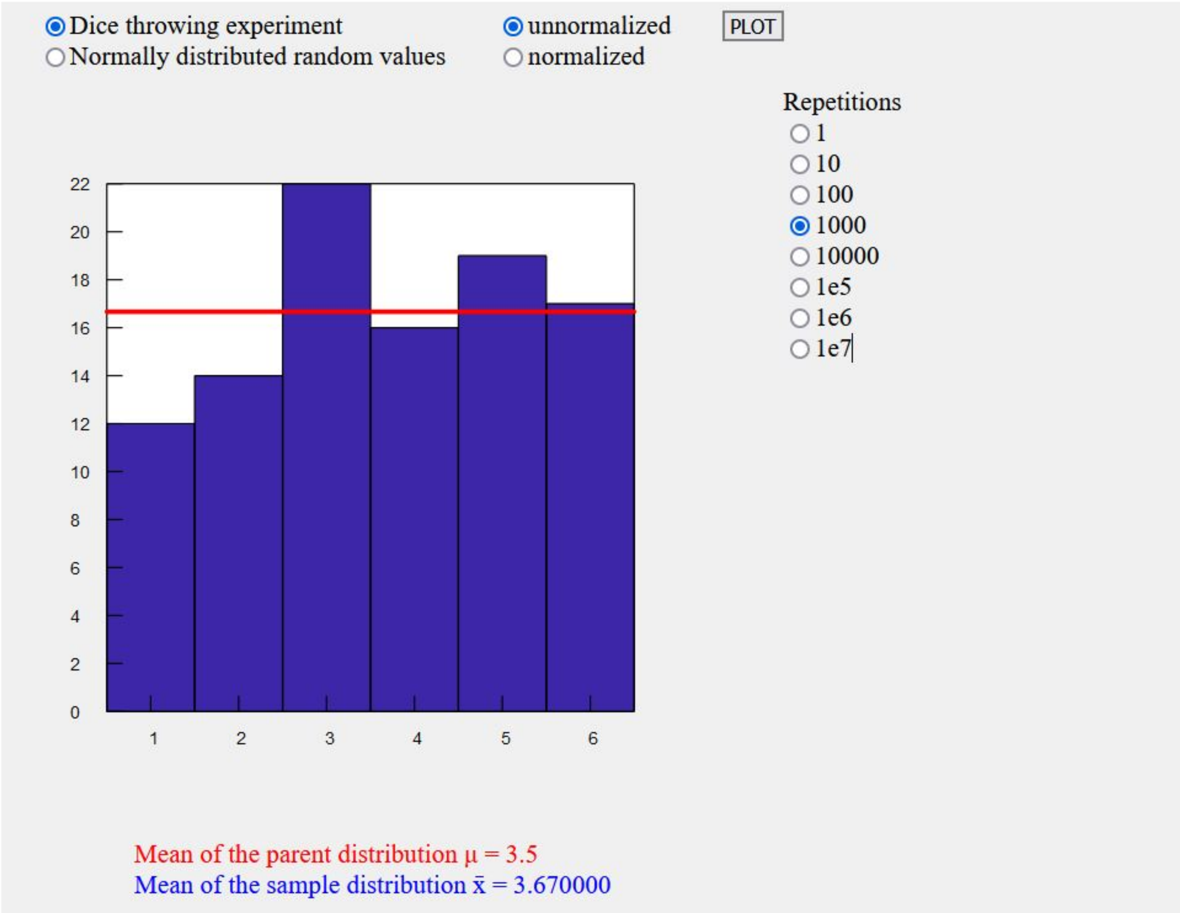
Tarvittavat välineet: Jokaisella ryhmällä tulee olla seuraavat välineet:

- 6-sivuinen noppa
- Karkkipussi, jossa noin 50 karkkia, joita ainakin kahta eri lajia
- Tietokone, jossa Geogebra, Excel tai vastaava

Toteutus:

1. Nopanheiton silmälukujen (tasajakauma) odotusarvon tutkiminen

- Seuraavaan on Excel-pohja liitteessä 1, jonkä lisäksi on ohjeissa kerrottu myös vaadittavat Excel-komennot.
- Kerätään ryhmissä dataa heittämällä noppia 100 kertaa per ryhmä, jonka jälkeen
  - ryhmissä lasketaan ja kirjoitetaan seuraavat tulokset taulukkoon (esim. Excel), eli
  - nopanheiton silmäluku X ja
  - lasketaan niiden frekvenssi f, eli kuinka monta ykköstä, kakkosta, kolmosta, nelosta, vitosta ja kutosta saadaan (COUNTIF(alue;silmäluku), LASKEJOS(alue; silmäluku) missä silmäluku on 1,2,3,4,5 tai 6 (frekvenssi f)). Tämän jälkeen



KUVA 1. Nopanheittosimulaattorin tuottama kuva.

  - lasketaan $X \cdot f$ jokaiselle eri silmäluvulle ja lopulta kaikkien näiden summa jaettuna heittojen kokonaismäärällä, eli keskiarvo.
- Veratillaan dataa, esim. ryhmän tulos vs. kaikkien ryhmien tulos tai itseheitetyt vs. koneen heittämät (RANDBETWEEN(1;6), SATUNNAISLUKU.VÄLILTÄ(1;6)).
- Lisätehtävä: Opiskelijoiden itse tuottaman satunnaisdatan ja kokeellisen datan vertailu: Opiskelijat listaavat lukuja satunnaisesti, jonka jälkeen tämä aineisto analysoidaan ja verrataan sitä kokeellisesti saatuihin tuloksiin. Millainen jakauma saadaan ja miten tunnusluvut eroavat? Kuinka pitkä on pisin tietyn silmäluvun sarja (esim. neljä kutosta peräkkäin) kokeellisessa datassa? Entä opiskelijoiden tuottamassa datassa?
- Kuvassa 1 esitelty Juha-Matti Huuskon kehittämä nopanheittosimulaattori (englanniksi, saa käyttää ja muokata vapaasti) löytyy sivulta `http://integraali.com/noppa-ja-normaali.html`. Jos heittokertoja on riittävästi, poikkeama 6-tahoisen nopan heittojen odotusarvosta 3,5 on luokkaa yksi per heittokertojen lukumäärän neliöjuuri. Vaativammilla teoreettisilla työkaluilla tämän voi myös perustella, jolloin saadaan suurten lukujen laki

$$\sum_{i=1}^{n} X_i = n\mathbb{E}X + O(\sqrt{n}),$$

josta otoskoolla jakamalla saadaan

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}X + O\left(\frac{1}{\sqrt{n}}\right),$$

missä X_i on heittokerran i tulos, n heittokertojen lukumäärä, $\mathbb{E}X$ satunnaismuuttujan X odotusarvo ja $|O(g(n))| \leq C \cdot g(n)$, missä C on positiivinen vakio ja g jokin funktio. Huomataan, että jälkimmäisessä yhtäsuuruudessa otoskoon kasvaessa virhe pienenee. Lisäksi suurten lukujen laki ja seuraava todistus pätevät tässä muodossa monille muillekin satunnaismuuttujille.

2. Suurten lukujen lain todistus

Huom. käsittelyn tasoa voi säätää opiskelijoiden mielenkiinnon ja perehtyneisyyden mukaan. Olkoot X_i toisistaan riippumattomia satunnaismuuttujia kaikilla indekseillä i. Todistus etenee seuraavasti

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X\right| > \varepsilon\right) &\leq \frac{\mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{n}{n}\mathbb{E}X\right)^2\right)}{\varepsilon^2} \\
&= \frac{1}{n^2\varepsilon^2}\mathbb{E}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X)\right)^2 \\
&= \frac{1}{n^2\varepsilon^2}\sum_{i,j=1}^{n}\mathbb{E}\left[(X_i - \mathbb{E}X)(X_j - \mathbb{E}X)\right] \quad \text{summattava on nolla kun } i \neq j \\
&= \frac{1}{n^2\varepsilon^2}n\mathbb{E}(X - \mathbb{E}X)^2,
\end{aligned}$$

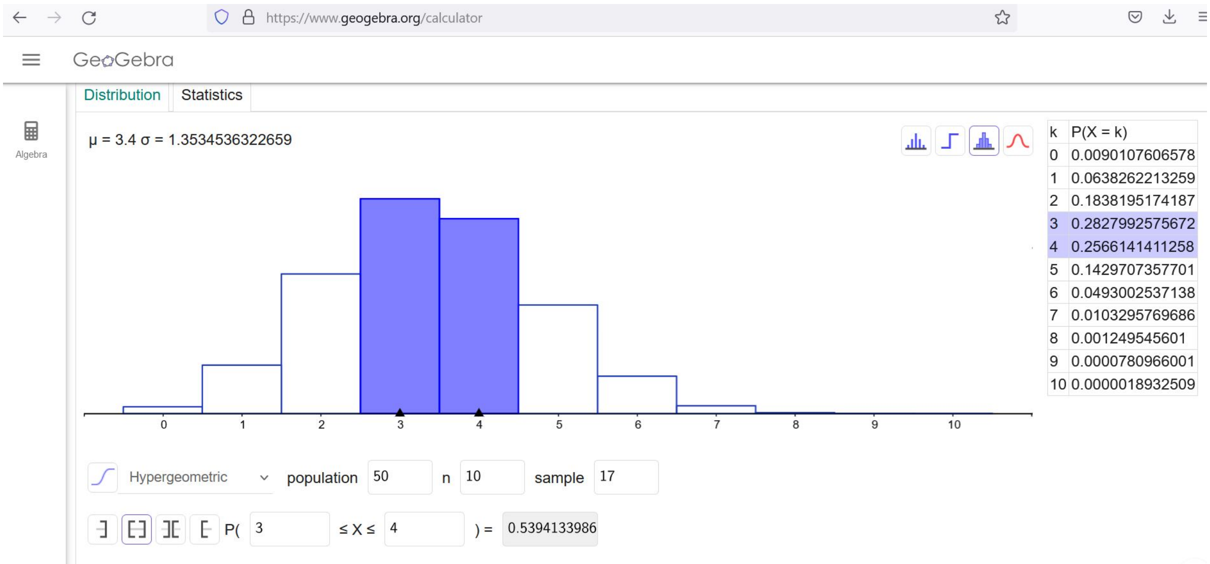
missä ensimmäisellä rivillä käytetään Markovin epäyhtälöä. Toiseksi viimeisellä rivillä taas summattava on nolla, kun $i \neq j$, sillä X_i ja X_j ovat toisistaan riippumattomia, joten $\mathbb{E}(X_iX_j) = (\mathbb{E}X)^2$. Viimeisellä rivillä odotusarvo on jokin vakio, toisinsanoen se ei riipu otoskoosta n. Huomaa, että tarkoituksena on laskea, että todennäköisyys $\mathbb{P}$ on nolla, kun otoskoko n kasvaa rajatta. Tämä pätee, sillä $n/n^2 = 1/n \to 0$, kun $n \to \infty$. □

3. Erilaisten karkkien lukumäärän (hypergeometrinen jakauma) odotusarvon tutkiminen

- Jokaiselle ryhmälle yhteensä 50 esinettä (karkkipussi), joista saa muodostettua kaksi eri kategoriaa tai laatua. Esimerkiksi hedelmä/salmiakki tai hyvä/paha.
- Tutkitaan empiirisesti, missä suhteessa tehdas tuottaa kahta eri kategoriaa. Huom: Pussien sisältöä ei tunneta.
- Jokaisessa ryhmässä nostetaan 10 karkkia pussista, jonka jälkeen lasketaan kuinka monta kumpaakin laatua tuli, eli millainen osuus tyyppiä 1 (salmiakki) ja tyyppiä 2 (hedelmä) saatiin.
- Empiirinen arvio on saatu tutkimalla karkkien suhteita ja mahdollinen virhearvio saadaan Markovin epäyhtälön avulla.
- Lasketaan lopuksi tehdaskeskiarvo, eli miten karkit ovat näissä pusseissa jakautuneet.
- Lisätehtävä: Onko tehdasarvoksi laskettu arvo oikea? Mitkä asiat voivat siihen vaikuttaa?
- Tehtävään löytyy Excel-pohja liitteestä 1.
- Mahdollisia tutkimuskysymyksiä:
  - Kuinka hyvin voidaan päätellä kymmenestä karkista, kuinka paljon koko pussissa on salmiakkeja? Huom. Vertaa koko ryhmän saamia arvoja.
  - Voidaanko päätellä, kuinka paljon salmiakkikarkkeja on yleensä pusseissa, yhdistämällä kaikkien tulokset?
- Tilanne on toistokoe ilman takaisinpanoa, jota mallinnetaan hypergeometrisella jakaumalla. Hypergeometrisen jakauman odotusarvo on
$$E(X) = \frac{nM}{N},$$
missä N on perusjoukon alkioiden lukumäärä (ts. yhden pussin karkkien lukumäärä), M määrätyn osajoukon alkioiden lukumäärä (ts. paljonko esimerkiksi salmiakkia) ja n on otettujen karkkien lukumäärä (nyt 10).

4. Hypergeometrinen jakauman tutkiminen ja visualisointi

Geogebrassa on todennäköisyyslaskuri-työkalu (probability-työkalu). Tehtävään löytyy esimerkkitiedosto, jossa: jakauma on Hypergeometrinen, populaatio 50, $n = 10$ ja otos on 17, eli esimerkiksi salmiakkien lukumäärä koko pussissa. Tuloksena syntyy kuvan 2 kaltainen kuva. Kysymys: Vastaako kuva aiemman empiirisen tutkimuksen tuloksia ja kuinka jakauman muoto muuttuu, jos lukuarvoja muuttaa?



Kuva 2. GeoGebran tuottama kuva.

5. Jatkopohdintoja ja kysymyksiä opiskelijoille

- Mikä merkitys otoksella on saatuun dataan? Onko otoksen ajankohdalla tai muilla taustamuuttujilla vaikutusta saatuihin tuloksiin? Ovatko otokset aidosti riippumattomia?
- Lisätieto: suurten lukujen laki liittyy olennaisesti myös vahvempaan ominaisuuteen, keskeiseen raja-arvolauseeseen, jonka mukaan, kun tarkastellaan riittävän suurta joukkoa sellaisia riippumattomia satunnaismuuttujia, joilla on olemassa odotusarvo ja varianssi, niin tietyillä lisäoletuksilla niiden keskiarvo noudattaa likimain normaalijakaumaa.