Teemu Salonen

MONTY HALLIN ONGELMA

ratkaisu ja siihen liittyvät todennäköisyydet

Tekniikan ja luonnontieteiden tiedekunta
Kandidaatintyö
Elokuu 2020

TIIVISTELMÄ

Teemu Salonen: Monty Hallin ongelma
Kandidaatintyö
Tampereen yliopisto
Tekniikka ja luonnontieteet, TkK
Elokuu 2020

Monty Hallin ongelma on hyvin tunnettu todennäköisyyslaskennan ongelma. Tässä työssä tutustutaan Monty Hallin ongelmaan ja todennäköisyyslaskennan perusteisiin, joita käytetään myöhemmin työssä Monty Hallin ongelman ratkaisemiseen. Työssä sovelletaan myös Monty Hallin ongelmaa kavattamalla ovien määrää. Työn tavoitteena on ratkaista Monty Hallin ongelma useilla erilaisilla todennäköisyyslaskennan menetelmillä, löytää ratkaisut sovellettuihin ongelmiin sekä löytää yleinen ratkaisu usean oven ongelmille.

Työn alkuosassa keskitytään todennäköisyyslaskentaan. Todennäköisyyslaskenta osuudessa käsitellään ensin todennäköisyyslaskennan aksioomia, jotka ovat todennäköisyyslaskennan perusta. Tämän jälkeen käydään läpi todennäköisyyden laskemista klassisen todennäköisyyden ja ehdollisen todennäköisyyden avulla. Tässä työssä ehdolliseen todennäköisyyteen liittyy olennaisesti Bayesin teoreema, jota käytetään myöhemmin Monty Hallin ongelman ratkaisemiseen. Osian lopussa tutustuaan lyhyesti myös todennäköisyyden frekvenssitulkintaan, jota hyödynnetään, kun simuloidaan Monty Hallin ongelmaa ja sen sovelluksia.

Työssä esitellään Monty Hallin ongelma sekä neljä erilaista ratkaisua tähän ongelmaan. Ratkaisuissa hyödynnetään työssä aiemmin käsiteltyjä todennäköisyyslaskennan lauseita, määritelmiä ja menetelmiä. Tärkeitä työkaluja ongalman ratkaisuissa ovat todennäköisyyden klassinen määritelmä, Bayesin teoreema sekä MATLAB-funktioden avulla ongelman simuloiminen.

Työn kolmannessa käsittelyosassa sovelletaan Monty Hallin ongelmaa kasvattamalla ovien määrää. Ensin tarkastellaan tilanteita, joissa ongelmaan on lisätty neljäs ovi, ja ratkaistaan nämä uudet ongelmat käyttäen samoja menetelmiä kuin aiemmin. Ratkaisuissa korostuvat klassinen todennäköisyys ja ongelman simulointi. Tämän jälkeen ovien määrää kasvatetaan kymmeneen ja selvitetään, miten tämä vaikuttaa todennäköisyyksiin. Lopuksi kehitetään yleinen ratkaisu, jolla voidaan ratkaista kaikki usean oven ongelmat.

Avainsanat: Monty Hallin ongelma, todennäköisyyslaskenta, klassinen todennäköisyys, ehdollinen todennäköisyys, Bayesin teoreema

SISÄLLYSLUETTELO

LYHENTEET JA MERKINNÄT

$\mathbb{N}$	luonnolliset luvut
Ω	perusjoukko
$\mathbb{R}$	reaaliluvut
TAU	Tampereen yliopisto (engl. Tampere University)
$P(A)$	todennäköisyys tapahtumalle A

1 JOHDANTO

"Let's Make a Deal" -televisio-ohjelma esitettiin Yhdysvaltojen televisioissa ensimmäistä kertaa joulukuussa vuonna 1963. Tässä Stefan Hatosin ja Monty Hallin kehittämässä kisailuohjelmassa yleisöstä valitut kilpailijat pääsivät ostamaan, myymään ja vaihtamaan kaikkea mahdollista, mitä ohjelman suunnittelijat olivatkaan keksineet. Ohjelma oli hyvin suosittu 1960- ja 1970-luvuilla. Eräs tunnetuimmista ohjelman tapahtumista oli eräänlainen kolmen oven ongelma, joka nimettiin ohjelman juontajan mukaisesti Monty Hallin ongelmaksi. [11]

Monty Hallin ongelma on hyvin kuuluisa todennäköisyyslaskennan ongelma. Siinä kilpailija valitsee yhden kolmesta ovesta, joiden takana on kuitenkin vain yksi palkinto. Tämän jälkeen juontaja avaa toisen jäljelle jääneistä ovista ja antaa kilpailijalle mahdollisuuden vaihtaa ovea toiseen yhä suljettuna olevaan oveen. Tällöin kilpailijan ongelmana on Monty Hallin ongelmana tunnettu päätös siitä, kannattaako vaihtaa ovea vai ei. Työssä tutustutaan todennäköisyyslaskentaan ja ratkaistaan sen avulla Monty Hallin ongelma. Monty Hallin ongelma sekä sen ratkaisu ovat kuitenkin monelle tuttuja. Tästä syystä työssä myös sovelletaan ongelmaa lisäämällä ovia. Työssä ratkaistaan myös nämä sovellukset todennäköisyyslaskennan avulla.

Työssä tutustutaan ensin lyhyesti todennäköisyyslaskentaan ja siihen liittyviin käsitteisiin ja määritelmiin, joita ongelman ratkaisuun tarvitaan. Erityisen tärkeitä käsitteitä ovat klassinen todennäköisyys ja ehdollinen todennäköisyys. Kolmannessa luvussa käsitellään tarkemmin Monty Hallin ongelmaa ja esitetään sen ratkaisu usealla eri tavalla. Neljännessä luvussa käsitellään ongelman sovelluksia ja esitetään niihin ratkaisuja. Viidennessä luvussa kootaan tiivistetysti yhteen käsitellyt ongelmat ja niiden ratkaisut.

2 TODENNÄKÖISYYSLASKENTA

Ennen kuin Monty Hallin ongelmaa voidaan tutkia tarkemmin, on ensin tutustuttava hieman todennäköisyyslaskentaan, jotta voidaan paremmin ymmärtää ongelma ja ratkaista se.

2.1 Todennäköisyyslaskennan aksioomat

Todennäköisyyslaskennassa lähdetään liikkeelle perusjoukosta, jota usein merkitään Ω:lla. Perusjoukon alkioita ovat kaikki mahdolliset yksittäiset tapahtumat. Tapahtuma A voi olla mikä tahansa perusjoukon osajoukko. Todennäköisyyttä tapahtumalle A merkitään $P(A)$.

Todennäköisyydet noudattavat Kolmogorovin aksioomia:

- **Esimmäinen aksiooma** Tapahtuman A todennäköisyys on aina vähintään yhtä suuri kuin nolla ja korkeintaan yhtä suuri kuin yksi eli

$$0 \leq P(A) \leq 1.$$

- **Toinen aksiooma** Perusjoukon todennäköisyys on aina yksi eli

$$P(\Omega) = 1.$$

- **Kolmas aksiooma** Tapahtumat A ja B ovat erillisiä joukkoja, jos molemmat eivät voi tapahtua samanaikaisesti. Tällöin tapahtumien yhdisteen todennäköisyys on tapahtumien todennäköisyyksien summa eli

$$P(A \cup B) = P(A) + P(B). \tag{2.1}$$

Nämä aksioomat ovat todennäköisyyslaskennan perusta. [2]

2.2 Klassinen todennäköisyys

Todennäköisyyksiä käsiteltäessä yleensä halutaan laskea todennäköisyyksiä. Yksinkertainen tapa laskea todennäköisyyksiä on hyödyntää klassista todennäköisyyttä.

Määritelmä 2.1. Olkoon perusjoukko Ω äärellinen joukko alkioita, joiden toteutuminen on yhtä todennäköistä ja olkoon tapahtuma A jokin perusjoukon osajoukoista. Tällöin *todennäköisyyden klassisen määritelmän* mukaan

$$P(A) = \frac{|A|}{|\Omega|}, \tag{2.2}$$

missä $|\Omega|$ on kaikkien alkioiden määrä perusjoukossa ja $|A|$ on niiden perusjoukon alkioiden lukumäärä, jotka toteuttavat tapahtuman A. [10]

Esimerkki 2.2. Heitetään tavallista kuusisivuista noppaa. Tällöin perusjoukko sisältää kaikki mahdolliset yksittäiset tapahtumat eli $\Omega = \{1, 2, 3, 4, 5, 6\}$. Valitaan, että tapahtuma A = ”heitosta saadaan parillinen luku”. Tällöin tapahtuma $A = \{2, 4, 6\}$ ja

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

2.3 Ehdollinen todennäköisyys ja riippuvuudet

Tarkasteltaessa tapahtumien todennäköisyyksiä on tärkeä tietää niiden väliset riippuvuussuhteet. Tapahtumat voivat olla joko toisistaan riippuvia tai riippumattomia.

Määritelmä 2.3. Oletetaan, että tapahtuman B todennäköisyys on suurempi kuin nolla. Tällöin *ehdollinen todennäköisyys* tapahtumalle A sillä ehdolla, että tapahtuma B tapahtuu, määritellään

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{2.3}$$

jossa $P(A \cap B)$ on tapahtumien leikkauksen todennäköisyys eli todennäköisyys sille, että molemmat tapahtumat tapahtuvat. [10]

Määritelmä 2.4. Tapahtumia A ja B kutsutaan toisistaan *riippumattomiksi*, jos

$$P(A \cap B) = P(A)P(B). \tag{2.4}$$

Todistetaan, että tapahtumien leikkaus saadaan lausekkeen (2.4) muotoon, kun tapahtumat ovat toisistaan riippumattomia.

Todistus. Lähdetään liikkeelle ehdollisen todennäköisyyden määritelmästä (Määritelmä 2.3). Lausekkeen (2.3) mukaan

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Tästä saadaan johdettua lauseke tapahtumien leikkauksen todennäköisyydelle, kun kerrotaan lausekkeen (2.3) molemmat puolet tapahtuman B todennäköisyydellä, jolloin saadaan

$$P(A \cap B) = P(A|B)P(B), \tag{2.5}$$

mikä on todennäköisyys tapahtumien leikkaukselle.

Jotta tapahtumat ovat toisistaan riippumattomia, toisen tapahtuman toteutuminen ei voi vaikuttaa toisen tapahtuman todennäköisyyteen. Tällöin voidaan sanoa, että $P(A|B) = P(A)$ [7]. Näin ollen lausekkeeseen (2.5) voidaan sijoittaa $P(A)$ ehdollisen todennäköisyyden paikalle, jolloin saadaan

$$P(A \cap B) = P(A)P(B),$$

mikä on sama, kuin riippumattomuuden määritelmä (Määritelmä 2.4).

$\square$

Lausekkeen (2.4) avulla voidaan helposti laskea todennäköisyyksiä tilanteissa, joissa käsitellään useampaa peräkkäistä toisistaan riippumatonta tapahtumaa. Toisistaan riippuvien tapahtumien kohdalla on käytettävä yleistä muotoa (2.5).

Esimerkki 2.5. Tavallisesta 52 pelikortin korttipakasta nostetaan kaksi korttia. Millä todennäköisyydellä molemmat ovat punaisia, jos a) pelikortit nostetaan yksi kerrallaan siten, että ensiksi nostettu kortti sekoitetaan takaisin pakkaan ennen toisen kortin nostamista? b) molemmat kortit nostetaan yhtä aikaa?

a) Koska pelikortti laitetaan takaisin pakkaan noston jälkeen, nostot ovat toisistaan riippumattomia tapahtumia. Klassisen todennäköisyyden mukaan molemmille nostoille $P(\text{Punainen}) = \frac{26}{52} = \frac{1}{2}$. Tällöin lausekkeen (2.4) mukaan

$$P(\text{Molemmat punaisia}) = P(\text{Punainen})P(\text{Punainen}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

b) Koska pelikorttia ei laiteta takaisin noston jälkeen, jälkimmäisen noston todennäköisyydet riippuvat ensimmäisestä nostosta. Merkitään sitä, että nostaa ensimmäisellä nostolla punaisen kortin tapahtumaksi A ja sitä, että nostaa toisella nostolla punaisen tapahtumaksi B. Tällöin ensimmäisen noston todennäköisyys on edelleen sama eli $P(A) = \frac{1}{2}$. Lasketaan todennäköisyys tapahtumalle B sillä ehdolla, että tapahtuma A tapahtuu. Tällöin pakasta puuttuu yksi punainen kortti eli saadaan $P(B|A) = \frac{25}{51}$. Lausekkeen (2.3) mukaan

$$P(B \cap A) = P(B|A)P(A) = \frac{25}{51} \times \frac{1}{2} = \frac{25}{102} \approx 0.245.$$

2.4 Bayesin teoreema

Bayesin teoreemaa voidaan hyödyntää haluttaessa laskea tiettyjä ehdollisia todennäköisyyksiä.

Lause 2.6. *(Bayesin teoreema) Olkoot erilliset tapahtumat B_1, B_2,...B_n sellaisia, että niiden yhdiste muodostaa koko perusjoukon ja $n \subset N$. Kun $P(A) > 0$, niin kaikille $i = 1, 2, \ldots\ n$*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}. \tag{2.6}$$

Lauseketta (2.6) voidaan yksinkertaistaa muotoon

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}, \tag{2.7}$$

mikä toimii kaikilla $i = 1, 2, \ldots\ n$. Kutsutaan tätä Bayesin teoreeman yksinkertaiseksi muodoksi.

Bayesin teoreemassa todennäköisyyttä $P(B_i)$ kutsutaan prioritodennäköisyydeksi. Prioritodennäköisyys tarkoittaa tapahtuman todennäköisyyttä ennen kuin tiedetään muiden tapahtumien lopputuloksia. Prioritodennäköisyydestä $P(B_i)$ saadaan posterioritodennäköisyys $P(B_i|A)$, kun tilanteesta saadaan lisää tietoa. Posterioritodennäköisyys tarkoittaa tapahtuman todennäköisyyttä sen jälkeen, kun tilanteesta on saatu lisätietoa, joka huomioidaan posterioritodennäköisyyttä laskettaessa. [6]

Esimerkki 2.7. Kahdessa korissa on värillisiä palloja. Ensimmäisessä korissa on kaksi sinistä ja kaksi punaista palloa, kun taas toisessa korissa on kolme punaista palloa ja yksi sininen pallo. Kori valitaan ensin sattumanvaraisesti, jonka jälkeen korista nostetaan pallo sokkona. Millä todennäköisyydellä pallo nostettiin toisesta korista, kun tiedetään, että nostettu pallo on sininen.

- Olkoon tapahtuma A ="*Nostetaan sininen pallo*"ja tapahtuma B ="*Valitaan toinen kori*". Todennäköisyyden klassisen määritelmän (Määritelmä 2.1) avulla saadaan laskettua prioritodennäköisyys

  $$P(B) = \frac{\text{haluttujen korien määrä}}{\text{kaikkien korien määrä}} = \frac{1}{2}.$$

  Tämä on todennäköisyys ennen, kuin tiedetään nostetun pallon olevan sininen.

  Kysytty posterioritodennäköisyys saadaan laskettua Bayesin teoreeman (Lause 2.7) avulla. Ensin tarvitsee kuitenkin laskea todennäköisyydet $P(A)$ ja $P(A|B)$. Tapahtumalle A saadaan laskettua ehdollinen todennäköisyys klassisen todennäköisyyden avulla, kun tiedetään, että pallo nostetaan jälkimmäisestä korista. Tällöin saadaan

  $$P(A|B) = \frac{1}{4}.$$

  Tapahtuman A prioritodennäköisyys saadaan hyödyntämällä todennäköisyyden klassista määritelmää (Määritelmä 2.1), Kolmogorovin kolmatta aksioomaa, eli lauseketta (2.1) sekä lauseketta (2.5), joka on lauseke tapahtumien leikkauksen todennäköisyydelle. Tapahtuman A toteutumiselle saadaan todennäköisyys

  $$P(A) = \frac{1}{2} \times \frac{2}{4} + \frac{1}{2} \times \frac{1}{4} = \frac{3}{8}.$$

  Sijoitetaan lasketut todennäköisyydet lausekkeeseen (2.7), jolloin saadaan

  $$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{\frac{1}{4} \times \frac{1}{2}}{\frac{3}{8}} = \frac{1}{3}.$$

  Tapahtuman B posterioritodennäköisyys on siis $\frac{1}{3}$.

Todistetaan seuraavaksi Bayesin teoreeman molemmat muodot. [10]

Todistus. Lähdetään liikkeelle ehdollisen todennäköisyyden määritelmästä (Määritelmä 2.3) tapahtumille A ja B_i. Lausekkeen (2.3) mukaan

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)}.$$

Sijoitetaan tämän lausekkeen osoittajaan tapahtumien leikkauksen todennäköisyyden lauseke (2.5), jolloin saadaan

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)},$$

mikä on jo sama kuin Bayesin teoreeman yksinkertainen muoto, eli lauseke (2.7). Bauerin ja Buckelin [1] mukaan tapahtuman A todennäköisyys voidaan vielä esittää ehdollisten todennäköisyyksien avulla

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{2.8}$$

Tämä johtuu siitä, että tapahtumat B_1, B_2,...B_n valittiin aiemmin niin, että niiden yhdiste muodostaa koko perusjoukon. Tällöin

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i),$$

mikä voidaan esittää lausekkeen (2.5) mukaan muodossa

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Tällöin päästään lausekkeeseen (2.8). Tämä lauseke voidaan nyt sijoittaa Bayesin teoreeman yksinkertaiseen muotoon, eli lausekkeeseen (2.7), tapahtuman A todennäköisyyden paikalle, jolloin saadaan

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)},$$

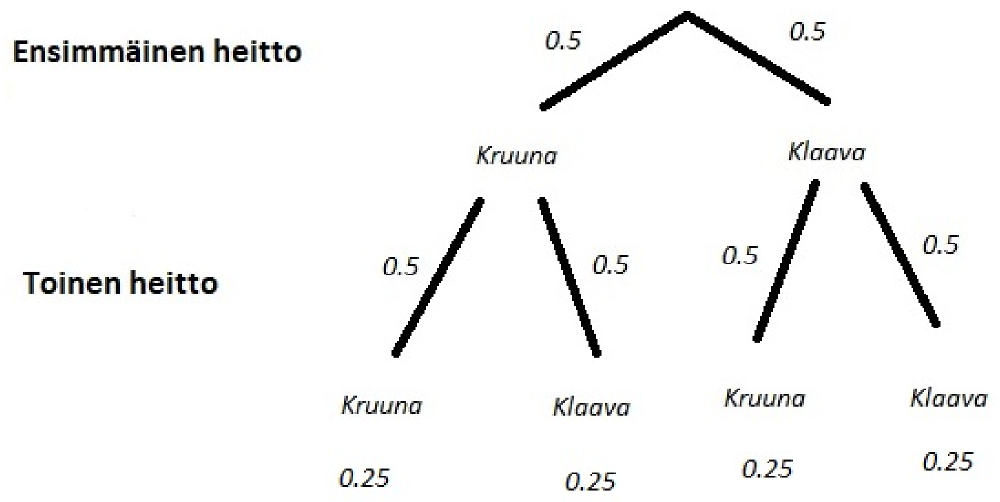
mikä on Bayesin teoreema. [3] $\square$

2.5 Todennäköisyyden puumalli

Todennäköisyyden puumalli on malli, joka havainnollistaa tapauksia, joissa käsitellään useampia tapahtumia ja helpottaa niiden todennäköisyyksien laskemista. Mallin ideana on, että jokaista mahdollista tapahtumasarjaa kuvataan yhtenä oksana. Yhden oksan todennäköisyys voidaan tämän jälkeen laskea käyttämällä lauseketta (2.5). Haluttu todennäköisyys saadaan laskemalla yhteen kaikkien haluttuun lopputulokseen johtavien tapahtumasarjojen todennäköisyydet. Havainnollistetaan puumallia esimerkillä. [8]

Esimerkki 2.8. Heitetään kolikkoa kaksi kertaa. Millä todennäköisyydellä tuloksena saadaan yksi kruuna ja yksi klaava?

- Yhdellä heitolla kruunan ja klaavan todennäköisyydet saadaan klassisen todennäköisyyden määritelmästä (Määritelmä 2.1). Näin saadaan $P(\text{kruuna}) = \frac{1}{2} = 0.5$ ja vastaavasti $P(\text{klaava}) = \frac{1}{2} = 0.5$. Nyt voidaan muodostaa tilannetta kuvaava puumalli. Tämä on esitetty kuvassa 2.1.



Kuva 2.1. *Tilanteesta on muodostettu puumalli, jossa näkyy kaikkien tapahtumasarjojen todennäköisyydet*

Kuten kuvasta 2.1 nähdään, tilanteessa muodostuu neljä oksaa, eli tapahtumasarjaa, joiden todennäköisyydet on saatu lausekkeen (2.5) avulla. Jokaisen oksan todennäköisyys on 0.25 ja näistä haluttu lopputulos toteutuu kahdessa. Tällöin

$$P(\text{yksi kruuna ja yksi klaava}) = 0.25 + 0.25 = 0.5 \quad .$$

2.6 Tilastollinen todennäköisyys

Todennäköisyyttä voidaa käsitellä myös tilastojen avulla. Tässä hyödynnetään todennäköisyyden frekvenssitulkintaa. Frekvenssitulkinnassa käsitellään kokeita, joita toistetaan useita kertoja. Tämän jälkeen kokeesta voidaan laskea suhteellinen frekvenssi jollekkin kokeen tuloksista. [9]

Määritelmä 2.9. Olkoon A jokin tapahtuma toistokokeessa. Tällöin *suhteellinen frekvenssi* $r(A)$ tapahtumalle A voidaan esittää

$$r(A) = \frac{f(A)}{n}, \tag{2.9}$$

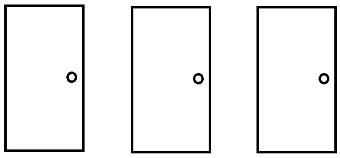
missä n on toistojen lukumäärä ja $f(A)$ on niiden toistojen lukumäärä, jolloin tapahtuma A toteutuu.

Frekvenssitulkinnassa todennäköisyys määritellään suhteellisen frekvenssin raja-arvona. [9] Todennäköisyyttä voidaan siis arvioida suhteellisena frekvenssinä ja lisäämällä kokeen toistojen määrää, sadaan parempi arvio todennäköisyydestä.

3 MONTY HALLIN ONGELMA

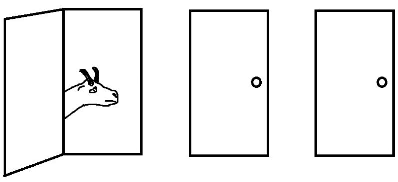
3.1 Ongelman kuvaus

Monty Hallin ongelma on nimetty televisiojuontaja Monty Hallin mukaan. Ongelma on samankaltainen kuin hänen juontamassaan kilpailuohjelmassa nimeltä "Let's make a deal". Ohjelmassa kilpailija sai tehtäväkseen valita yhden kolmesta suljetusta ovesta, kuten kuvassa 3.1. Kahden oven takana oli vuohi ja yhden oven takana palkinto, eli auto. Kilpailija ei tiedä, mitä minkäkin oven takana on, mutta juontaja tietää. [4]



Kuva 3.1. *Kolme ovea, joista kilpailija valitsee yhden.*

Kun kilpailija on valinnut oven, juontaja avaa toisen ovista, joita kilpailija ei valinnut. Juontajan on aina valittava ovi, jonka takana on vuohi ja jos molempien ovien takana on vuohi, hän valitsee oven satunnaisesti. Tätä tilannetta havainnollistaa kuva 3.2.



Kuva 3.2. *Juontaja avaa oven, jonka takana on vuohi.*

Tämän jälkeen kilpailija saa halutessaan vaihtaa toiseen suljettuna olevaan oveen. Tätä valintaa vaihtamisen ja aluksi valitsemansa oven pitämisen välillä kutsutaan Monty Hallin ongelmaksi.

3.2 Ongelman ratkaisuja

Seuraavaksi yritetään ratkaista Monty Hallin ongelma aiemmin käsitellyillä menetelmillä, kuten esimerkiksi klassisen todennäköisyyden avulla ja Bayesin teoreeman avulla.

3.2.1 Ongelman ratkaisu klassisen todennäköisyyden avulla

Ongelma voidaan ratkaista yksinkertaisesti soveltamalla klassista todennäköisyyttä. Klassisen todennäköisyyden avulla voidaan laskea todennäköisyys sille, että palkinto on kilpailijan valitseman oven takana. Ovia on kolme, joten klassisen todennäköisyyden määritelmän (Määritelmä 2.1) mukaan

$$P(\text{Palkinto kilpailijan valitseman oven takana}) = \frac{\text{valittujen ovien määrä}}{\text{kaikkien ovien määrä}} = \frac{1}{3}.$$

Vastaavasti saadaan laskettua

$$P(\text{Palkinto jonkun muun oven takana}) = \frac{2}{3}.$$

Valinnan jälkeen juontaja avaa toisen ovista, joita ei valittu, ja paljastaa vuohen sen takaa. Tämä ei muuta kilpailijan valitseman oven todennäköisyyttä, vaikka näin voisikin luulla. Tällöin kilpailijan valitseman oven voittotodennäköisyys on yhä $\frac{1}{3}$. Kahden valitsematta jääneen oven todennäköisyys on aluksi $\frac{2}{3}$. Myös tämä todennäköisyys säilyy vaikka toinen ovista avataan. Koska tiedetään, että palkinto ei voi olla avatun oven takana, on toisen oven todennäköisyyden oltava $\frac{2}{3}$, sillä tämä on myös ovien yhteenlaskettu voittotodennäköisyys. Näin ollen siis oven vaihtaminen kannattaa.

3.2.2 Ongelman ratkaisu Bayesin teoreeman avulla

Ratkaistaan Monty Hallin ongelma hyödyntäen Bayesin teoreemaa ja erityisesti sen yksinkertaista muotoa, lauseketta (2.7). Oletaan, että kilpailija valitsee ensimmäisen oven. Jotta Bayesin teoreemaa voidaan käyttää, tarvitaan tapahtumat A ja B. Määritellään $B =$"*Auto on ensimmäisen oven takana*" ja $A =$"*Juontaja avaa kolmannen oven*". Klassisen todennäköisyyden määritelmän (Määritelmä 2.1) avulla voidaan laskea tarvittavat todennäköisyydet, jotta voidaan hyödyntää Bayesin teoreemaa. Todennäköisyys tapahtumalle B voidaan laskea

$$P(B) = \frac{1}{3},$$

sillä ovia on kolme ja kaikki näistä ovat alussa yhtä todennäköisiä. [5]

Tapahtuman A todennäköisyys riippuu siitä, minkä oven takana auto on. Jos auto on toisen oven takana, juontaja avaa varmasti kolmannen oven, sillä hän ei voi avata kilpailijan ovea. Jos auto on kolmannen oven takana, juontaja ei voi avata sitä, sillä juontajan on avattava ovi, jonka takana on vuohi. Jos auto on ensimmäisen oven takana, juontaja valitsee satunnaisesti joko toisen tai kolmannen oven. Kuten aiemmin laskettiin, auton todennäköisyys olla tietyn oven takana on $\frac{1}{3}$.

Tällöin todennäköisyys tapahtumalle A voidaan laskea

$$P(A) = \frac{1}{3} \times 1 + \frac{1}{3} \times 0 + \frac{1}{3} \times \frac{1}{2} = \frac{1}{2}.$$

Lopuksi lasketaan todennäköisyys tapahtumalle A, sillä ehdolla, että tapahtuma B tapahtuu. Tällöin auto on kilpailijan valitseman oven takana ja juontaja valitsee satunnaisesti toisen kahdesta ovesta, jolloin

$$P(A|B) = \frac{1}{2}.$$

Lasketut todennäköisyydet voidaan sijoittaa lausekkeeseen (2.7), jolloin saadaan

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

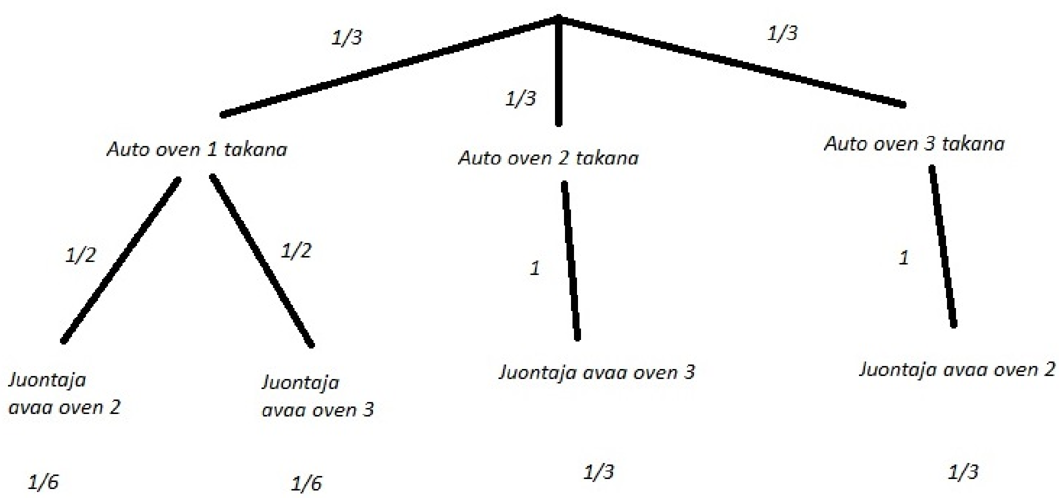
Nyt tiedetään todennäköisyys, että auto on kilpailijan valitseman oven takana juontajan avattua kolmannen oven. Koska suljettuja ovia on enää vain kaksi, Kolmogorovin toisen aksiooman perusteella toisen oven todennäköisyys voittaa on

$$1 - \frac{1}{3} = \frac{2}{3}$$

ja tästä syystä oven vaihtaminen kannattaa. [5]

3.2.3 Ongelman ratkaisu puumallin avulla

Ongelma voidaan myös käymällä läpi kaikki mahdolliset lopputulokset ja niiden todennäköisyydet. Tässä voimme käyttää apuna puumallia 2.5. Puumalliin vaadittavat todennäköisyydet laskettiin jo edellisessä ratkaisussa 3.2.2, joten niitä ei tarvitse laskea enää uudelleen. Kuvassa 3.3 on esitetty puumalli Monty Hallin ongelman tilanteesta hyödyntäen aiemmin laskettuja todennäköisyyksiä.



Kuva 3.3. *Tilanteesta on muodostettu puumalli, jossa näkyy kaikkien tapahtumasarjojen todennäköisyydet*

Kuvasta 3.3 nähdään, että kahdessa lopputuloksessa vaihtamalla voittaa, kun taas kahdessa tapauksessa, joissa auto on ensimmäisen oven takana, häviää. Todennäköisyys, että vaihtamalla voittaa

$$P(\text{Vaihtamalla oven voittaa}) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

Vastaavasti todennäköisyys, että voittaa pitämällä alussa valitsemansa oven

$$P(\text{Pitämällä oven voittaa}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

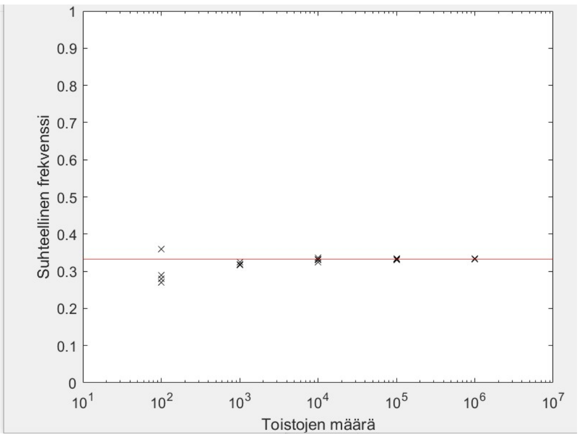
Tästä nähdään, että oven vaihtaminen kannattaa.

3.2.4 Ongelman ratkaisu simuloimalla

Ongelma voidaan ratkaista simuloimalla. Kun tilannetta simuloidaan monta kertaa, todennäköisyyden frekvenssitulkinnan mukaan tapahtuman suhteellinen frekvenssi lähestyy sen todennäköisyyttä. Tilannetta sovelletaan MATLAB-ohjelmalla. Simulointia varten on luotu kaksi funktiota, joista toinen simuloi tilanteen yhden kerran ja toinen toistaa tätä tilannetta, sekä esittää sen tulokset. Nämä simulointiin käytetyt MATLAB-funktiot löytyvät liitteestä A.

Simulointi tehtiin usealla eri toistomäärällä ja jokaisella toistomäärällä tehtiin viisi toistoa, jotta voidaan samalla havainnollistaa, kuinka toistomäärien kasvattaminen pienentää tulosten hajontaa. Ensin simulointiin tilannetta, jossa kilpailija ei halua vaihtaa ovea. Simuloinnin tulokset on esitetty kuvassa 3.4. Kuvassa nähdään numeroarvot suhteellisista frekvensseistä, sekä kuvaaja, jossa suhteelliset frekvenssit ovat esitettynä toistokertojen funktiona. Lisäksi kuvaajassa on punaisella esitetty jo aiemmin klassisella tavalla laskettu todennäköisyys.

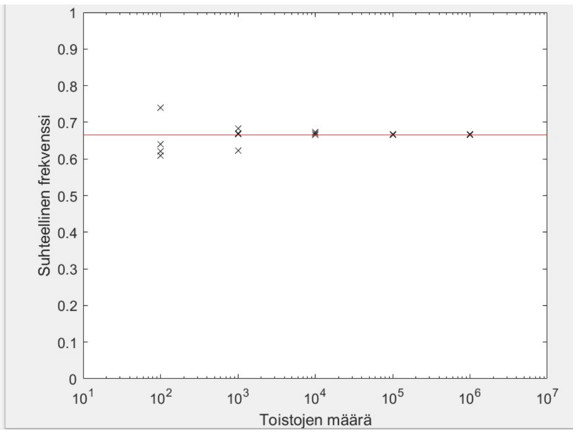
```
>> simulointi(3,1,0,5)
Sadalla toistolla suhteelliset frekvenssit olivat:
    0.3600    0.2700    0.2900    0.2800    0.3600

Tuhannella toistolla suhteelliset frekvenssit olivat:
    0.3190    0.3160    0.3190    0.3240    0.3190

Kymmenellätuhannella toistolla suhteelliset frekvenssit olivat:
    0.3301    0.3355    0.3248    0.3315    0.3318

Sadallatuhannella toistolla suhteelliset frekvenssit olivat:
    0.3345    0.3305    0.3328    0.3339    0.3334

Miljoonalla toistolla suhteelliset frekvenssit olivat:
    0.3332    0.3333    0.3334    0.3332    0.3329

>>
```



Kuva 3.4. *Simuloinnin tulokset, kun kilpailija ei vaihda ovea.*

Kuten kuvasta 3.4 nähdään, toistojen määrän kasvaessa suhteellinen frekvenssi lähestyy aiemmin laskettua todennäköisyyttä $\frac{1}{3}$. Kuvassa 3.5 on esitetty simuloinnin tulokset tilanteessa, jossa kilpailija vaihtaa ovea.

```
>> simulointi(3,1,1,5)
Sadalla toistolla suhteelliset frekvenssit olivat:
    0.7400    0.6400    0.6400    0.6200    0.6100

Tuhannella toistolla suhteelliset frekvenssit olivat:
    0.6670    0.6230    0.6830    0.6840    0.6690

Kymmenellätuhannella toistolla suhteelliset frekvenssit olivat:
    0.6697    0.6693    0.6731    0.6654    0.6743

Sadallatuhannella toistolla suhteelliset frekvenssit olivat:
    0.6657    0.6648    0.6660    0.6661    0.6681

Miljoonalla toistolla suhteelliset frekvenssit olivat:
    0.6665    0.6655    0.6658    0.6664    0.6672

>>
```



Kuva 3.5. *Simuloinnin tulokset, kun kilpailija vaihtaa oven.*

Myös tässä tilanteessa suhteellinen frekvenssi lähenee aiemmin laskettua todennäköisyyttä $\frac{2}{3}$, kun toistoja on riittävän paljon. Näin ollen myös simuloinnin perusteella voidaan sanoa, että vaihtaminen kannattaa.

4 MONTY HALLIN ONGELMAN SOVELLUKSIA

Monty Hallin ongelmassa tarkastellaan voiton todennäköisyyksiä tilanteessa, jossa ovia on kolme. Vastaavan tyyppiseen ongelmaan on mahdollista törmätä jossakin muussa yhteydessä ja tällöin tilanne voi poiketa jonkin verran Monty Hallin ongelmasta. Tarkastellaan seuraavaksi todennäköisyyksiä tilanteissa, joissa on kasvatettu ovien määrää Monty Hallin ongelman tilanteesta.

4.1 Neljän oven ongelma

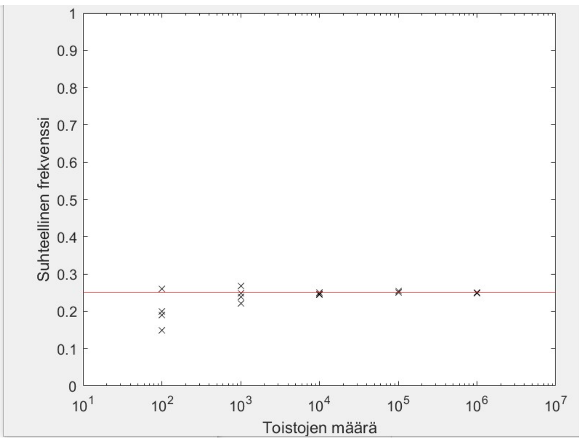
Kun aiemmin käsiteltyyn tilanteeseen lisätään neljäs ovi, voi tilanteessa muuttua muutakin. Avaako juontaja nyt molemmat ovet, joiden takana ei ole palkintoa vai avataanko edelleen vain yksi ovi, jolloin kilpailijan on ovea vaihdettaessa valittava toinen kahdesta jäljellä olevasta ovesta?

Tarkastellaan ensin tilannetta, jossa kilpailija ei vaihda ovea. Todennäköisyyden klassisen määritelmän (Määritelmä 2.1) avulla saadaan todennäköisyys, että palkinto on kilpailijan valitseman oven takana

$$P(\text{Palkinto kilpailijan valitseman oven takana}) = \frac{\text{valittujen ovien määrä}}{\text{kaikkien ovien määrä}} = \frac{1}{4}.$$

Todennäköisyys tälle tapahtumalle ei muutu ovia avattaessa, joten vaihtamatta ovea todennäköisyys voittaa on $\frac{1}{4}$. Tämä voidaan osoittaa myös simuloimalla, kuten osiossa 3.2.4. Simulointiin käytetyt MATLAB-funktiot A toimivat myös usean oven tilanteessa. Kuvassa 4.1 on esitetty tulokset kun simuloidaan tilannetta, jossa on neljä ovea ja kilpailija ei halua vaihtaa oveaan.

```
>> simulointi(4,2,0,5)
Sadalla toistolla suhteelliset frekvenssit olivat:
    0.2600    0.2600    0.2000    0.1500    0.1900

Tuhannella toistolla suhteelliset frekvenssit olivat:
    0.2480    0.2210    0.2690    0.2480    0.2380

Kymmenellätuhannella toistolla suhteelliset frekvenssit olivat:
    0.2454    0.2504    0.2502    0.2460    0.2439

Sadallatuhannella toistolla suhteelliset frekvenssit olivat:
    0.2501    0.2506    0.2497    0.2535    0.2538

Miljoonalla toistolla suhteelliset frekvenssit olivat:
    0.2498    0.2501    0.2493    0.2504    0.2501

>>
```



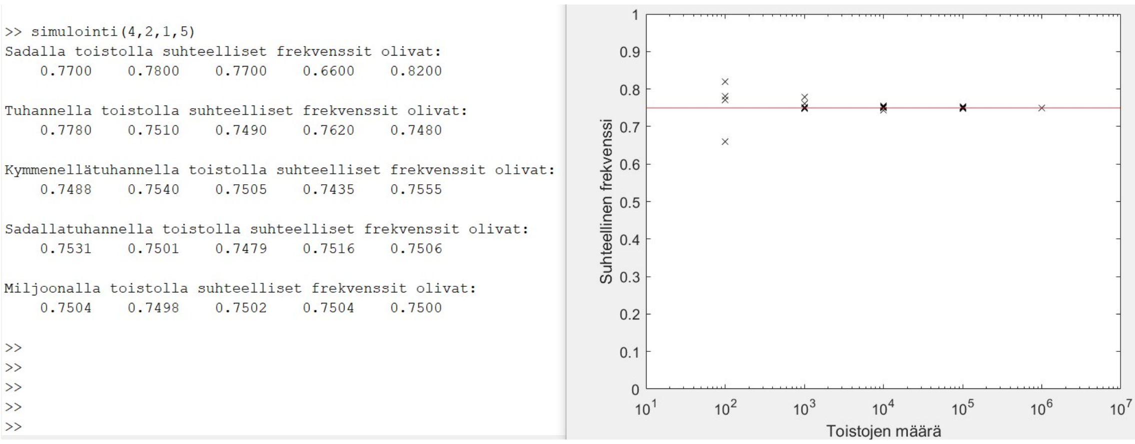
Kuva 4.1. *Simuloinnin tulokset neljällä ovella, kun kilpailija ei vaihda ovea.*

Kuvasta 4.1 nähdään, että simuloimalla päädytään samaan tulokseen $\frac{1}{4}$.

Kilpailijan halutessa vaihtaa ovea, avattavien ovien määrä vaikuttaa todennäköisyyteen. Käsitellään ensin tilannetta, jossa molemmat mahdollisista avattavista ovista avataan. Todennäköisyys, että auto on jonkun muun, kuin kilpailijan valitseman oven takana on

$$P(\text{Palkinto jonkun muun oven takana}) = \frac{\text{ei valittujen ovien määrä}}{\text{kaikkien ovien määrä}} = \frac{3}{4}.$$

Kun näistä kolmesta ovesta kaksi avataan, palkinto ei voi olla kummankaan niistä takana. Tällöin jäljelle jäävä avaamaton ovi saa koko todennäköisyyden $\frac{3}{4}$, eli tässä tilanteessa vaihtaminen on vieläkin kannattavampaa kuin kolmen oven tapauksessa. Kuvassa 4.2 on esitetty tulokset, kun simuloidaan tilannetta, jossa kilpailija vaihtaa ovea ja kaksi ovea avataan.



Kuva 4.2. *Simuloinnin tulokset neljällä ovella, joista kaksi avataan, ja kilpailija vaihtaa ovea.*

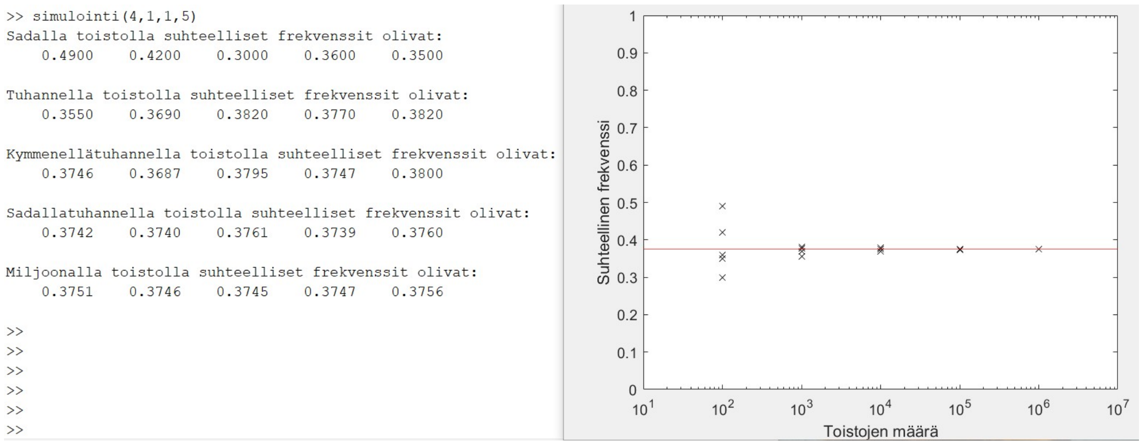
Kuvasta 4.2 nähdään, että suurilla toistomäärillä tehdyn simuloinnin tulokset vastaavat hyvin laskettuja todennäköisyyksiä.

Tarkastellaan vielä lopuksi tilannetta, jossa vain yksi ovi avataan. Tässä tilanteessa kolmelle ovelle, joita kilpailija ei valitse saadaan jälleen todennäköisyys $\frac{3}{4}$. Nyt ovia kuitenkin avataan vain yksi, jolloin jäljelle jää kaksi ovea, joihin kilpailija voi vaihtaa alussa valitsemansa oven. Alussa saatu todennäköisyys $\frac{3}{4}$ jakautuu nyt tasan kahdelle jäljellä olevalle ovelle, jolloin voiton todennäköisyydeksi saadaan

$$P(\text{kilpailija voittaa vaihtamalla ovea}) = \frac{\frac{3}{4}}{2} = \frac{3}{8}$$

.

Tämä todennäköisyys on yhä suurempi, kuin tilanteessa, jossa kilpailija ei vaihda ovea, eli vaihtaminen kannattaa Tässä tilanteessa ero on kuitenkin huomattavasti huomattavasti pienempi, kuin tilanteessa jossa avattiin kaksi ovea. Kuvassa 4.3 on esitetty tulokset, kun simuloidaan tätä tilannetta.



Kuva 4.3. *Simuloinnin tulokset neljällä ovella, joista avataan yksi, ja kilpailija vaihtaa ovea.*

Näiden kolmen käsitellyn tilanteen perusteella voidaan sanoa, että vaihtaminen kannattaa aina myös neljän oven ongelmassa. Kun ovia avattiin kaksi, saatiin vielä suurempi todennäköisyys voittaa vaihtamalla, kuin kolmen oven tapauksessa, mutta avattaessa vain yksi ovi, vaihtamisen todennäköisyys oli huomattavasti pienempi.

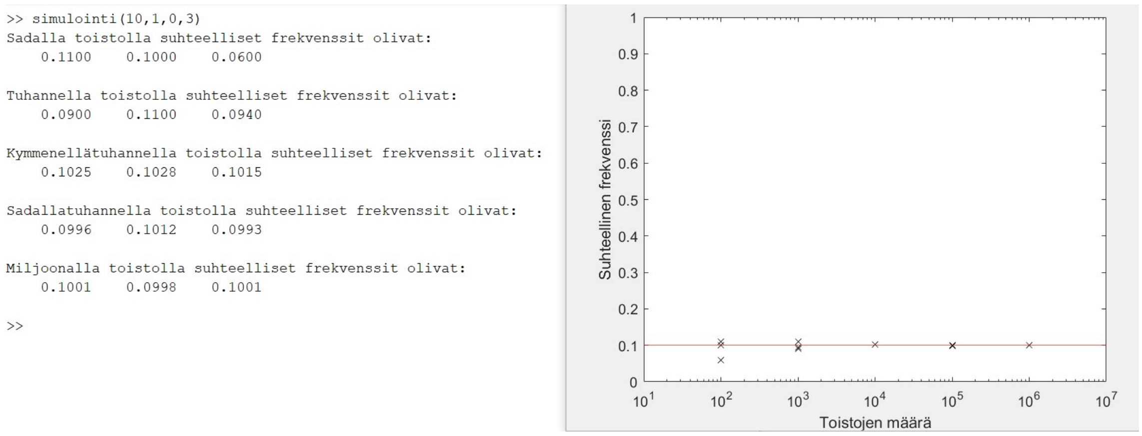
4.2 Kymmenen oven ongelma

Äsken käsiteltiin neljän oven ongelmaa. Mitä tapahtuu todennäköisyyksille, jos ovien määrää kasvatetaankin kymmeneen? Kuten jo neljän oven ongelmassa huomattiin, muodostuu useita erilaisia tilanteita, kun muutetaan avattavien ovien määrää. Kymmenen oven ongelmassa erilaisi tapauksia muodostuu jo niin monta, ettei kaikkia kannata erikseen tarkastella. Otetaan tarkasteluun kolme erilaista tilannetta.

Ensimmäisessä tilanteessa kilpailija ei vaihda ovea. Tähän saadaan laskettua todennäköisyys todennäköisyyden klassisen määritelmän (Määritelmä 2.1) avulla. Kilpailija voittaa vain, jos palkinto on kilpailijan valitseman oven takana. Tälle saadaan todennäköisyys

$$P(\text{Palkinto kilpailijan valitseman oven takana}) = \frac{\text{valittujen ovien määrä}}{\text{kaikkien ovien määrä}} = \frac{1}{10}.$$

Tilannetta on myös simuloitu MATLAB-funktioiden A avulla. Aiemmasta poiketen, kymmenen oven ongelman tilanteissa toistoja tehtiin vain kolme kertaa, sillä funktiot toimivat hitaammin, kun ovien määrää kasvatetaan. Simuloinnin tulokset ovat esitettynä kuvassa 4.4.

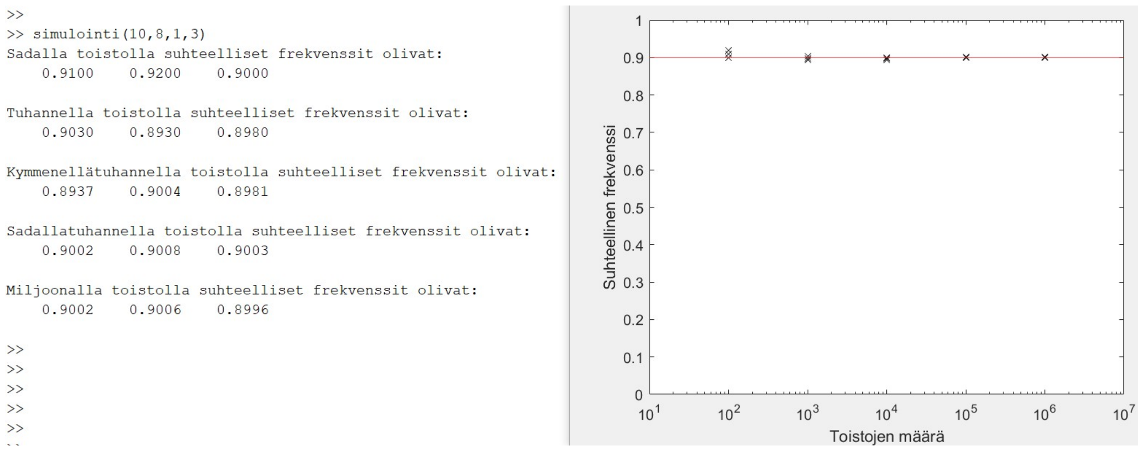


Kuva 4.4. *Simuloinnin tulokset kymmenellä ovella, kun kilpailija ei vaihda ovea.*

Kun ovien määrä kasvaa, todennäköisyys voittaa ilman vaihtamista pienenee. Tarkastellaan seuraavaksi kahta tilannetta, joissa kilpailija haluaa vaihtaa ovea. Todennäköisyys, että palkinto on on jonkun muun kuin kilpailijan valitseman oven takana, saadaan todennäköisyyden klassisen määritelmän (Määritelmä 2.1) avulla

$$P(\text{Palkinto jonkun muun oven takana}) = \frac{\text{ei valittujen ovien määrä}}{\text{kaikkien ovien määrä}} = \frac{9}{10}.$$

Tarkastellaan ensin tilannetta, jossa juontaja avaa kahdeksan ovea, eli suurimman määrän, jota hän voi avata kymmenen oven ongelmassa. Ovien avaamisen jälkeen jäljelle jää vain yksi ovi kilpailijan alussa valitseman oven lisäksi. Tämä ovi saa nyt koko todennäköisyyden $\frac{9}{10}$, joka oli muilla kuin kilpailijan valitsemalla ovella. Simuloidaan tätä tilannetta samalla tavalla kuin aiemminkin, jolloin saadaan kuvan 4.5 mukaiset tulokset.



Kuva 4.5. *Simuloinnin tulokset kymmenellä ovella, joista kahdeksan avataan ja kilpailija vaihtaa ovea.*

Kuten kuvasta 4.5 huomataan, myös simuloimalla päädytään samaan tulokseen $\frac{9}{10}$. Tämän perusteella voidaan sanoa, että tässäkin tilanteessa vaihtaminen kannattaa.

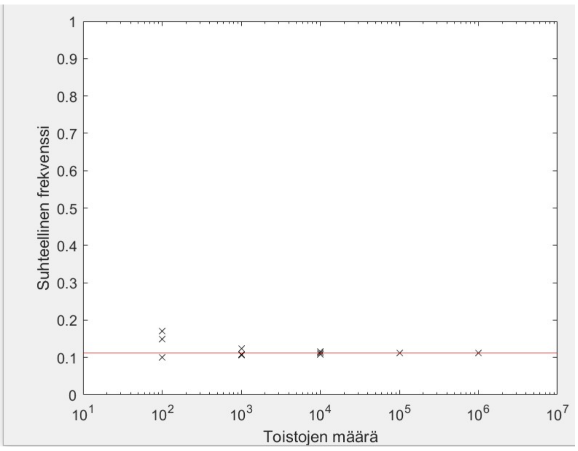
Viimeisenä tilanteena kymmenen oven ongelmasta käsitellään tilanne, jossa juontaja avaa vain yhden oven. Oven avaamisen jälkeen on vielä kahdeksan ovea, joihin kilpailija voi vaihtaa alussa valitsemansa oven. Aiemmin muille kuin kilpailijan valitsemalle ovelle laskettu todennäköisyys

$\frac{9}{10}$ jaetaan nyt tasan jäljellä olevien kahdeksan oven kesken. Tällöin todennäköisyys, että kilpailija voittaa vaihtamalla ovea on

$$P(\text{kilpailija voittaa vaihtamalla ovea}) = \frac{\frac{9}{10}}{8} = \frac{9}{80} = 0.1125.$$

Kuvassa 4.6 on esitetty tulokset, kun simuloidaan kyseistä tilannetta.

```
>>
>> simulointi(10,1,1,3)
Sadalla toistolla suhteelliset frekvenssit olivat:
    0.1700    0.1000    0.1500

Tuhannella toistolla suhteelliset frekvenssit olivat:
    0.1230    0.1080    0.1060

Kymmenellätuhannella toistolla suhteelliset frekvenssit olivat:
    0.1081    0.1113    0.1169

Sadallatuhannella toistolla suhteelliset frekvenssit olivat:
    0.1121    0.1122    0.1116

Miljoonalla toistolla suhteelliset frekvenssit olivat:
    0.1123    0.1128    0.1120

>>
>>
>>
>>
>>
```



Kuva 4.6. *Simuloinnin tulokset kymmenellä ovella, joista yksi avataan ja kilpailija vaihtaa ovea.*

Kuvan 4.6 perusteella nähdään, että myös simuloimalla päädytään samaan tulokseen. Saatu todennäköisyys on jälleen suurempi, kuin kilpailijan alussa valitsemalla ovella, joten vaihtaminen kannattaa. Nyt todennäköisyydet ovat kuitenkin hyvin lähellä toisiaan, joten vaihtamisella saavutettava etu ei ole enää kovin suuri.

4.3 Ongelman yleinen ratkaisu

Kuten aiemmin huomattiin, muuttamalla ovien kokonaismäärää sekä avattavien ovien määrää, saadaan muodostettua useita erilaisia tilanteita. Ei ole kuitenkaan järkevää käydä jokaista yksittäistä tilannetta erikseen läpi, vaan on järkevämpää yrittää löytää ongelmalle yleinen ratkaisu, jolla saadaan ratkaistua todennäköisyys jokaisessa tilanteessa. Tilanteille, joissa kilpailija haluaa vaihtaa ovea sekä tilanteille, joissa kilpailija ei halua vaihtaa ovea, joudutaan kuitenkin ratkaisemaan erilliset ratkaisut.

Lähdetään liikkeelle merkinnöistä. Merkitään ovien määrää kirjaimella n ja avattavien ovien määrää kirjaimella m. Käsitellään ensin tilannetta, jossa kilpailija ei vaihda ovea. Tällöin kilpailija voittaa vain jos palkinto on hänen valitsemansa oven takana. Tälle tapahtumalle saadaan todennäköisyyden klassisen määritelmän (Määritelmä 2.1) mukaan todennäköisyys

$$P(\text{Palkinto kilpailijan valitseman oven takana}) = \frac{\text{valittujen ovien määrä}}{\text{kaikkien ovien määrä}} = \frac{1}{n}. \tag{4.1}$$

Tämä on lauseke kilpailijan voiton todennäköisyydelle tilanteissa, jossa kilpailija ei vaihda ovea.

Seuraavaksi käsitellään tilannetta, jossa kilpailija vaihtaa ovea. Todennäköisyys, että palkinto on jonkun muun kuin kilpailijan valitseman oven takana saadaan vastaavasti

$$P(\text{Palkinto jonkun muun oven takana}) = \frac{\text{ei valittujen ovien määrä}}{\text{kaikkien ovien määrä}} = \frac{n-1}{n}.$$

Tämän jälkeen ovia avataan m kappaletta. Ovien avauksen jälkeen on oltava vähintään kaksi avaamatonta ovea, joista toinen on kilpailijan alussa valitsema ovi. Voidaan siis sanoa, että $m \leq n-2$. Kun ovet on avattu, valitsematta jääneillä ovilla ollut todennäköisyys jakautuu tasan jäljellä olevien suljettujen valitsematta jääneiden ovien kesken, kun taas kilpailijan valitseman oven todenäköisyys pysyy samana. Kun vähennetään alussa valittu ovi sekä avatut ovet, ovia, joihin kilpailija voi vaihtaa ovensa, on jäljellä $n-m-1$ kappaletta. Tällöin ovea vaihtamalla voittamisen todennäköisyydeksi saadaan

$$P(\text{kilpailija voittaa vaihtamalla ovea}) = \frac{\frac{n-1}{n}}{n-m-1} = \frac{n-1}{n(n-m-1)}. \tag{4.2}$$

Tämä on yleinen muoto voittamisen todennäköisyydelle tilanteissa, jossa kilpailija vaihtaa ovea. Nyt on vielä selvitettävä, onko todennäköisyys vaihtamalla ovea aina suurempi, kuin todennäköisyys voittaa silloin kun ovea ei vaihdeta. Oletetaan, että todennäköisyys on suurempi, kun ovea vaihdetaan, ja yritetään todistaa tämä oletus.

Todistus. Aloitetaan merkitsemällä että, lauseke (4.1) on pienempää kuin lauseke (4.2)

$$\frac{1}{n} < \frac{n-1}{n(n-m-1)}.$$

Muutetaan lausekkeet samannimisiksi laventamalla vasenta puolta

$$\frac{(n-m-1)}{n(n-m-1)} < \frac{n-1}{n(n-m-1)}.$$

Koska n sekä m ovat positiivisia kokonaislukuja ja $m \leq n-2$, nimittäjä on aina vähintään yksi. Epäyhtälön molemmat puolet ovat samannimisiä, joten voidaan verrata vain osoittajia, jolloin epäyhtälö saadaan muotoon

$$n-m-1 < n-1.$$

Vähennetään molemmilta puolilta $n-1$, jolloin saadaan

$$-m < 0.$$

Koska avattavien ovien lukumäärän on oltava positiivinen kokonaisluku, epäyhtälö on tosi. Ovea kannattaa siis aina vaihtaa, sillä tällöin voittaminen on todennäköisempää kaikissa tilanteissa riippumatta ovien tai avattavien ovien määrästä. $\square$

5 YHTEENVETO

Tässä työssä käsiteltiin Monty Hallin ongelmaa ja sen ratkaisua. Monty Hallin ongelmaa varten käytiin läpi todennäköisyyslaskennan perusteita, joita käytettiin työkaluna ongelman ratkaisemiseksi. Työssä käytiin läpi todennäköisyyslaskentaa riittävässä määrin, jotta Monty Hallin ongelma saatiin ratkaistua usealla eri tavalla kuten työn tarkoituksena oli. Ratkaisuissa hyödynnettiin erityisesti klassista todennäköisyyttä, Bayesin teoreemaa sekä ongelman simulointia MATLAB-funktioilla. Kaikilla menetelmillä päädyttiin tulokseen, että ovea kannattaa aina vaihtaa, sillä tällöin todennäköisyys voittaa kaksinkertaistuu.

Tämän lisäksi Monty Hallin ongelmaa sovellettiin kasvattamalla ovien määrää. Nämäkin ongelmat ratkaistiin todennäköisyyslaskennan avulla. Neljän ja kymmenen oven ongelmia ratkaistiin hyödyntämällä klassista todennäköisyyttä ja simulointia. Näissäkin ongelmissa todettiin oven vaihtamisen olevan kannattavaa. Osassa tilanteista vaihtamalla todennäköisyys voittaa saattoi kasvaa moninkertaisesti, kun taas osassa tilanteista ero oli hyvinkin pieni. Huomattiin, että tämä riippui ovien kokonaismäärän lisäksi myös avattavien ovien määrästä.

Eräs työn tärkeimmistä asioista Monty Hallin ongelman ratkaisujen ohella oli yleisen ratkaisun muodostaminen Monty Hallin ongelmaa vastaaville usean oven ongelmille. Monty Hallin ongelma on tunnettu todennäköisyyslaskennan ongelma ja samalla myös sen ratkaisu on monelle tuttu. Kuitenkin harva ihminen kohtaa elämässään vastaavaa tilannetta, jossa hän tekee valinnan kolmen oven välillä niin, että mahdollisuutena on voittaa auto vuohen sijaan. Vastaavan tyyppiseen ongelmaan voi kuitenkin törmätä tai sen voi itse muodostaa esimerkiksi pelikorteilla. Kun ongelma otetaan pois sen alkuperäisestä kontekstistaan, voivat myös muut asiat muuttua ongelmassa. Jos ongelma toteutetaan esimerkiksi pelikorteilla havainnollistaen, voi korttien määrä muuttua esimerkiksi viiteen. Tällöin yleinen ratkaisu on apuna vastaavissa ongelmissa. Vaikka lauseketta yleiselle ratkaisulle ei muistaisikaan, monesti riittää kun muistaa, että oven tai esimerkiksi juuri pelikortin vaihtaminen kasvattaa todennäköisyyttä voittaa tämän tyyppisissä ongelmissa.

LÄHTEET

[1] H. Bauer ja R. B. Buckel. *Probability theory*. Walter de Gruyter, 1996.

[2] C. A. Gorini. *Master Math : Probability*. Cengage Learning, 2011.

[3] A. Klenke. *Probability Theory A Comprehensive Course*. Springer, 2014.

[4] *Monty Hall Problem*. URL: http://mathworld.wolfram.com/MontyHallProblem.html (viitattu 08. 05. 2020).

[5] *Monty Hall Problem: Solution Explained Simply*. URL: https://www.statisticshowto.com/probability-and-statistics/monty-hall-problem/ (viitattu 21. 06. 2020).

[6] *Probability concepts explained: Bayesian inference for parameter estimation.* URL: https://towardsdatascience.com/probability-concepts-explained-bayesian-inference-for-parameter-estimation-90e8930e5348 (viitattu 30. 07. 2020).

[7] *probability theory*. 2020. URL: https://academic-eb-com.libproxy.tuni.fi/levels/collegiate/article/probability-theory/109439 (viitattu 29. 04. 2020).

[8] *Probability Tree Diagrams: Examples, How to Draw in Easy Steps*. URL: https://www.statisticshowto.com/how-to-use-a-probability-tree-for-probability-questions/ (viitattu 17. 06. 2020).

[9] T. Rudas. *Probability theory a primer*. Thousand Oaks: Sage Publications, 2004.

[10] R. B. Schinazi. *Probability with Statistical Applications*. Boston: Birkhäuser Boston, 2012.

[11] *Show info*. URL: http://www.letsmakeadeal.com/showinfo.htm (viitattu 25. 05. 2020).

A MATLAB-FUNKTIOT MONTY HALLIN ONGELMAN SIMULOINTIIN

MATLAB-funktio, jota käytettiin työssä simuloimaan Monty Hallin ongelmaa ja sen sovelluksia on esitetty kuvassa A.1. Tällä funktiolla voidaan simuloida Monty Hallin ongelmaa halutulla määrällä ovia kun oletetaan, että kilpailija valitsee aina oven 1. Ovia on oltava vähintään kolme ja ovia voidaan avata korkeintaan kaksi vähemmän kuin ovia on yhteensä. Jos halutaan, että ovea ei vaihdeta, laitetaan kohtaan "vaihto"0, kun taas laittamalla siihen 1, kilpailija vaihtaa ovea sattumanvaraisesti. Funktio palauttaa oven, jonka takana on auto, numeron sekä kilpailijan oven numeron.

```
function [auto,uusiovi] = montyhall(ovienlkm,oviaavataan,vaihto)
auto=randi(ovienlkm);
ovet=[2:ovienlkm];
kaikkiovet=[2:ovienlkm];
if auto==1
    avataan=randsample(ovet,oviaavataan);
else
    ovet(ovet==auto)=[];
    if ovienlkm==3
        avataan=ovet;
    else
        avataan=randsample(ovet,oviaavataan);
    end
end
if vaihto==1
    doorsleft=size(kaikkiovet,2)-size(avataan,2);
    if doorsleft==1
        for j=avataan
            poista=find(kaikkiovet==j);
            kaikkiovet(poista)=[];
        end
        uusiovi=kaikkiovet;
    else
        for j=avataan
            poista=find(kaikkiovet==j);
            kaikkiovet(poista)=[];
        end
        uusiovi=randsample(kaikkiovet,1);
    end
else
    uusiovi=1;
end
end
```

Kuva A.1. *Funktio Monty Hallin ongelman simulointiin*

MATLAB-funktio, joka toisti simulointia ja esitti sen tulokset on esitetty kuvissa A.2 ja A.3. Tämä funktio toistaa Monty Hall simulointia useaan kertaan ja kertoo sen tulokset sanallisesti. Usean toiston tapauksessa tulokset esitetään myös kuvaajassa. Tuloksena saadaan suhteelliset frekvenssit voittamiselle.

```
1  function [] = simulointi(ovienlkm,oviaavataan,vaihto,toistoja)
2  sata=ones(toistoja,1)*100;
3  tuhat=ones(toistoja,1)*1000;
4  kymtuhat=ones(toistoja,1)*10000;
5  satatuhat=ones(toistoja,1)*100000;
6  miljoona=ones(toistoja,1)*1000000;
7  tulossata=ones(toistoja,1);
8  tulostuhat=ones(toistoja,1);
9  tuloskymtuhat=ones(toistoja,1);
10 tulossatatuhat=ones(toistoja,1);
11 tulosmiljoona=ones(toistoja,1);
12 for r=1:toistoja
13     montysata=0;
14     montytuhat=0;
15     montykymtuhat=0;
16     montysatatuhat=0;
17     montymiljoona=0;
18     for re=1:1:100
19         [ovi,uusiovi]=montyhall(ovienlkm,oviaavataan,vaihto);
20         if ovi==uusiovi
21             montysata=montysata+1;
22         end
23     end
24     for re=1:1:1000
25         [ovi,uusiovi]=montyhall(ovienlkm,oviaavataan,vaihto);
26         if ovi==uusiovi
27             montytuhat=montytuhat+1;
28         end
29     end
30     for re=1:1:10000
31         [ovi,uusiovi]=montyhall(ovienlkm,oviaavataan,vaihto);
32         if ovi==uusiovi
33             montykymtuhat=montykymtuhat+1;
34         end
35     end
36     for re=1:1:100000
37         [ovi,uusiovi]=montyhall(ovienlkm,oviaavataan,vaihto);
38         if ovi==uusiovi
39             montysatatuhat=montysatatuhat+1;
40         end
41     end
42     for re=1:1:1000000
43         [ovi,uusiovi]=montyhall(ovienlkm,oviaavataan,vaihto);
44         if ovi==uusiovi
45             montymiljoona=montymiljoona+1;
46         end
47     end
```

Kuva A.2. *Rivit 1-47 simuloinnissa käytetystä funktiosta*

```
48     if toistoja>1
49         tulossata(r)=montysata/100;
50         tulostuhat(r)=montytuhat/1000;
51         tuloskymtuhat(r)=montykymtuhat/10000;
52         tulossatatuhat(r)=montysatatuhat/100000;
53         tulosmiljoona(r)=montymiljoona/1000000;
54     else
55         disp(['Sadalla toistolla voitti ',num2str(montysata),' kertaa.'])
56         disp(['Tällöin suhteellinen frekvenssi on ',num2str(montysata/100),' .'])
57         disp(['Tuhannella toistolla voitti ',num2str(montytuhat),' kertaa.'])
58         disp(['Tällöin suhteellinen frekvenssi on ',num2str(montytuhat/1000),' .'])
59         disp(['Kymmenellätuhannella toistolla voitti ',num2str(montykymtuhat),' kertaa.'])
60         disp(['Tällöin suhteellinen frekvenssi on ',num2str(montykymtuhat/10000),' .'])
61         disp(['Sadallatuhannella toistolla voitti ',num2str(montysatatuhat),' kertaa.'])
62         disp(['Tällöin suhteellinen frekvenssi on ',num2str(montysatatuhat/100000),' .'])
63         disp(['Miljoonalla toistolla voitti ',num2str(montymiljoona),' kertaa.'])
64         disp(['Tällöin suhteellinen frekvenssi on ',num2str(montymiljoona/1000000),' .'])
65     end
66 end
67 if toistoja>1
68     semilogx(sata,tulossata,'xk')
69     hold on
70     semilogx(tuhat,tulostuhat,'xk')
71     semilogx(kymtuhat,tuloskymtuhat,'xk')
72     semilogx(satatuhat,tulossatatuhat,'xk')
73     semilogx(miljoona,tulosmiljoona,'xk')
74     if vaihto==0
75         yline(1/ovienlkm,'Color','red','linewidth',0.5);
76     else
77         tod=(1-(1/ovienlkm))/(ovienlkm-oviaavataan-1);
78         yline(tod,'Color','red','linewidth',0.5);
79     end
80     xlim([10 10^7])
81     ylim([0 1])
82     xlabel('Toistojen määrä')
83     ylabel('Suhteellinen frekvenssi')
84     hold off
85     disp(['Sadalla toistolla suhteelliset frekvenssit olivat:'])
86     disp(tulossata')
87     disp(['Tuhannella toistolla suhteelliset frekvenssit olivat:'])
88     disp(tulostuhat')
89     disp(['Kymmenellätuhannella toistolla suhteelliset frekvenssit olivat:'])
90     disp(tuloskymtuhat')
91     disp(['Sadallatuhannella toistolla suhteelliset frekvenssit olivat:'])
92     disp(tulossatatuhat')
93     disp(['Miljoonalla toistolla suhteelliset frekvenssit olivat:'])
94     disp(tulosmiljoona')
95 end
```

Kuva A.3. *Rivit 48-95 simuloinnissa käytetystä funktiosta*