1A Todennäköisyyden laskusäännöt

Tässä harjoituksessa on muista poiketen 4 tuntitehtävää. Kotitehtäviä seuraavaa kertaa varten on 2 kpl, kuten yleensäkin. Tehtävissä vihreällä merkityt osat ovat lisävihjeitä.

Ratkaisuissa vihreällä merkityt osat lisäinformaatiota. Ne eivät ole osa vaadittua *ratkaisua*, mutta niiden sisältö kuuluu kurssin *oppisisältöön* eli ne kannattaa lukea.

Kun sanomme, että alkio valitaan tai poimitaan n-alkioisesta joukosta *umpimähkään*, tarkoitamme, että jokaisella joukon alkiolla on *sama* todennäköisyys $1/n$ tulla valituksi.

Tuntitehtävät

1A1 (Kaksi nopanheittoa) Kun tavallista noppaa heitetään kaksi kertaa, voidaan satunnaisilmiöön liittyvä perusjoukko määritellä kaavalla

$$\begin{aligned} S &= \{(x,y) : x = 1,2,3,4,5,6 \ \text{ ja } \ y = 1,2,3,4,5,6\} \\ &= \{(1,1),(1,2),(1,3),\dots,(6,6)\}, \end{aligned}$$

missä $x =$ ensimmäisen heiton tulos ja $y =$ toisen heiton tulos. Väritä perusjoukkoa kuvaavaan 6×6 -ruudukkoon seuraavat tapahtumat ja määritä niiden todennäköisyydet:

(a) $A = \{(x,y) \in S : x = 1\}$,

(b) $B = \{(x,y) \in S : y \geq 4\}$,

(c) $C = \{(x,y) \in S : x + y = 7\}$,

(d) $A \cup C =$ joukkojen A ja C yhdiste,

(e) $B \cap C =$ joukkojen B ja C leikkaus,

(f) $C \setminus A =$ joukkojen C ja A erotus.

Tässä tilanteessa kaikki 36 toteumaa ovat *yhtä todennäköiset*, joten tapahtuman todennäköisyys voidaan määrittää helposti, kun tiedetään montako toteumaa (taulukon ruutua) siihen sisältyy.

Ratkaisu.

(a) Tapahtuman A kuva on pystypalkki, joka sisältää sarakkeen $x = 1$. Sarakkeessa on 6 ruutua, joten $\mathrm{P}(A) = 6 \times \frac{1}{36} = 1/6$.

(b) Tapahtuman B kuva on vaakapalkki, jota vastaa rivit $y = 4$, $y = 5$ ja $y = 6$. Vaakapalkissa on 18 ruutua, joten $\mathrm{P}(B) = 18 \times \frac{1}{36} = 1/2$.

(c) Tapahtuman C kuva käsittää ruudut $(1,6),(2,5),\dots,(6,1)$. Näitä on 6 kpl, joten $\mathrm{P}(C) = 6 \times \frac{1}{36} = 1/6$.

Huomautus. Tässä kannattaa miettiä, miten halutut ruudut voi löytää helposti, systemaattisesti ja luotettavasti. Tietysti voisi käydä läpi kaikki 36 ruutua ja jokaisessa kokeilla erikseen, toteutuuko ehto eli onko $x + y = 7$. Helpompi tapa: Jos kokeillaan x:lle jotakin arvoa, niin voidaan päätellä, mikä arvo y:llä on tällöin pakko olla jotta ehto toteutuisi. Nimittäin, jotta toteutuu $x + y = 7$, on tietysti oltava $y = 7 - x$.

(d) Tapahtuman $A \cup C$ kuva saadaan värittämällä tapahtuman A pystypalkki ja tapahtuman C vaakapalkki samalla värillä. Tämä tapahtuma käsittää 11 ruutua, joten $\mathrm{P}(A \cup C) = 11 \times \frac{1}{36} = 11/36$.

(e) Tapahtuman $B \cap C$ kuva saadaan värittämällä samaan kuvaan haalealla joukot B ja C; ja sitten värittämällä tummalla ne ruudut, jotka kuuluvat molempiin joukkoihin. Tämä leikkausjoukko käsittää ruudut $(3,4)$, $(2,5)$ ja $(1,6)$, joten $\mathrm{P}(B \cap C) = 3 \times \frac{1}{36} = 1/12$.

(f) Tapahtuman $C \setminus A$ kuva saadaan värittämällä ensin joukko C ja sitten kumittamalla siitä pois joukon A alkiot. Tämä kuva sisältää ruudut $(2,5), (3,4), (4,3), (5,2), (6,1)$, joten $\mathrm{P}(C \setminus A) = 5 \times \frac{1}{36} = \frac{5}{36}$.

1A2 (Maanäytteet) Kaatopaikan läheisestä maaperästä satunnaisesti kerätyistä näytteistä löydetään todennäköisyydellä 0.30 haitallinen määrä lyijyä, ja todennäköisyydellä 0.40 haitallinen määrä arseeniyhdisteitä. Todennäköisyydellä 0.10 löydetään molempia.

Huom: "Löydetään lyijyä" on eri asia kuin "löydetään vain lyijyä". Jos löydetään lyijyä, saatetaan löytää *myös* arseeniyhdisteitä, tai sitten ei.

Laske seuraavien tapahtumien todennäköisyydet.

(a) Löydetään lyijyä mutta ei arseeniyhdisteitä.

(b) Löydetään arseeniyhdisteitä mutta ei lyijyä.

(c) Löydetään ainakin jompaakumpaa ainetta.

(d) Ei löydetä kumpaakaan.

Ratkaisu. Merkitään

$$\begin{aligned} L &= \text{“löydetään lyijyä”}, \\ A &= \text{“löydetään arseeniyhdisteitä”}. \end{aligned}$$

Huomataan, että on neljä mahdollisuutta sen suhteen, mitä aineita löytyy. Voidaan löytää molempia ($L \cap A$), vain lyijyä ($L \cap A^c$), vain arseeniyhdisteitä ($L^c \cap A$), tai ei kumpaakaan ($L^c \cap A^c$). Nämä voi piirtää esim. 2×2 taulukoksi, niin että tapahtumat L ja L^c ovat vaakarivejä, ja A ja A^c pystysarakkeita. Tehtävänannosta tiedämme, että $\mathrm{P}(L) = 0.30$, $\mathrm{P}(A) = 0.40$ ja $\mathrm{P}(L \cap A) = 0.10$.

(a) Tapahtuma L on kahden *toisensa poissulkevan* tapahtuman yhdiste, nimittäin $L = (L \cap A) \cup (L \cap A^c)$. Toisin sanoen, tapahtuma L voi tapahtua joko niin, että löydetään molempia aineita, tai niin, että löydetään vain lyijyä. Poissulkevien tapahtumien summasäännön mukaan

$$\mathrm{P}(L) = \mathrm{P}(L \cap A) + \mathrm{P}(L \cap A^c)$$

eli

$$0.30 = 0.10 + \mathrm{P}(L \cap A^c),$$

josta voidaan ratkaista

$$\mathrm{P}(L \cap A^c) = \mathrm{P}(L) - \mathrm{P}(L \cap A) = 0.30 - 0.10 = 0.20.$$

Huom. Tapahtuman $L \cap A^c$ eli "vain lyijyä" voi kirjoittaa myös joukkoerotuksena $L \setminus A$.

(b) Kuten edellisessä kohdassa,

$$P(A) = P(A \cap L) + P(A \cap L^c),$$

josta voidaan ratkaista

$$P(A \cap L^c) = P(A) - P(A \cap L) = 0.40 - 0.10 = 0.30.$$

Huom. Tapahtuma $A \cap L$ on sama tapahtuma kuin $L \cap A$, eli “löydetään molempia”. Joukkojen leikkaus on symmetrinen operaatio.

(c) Kyseessä on tapahtumien yhdiste $L \cup A$. Sen todennäköisyys voidaan laskea monella tavalla. Yksi tapa on yleinen summasääntö

$$P(L \cup A) = P(L) + P(A) - P(L \cap A) = 0.30 + 0.40 - 0.10 = 0.60.$$

Toinen vaihtoehto olisi laskea yhteen kolmen poissulkevan tapahtuman “vain lyijyä”, “vain arseeniyhdisteitä” ja “molempia” todennäköisyydet $0.20 + 0.30 + 0.10 = 0.60$.

(d) Kyseessä on edellisen kohdan tapahtuman vastakohta (komplementti), joten todennäköisyys on $1 - 0.60 = 0.40$.

1A3 (Sininen taksi) Kaupungissa liikennöi sata taksia, joista yksi on sininen ja loput vihreitä. Eräänä yönä yksi takseista ajaa polkupyöräilijän päälle ja pakenee paikalta. Silminnäkijän lausunnon mukaan yliajanut taksi oli sininen, minkä johdosta poliisi vangitsee sinisen taksin kuljettajan epäiltynä törkeästä liikenneturvallisuuden vaarantamisesta. Aiempien tutkimusten perusteella tiedetään, että vastaavissa olosuhteissa silminnäkijät näkevät sinisen auton sinisenä 90% todennäköisyydellä ja vihreän auton sinisenä 8% todennäköisyydellä.

Mikä on todennäköisyys, että yliajoon syyllistynyt taksi oli sininen? Tulisiko sinisen taksin kuljettaja tuomita tämän näytön perusteella?

Ratkaisu. Merkitään

- T_0 = “yliajanut taksi oli vihreä”,
- T_1 = “yliajanut taksi oli sininen”,
- H_0 = “silminnäkijä kertoi nähneensä vihreän taksin”,
- H_1 = “silminnäkijä kertoi nähneensä sinisen taksin”.

Tapahtumien T_0 ja T_1 prioritodennäköisyydet (ennen silminnäkijän lausuntoja) ovat $P(T_0) = 0.99$ ja $P(T_1) = 0.01$. Lisäksi tiedetään, että $P(H_1 \mid T_0) = 0.08$ ja $P(H_1 \mid T_1) = 0.90$. Osituskaavan perusteella

$$\begin{aligned} P(H_1) &= P(H_1 \mid T_0)\, P(T_0) + P(H_1 \mid T_1)\, P(T_1) \\ &= 0.08 \times 0.99 + 0.90 \times 0.01 \\ &= 0.0882. \end{aligned}$$

Näin ollen Bayesin kaavan mukaan tapahtuman T_1 posterioritodennäköisyys (silminnäkijän lausunnon jälkeen) on

$$\mathrm{P}(T_1 \,|\, H_1) = \frac{\mathrm{P}(H_1 \,|\, T_1)\,\mathrm{P}(T_1)}{\mathrm{P}(H_1)} = \frac{0.90 \times 0.01}{0.0882} \approx 0.102.$$

Ylläolevien tietojen valossa on todennäköisempää, että yliajanut taksi oli vihreä ($\approx 89.8\%$) kuin sininen ($\approx 10.2\%$). Tämän todistusaineiston pohjalta ei ole perusteita tuomita sinisen taksin kuljettajaa.

1A4 (Otanta kahdella tavalla) Pienessä kylässä on 120 asukasta, joista 20 on maanomistajia. Numeroimme asukkaat kokonaisluvuin $1, 2, \ldots, 120$ siten, että maanomistajilla on numerot $1, 2, \ldots, 20$.

(a) Valitaan satunnainen kylän asukas umpimähkään. Mikä on todennäköisyys, että valittu asukas on maanomistaja? Merkitään tätä lukua jatkossa p:llä.

(b) **Otanta takaisinpanolla.** Toistamme prosessin "valitse kylän asukas umpimähkään" kolme kertaa. Joka kerta valitaan kaikista asukkaista, joten sama asukas saatetaan valita uudestaankin. Olkoon valittujen asukkaiden jono valintajärjestyksessä (i, j, k). Montako erilaista jonoa on olemassa? Merkitään tätä lukua jatkossa N:llä.

(c) Edellisen kohdan tapauksessa, montako erilaista jonoa on, joissa kaikki kolme asukasta ovat maanomistajia? Merkitään tätä lukua A:lla.

(d) Laske A/N viidellä desimaalilla. Tämä on tn sille, että tulimme valinneeksi kolme maanomistajaa. Vertaa tätä lukua lukuun p^3 ja selitä havaintosi. Vihje: Vertaa tilannetta nopanheittoon.

(e) **Otanta ilman takaisinpanoa.** Valitaan kolme asukasta seuraavasti: Ensimmäinen valitaan umpimähkään kaikista asukkaista. Toinen valitaan jäljellä olevista umpimähkään, ja kolmas taas jäljellä olevista umpimähkään. Montako erilaista jonoa voidaan valita? Merkitään tätä lukua M:llä.

(f) Edellisen kohdan tapauksessa, montako erilaista jonoa on, joissa kaikki kolme asukasta ovat maanomistajia? Merkitään tätä lukua B:llä.

(g) Laske B/M viidellä desimaalilla. Tämä on tn sille, että tulimme valinneeksi kolme maanomistajaa. Vertaa aiempiin tuloksiin ja selitä.

Vihje: Järjestetyt jonot (luento 1A, Leskelä §1.8).

Ratkaisu.

(a) Kullakin asukkaalla on tn $1/120$ tulla valituksi, joten additiivisuuden nojalla tn valita joku maanomistajista on $20 \cdot (1/120) = 20/120 = 1/6$.

(b) Valinnalle i on 120 vaihtoehtoa. Valinnalle j on taas 120 ja valinnalle k 120, joten $N = 120^3 = 1728000$.

(c) Valinnalle i on 20 vaihtoehtoa (kylän maanomistajat). Valinnalle j taas 20 ja valinnalle k 20, joten $A = 20^3 = 8000$.

(d) $A/N = 8000/1728000 \approx 0.00463$. Tämä on sama kuin p^3, koska jokainen valinta vastaa matemaattisesti nopanheittoa, jossa tn osua maanomistajaan on 1/6.

(e) Valinnalle i on 120 vaihtoehtoa, *sen jälkeen* valinnalle j 119 vaihtoehtoa, ja sitten valinnalle k 118 vaihtoehtoa. Vaikka j:n valintamahdollisuudet ($k$etkä 119 asukasta ovat jäljellä) riippuvat siitä kenet valittiin ensimmäisenä, niitä on jokaisessa tapauksessa *yhtä monta* eli 119. Siten $M = 120 \cdot 119 \cdot 118 = 1685040$.

(f) Valinnalle i on 20 vaihtoehtoa. Sen jälkeen j:lle on 19 vaihtoehtoa (jäljellä olevat maanomistajat), ja sitten k:lle 18 vaihtoehtoa. Siten $B = 20 \cdot 19 \cdot 18 = 6840$.

(g) $B/M = 6840/1685040 \approx 0.00406 = 0.406\%$. Tämä on *pienempi* kuin p^3, koska jos on jo poimittu 1 tai 2 maanomistajaa, niin jäljellä olevassa kyläläisten joukossa on *suhteellisesti vähemmän* maaomistajia kuin alun perin (vähemmän kuin 1/6). Itse asiassa B/M voitaisiin laskea myös todennäköisyyden tulosäännöllä

$$\frac{20}{120} \cdot \frac{19}{119} \cdot \frac{18}{118}.$$

Otanta *takaisinpanolla* ja *ilman takaisinpanoa* on tilastotieteellistä terminologiaa, kun poimitaan satunnainen otos jostakin populaatiosta. Populaatiota voidaan ajatella laatikkona, jossa on palloja. Kun laatikosta on poimittu yksi pallo, se joko *pannaan takaisin* laatikkoon niin että seuraava poiminta tapahtuu jälleen samasta populaatiosta; tai palloa *ei panna takaisin* eli seuraava poiminta tapahtuu jäljellä olevasta pienemmästä populaatiosta.

Kotitehtävät

1A5 (Peliautomaatit) Kasinolla on kahdenlaisia peliautomaatteja, jotka näyttävät ulospäin täysin samalta. Tyypin A peliautomaateissa voittotodennäköisyys on 5% ja tyypin B peliautomaateissa 10%. Tiedetään, että 80% kasinon peliautomaateista on tyyppiä A ja 20% tyyppiä B. Peluri valitsee kasinolta satunnaisen peliautomaatin ja pelaa sitä yhden pelin verran.

(a) Mikä on todennäköisyys, että hän saa voiton?

(b) Jos peluri *saa* voiton, mikä on todennäköisyys, että automaatti on tyyppiä A? Entä tyyppiä B? Selitä luvut arkijärjellä.

(c) Jos peluri *ei saa* voittoa, mikä on todennäköisyys, että automaatti on tyyppiä A? Entä tyyppiä B? Selitä luvut arkijärjellä.

Arviointiohje. Kustakin kohdasta 2/3 pistettä, pyöristetään ylöspäin.

Ratkaisu.

(a) Merkitään

$$\begin{aligned} A &= \text{“valittu automaatti on tyyppiä A”}, \\ B &= \text{“valittu automaatti on tyyppiä B”}, \\ V &= \text{“peluri saa voiton”}, \\ V^c &= \text{“peluri ei saa voittoa”}. \end{aligned}$$

Tehtävän asettelun mukaan

$$\begin{aligned} \mathrm{P}(V \mid A) &= 0.05, \\ \mathrm{P}(V \mid B) &= 0.10, \\ \mathrm{P}(A) &= 0.80, \\ \mathrm{P}(B) &= 0.20. \end{aligned}$$

Koska tapahtumat A ja B muodostavat osituksen, saadaan kysytty todennäköisyys osituskaavasta

$$\begin{aligned} \mathrm{P}(V) &= \mathrm{P}(V \mid A)\,\mathrm{P}(A) + \mathrm{P}(V \mid B)\,\mathrm{P}(B) \\ &= 0.05 \times 0.80 + 0.10 \times 0.20 \\ &= 0.04 + 0.02 = 0.06 = 6\%. \end{aligned}$$

(b) Bayesin kaavan ja (a)-kohdan avulla on

$$\mathrm{P}(A \mid V) = \frac{\mathrm{P}(A)\,\mathrm{P}(V \mid A)}{\mathrm{P}(V)} = \frac{0.80 \cdot 0.05}{0.06} \approx 0.667 = 66.7\%.$$

Samaan tapaan laskemalla $\mathrm{P}(B \mid V) \approx 0.333$. Tämän voi päätellä myös siitä, että tapahtumista A ja B tapahtuu varmasti jompikumpi, joten $\mathrm{P}(B \mid V) = 1 - \mathrm{P}(A \mid V)$.

Selitys arkijärjellä: Satunnaisesti valittu automaatti on B-tyyppiä todennäköisyydellä 20%, mutta voiton saaminen suurentaa todennäköisyyttä sille, että kyseessä on B-tyyppi, sillä B-tyypissä voitot ovat todennäköisempiä.

(c) Koska $\mathrm{P}(V) = 0.06$, on $\mathrm{P}(V^c) = 0.94$. Lisäksi $\mathrm{P}(V^c \mid A) = 0.95$, koska tyypin A automaateissa tulee voitto todennäköisyydellä 0.05, ja siis ei voittoa todennäköisyydellä 0.95.

Lasketaan kuten edellisessä kohdassa.

$$\mathrm{P}(A \mid V^c) = \frac{\mathrm{P}(A)\,\mathrm{P}(V^c \mid A)}{\mathrm{P}(V^c)} = \frac{0.80 \cdot 0.95}{0.94} \approx 0.809 = 80.9\%.$$

Vastaavasti $\mathrm{P}(B \mid V^c) \approx 1 - 0.809 = 0.191 = 19.1\%$.

Selitys arkijärjellä: Kun ei saatu voittoa, niin B-tyypin (jossa voitot ovat todennäköisempiä) todennäköisyys laski hiukan alkuperäisestä 20%:sta.

1A6 (Tunnisteet) Eräällä planeetalla on 5^{10} asukasta, joista jokainen on identifioitu 10-merkkisen tunnisteen avulla. Tunnisteiden jokainen merkki on joko A, B, C, D tai E (jolloin erilaisia tunnisteita on olemassa yhteensä 5^{10}) ja joka asukkalla on eri tunniste (jolloin kaikki tunnisteet ovat siis tällä hetkellä käytössä). Kaksi esimerkkiä tunnisteista: CEAADEEBBB, AABABCAADE.

(a) Kuinka monen asukkaan tunniste on palindromi? (eli sama luettuna etu- ja takaperin, esim. AABCAACBAA)

(b) Kuinka monen asukkaan tunnisteessa mitkään kaksi peräkkäistä merkkiä eivät ole samoja? (esim. ABADABACAD on tällainen, kun taas AACBDBABCB ei ole)

Valitaan satunnainen planeetan asukas (valittu asukas on siis yhtä suurella todennäköisyydellä mikä tahansa planeetan asukkaista). Millä todennäköisyydellä,

(c) hänen tunnisteensa on palindromi?

(d) hänen tunnisteensa mitkään kaksi peräkkäistä merkkiä eivät ole samoja?

Arviointiohje. 0.5 pistettä per kohta, mikäli kyseinen kohta on oikein. Tehtävän kokonaispistemäärä muodostetaan pyöristämällä ylöspäin lähimpään kokonaislukuun.

Ratkaisu.

(a) Tarkastelemalla vain tunnisteiden ensimmäistä viittä merkkiä huomataan, että 10-merkkisiä palindromeja on täsmälleen yhtä monta kuin kirjaimista A—E muodostettavia erilaisia 5-merkkisiä jonoja.

Toisin sanoen, 10-merkkisen palindromin 5 viimeistä merkkiä määräytyvät täysin sen 5 ensimmäisen merkin perusteella (jotta lopputulos tosiaan olisi palindromi), jolloin kaikki 10-merkkiset palindromit saadaan rakennettua niin, että valitaan mikä tahansa 5-merkkinen jono ja kirjoitetaan se ja sen käänteinen versio peräkkäin (esim. AABAC -> AABACCABAA).

(Huomaa, että kaksi erilaista 5-merkkistä jonoa eivät myöskään ikinä tuota samaa 10-merkkistä palindromia, jolloin olemme siis muodostaneet *bijektion* näiden kahden joukon välille.)

Koska kirjaimista A–E muodostettavia 5-merkkisiä jonoja on olemassa yhteensä $5^5 = 3125$, (ensimmäinen merkki voidaan valita 5 tavalla, toinen merkki voidaan valita 5 tavalla, kolmas merkki...) on myös tunnisteita jotka ovat palindromeja yhteensä **3125** erilaista.

(b) Kaikki ehdon täyttävät tunnisteet voidaan rakentaa seuraavasti:

- Valitaan tunnisteen ensimmäinen merkki. Tälle on 5 eri vaihtoehtoa.
- Valitaan tunnisteen toinen merkki. Se ei saa olla sama kuin ensimmäinen, eli vaihtoehtoja on 4.
- Valitaan tunnisteen kolmas merkki. Se ei saa olla sama kuin toinen, eli vaihtoehtoja on 4.
- ...
- Valitaan tunnisteen kymmenes merkki. Se ei saa olla sama kuin yhdeksäs, eli vaihtoehtoja on 4.

Edellisellä prosessilla on yhteensä $5 \cdot 4 \cdot 4 \cdot \ldots \cdot 4 = 5 \cdot 4^9 =$ **1 310 720** erilaista lopputulosta. Siis näin monta on tunnuksia, joissa ei ole kahta samaa merkkiä peräkkäin.

(c) Kaikkiaan tunnisteita on $5^{10} = 9\ 765\ 625$. Kysytty todennäköisyys on palindromien (ks. a-kohta) osuus kaikista tunnisteista

$$\frac{3\ 125}{9\ 765\ 625} \approx 0.032\%.$$

(d) Kysytty todennäköisyys on ehdon täyttävien tunnusten (ks. b-kohta) osuus kaikista tunnisteista

$$\frac{1\ 310\ 720}{9\ 765\ 625} \approx 13.4\%.$$