Luku 2

Todennäköisyyslaskenta ja kombinatoriikka

Tässä luvussa käsitellään lähinnä vain äärellisiä ja numeroituvasti äärettömiä otosavaruuksia Ω. Lopuksi esitetään todennäköisyyden aksioomat, jotka soveltuvat myös silloin, kun Ω ei ole numeroituva.

Jos äärellisen otosavaruuden $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ jokainen alkeistapaus on yhtä mahdollinen, niin jakaumafunktio on

$$p_i = P(\omega_i) = \frac{1}{n}, \quad 1 \leq i \leq n$$

missä n on alkeistapausten lukumäärä. Silloin saadaan '*suotuisat per kaikki*' -sääntö eli klassinen todennäköisyyden määritelmä

$$P(A) = \sum_{\omega_i \in A} p_i = \sum_{\omega_i \in A} \frac{1}{n} = \frac{|A|}{n},$$

missä $|A|$ on A:n alkioiden lukumäärä.

Todennäköisyyden aksioomat

Kutsumme tästä lähtien joukkoa *numeroituvaksi,* jos se on *äärellinen* tai *numeroituvasti ääretön.* Oletamme tässä, että otosavaruus Ω on numeroituva, jotta voisimme esittää todennäköisyyden aksioomat mahdollisimman yksinkertaisessa muodossa. Todennäköisyys luonnehditaan Ω:n osajoukkojen joukossa määriteltynä funktiona (ks. Määritelmä 2.1, s. 23).

Aksioomat Todennäköisyys(mitta) P on otosavaruudessa Ω määritelty Ω:n osajoukkojen kuvaus eli funktio, joka toteuttaa seuraavat kolme aksioomaa:

1. Funktion P arvo on kaikilla osajoukoilla $A \subset \Omega$ epänegatiivinen:
$$P(A) \geq 0.$$
2. Minkä tahansa kahden toisensa poissulkevan tapahtuman A ja B unionin $A + B$ kuva on A:n ja B:n kuvien summa:
$$P(A + B) = P(A) + P(B), \qquad \text{kun } A \cap B = \emptyset.$$
3. Koko otosavaruuden Ω kuva on 1 eli
$$P(\Omega) = 1.$$

2.1 Todennäköisyyden ominaisuuksia

Oletetaan aksioomat 1–3. Niistä voidaan johtaa mm. seuraavat todennäköisyysmitan P ominaisuudet:

4. $P(\emptyset) = 0.$
5. $P(A^c) = 1 - P(A)$ kaikilla $A \subset \Omega$.
6. Jos $A \subset B \subset \Omega$, niin $P(A) \leq P(B)$ ja $P(B - A) = P(B) - P(A)$. Seuraus: $P(A) \leq 1$ kaikilla $A \subset \Omega$.
7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. (Lause 2.4, s.20)
8. $P(A \cup B) \leq P(A) + P(B)$.

Lisäominaisuuksia:

9. Mille tahansa kahdelle tapahtumalle A ja B pitää paikkansa, että
$$P(A) = P(A \cap B) + P(A \cap B^c)$$
$$P(B) = P(B \cap A) + P(B \cap A^c).$$
10. $P(A \cap B) = P(A) + P(B) - 1$. (Bonferronin epäyhtälö, s.21)

11. Soveltamalla yhteenlaskulausetta saadaan (Lause 2.5, s. 21)

$$
\begin{aligned}
P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) \\
&\quad - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3).
\end{aligned}
$$

Todennäköisyyden additiivisuus (Ominaisuus 4) voidaan suoraviivaisesti yleistää useammalle kuin kahdelle erilliselle joukolle.

Lause 2.1 *Olkoot A_1, A_2, ..., A_n parittain pistevieraat (erilliset) Ω:n osajoukot eli tapahtumat (ts. $A_i \cap A_j = \emptyset$, kun $i \neq j$). Silloin*

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

Itse asiassa additiivisuus yleistyy myös ärettömän monelle parittain erilliselle tapahtumalle A_1, A_2, A_3, ... Silloin

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots .$$

Jos A_1, A_2, ..., A_n ovat parittain erilliset (ts. $A_i \cap A_j = \emptyset$, kun $i \neq j$) ja $\Omega = A_1 \cup A_2 \cup \cdots \cup A_n$, niin joukkokokoelma $A_1, A_2, \ldots, A_n$ on *otosavaruuden Ω ositus.*

Lause 2.2 *Olkoon kokoelma $A_1, A_2, \ldots, A_n$ otosavaruuden Ω ositus ja $E \subset \Omega$ on jokin tapahtuma. Silloin*

$$P(E) = \sum_{i=1}^{n} P(E \cap A_i).$$

Seuraus 2.1 *Mille tahansa kahdelle tapahtumalle A ja B pitää paikkansa, että*

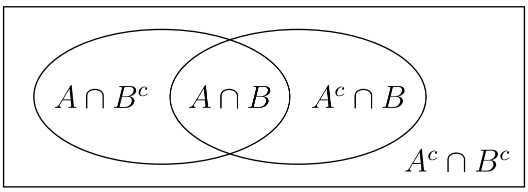
$$P(A) = P(A \cap B) + P(A \cap B^c).$$

Lauseen 4.1 kohta 4 voidaan yleistää myös joukoille, jotka eivät ole erillisiä. Tällöin saadaan seuraava *yhteenlaskulause.*

Lause 2.3 *Jos $A \subset \Omega$ ja $B \subset \Omega$, niin*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{2.1.1}$$

Todistus. Kuten Kuvio 4.1 osoittaa, joukot $A \cap B^c$, $A \cap B$, $A^c \cap B$ muodos-



Kuvio 2.1. Tapahtuman $A \cup B$ ositus.

tavat tapahtuman $A \cup B$ osituksen. Siksi

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B). \tag{2.1.2}$$

Lauseen 4.2 mukaan vastaavasti

$$\begin{aligned} P(A) &= P(A \cap B^c) + P(A \cap B) \\ P(B) &= P(A^c \cap B) + P(A \cap B), \end{aligned}$$

joten

$$P(A) + P(B) = P(A \cap B^c) + P(A^c \cap B) + 2\,P(A \cap B). \tag{2.1.3}$$

Kun identiteetistä (4.1.3) vähennetään puolittain $P(A \cap B)$, saadaan lauseke

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

jonka oikea puoli on (4.1.2):n mukaan $P(A \cup B)$. Näin yhteenlaskulause on todistettu. □

Tämä todennäköisyyksien yhteenlaskulause voidaan edelleen yleistää mielivaltaisen monelle tapahtumalle. Esitämme aluksi yleistyksen, kun tapahtumia on kolme. Yleinen tapaus saadaan samalla periaatteella, mutta se esitetään vasta myöhemmin.

Lause 2.4 *Jos A_1, A_2 ja A_3 ovat Ω:n osajoukkoja (tapahtumia), niin*

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = {} & P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) \\ & - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3). \end{aligned} \tag{2.1.4}$$

Bonferronin epäyhtälö. Koska $P(A \cup B) \leq 1$, seuraa Lauseesta 4.4 epäyhtälö

$$P(A \cap B) \geq P(A) + P(B) - 1. \tag{2.1.5}$$

Epäyhtälöä 4.1.5 sanotaan *Bonferronin epäyhtälöksi.*

Esimerkki 2.1 Bonferronin epäyhtälö saattaa olla käyttökelpoinen silloin, kun ei pystytä laskemaan todennäköisyyttä $P(A \cap B)$ tarkasti, mutta tunnetaan todennäköisyydet $P(A)$ ja $P(B)$. Olkoon esimerkiksi $P(A) = P(B) = 0.95$. Silloin

$$P(A \cap B) \geq 0.95 + 0.95 - 1 = 0.90.$$

Jos $P(A) + P(B) < 1$, niin alaraja (4.1.5):ssa on negatiivinen ja epäyhtälö pitää triviaalisti paikkansa. □

2.2 Symmetriaan perustuva todennäköisyys

Jos äärellisen otosavaruuden $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ jokainen alkeistapaus on yhtä mahdollinen, niin jakaumafunktio on

$$p_i = P(\omega_i) = \frac{1}{n}, \quad 1 \le i \le n$$

missä n on alkeistapausten lukumäärä. Silloin jokaisen tapahtuman todennäköisyys on yksinkertaisesti

$$P(A) = \sum_{\omega_i \in A} p_i = \sum_{\omega_i \in A} \frac{1}{n} = \frac{|A|}{n}, \tag{2.2.1}$$

missä $|A|$ on A:n alkioiden lukumäärä. Tapahtuman A todennäköisyys saadaan siis jakamalla A:n alkeistapausten lukumäärä kaikkien alkeistapahtumien lukumäärällä n. Tätä '*suotuisat per kaikki*' -sääntöä kutsutaan myös klassiseksi todennäköisyyden määritelmäksi.

Esimerkki 2.2 Heitetään harhatonta noppaa. Silloin eri silmälukuja voidaan pitää yhtä mahdollisina ja jakaumafunktio on perusteltua määritellä $p_i = \frac{1}{6}$, $i = 1, \ldots, 6$ otosavaruudessa $\Omega = \{1, 2, 3, 4, 5, 6\}$. Jos heitetään kahta noppaa, voidaan symmetrisiksi alkeistapauksiksi valita järjestetyt parit

$$(1,1), (1,2), (1,3), \ldots, (6,6).$$

Siinä tulokset on annettu muodossa (1. nopan silmäluku, 2. nopan silmäluku). Tämän satunnaiskokeen otosavaruus on siis

$$\Omega = \{\, (i,j) \mid i,j \in \{1,2,3,4,5,6\} \,\}.$$

Koska $|\Omega| = 36$, niin $P\big(\{(i,j)\}\big) = \frac{1}{36}$ kaikilla $i, j \in \{1,2,3,4,5,6\}$.

Väitetään, että ranskalainen aatelismies ja uhkapeluri Chevalier de Méré havaitsi kokeellisesti seuraavan tuloksen:

(i) Heitettäessä noppaa 4 kertaa kannattaa lyödä vetoa siitä, että saadaan ainakin yksi kuutonen.

(ii) Heitettäessä kahta noppaa 24 kertaa *ei kannata* lyödä vetoa siitä, että saadaan ainakin yksi kuutospari.

De Méré huomasi jäävänsä pitkässä pelisarjassa häviölle lyödessään vetoa kuutosparin puolesta. Hän ei kuitenkaan pystynyt teoreettisesti selittämään havaintoaan (de Mérén ongelma) ja niinpä hän kääntyi ranskalaisen filosofin ja matemaatikon Pascalin puoleen (n. 1650). De Mérén ongelman uskotaan antaneen alkusysäyksen kuuluisaan Pascalin ja Fermatin väliseen kirjeenvaihtoon, joka johti todennäköisyyslaskennan syntyyn. □

Klassisen määritelmän mukaan tapahtuman A todennäköisyys saadaan jakamalla joukon A alkioiden lukumäärä kaikkien alkeistapausten lukumäärällä. Vaikka tehtävä on periaatteessa helppo, se voi käytännössä osoittautua yllättävän hankalaksi. Lukumäärien laskemisen helpottamiseksi esitämme seuraavassa joitain kombinatoriikan periaatteita ja tuloksia.

2.3 Aksiomaattinen lähestymistapa

Kutsumme tästä lähtien joukkoa *numeroituvaksi,* jos se on *äärellinen* tai *numeroituvasti ääretön.* Oletamme tässä, että otosavaruus Ω on numeroituva, jotta voisimme esittää todennäköisyyden aksioomat mahdollisimman yksinkertaisessa muodossa. Todennäköisyys luonnehditaan Ω:n osajoukkojen joukossa määriteltynä funktiona.

Määritelmä 2.1 Todennäköisyys(mitta) P on otosavaruudessa Ω määritelty Ω:n osajoukkojen kuvaus eli funktio, joka toteuttaa seuraavat kolme aksioomaa:

(i) Funktion P arvo on kaikilla osajoukoilla $A \subset \Omega$ epänegatiivinen:

$$P(A) \geq 0.$$

(ii) Minkä tahansa kahden toisensa poissulkevan tapahtuman A ja B unionin $A + B$ kuva on A:n ja B:n kuvien summa:

$$P(A + B) = P(A) + P(B), \qquad \text{kun } A \cap B = \emptyset.$$

(iii) Koko otosavaruuden Ω kuva on 1 eli

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

2.4 Kombinatoriikkaa

2.4.1 Summa- ja tuloperiaate

Olkoot kokeiden $\mathcal{E}_1$ ja $\mathcal{E}_2$ otosavaruudet Ω_1 ja Ω_2. Silloin kokeiden tulosvaihtoehtojen lukumäärät ovat $|\Omega_1|$ ja $|\Omega_2|$. Merkitään $|\Omega_1| = n_1$ ja $|\Omega_2| = n_2$.

Summaperiaate. Tehdään joko koe $\mathcal{E}_1$ tai $\mathcal{E}_2$. Silloin mahdollisten tulosten lukumäärä on $n_1 + n_2$.

Tuloperiaate. Tehdään ensin koe $\mathcal{E}_1$ ja sitten $\mathcal{E}_2$. Silloin yhdistetyn kokeen $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$ tulosvaihtoehtojen lukumäärä on $n_1 n_2$.

Esimerkki. Heitetään esimerkiksi kahta noppaa. Silloin tulosvaihtoehtojen lukumäärä on $6^2 = 36$.

Yleinen tulosääntö. Tehdään peräkkäin r monivalintaa.

1. valinta voidaan tehdä n_1:llä tavalla ($\mathcal{E}_1$:llä n_1 tulosvaihtoehtoa),

2. valinta voidaan tehdä n_2:llä tavalla, jne., ja

r. valinta voidaan tehdä n_r:llä tavalla.

Silloin valintatulosten kokonaismäärä on $n_1 n_2 \ldots n_r$.

Esimerkki. Ravintolassa on tarjolla 5 alku-, 4 pää- ja 6 jälkiruokalajia. Silloin kolmen ruokalajin aterioita on $5 \times 4 \times 6 = 120$.

Otantamallit

Valitaan uurnasta r palloa. Uurnassa on n erilaista palloa: $\{1, 2, \ldots, n\}$.

1. **Järjestetty otos palauttaen.** Valitaan r palloa palauttaen ja valintajärjestys otetaan huomioon. Tulosäännön mukaan saadaan n^r *järjestettyä r-otosta.*

2. **Järjestetty otos palauttamatta.** Valitaan r palloa ($r \leq n$) palauttamatta ja valintajärjestys otetaan huomioon (*r-permutaatio*). Tulosäännön mukaan r-permutaatioiden lukumäärä on

$$n^{(r)} \triangleq n(n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!}, \text{ missä}$$

kertoma $n! = n^{(n)} = n(n-1)\ldots 2 \cdot 1$ on n:n alkion permutaatioiden lukumäärä.

3. **Otanta palauttamatta, kun järjestystä ei oteta huomioon.** Valitaan r palloa ($r \leq n$) palauttamatta ja järjestystä ei panna merkille. Tulosäännön mukaan r-kombitaatioiden (r-osajoukkojen) lukumäärä on
$$\binom{n}{r} = \frac{n^{(r)}}{r!} = \frac{n!}{r!(n-r)!},$$
mikä luetaan n r:n yli. $\binom{n}{r}$ on binomikerroin.

   Esimerkki. Uurnassa n_1 valkoista ja n_2 mustaa palloa ja $n = n_1 + n_2$. Pallot tunnistetaan vain värin perusteella. Kuinka monta tunnistettavasti erilaista permutaatiota (järjestystä) saadaan? Vastaus: $\binom{n}{n_1}$.

   Multinomikerroin. Uurnassa on erivärisiä palloja (k väriä): n_1 väriä 1, n_2 väriä 2 jne. n_k väriä k, missä $n = n_1 + n_2 + \cdots + n_k$. Pallot tunnistetaan vain värin perusteella. Kuinka monta tunnistettavasti erilaista permutaatiota (järjestystä) saadaan? Vastaus:
$$\frac{n!}{n_1!\, n_2! \cdots n_k!} = \binom{n}{n_1\ n_2\ \ldots\ n_k}.$$

4.

5. **Otanta palauttaen, kun järjestystä ei oteta huomioon.** Valitaan r palloa palauttamaen ja järjestystä ei panna merkille. Valitaan uurnasta pallo r kertaa ja ennen palauttamista pannaan merkille numero (ei numeroiden järjestystä). Montako tunnistettavasti erilaista valintaa saadaan? Vastaus:
$$\binom{r+n-1}{n-1} = \binom{r+n-1}{r}.$$

Allokointimallit

Vaihtoehtoinen tapa on ajatella, että sijoitetaan r palloa (numeroitu: $1, 2, \ldots, r$) n:ään laatikkoon (numeroitu: $1, 2, \ldots, n$). Edellä käsitellyt otantamallit liittyvät seuraaviin neljään tilanteeseen:

1. Jokainen laatikko voi sisältää palloja minkä tahansa määrän $1 \leq i \leq r$. Pallojen numerot havaitaan. Erilaisten asetelmien lukumäärä on n^r.

2. Mikään laatikko ei saa sisältää enempää kuin yhden pallon ($r \leq n$) ja pallojen numerot havaitaan. Erilaisia asetelmia on $n^{(r)}$.

3. Mikään laatikko ei saa sisältää enempää kuin yhden pallon ($r \leq n$) ja pallojen numeroita ei havainnoida. Erilaisia asetelmia on $\binom{n}{r}$.

4. Jokainen laatikko voi sisältää palloja minkä tahansa määrän $1 \leq i \leq r$. Pallojen numeroita ei havainnoida. Erilaisia asetelmia on $\binom{r+n-1}{n-1}$.

2.4.2 Valinta järjestyksessä

Tarkastellaan ensin *valintaa palauttaen.* Olkoon perusjoukon Ω alkioiden lukumäärä $|\Omega| = n$ ja voimme siis ajatella, että alkiot on numeroitu 1:stä n:ään. Valitaan Ω:sta peräkkäin r alkiota ja jokainen valittu alkio palautetaan takaisin Ω:aan ennen seuraavaa valintaa. Valinnan tuloksena saatua järjestettyä jonoa kutsutaan *järjestetyksi r-otokseksi* $(a_1, a_2, \ldots, a_r)$, jossa jokainen alkio $1 \leq a_j \leq n$. Järjestetyssä r-otoksessa sama alkio voi siis toistua monta kertaa. Tehdään esimerkiksi järjestetty 3-otos joukosta $A = \{a, b\}$. Silloin kaikki mahdolliset järjestetyt 3-otokset ovat *aaa*, *aab*, *aba*, *abb*, *baa*, *bab*, *bba*, *bbb*. Järjestettyjen 3-otosten lukumäärä on tuloperiaatteen mukaan $2^3 = 8$. Samalla tavalla tuloperiaatteesta seuraa, että Ω:sta valittujen järjestettyjen r-otosten lukumäärä on n^r.

Tarkastellaan nyt *valintaa palauttamatta.* Jos Ω:sta valitaan järjestyksessä r alkiota ($r \leq n$) palauttamatta, saadaan järjestetty r-otos, jossa sama alkio voi esiintyä vain kerran. Tällaista järjestettyä r-otosta kutsutaan Ω:n *r-permutaatioksi.* Esimerkiksi joukon $B = \{a, b, c, d\}$ 2-permutaatiot ovat

$$ab,\ ba,\ ac,\ ca,\ ad,\ da,\ bc,\ cb,\ bd,\ db,\ cd,\ dc,$$

joiden lukumäärä on tuloperiaatteen nojalla $4 \cdot 3 = 12$. Yleisesti r-permutaatioiden lukumäärä Ω:sta on

$$n^{(r)} = n(n-1)(n-2)\cdots(n-r+1), \qquad 0 < r \leq n.$$

Merkintä $n^{(r)}$ luetaan ”n:n r-kertoma”. Kun $r = n$, saadaan *joukon Ω n-permutaatio,* jota kutsutaan yksinkertaisesti joukon *permutaatioksi.* Permutaatio on siis joukon alkioiden järjestetty jono. Joukon Ω permutaatioiden lukumäärä on siis

$$n^{(n)} = n(n-1)(n-2)\cdots 2 \cdot 1$$

ja sitä merkitään $n!$ ja luetaan ”n-kertoma”.

Esimerkki 2.3 (Syntymäpäiväongelma) Kutsuilla on r henkilöä. Henkilöiden syntymäpäivät muodostavat r:n päivämäärän jonon, jossa sama päivämäärä voi toistua. Vuoden päivien lukumäärä $n = 365$, jos karkausvuotta ei oteta huomioon. Oletetaan, että kaikki mahdolliset 365^r syntymäpäiväjonoa ovat yhtä todennäköiset. Mikä on todennäköisyys, että ainakin kahdella henkilöllä on sama syntymäpäivä? Ensinnäkin 365:n päivän r-permutaatioiden lukumäärä on $365^{(r)}$, mikä on siis kaikkien r:n pituisten eri syntymäpäivistä muodostettujen jonojen lukumäärä. Todennäköisyys, että kaikilla on eri syntymäpäivä, on kaavan (4.2.1) mukaan

$$P(\text{'Eri syntymäpäivät'}) = \frac{365^{(r)}}{365^r}.$$

Silloin todennäköisyys, että ainakin kahdella sama syntymäpäivä on

$$1 - \frac{365^{(r)}}{365^r}.$$

□

2.4.3 Osajoukon valinta

Kun Ω:sta valitaan r alkiota ($r \le n$) palauttamatta, saadaan Ω:n osajoukko. Nyt ei siis kiinnitetä huomiota alkioiden järjestykseen, vaan ainoastaan siihen, mitkä alkiot osajoukkoon kuuluvat. Joukon r:n alkion osajoukkoa kutsutaan joukon r-kombinaatioksi. Esimerkiksi joukon $B = \{a, b, c, d\}$ 2-kombinaatiot ovat

$$\{a, b\},\ \{a, c\},\ \{a, d\},\ \{b, c\},\ \{b, d\},\ \{c, d\}.$$

Jokaista 2-kombinaatiota kohti on olemassa kaksi 2-permutaatiota. Esimerkiksi 2-kombinaatioon $\{a, b\}$ liittyvät 2-permutaatiot ovat ab, ba. Koska 2-permutaatiota on $4 \cdot 3 = 12$ kappaletta, niin 2-kombinaatiota on $\frac{4\cdot 3}{2} = 6$ kappaletta. Ja yleisesti: Koska r-permutaatioden lukumäärä jokaista r-kombinaatiota kohti on $r!$ ja r-permutaatioden lukumäärä on $n^{(r)}$, niin r-kombinaatioiden lukumäärä on

$$\frac{n^{(r)}}{r!} = \frac{n!}{r!\,(n-r)!},$$

jota merkitään $\binom{n}{r}$ ja se luetaan "n r:n yli".

Joukossa $A = \{a, a, a, a, b, b, c, c, c, d\}$ on 4 a-kirjainta, 2 b:tä, 3 c:tä ja yksi d. Kuinka monta erilaista 10-kirjaimista sanaa näistä kirjaimista voidaan muodostaa? Sanassa on kirjaimille 10 eri paikkaa ja jokainen kirjain voidaan sijoittaa johonkin 10:stä mahdollisesta paikasta. Ensiksikin a-kirjaimien paikka voidaan valita $\binom{10}{4}$ tavalla, jäljelle jääneisiin 6:een paikkaan voidaan b sijoittaa $\binom{6}{2}$ tavalla, sen jäkeen c $\binom{4}{3}$ tavalla ja lopuksi d:lle jää $\binom{1}{1} = 1$ paikka. Kertolaskuperiaatteen mukaan kaikkien mahdollisten sanojen lukumäärä on

$$\binom{10}{4}\binom{6}{2}\binom{4}{3}\binom{1}{1} = \frac{10!}{4!\,2!\,3!\,1!} = 12600.$$

Olkoon joukossa n alkiota, joista n_1 kuuluu 1. ryhmään, n_2 2. ryhmään ja lopulta n_k alkiota k. ryhmään, joten $n = n_1 + n_2 + \cdots + n_k$. Joukosta valitaan peräkkäin palauttamatta alkioita kunnes kaikki on valittu. Kuinka monta tunnistettavasti erilaista alkiojonoa voidaan saada? Nyt ajatellaan, että kunkin ryhmän alkiot ovat keskenään samanlaisia. Emme voi siis tunnistaa erilaisia ryhmän sisäisiä järjestyksiä. Vastaus saadaan samalla tavalla kuin edellisessä esimerkissä, joten valintojen lukumäärä on

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k} = \frac{n!}{n_1!\,n_2!\cdots n_k!}.$$

Tätä lauseketta sanotaan *multinomikertoimeksi* ja sitä merkitään

$$\frac{n!}{n_1!\,n_2!\cdots n_k!} = \binom{n}{n_1\ n_2\ \ldots\ n_k}. \tag{2.4.1}$$

Kun $k = 2$, saadaan erikoistapauksena binomikerroin

$$\binom{n}{n_1\ n_2} = \frac{n!}{n_1!\,n_2!} = \frac{n!}{n_1!\,(n-n_1)!} = \binom{n}{n_1}.$$

2.4.4 Otanta palauttaen, kun järjestystä ei oteta huomioon

Valitaan r palloa uurnasta, jossa on k erilaista (esimerkiksi eriväristä) palloa. Jokaisessa valinnassa rekisteröidään pallon väri ja pallo palautetaan uurnaan ennen seuraavaa valintaa. Olkoon uurnassa esimerkiksi 3 erilaista palloa: $\odot$, $\ominus$, $\oplus$. Valitaan uurnasta 3 palloa palauttaen ($r = k = 3$).

Taulukko 2.1. Erilaiset valinnat palauttaen, kun $r = k = 3$.

Tulos	$\odot$	$\ominus$	$\oplus$	
$\odot\odot\odot$	$\ast\ast\ast$			$\ast\ast\ast\vert\vert$
$\odot\odot\ominus$	$\ast\ast$	$\ast$		$\ast\ast\vert\ast\vert$
$\odot\odot\oplus$	$\ast\ast$		$\ast$	$\ast\ast\vert\vert\ast$
$\odot\ominus\ominus$	$\ast$	$\ast\ast$		$\ast\vert\ast\ast\vert$
$\odot\ominus\oplus$	$\ast$	$\ast$	$\ast$	$\ast\vert\ast\vert\ast$
$\odot\oplus\oplus$	$\ast$		$\ast\ast$	$\ast\vert\vert\ast\ast$
$\ominus\ominus\ominus$		$\ast\ast\ast$		$\vert\ast\ast\ast\vert$
$\ominus\ominus\oplus$		$\ast\ast$	$\ast$	$\vert\ast\ast\vert\ast$
$\ominus\oplus\oplus$		$\ast$	$\ast\ast$	$\vert\ast\vert\ast\ast$
$\oplus\oplus\oplus$			$\ast\ast\ast$	$\vert\vert\ast\ast\ast$

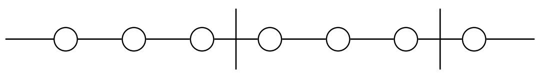
Jokaisen valinnan jälkeen pistetään merkki ”$\ast$” kyseisen pallon kohdalle. Kaikkien valintojen jälkeen meillä on 3 (r) merkkiä. Huomaa, että r voi olla suurempi kuin k, vaikka esimerkissä $r = k = 3$. Taulukon 4.1 viimeisellä sarakkeella pallojen valinta on esitetty merkkien ”$\ast$” ja ”$\vert$” jonona ilmeisellä tavalla. Jonossa on yhteensä 5 ($= r + k - 1$) merkkiä. Kuinka monella tavalla 2 ”$\vert$”-merkkiä voi jakaa 3 ”$\ast$”-merkkiä ryhmiin? Vastaus on $\frac{5!}{3!\,2!} = \binom{5}{2} = 10$. Vastaavasti yleisessä tapauksessa erilaisten tulosvaihtoehtojen lukumäärä on

$$\binom{k + r - 1}{r} = \binom{k + r - 1}{k - 1}.$$

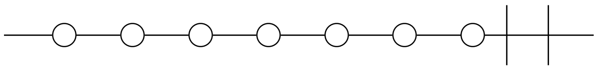
Esimerkki 2.4 Tarkastellaan yhtälöä

$$x_1 + x_2 + \cdots x_k = r, \tag{2.4.2}$$

missä k ja r ovat annettuja positiivisia kokonaislukuja ja muuttujat x_1, x_2, $\ldots$, x_k voivat saada arvoikseen epänegatiivisia kokonaislukuja. Montako erilaista ratkaisua yhtälöllä (4.4.2) on? Olkoon esimerkiksi $r = 7$ ja $k = 3$. Tarkastellaan kysymystä 'helmitaululla',



jossa on esitetty ratkaisu $x_1 = 3$, $x_2 = 3$, $x_3 = 1$. Ratkaisu $x_1 = 7$, $x_2 = 0$, $x_3 = 0$ on helmitaululla muotoa



Helmitaululla on 7 helmeä ($r = 7$) ja 2 jakoviivaa ($k - 1 = 2$) eli yhteensä 9 ($k + r - 1 = 9$) objektia. Jakoviivat voidaan sijoittaa helmitaululla $\binom{9}{2} = 36$ tavalla. Analogisesti voimme päätellä, että yhtälön (4.4.2) epänegatiivisten kokonaislukuratkaisujen lukumäärä on $\binom{k+r-1}{k-1}$. □

2.4.5 Kombinatoriikan merkintöjä ja identiteettejä

Olkoon r epänegatiivinen kokonaisluku ja n reaaliluku. Nyt n:n r-kertoma määritellään kaikille reaaliluvuille samalla tavalla kuin epänegatiivisille kokonaisluvuille:

$$\begin{aligned} n^{(r)} &= n(n-1)(n-2)\cdots(n-r+1), \qquad r > 0; \\ n^{(0)} &= 1. \end{aligned} \tag{2.4.3}$$

Jos n on epänegatiivinen kokonaisluku, niin $n^{(r)}$ on n:n alkion joukon $\{1, 2, \ldots, n\}$ kaikkien r:n ($r \leq n$) kokoisten *järjestettyjen osajoukkojen lukumäärä.* Eritysesti n-kertoma on

$$n! = n^{(n)} = n(n-1)(n-2)\cdots 2 \cdot 1$$

ja 0-kertoma määritellään $0! = 0^{(0)} = 1$.

Olkoon r epänegatiivinen kokonaisluku ja n reaaliluku. Määritellään

$$\binom{n}{r} = \begin{cases} \dfrac{n^{(r)}}{r!} = \dfrac{n!}{(n-r)!\,r!}, & r \geq 0; \\ 0, & r < 0. \end{cases} \tag{2.4.4}$$

Huomaa, että yllä esitetyt $n^{(r)}$ ja $\binom{n}{r}$ on määritelty kaikilla reaaliluvuilla $n \in \mathbb{R}$. Lausekkeille esitettiin kombinatorinen tulkinta, kun n on positiivinen kokonaisluku.

Esimerkki 2.5 Kertoman ja binomikertoimen laskuesimerkkejä:

$$\begin{aligned} 3^{(5)} &= 3 \cdot 2 \cdot 1 \cdot 0 \cdot (-1) = 0 \\ (0.5)^{(4)} &= 0.5 \cdot (-0.5)(-1.5)(-2.5) = -0.9375 \\ \binom{3}{-1} &= 0 \qquad \text{määritelmän perusteella} \\ \binom{n}{0} &= \frac{n^{(0)}}{0!} = 1 \qquad \text{kaikilla } n \in \mathbb{R} \\ \binom{0.5}{4} &= \frac{0.5^{(4)}}{4!} = \frac{0.5 \cdot (-0.5)(-1.5)(-2.5)}{6} = -\frac{5}{128} \\ \binom{-2}{3} &= \frac{-2^{(3)}}{3!} = \frac{(-2)(-3)(-4)}{6} = -4. \end{aligned}$$

□

Määritelmien perusteella on suoraviivaista todeta, että

$$\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r},$$

$$r\binom{n}{r} = n\binom{n-1}{r-1}. \tag{2.4.5}$$

Jos s on epänegatiivinen kokonaisluku, niin silloin

$$r^{(s)}\binom{n}{r} = n^{(s)}\binom{n-s}{r-s}. \tag{2.4.6}$$

Stirlingin kaava

$$n! \approx \sqrt{2\pi n} \cdot n^n \mathrm{e}^{-n}. \tag{2.4.7}$$

antaa kertomalle hyvän likiarvon.

2.4.6 Binomilause, hypergeometrinen identiteetti ja multinomilause

Lause 2.5 (Binomilause) *Olkoon n mikä tahansa positiivinen kokonaisluku. Silloin*

$$(1+t)^n = \sum_{r=0}^{n} \binom{n}{r} t^r \tag{2.4.8}$$

kaikilla reaaliluvuilla $t \in \mathbb{R}$.

Kertoimia $\binom{n}{r}$ kutsutaan binomikertoimiksi. Binomisarja on binomilauseen yleistys.

Lause 2.6 (Binomisarja) *Olkoon $\alpha \in \mathbb{R}$ nollasta poikkeava reaaliluku. Silloin sarja*

$$(1+t)^\alpha = \sum_{r=0}^{\infty} \binom{\alpha}{r} t^r \tag{2.4.9}$$

suppenee kaikilla $|t| < 1$ ja hajaantuu, kun $|t| > 1$. Jos $\alpha > -1$, niin sarja suppenee myös pisteessä $t = +1$, ja jos $\alpha > 0$, niin sarja suppenee myös pisteessä $t = -1$.

Lause 2.7 (Hypergeometrinen identiteetti) *Olkoot a ja b reaalilukuja ja n positiivinen kokonaisluku. Silloin*

$$\sum_{r=0}^{\infty} \binom{a}{r}\binom{b}{n-r} = \binom{a+b}{n}. \tag{2.4.10}$$

Lause 2.8 (Multinomilause) *Olkoon annettu positiivinen kokonaisluku n ja reaaliluvut t_1, t_2, ..., t_k. Silloin*

$$(t_1 + t_2 + \cdots + t_k)^n = \sum\sum\cdots\sum \binom{n}{n_1 \; n_2 \; \ldots \; n_k} t_1^{n_1} t_2^{n_2} \cdots t_k^{n_k}, \tag{2.4.11}$$

missä summa käy yli kaikkien sellaisten epänegatiivisten kokonaislukujen n_1, n_2, ..., n_k, että $n_1 + n_2 + \cdots + n_k = n$.

2.4.7 Gammafunktio

Gammafunktio Γ määritellään integraalina

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \mathrm{e}^{-x} \, \mathrm{d}x \tag{2.4.12}$$

kaikilla positiivisilla reaaliluvuilla α. Epäoleellinen integraali (4.4.12) suppenee, kun $\alpha > 0$. Tällä funktiolla on tärkeitä sovelluksia ei vain tilastotieteessä ja todennäköisyyslaskennassa, vaan yleisemminkin sovelletussa matematiikassa.

Kun $\alpha > 1$, voidaan (4.4.12) lausua osittaisintegraalina

$$\int_0^\infty x^{\alpha-1} \mathrm{e}^{-x} \, \mathrm{d}x = \Big/_0^\infty (-x^{\alpha-1} \mathrm{e}^{-x}) + (n-1) \int_0^\infty x^{\alpha-2} \mathrm{e}^{-x} \, \mathrm{d}x,$$

josta seuraan relaatio

$$\Gamma(\alpha + 1) = \alpha \, \Gamma(\alpha). \tag{2.4.13}$$

Jos $\alpha = n$ on positiivinen kokonaisluku, niin

$$\Gamma(n+1) = n \, \Gamma(n) = n(n-1) \cdots 2 \cdot 1 \cdot \Gamma(1) = n! \, \Gamma(1).$$

Koska

$$\Gamma(1) = \int_0^\infty \mathrm{e}^{-x} \, \mathrm{d}x = \Big/_0^\infty (-\mathrm{e}^{-x}) = 1,$$

niin positiivisilla kokonaisluvuilla

$$\Gamma(n+1) = n!$$

Tämän ominaisuuden perusteella gammafunktiota kutsutaan joskus *yleistetyksi kertomaksi.* Voidaan myös osoittaa, että $\Gamma\big(\frac{1}{2}\big) = \sqrt{\pi}$, joten

$$\Gamma\left(n + \frac{1}{2}\right) = \left(n - \frac{1}{2}\right)\left(n - \frac{3}{2}\right) \cdots \frac{1}{2}\sqrt{\pi},$$

kun n on positiivinen kokonaisluku.

2.5 Satunnaismuuttuja

Satunnaiskokeiden tulokset esitetään tavallisesti numeeristen muuttujien avulla. Nämä muuttujat luonnehtivat tarkasteltavan satunnaiskokeen tuloksia. Tilastollisen tarkastelun kannalta onkin oleellista osata määritellä 'oikeat' muuttujat.

Määritelmä 2.2 Olkoon Ω jonkin satunnaiskokeen otosavaruus. *Satunnaismuuttuja* (SM) X on kuvaus (funktio) Ω:lta reaalilukujen joukkoon $\mathbb{R}$.

Satunnaismuuttujia merkitään isoilla kirjaimilla X, Y, Z, ... Voimme kirjoittaa

$$X\colon \Omega \to \mathbb{R},$$

missä $X(\omega)$ on reaaliluku. Satunnaismuuttuja X liittää siis jokaiseen alkeistapaukseen $\omega \in \Omega$ yhden ja vain yhden reaaliluvun $X(\omega) \in \mathbb{R}$. Satunnaismuuttujien $X, Y, Z, \ldots$ arvoja merkitään pienillä kirjaimilla x, y, z, ... Merkitään siis $X(\omega) = x$. Jos X:n arvojen muodostama joukko $S \subset \mathbb{R}$ (arvojoukko) on numeroituva (äärellinen tai ääretön), niin X on *diskreetti.* Jos otosavaruus Ω on numeroituva, niin Ω:lla määritellyt satunnaismuuttujat ovat välttämättä diskreettejä. Seuraavassa tarkastellaan diskreettejä satunnaismuuttujia.

Esimerkki 2.6 Heitetään harhatonta lanttia 3 kertaa. Satunnaismuuttuja X on 'kruunujen lukumäärä'. Merkitään R = 'kruunu' ja L = 'klaava'. Silloin

ω:	RRR	RRL	RLR	RLL	LRR	LRL	LLR	LLL
$X(\omega)$:	3	2	2	1	2	1	1	0

Silloin esimerkiksi $X(\text{RRL}) = X(\text{RLR}) = 2$. □

Esimerkki 2.7 Mielipidekyselyssä tiedustellaan 100:lta satunnaisesti valitulta suomalaiselta, millainen kanta heillä on Suomen NATO-jäsenyyteen. Mahdolliset kannanotot ovat: kannattaa (K), ei kantaa (E) ja vastustaa (V). Mahdollisten vastausten lukumäärä eli otosavaruuden koko on silloin 3^{100}. Jos olemme kuitenkin kiinnostuneita, esimerkiksi 'kannattajien lukumäärästä' X, niin silloin X:n mahdollinen arvojoukko on $\{0, 1, \ldots, 100\}$, jonka alkioiden lukumäärä on 101. Alkeistapaus ω on 100:n pituinen tyyppiä "KEVVVE...EV" oleva jono. Satunnaismuuttujan X arvo $X(\omega)$ on alkeistapauksesta ω laskettu kannattajien lukumäärä, esimerkiksi 36. □

Olemme jo edellä implisiittisesti soveltaneet satunnaismuuttujan käsitettä. Nopan ja lantin heittoon sekä korttipakkaan liittyvillä satunnaiskokeilla on perinteisesti havainnollistettu todennäköisyyslaskennan käsitteitä. Taulukossa 4.2 on esitetty muutamia tuttuja satunnaismuuttujia.
Määritelmässä 4.2 olemme todenneet, että satunnaismuuttujan arvot ovat reaalilukuja. Näin ei aina välttämättä ole. Esimerkiksi Taulukon 4.2 satunnaismuuttujan W arvo on valitun kortin 'maa'. Nämä arvot voidaan kuitenkin

Taulukko 2.2. Joitakin satunnaismuuttujia ja niiden arvoalueet.

Satunnais-muuttuja	Kuvaus	Arvojoukko S
X	Nopan silmäluku	$\{1, 2, 3, 4, 5, 6\}$
Y	Kruunujen lukumäärä 3:ssa lantin heitossa	$\{0, 1, 2, 3\}$
Z	Heittojen lukumäärä kunnes saadaan 1. kruunu	$\{1, 2, 3 \ldots\}$
W	Korttipakasta satunnaisesti valitun kortin maa	$\{\spadesuit, \heartsuit, \clubsuit, \diamondsuit\}$

aina tarvittaessa koodata numeerisesti. Joissain yhteyksissä tarkastelemme esimerkiksi satunnaispareja, satunnaisjonoja tai satunnaisjärjestyksiä. Näihin satunnaismuuttujan yleistyksiin palataan tuonnempana.

Huomautus 2.1 Jos X ja Y ovat satunnaismuuttujia, niin

$$aX, \qquad X+Y, \qquad X-Y, \qquad XY \qquad \text{ja} \qquad \frac{X}{Y} \quad (Y \neq 0)$$

ovat satunnaismuuttujia, missä a on reaaliluku. Nämä tulokset seuraavat siitä, että satunnaismuuttuja on funktio.

Matematiikan analyysin kursseilla opitun perusteella tiedämme, että *funktion funktio* on edelleen funktio:

$$x \to \sin(\log x) \qquad \text{tai} \qquad x \to f[h(x)] = (f \circ h)(x).$$

Yhdistetty satunnaismuuttuja on siis edelleen satunnaismuuttuja. Jos W (Taulukko 4.2) on esimerkiksi satunnaisesti valitun kortin maa ja V maiden joukossa $S = \{\spadesuit, \heartsuit, \clubsuit, \diamondsuit\}$ määritelty väri, niin satunnaismuuttujan *kortin väri* $V(W) = V[W(\omega)]$ arvoalue on $S_V = \{\text{musta, punainen}\}$. Korttipakan kortit (52 kpl) muodostavat alkeistapahtumien joukon ω. Olkoon Y kruunujen lukumäärä 3:ssa lantin heitossa (Taulukko 4.2). Silloin esimerkiksi

$$g(Y) = Y - \frac{3}{2} \qquad \text{tai} \qquad h(Y) = \left(Y - \frac{3}{2}\right)^2$$

ovat satunnaismuuttujia.

Kahden tai useamman satunnaismuuttujan funktio on edelleen satunnaismuuttuja. Jos siis X ja Y ovat satunnaismuuttujia, niin

$$\omega \to h[X(\omega), Y(\omega)]$$

määrittelee satunnaismuuttujan, kun kahden muuttujan funktio h on määritelty arvojoukossa $\{\,(X(\omega), Y(\omega)) \mid \omega \in \Omega\,\} \subset \mathbb{R}^2$. Tätä satunnaismuuttujaa merkitään lyhyesti $h(X, Y)$.

Määritelmä 2.3 (Indikaattorifunktio) Olkoon A tapahtuma otosavaruudessa Ω. Tapahtuman A indikaattorifunktio I_A saa arvon 0 tai 1 seuraavasti:

$$I_A(\omega) = \begin{cases} 1, & \text{jos } \omega \in A; \\ 0, & \text{jos } \omega \notin A. \end{cases}$$

Jos tapahtuma A sattuu, niin $I_A = 1$, muutoin $I_A = 0$. Indikaattorifunktio on satunnaismuuttuja ja

$$P(I_A = 1) = P(A) \qquad \text{ja} \qquad P(I_A = 0) = P(A^c) = 1 - P(A).$$

Voimme käyttää indikaattorifunktiota vaikkapa lukumäärien laskemiseen. Heitetään lanttia n kertaa ja olkoon X_k tapahtuman 'kruunu k. heitossa' $(1 \le k \le n)$ indikaattorifunktio. Silloin satunnaismuuttuja

$$X = X_1 + X_2 + \cdots + X_n \tag{2.5.1}$$

on kruunujen lukumäärä n:ssä heitossa, koska summa on ykkösten (kruunujen) lukumäärä n:ssä heitossa.

2.5.1 Satunnaismuuttujan jakauma

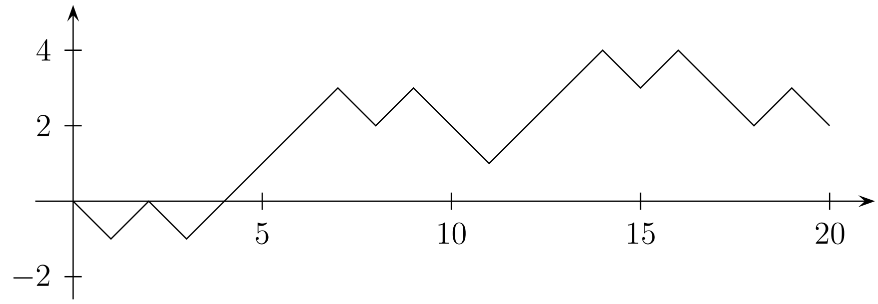
Olkoon Y kruunujen lukumäärä 3:ssa lantin heitossa (Esimerkki 4.6). Silloin satunnaismuuttujaa koskevat väittämät, kuten 'täsmälleen yksi kruunu' $\equiv$ "$Y = 1$" tai 'korkeintaan 2 kruunua' $\equiv$ "$Y \le 2$" määrittelevät tapahtuman. Tapahtumat voidaan silloin kirjoittaa muodossa "$Y \in A$". Jos $S = \{0, 1, 2, 3\}$, $A = \{1\}$ ja $B = \{0, 1, 2\}$, niin "$Y = 1$" $\equiv$ "$Y \in A$" ja "$Y \le 2$" $\equiv$ "$Y \in B$". Jatkossa tulemme pääsääntöisesti tarkastelemaan *satunnaismuuttujien määrittämiä tapahtumia.*

Tapahtuman todennäköisyyttä merkitään $P(Y \in B)$ tai lyhyesti $P(B)$. Merkintä $P(Y \in B)$ osoittaa, että tapahtuma on määritelty satunnaismuuttujan Y avulla. Koska B voi olla mikä tahansa Y:n arvoalueen S_Y osajoukko, todennäköisyydet $P(Y \in B)$ määrittelevät *satunnaismuuttujan jakauman.* Jos $y \in S_Y$, silloin Y:n arvon $Y = y$ todennäköisyys on $P(Y = y)$. Vastaavasti tapahtuman "$Y \in B$" todennäköisyys on

$$P(Y \in B) = \sum_{y \in B} P(Y = y).$$

Esimerkki 2.8 (Satunnaiskävely, Random Walk) Pekka ja Paavo pelaavat "kruunua ja klaavaa". Tässä pelissä heitetään peräkkäin lanttia n kertaa – tässä esimerkissä $n = 20$. Aina kun tulee kruunu (R), Pekka voittaa euron Paavolta. Kun tulee klaava (L), Pekka häviää euron Paavolle. Kuviossa 4.2 esitetyn pelin tulos $(n = 20)$ on

L R L R R R R L R L L R R R L R L L R L.



Kuvio 2.2. ”Kruunu ja klaava” -pelin tuloksen kehitys, kun pelin pituus on 20 heittoa.

Pekka voittaa 2 euroa.

Mikä on todennäköisyys, että Pekka voittaa s euroa, kun $n = 20$ ($-20 \leq s \leq 20$)? On helppo nähdä, että mahdollinen voitto on parillinen. Voitto S voidaan määritellä satunnaismuuttujien X_i ($i = 1, 2, \ldots, 20$) summana:

$$S_{20} = X_1 + X_2 + \cdots + X_{20},$$

missä

$$X_i = \begin{cases} 1, & \text{kun kruunu } i. \text{ heitossa;} \\ -1, & \text{kun klaava } i. \text{ heitossa.} \end{cases}$$

Minkä voiton $S_{20} = s$ todennäköisyys on suurin (pienin)?

On mielenkiintoista tarkastella myös sitä, kuinka usein Pekka on voitolla pelin aikana. Jos pelaajat ovat tasoissa (voitto 0), määrittelemme, että Pekka on johdossa, jos hän oli edellisellä heitolla johdossa. Jos Pekka oli tappiolla edellisellä heitolla ja pääsi tasoihin, sovimme, että hän on edelleen tappiolla. Jokainen peli tuottaa vastaavan kuvaajan kuin Kuviossa 4.2. Kuvaajassa on yhdistetty pisteet $(0, 0)$, $(1, S_1)$, $(2, S_2)$, $\ldots$, $(20, S_{20})$.

Tällaista prosessia kutsutaan satunnaiskävelyksi (random walk). Eräs tapa havainnollistaa satunnaiskävelyä on ajatella, että satunnaiskävelijä RW (Random Walker) lähtee origosta (itään) ja astuu sekunnissa askeleen oikealle (etelään) tai vasemmalle (pohjoiseen). Esimerkiksi Kuviossa 4.2 kuvaaja kulkee pisteen $(5, 1)$ kautta. RW on 5 sekunnin kävelyn jälkeen yhden askeleen pohjoiseen $x-$akselista. On helppo todeta, että kaikkien mahdollisten pelin kulkujen lukumäärä on 2^{20}. Koska raha on harhaton ja heitot ovat toisistaan riippumattomat, kaikki 2^{20} pelin kulkua ovat yhtä todennäköiset. $\square$

2.5.2 Kertymäfunktio

Edellä olemme käsitelleet otosavaruuden osajoukkojen eli tapahtumien todennäköisyyksiä. Nämä joukot määritellään tavallisesti satunnaismuuttujien avulla. Esimerkiksi

$$\{a \leq X \leq b\} = \{\, \omega \mid a \leq X(\omega) \leq b \,\}, \tag{2.5.2}$$

missä X on satunnaismuuttuja, a ja b ovat annettuja vakioita. Koska otosavaruus on numeroituva, kaikilla muotoa (4.5.2) olevilla joukoilla on todennäköisyys ja sitä merkitään

$$P(a \leq X \leq b) = P(\{\, \omega \mid a \leq X(\omega) \leq b \,\}).$$

Jos $A \subset \mathbb{R} = (-\infty, \infty)$ on jokin reaalilukujoukko, niin merkitään

$$P(X \in A) = P(\{\, \omega \mid X(\omega) \in A \,\}). \tag{2.5.3}$$

Joukko A voi olla esimerkiksi suljettu väli $[a, b]$, avoin väli (a, b) puoliavoin väli $(a, b]$ tai $[a, b)$, ääretön väli $(-\infty, b]$ tai $[a, \infty)$, usean välin yhdiste tai kokonaislukujen $\{k, k+1, \ldots, k+n\}$ joukko. Yhden pisteen x joukko $\{x\}$ on tärkeä erikoistapaus. Olkoon X esimerkiksi henkilön ikä, kun Ω on suomalaisten joukko. Silloin $\{X = 20\} = \{\, \omega \mid X(\omega) = 20 \,\}$ on 20-vuotiaiden suomalaisten joukko. Todennäköisyys on

$$P(X = 20) = P(X \in \{20\}) = P(\{\, \omega \mid X(\omega) = 20 \,\}).$$

Olkoon X:n arvojoukko $S_X = \{x_1, x_2, \ldots, x_n, \ldots, \}$, missä X:n arvot on lueteltu jossain järjestyksessä. Määritellään

$$p_n = P(X = x_n),$$

kun $x_n \in S_X$. Jos $x \notin S_X$, niin $P(X = x) = 0$. Toisaalta voi olla, että $P(X = x_i) = 0$ jollakin X:n arvolla $x_i \in S_X$. Jos tunnemme kaikki todennäköisyydet p_n, on ilmeistä, että voimme laskea kaikki satunnaismuuttujaan X liittyvät todennäköisyydet. Silloin todennäköisyydet (4.5.2) ja (4.5.3) ovat

$$P(a \leq X \leq b) = \sum_{a \leq x_n \leq b} p_n \qquad \text{ja} \qquad P(X \in A) = \sum_{x_n \in A} p_n.$$

Kun A on ääretön väli $(-\infty, x]$, niin jokaista reaalilukua x kohti voidaan määritellä funktio

$$F_X(x) = P(X \leq x) = \sum_{x_n \leq x} p_n.$$

Tätä funktiota F_X kutsutaan X:n *kertymäfunktioksi* (kf). Jokaiseen satunnaismuuttujaan X liittyy siis kertymäfunktio, jota merkitään $F_X(x)$.

Määritelmä 2.4 (Kertymäfunktio) Satunnaismuuttujan X kertymäfunktio $F_X(x)$ on sellainen kuvaus $F_X \colon \mathbb{R} \to [0, 1]$, että

$$F_X(x) = P(X \leq x), \qquad \text{kaikilla } x \in \mathbb{R}.$$

Merkintä $P(X \leq x)$ on lyhennys merkinnästä $P(\{X \leq x\})$, missä $P(\{X \leq x\}) = P(\{\, \omega \mid X(\omega) \leq x \,\})$. Merkitään X:n arvoaluetta $S = X(\Omega) = \{\, x \mid X(\omega),\ \omega \in \Omega \,\}$.

Esimerkki 2.9 Esimerkissä 4.6 heitettiin harhatonta lanttia 3 kertaa. Satunnaismuuttuja X on 'kruunujen lukumäärä' ja X:n arvojoukko $S = \{0, 1, 2, 3\}$. Nyt

$$\begin{aligned}
\{\,\omega \mid X = 0\,\} &= \{\mathrm{LLL}\},\\
\{\,\omega \mid X = 1\,\} &= \{\mathrm{RLL}, \mathrm{LRL}, \mathrm{LLR}\},\\
\{\,\omega \mid X = 2\,\} &= \{\mathrm{RRL}, \mathrm{LRR}, \mathrm{RLR}\},\\
\{\,\omega \mid X = 3\,\} &= \{\mathrm{RRR}\},
\end{aligned}$$

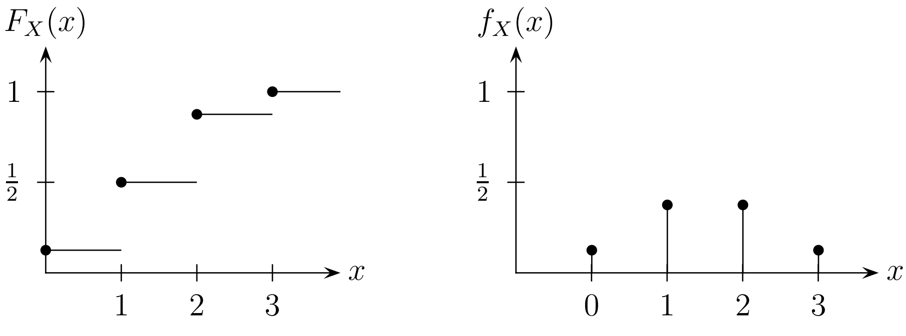
missä kaikki alkeistapaukset ovat yhtä todennäköisiä. Silloin

$$\begin{aligned}
F_X(0) &= P(X \le 0) = P(\{\mathrm{LLL}\}) = 1/8,\\
F_X(1) &= P(X \le 1) = P(\{\mathrm{LLL}, \mathrm{RLL}, \mathrm{LRL}, \mathrm{LLR}\}) = 4/8,\\
F_X(2) &= P(X \le 2)\\
&= P(\{\mathrm{LLL}, \mathrm{RLL}, \mathrm{LRL}, \mathrm{LLR}, \mathrm{RRL}, \mathrm{LRR}, \mathrm{RLR}\}) = 7/8,\\
F_X(3) &= P(X \le 3)\\
&= P(\{\mathrm{LLL}, \mathrm{RLL}, \mathrm{LRL}, \mathrm{LLR}, \mathrm{RRL}, \mathrm{LRR}, \mathrm{RLR}, \mathrm{RRR}\}) = 1.
\end{aligned}$$

Satunnaismuuttujan X kertymäfunktio on siis

$$F_X(x) = \begin{cases} 0, & \text{kun } x < 0;\\ \frac{1}{8}, & \text{kun } 0 \le x < 1;\\ \frac{4}{8}, & \text{kun } 1 \le x < 2;\\ \frac{7}{8}, & \text{kun } 2 \le x < 3;\\ 1, & \text{kun } x \ge 3. \end{cases}$$

□



Kuvio 2.3. Satunnaismuuttujan X kertymäfunktion $F_X(x)$ ja todennäköisyysfunktion $f_X(x)$ kuvaajat.

Jos $x_1 \le x_2$, niin $\{X \le x_1\} \subset \{X \le x_2\}$ ja todennäköisyyden monotonisuusominaisuuden perusteella [Lause (4.1), kohta 3)] $P(X \le x_1) \le P(X \le x_2)$, joten kertymäfunktio $F(x)$ on *kasvava* (ei vähenevä). Seuraavassa lauseessa esitetään kertymäfunktion ominaisuudet.

Lause 2.9 *Funktio $F(x)$ on satunnaismuuttujan X kertymäfunktio, jos ja vain jos seuraavat kolme ehtoa pitävät paikkansa:*

1. $\lim_{x \to -\infty} F(x) = 0$ *ja* $\lim_{x \to \infty} F(x) = 1$.
2. $F(x)$ *on x:n kasvava funktio.*
3. $F(x)$ *on oikealta jatkuva, ts. kaikilla $x_0 \in \mathbb{R}$ on* $\lim_{x \to x_0+} F(x) = F(x_0)$ ($x \to x_0+$ *tarkoittaa, että x_0:aa lähestytään oikealta*).

Esimerkissä 4.9 esitetty kertymäfunktio on porrasfunktio. Pitää yleisestikin paikkansa, että diskreetin satunnaismuuttujan kertymäfunktio on porrasfunktio. Voimme sanoa, että satunnaismuuttuja on diskreetti satunnaismuuttuja, jos sen kertymäfunktio on porrasfunktio.

2.5.3 Diskreetin satunnaismuuttujan todennäköisyysfunktio

Olkoon X otosavaruudessa Ω määritelty diskreetti satunnaismuuttuja. Satunnaismuuttujan X todennäköisyysfunktio (tnf) $f_X(x)$ määritellään siten, että

$$f_X(x) = P(X = x). \tag{2.5.4}$$

Jos merkitään X:n arvoaluetta $S_X = X(\Omega) = \{\, x \mid X(\omega) = x,\ \omega \in \Omega \,\}$, niin todennäköisyysfunktio on kuvaus

$$f_X(x)\colon S_X \to [0, 1].$$

Huomattakoon, että $f_X(x)$ on määritelty kaikilla reaaliluvuilla, mutta $f_X(x) = 0$ aina, kun $x \notin S$. Diskreetin satunnaismuuttujan arvojoukko on numeroituva, joten arvoja on korkeintaan yhtä paljon kuin kokonaislukuja. Niinpä diskreettien satunnaismuuttuien arvojoukko on tavallisimmin jokin kokonaislukujen ja erityisesti positiivisten kokonaislukujen osajoukko.

Olkoon X diskreetti satunnaismuuttuja, jonka arvoalue on S_X. Silloin X:n todennäköisyysfunktio toeteuttaa seuraavat ehdot:

$$f_X(x) > 0, \qquad \text{kun} \quad x \in S_X; \tag{2.5.5}$$

$$f_X(x) = 0, \qquad \text{kun} \quad x \notin S_X; \tag{2.5.6}$$

$$\sum_{x \in S_X} f_X(x) = 1. \tag{2.5.7}$$

Esimerkki 2.10 Jatketaan esimerkiä 4.9, jossa heitettiin harhatonta lanttia 3 kertaa. Satunnaismuuttuja X on 'kruunujen lukumäärä'. Määritetään X:n todennäköisyysfunktio. Nyt

$$
\begin{aligned}
X^{-1}(0) &= \{\mathrm{LLL}\}, \\
X^{-1}(1) &= \{\mathrm{RLL}, \mathrm{LRL}, \mathrm{LLR}\}, \\
X^{-1}(2) &= \{\mathrm{RRL}, \mathrm{LRR}, \mathrm{RLR}\}, \\
X^{-1}(3) &= \{\mathrm{RRR}\},
\end{aligned}
$$

missä merkintä $X^{-1}(x) = \{\, \omega \mid X(\omega) = x \,\}$ on kaikkien sellaisten alkeistapausten ω joukko, jotka kuvautuvat pisteeseen x. Koska alkeistapaukset ovat yhtä todennäköisiä, satunnaismuuttujan arvojen todennäköisyydet ovat

$$
\begin{aligned}
P(X=0) &= P(X^{-1}(0)) = P(\{\mathrm{LLL}\}) = 1/8, \\
P(X=1) &= P(X^{-1}(1)) = P(\{\mathrm{RLL}, \mathrm{LRL}, \mathrm{LLR}\} = 3/8, \\
P(X=2) &= P(X^{-1}(2)) = P(\{\mathrm{RRL}, \mathrm{LRR}, \mathrm{RLR}\} = 3/8, \\
P(X=3) &= P(X^{-1}(3)) = P(\{\mathrm{RRR}\} = 1/8.
\end{aligned}
$$

Satunnaismuuttujan X todennäköisyysfunktio on siis

$$
f_X(x) = \begin{cases} \frac{1}{8}, & \text{kun } x = 0; \\ \frac{3}{8}, & \text{kun } x = 1; \\ \frac{3}{8}, & \text{kun } x = 2; \\ \frac{1}{8}, & \text{kun } x = 3; \\ 0, & \text{muutoin.} \end{cases}
$$

□

Jos X on diskreetti satunnaismuuttuja, niin X:n arvojoukko $S_X = \{x_1, x_2, \ldots\}$. Silloin X:n kertymäfunktio voidaan lausua todennäköisyysfunktion avulla seuraavasti:

$$
F_X(x) = \sum_{x_i \le x} f_X(x_i),
$$

joka on porrasfunktio. Porrasfuntion $F(x)$ 'hyppäykset' ovat pisteissä x_1, x_2, ... ja hyppäysten suuruudet ovat $f_X(x_1)$, $f_X(x_2)$, ... Jos esimerkiksi

$$
f(x) = \frac{1}{2^x}, \qquad x = 1, 2, \ldots,
$$

niin esimerkiksi

$$
F(3.5) = P(x \le 3.5) = f(1) + f(2) + f(3) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}.
$$

2.5.4 Diskreetti tasajakauma

Olkoon X satunnaismuuttuja, jonka arvojoukko on $S = \{x_1, x_2, \ldots, x_N\}$. Jos

$$
f(x_k) = \frac{1}{N} \qquad \text{kaikilla } k = 1, 2, \ldots, N,
$$

niin X noudattaa diskreettiä tasajakaumaa ja merkitään $X \sim \mathrm{Tasd}(x_1, x_2, \ldots, x_N)$. Hyvin usein X:n arvojoukko on $S = \{1, 2, \ldots, N\}$, jolloin merkitään $X \sim \mathrm{Tasd}(1, 2, \ldots, N)$. Esimerkiksi nopanheitossa silmäluvun X arvojoukko on $S = \{1, 2, 3, 4, 5, 6\}$ ja todennäköisyysfunktio

$$f(x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6.$$

2.6 Otanta palauttamatta

Tarkastellaan nyt koetta, jossa valitaan n alkiota N:n alkion joukosta ($n \leq N$), jota kutsutaan *populaatioksi.* Valintaprosessia kutsutaan *otannaksi.* Halutaan esimerkiksi tietää ennen presidentin vaaleja, mikä on ehdokkaiden kannatus. Kannatuksesta voi saada tietoa tiedustelemalla äänestäjiltä, ketä he aikovat äänestää. On käytännössä mahdotonta haastatella kaikkia potentiaalisia äänestäjiä. Siksi tehdään otos, eli valitaan vain osa mahdollisista äänestäjistä ja haastatellaan heidät. Populaation muodostavat siis äänioikeutut kansalaiset. Nimitämme menetelmää, jolla otos valitaan, *otantamenetelmäksi.* Tässä esityksessä tarkastellaan vain *yksinkertaista satunnaisotantaa* (YSO). YSO:ssa kaikki mahdolliset n:n kokoiset otokset ovat yhtä todennäköisiä. YSO:lla valittua otosta kutsutaan *yksinkertaiseksi satunnaisotokseksi.*

Yksinkertaisessa satunnaisotannassa *palauttamatta* otos valitaan siten, että kukin alkio voi tulla otokseen korkeintaan kerran. Valitaan n:n alkion otos N:stä. Ajatellaan alkiot valituiksi järjestyksessä. Silloin 1. alkio voidaan valita N:llä tavalla ja 2. alkio $(N-1)$:llä tavalla, koska toisen täytyy olla eri alkio kuin ensimmäinen, jne. Lopulta n. alkio voidaan valita $[N-(n-1)]$:llä tavalla. Kaikkien mahdollisten *järjestettyjen otosten* lukumäärä on

$$N(N-1)(N-2)\cdots(N-n+1) = N^{(n)}.$$

Otos on *valittu satunnaisesti,* jos jokainen $N^{(n)}$:stä järjestetystä otoksesta on yhtä todennäköinen. Silloin jokaisen järjestetyn otoksen todennäköisyys on $1/N^{(n)}$.

Koska kaikkien mahdollisten otosten eli osajoukkojen lukumäärä on $\binom{N}{n}$ ja otokset oletetaan yhtä todennäköisiksi, niin jokaisen otoksen todennäköisyys on

$$\frac{1}{\binom{N}{n}}, \qquad \text{missä otosten lukumäärä on } \binom{N}{n} = \frac{N^{(n)}}{n!}.$$

Otoksia on siis $\binom{N}{n}$ kappaletta ja YSO:ssa ne ovat yhtä todennäköisiä.

Esimerkki 2.11 Monissa korttipeleissä jaetaan n:n kortin käsi (otos) pakasta (populaatio), jossa on N korttia. Pakka on *hyvin sekoitettu,* jos pakan korttien kaikki $N!$ järjestystä ovat yhtä todennäköisiä. Oletetaan, että n korttia on jaettu hyvin sekoitetusta pakasta. Sellaisia pakan järjestyksiä, joissa nämä n korttia ovat tietyssä järjestyksessä (esimerkiksi pakan päällä), on $(N-n)!$ kappaletta. Todennäköisyys saada n korttia tietyssä järjestyksessä on

$$\frac{(N-n)!}{N!} = \frac{1}{N^{(n)}}.$$

Jokaisen järjestetyn otoksen todennäköisyys on siis $1/N^{(n)}$ ja jokaisen otoksen eli käden todennäköisyys on $1/\binom{N}{n}$.

Mikä on esimerkiksi todennäköisyys, että tavallisesta korttipakasta ($N = 52$) saadaan patasuora ($\spadesuit 1, \spadesuit 2, \spadesuit 3, \spadesuit 4, \spadesuit 5$)? Erilaisten viiden käsien lukumäärä on $\binom{52}{5} = 2598960$, joten patasuoran todennäköisyys on $1/2598960$. Jos lisäksi korttien pitää jaossa tulla annetussa järjestyksessä $(1, 2, 3, 4, 5)$, niin järjestettyjen otosten lukumäärä $52^{(5)} = 311875200$ ja järjestetyn otoksen ($\spadesuit 1, \spadesuit 2, \spadesuit 3, \spadesuit 4, \spadesuit 5$) todennäköisyys on $1/311875200$. $\square$

2.6.1 Hypergeometrinen jakauma

Oletetaan, että populaatiossa on $a + b$ alkiota – esimerkiksi väestössä on a miestä ja b naista tai tuotepopulaatiossa on a viallista ja b virheetöntä tuotetta. Valitaan populaatiosta n:n kokoinen satunnaisotos palauttamatta. Mikä on todennäköisyys, että otokseen tulee x kappaletta tyyppiä 1 olevia alkiota ja $n - x$ kappaletta tyyppiä 2? Tavanomainen todennäköisyyslaskennassa käytetty satunnaiskoe on pallojen valinta uurnasta. Uurnassa on a valkoista ja b mustaa palloa. Valitaan uurnasta satunnaisesti palauttamatta n palloa. Mikä on todennäköisyys, että otokseen tulee x valkoista palloa?

Taulukko 2.3. Valinta palauttamatta äärellisestä populaatiosta

	Tyyppi 1	Tyyppi 2	Yhteensä
Populaatio	a	b	$a+b$
Otos	x	$n-x$	n

Populaatiosta voidaan valita kaikkiaan $\binom{a+b}{n}$ yhtä todennäköistä n:n kokoista otosta palauttamatta. Koska a:sta tyyppiä 1 olevasta alkiosta voidaan valita x kappaletta $\binom{a}{x}$ tavalla ja $n - x$ alkiota b:sta $\binom{b}{n-x}$ tavalla, saadaan kaikkiaan $\binom{a}{x}\binom{b}{n-x}$ sellaista otosta, joissa on x kappaletta tyyppiä 1 ja $n - x$ kappaletta tyyppiä 2 olevaa alkiota. Olkoon nyt satunnaismuuttuja X tyyppiä 1 olevien alkioiden lukumäärä otoksessa. Silloin satunnaismuuttujan X todennäköisyysfunktio $f(x)$ on

$$f(x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}, \qquad x = 0, 1, 2, \ldots \tag{2.6.1}$$

Edellä esitetystä otanta-asetelmasta tietysti seuraa, että ehtojen $x \leq a$ ja $n - x \leq b$ täytyy olla voimassa. Jos ehdot eivät ole voimassa, niin $f(x) =$

0. Jakaumaa (4.6.1) kutsutaan *hypergeometriseksi jakaumaksi.* Se on tärkeä myös esimerkiksi joidenkin ns. tarkkojen testien konstruoinnissa.

2.6.2 Tarkistusotanta teollisuudessa

Mikään teollisuusprosessi ei ole täydellinen, siksi myös virheellisiä tuotteita on odotettavissa. Yrityksillä on käytössä erilaisia laadunvarmistusjärjestelmiä, jotta voitaisiin pitää yllä riittävän hyvä laatu. Virheelliset tuotteet olisi havaittava ja poistettava, jotta ne eivät joutuisi asiakkaalle saakka. Tietysti voitaisiin tarkistaa jokainen tuote riittävän tarkasti. Täydellinen tarkistus ei ole käytännössä yleensa realistinen – se ei ole taloudellisesti kannattavaa tai se on jopa mahdotonta, jos tarkistus esimerkiksi tuhoaa tuotteen. On siis yleensä käytettävä tarkistusotantaa.

Oletetaan, että tuotteet ovat joko virheellisiä tai hyväksyttäviä ja ne tulevat laadun tarkistukseen N:n tuotteen erissä. Valitaan jokaisesta erästä satunnaisesti n tuotetta tarkistukseen. Oletetaan, että löydetään x viallista. Jos x on suuri, todennäköisesti erässä on paljon viallisia ja erä pitäisi hylätä tai panna jatkotarkistukseen. Voimme käyttää päätössääntöä:

Hyväksy erä, jos $x \leq h$, muutoin hylkää erä (tai testaa lisää).

Nyt olisi valittava hyväksymisraja h mahdollisimman viisaasti. On tietysti mahdollista, että otoksessa $x > h$, vaikka viallisten lukumäärä v tuote-erässä ei olisikaan "liian" suuri. Toisaalta voi ehto $x \leq c$ toteutua, vaikka tuote-erässä olisi "liikaa" viallisia. Edellä mainitut päätäntävirheet ovat siis seuraavat:

1. lajin virhe – erä, jossa on vähän viallisia, hylätään;

2. lajin virhe – erä, jossa on paljon viallisia, hyväksytään.

Jos hyväksymisrajaa h kasvatetaan, pienenee 2. lajin virhe, mutta 1. lajin virhe kasvaa. Molempia virheitä voidaan pienentää samanaikaisesti, kasvattamalla otoskokoa n, mutta se taas nostaa tarkistuskustannuksia. Jotta h:n ja n:n arvot voitaisiin määrittää optimaalisesti, olisi tunnettava tarkistuskustannukset sekä 1. ja 2. lajin virheiden aiheuttamat kustannukset. Olisi myös tiedettävä virheelisten lukumäärän v jakauma yli tuote-erien.

2.7 Otanta palauttaen

Valitaan n:n alkion otos populaatiosta, jossa on N alkiota. Ajattelemme, että populaation alkiot on numeroitu juoksevasti $\{1, 2, \ldots, N\}$. Otannassa palauttaen populaation alkio voidaan valita otokseen useammin kuin kerran. On esimerkiksi mahdollista, että otokseen tulee sama alkio toistuvasti n kertaa. Voimme ajatella valinnan prosessina, jossa alkiot valitaan peräkkäin. Jokaisen valinnan jälkeen alkio palautetaan populaatioon, mutta sitä ennen saatu alkio merkitään muistiin. Silloin 1. alkio voidaan valita N:llä tavalla, 2. alkio myös N:llä tavalla ja lopulta n. alkio N:llä tavalla, koska edellisissä valinnoissa valitut voivat tulla uudestaan otokseen. Kaikkien mahdollisten palauttaen valittujen *järjestettyjen otosten* lukumäärä on siis N^n. Sanomme, että otos on *valittu satunnaisesti palauttaen,* jos kaikki mahdolliset N^n järjestettyä jonoa ovat yhtä todennäköiset. Näin valittu otos on *yksinkertainen satunnaisotos* (*YSO*) *palauttaen.*

Oletetaan esimerkiksi, että valitaan kolme numeroa palauttaen numeroista $0, 1, 2, \ldots, 9$. Silloin voidaan saada $10^3 = 1000$ yhtä mahdollista järjestettyä jonoa $000, 001, 002, \ldots, 999$. Osajoukko $\{1, 2, 3\}$ voidaan valita $3! = 6$ tavalla, joten otoksen $\{1, 2, 3\}$ todennäköisyys on 0.006. Otos $\{1, 1, 3\}$ voidaan saada 3:lla tavalla, koska järjestetyt jonot $(1, 1, 3)$, $(1, 3, 1)$ ja $(3, 1, 1)$ sisältävät samat alkiot. Otoksen $\{1, 1, 3\}$ todennäköisyys on 0.003. Otos $\{1, 1, 1\}$ saadaan vain yhdellä tavalla, joten sen todennäköisyys on 0.001. Otannassa palauttaen (järjestämättömät) otokset *eivät ole yhtä todennäköisiä* kuten otannassa palauttamatta.

Olkoon A_i tapahtuma, että valitaan i. alkio, $i = 1, 2, \ldots, N$. Koska valinnan (kokeen) tulos on varmasti yksi ja vain yksi tapahtumista A_1, A_2, $\ldots$, A_N, niin $\Omega = A_1 \cup A_2 \cup \cdots \cup A_N$ on kokeeseen (valinta palauttaen) liittyvän otosavaruuden ositus. Valinta toistetaan n kertaa. Oletetaan, että populaation i. alkio toistuu otoksessa n_i $(0 \leq n_i \leq n)$ kertaa $(i = 1, 2, \ldots, N)$. Silloin $\sum_{i=1}^{n} n_i = n$ ja erilaisten järjestettyjen otosten lukumäärä on tuloksen (4.4.1) mukaan

$$\binom{n}{n_1\ n_2\ \ldots\ n_N} = \frac{n!}{n_1!\, n_2! \cdots n_N!}.$$

Olkoon X_i alkion i toistojen lukumäärä otoksessa. Nyt siis jokaisen X_i:n arvoalue on $\{0, 1, 2, \ldots, n\}$ ja $X_1 + X_2 + \cdots + X_N = n$. Merkitään todennäköisyyttä $P(X_1 = n_1, X_2 = n_2, \ldots, X_N = n_N)$ yksinkertaisesti $P(n_1, n_2, \ldots, n_N)$, joka voidaan siis laskea kaavalla

$$P(n_1, n_2, \ldots, n_N) = \binom{n}{n_1\ n_2\ \ldots\ n_N} \frac{1}{N^n}.$$

Esimerkki 2.12 Valitaan populaatiosta $\{A_1, A_2, A_3\}$ $(N = 3)$ 5 kertaa $(n = 5)$ alkio palauttaen. Silloin $A_1A_1A_3A_1A_3$ on eräs mahdollinen tulosjono (otos palauttaen), missä $X_1 = 3$, $X_2 = 0, X_3 = 2$ ja $X_1 + X_2 + X_3 = 5$. Jonon $A_1A_1A_3A_1A_3$ todennäköisyys, samoin kuin jokaisen viiden pituisen järjestetyn otoksen, todennäköisyys on $1/3^5$. Koska erilaisia tulosjonoja, joissa $X_1 = 3$, $X_2 = 0$ ja $X_3 = 2$, on

$$\binom{5}{3\ 0\ 2} = \frac{5!}{3!\, 0!\, 2!} = 10,$$

niin

$$P(X_1 = 3, X_2 = 0, X_3 = 2) = \binom{5}{3\ 0\ 2} \frac{1}{3^5} = \frac{10}{243} = 0.04115.$$

Mikä on todennäköisyys, että n:n kokoiseen järjestettyyn otokseen tulee populaation n ensimmäistä alkiota ($n \leq N$) missä tahansa järjestyksessä? Kyseinen tapahtuma sattuu täsmälleen silloin, kun $X_1 = X_2 = \cdots = X_n = 1$ ja $X_{n+1} = \cdots = X_N = 0$. Tämän tapahtuman todennäköisyys on siis

$$P(1, 1, \ldots, 1, 0, \ldots, 0) = \frac{n!}{(1!)^n (0!)^{N-n}} \frac{1}{N^n} = \frac{n!}{N^n}.$$

$\square$

Otoksen kaikki alkiot erilaisia

Sellaisia järjestettyjä otoksia, joissa mikään alkio ei toistu, on

$$N^{(n)} = N(N-1) \cdots (N-n+1)$$

kappaletta. Jos otos valitaan palauttaen, niin todennäköisyys, että otoksessa mikään alkio ei toistu, on

$$P(\text{'Sama ei toistu'}) = \frac{N^{(n)}}{N^n} = \frac{N!}{(N-n)!\, N^n}. \tag{2.7.1}$$

On selvää, että todennäköisyys (4.7.1) on 0, jos $n > N$. Huomaa, että $N^{(n)} = 0$, jos $n > N$. Syntymäpäiväongelmassa (Esimerkki 4.3) $N = 365$ ja $n = r$.

Soveltamalla Stirlingin kaavaa (4.4.7) kertoimiin $N!$ ja $(N-1)!$ saadaan likiarvo

$$\frac{N!}{(N-n)!\, N^n} \approx \left(\frac{N}{N-n} \right)^{N-n+1/2} \mathrm{e}^{-n}. \tag{2.7.2}$$

Kun $N \to \infty$ ja n on kiinnitetty, niin lauseke (4.7.2) lähestyy ykköstä. Jos siis hyvin suuresta populaatiosta valitaan n alkiota ($n \ll N$) palauttaen, niin on hyvin epätodennäköistä, että sama alkio valitaan usemmin kuin kerran. Otanta palauttaen ja palauttamatta ovat käytännöllisesti katsoen jokseenkin identtiset, kun populaation koko N on paljon suurempi kuin otoskoko n.

2.8 Binomijakauma

Oletetaan, että populaatiossa on kahdenlaisia alkioita: a kappaletta tyyppiä A ja b kappaletta tyyppiä B. Valitaan populaatiosta n alkiota palauttaen. Mikä on todennäköisyys, että otokseen tulee x alkiota tyyppiä A ja $n-x$ alkiota tyyppiä B? Voimme käyttää vastavaa uurnamallia kuin hypergeometrisen jakauman yhteydessä. Uurnassa on a valkoista ja b mustaa palloa. Valitaan uurnasta satunnaisesti palauttaen n palloa. Mikä on todennäköisyys, että otokseen tulee x valkoista palloa? Koska otanta tehdään palauttaen, uurnan sisältö ei muutu.

Taulukko 2.4. Otanta palauttaen

	Tyyppi A	Tyyppi B	Yhteensä
Populaatio	a	b	$a+b$
Otos	x	$n-x$	n

Kaikkien mahdollisten n:n kokoisten yhtä todennäköisten järjestettyjen jonojen lukumäärä on $(a+b)^n$. Sellaisia järjestettyjä otoksia, joissa on ensin x kappaletta tyyppiä A olevia alkioita ja sitten $n-x$ tyyppiä B, on $a^x b^{n-x}$. Tyyppiä A olevan x:n alkion paikka n:n pituisessa jonossa voidaan valita $\binom{n}{x}$ tavalla. Otoksia (järjestämättömiä), joissa on x kappaletta tyyppiä A ja $n-x$ kappaletta tyyppiä B olevia alkioita, on $\binom{n}{x}a^x b^{n-x}$ kappaletta. Olkoon satunnaismuuttuja X tyyppiä A olevien alkioiden lukumäärä otoksessa. Silloin X:n todennäköisyysfunktio on

$$f(x) = \binom{n}{x}\frac{a^x b^{n-x}}{(a+b)^n}, \qquad x = 0, 1, 2, \ldots, n.$$

Merkitään tyyppiä A olevien alkioiden suhteellista osuutta $p = \frac{a}{a+b}$ ja $1-p = \frac{b}{a+b}$ on tyyppiä B olevien suhteellinen osuus. Nyt X:n todennäköisyysfunktio voidaan kirjoittaa sen tavallisimmassa esitysmuodossa

$$f(x) = \binom{n}{x}p^x(1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n. \tag{2.8.1}$$

Funktio (4.8.1) on *binomijakauman* todennäköisyysfunktio. Kun X noudattaa binomijakaumaa, merkitsemme $X \sim \mathrm{Bin}(n,p)$.

2.8.1 Binomijakauma hypergeometrisen jakauman likiarvona

Kun populaation koko on paljon suurempi kuin otoskoko, on tuloksen kannalta jokseenkin samantekevää, tehdäänkö otanta palauttaen vai palauttamatta. Kun $a+b$ on paljon suurempi kuin n (merkitään $a+b \gg n$), niin binomijakauma (4.8.1) on hypergeometrisen jakauman (4.6.1) hyvä likiarvo. Otanta palauttamatta voidaan luonnehtia hypergeometrisen jakauman avulla ja otanta palauttaen binomijakauman avulla.

Lause 2.10 *Jos $a+b \gg n$, niin*

$$\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}} \approx \binom{n}{x}p^x(1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, \tag{2.8.2}$$

missä $p = a/(a+b)$.

Koska binomitodennäköisyydet on helpompi laskea kuin hypergeometriset todennäköisyydet, voidaan relaatiota (4.8.2) käyttää laskennassa hyväksi, kun $a+b \gg n$. Tosin nykyisillä ohjelmilla on helppo laskea tarkat todennököisyydet suoraan hypergeometrisesta jakaumasta, vaikka $a+b$ on suuri.

2.9 Todennäköisyyden yleiset aksioomat

Numeroituvien otosavaruuksien tapauksessa on kaikkiin tapahtumiin helppo liittää todennäköisyydet alkeistapahtumien todennäköisyyksien avulla. Todennäköisyyden keskeiset ominaisuudet (Lause 4.1) seuraavat sitten helposti Määritelmästä 1.1. Jos tapahtumat A_1, A_2, …, A_n ovat parittain erilliset, niin Lauseen 4.2 mukaan

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

Additiivisuus voidaan todistaa (numeroituvan otosavaruuden Ω tapauksessa) myös äärettömän monelle parittain erilliselle tapahtumalle A_1, A_2, A_3, …(vrt. Pykälä 1.3.4). Silloin

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots .$$

Ylinumeroituvasti äärettömille otosavaruuksilla ei todennäköisyyttä voida määritellä alkeistapausten todennäköisyyksien avulla. Silloin todennäköisyys määritellään aksiomaattisesti sopivasti määritellyille otosavaruuden Ω osajoukoille.

Määritelmä 2.5 Olkoon Ω otosavaruus ja $P(A)$ on Ω:n sopivasti valituilla osajoukoilla $A \subset \Omega$ määritelty reaaliarvoinen funktio. Funktio $P(A)$ on todennäköisyysmitta, jos se toteuttaa seuraavat aksioomat:

1. $0 \leq P(A) \leq 1$.
2. $P(\emptyset) = 0 \quad \text{ja} \quad P(\Omega) = 1$.
3. Jos $A_1, A_2, A_3, \ldots$ on parittain pistevieraitten Ω:n osajoukkojen jono, niin
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Näistä aksioomista voidaan johtaa samat lauseet kuin numeroituvien otosavaruuksien tapauksessa. Yleisessä teoriassa Ω:n *kaikki osajoukot eivät ole tapahtumia.* Määrittelimme edellä todennäköisyyden Ω:n 'sopivasti valituille osajoukoille'. Niiden osajoukkojen, jotka ovat tapahtumia, täytyy muodostaa ns. *σ-algebra.*

Määritelmä 2.6 Kokoelma $\mathcal{F}$ on otosavaruuden Ω osajoukkojen muodostama σ-algebra, jos

1. $\Omega \in \mathcal{F}$.
2. Jos $A \in \mathcal{F}$, niin $A^c \in \mathcal{F}$.
3. Jos $A_1, A_2, \ldots \in \mathcal{F}$, niin $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Kolmikko $(\Omega, \mathcal{F}, P)$ on *todennäköisyysavaruus,* missä Ω on ei-tyhjä otosavaruus, $\mathcal{F}$ on σ-algebra ja $P\colon \mathcal{F} \to [0,1]$ on todennäköisyys(mitta). Tämän todennäköisyyden aksiomatisoinnin esitti venäläinen matemaatikko A. N. Kolmogorov (1903–87) vuonna 1929.

Todennäköisyyslaskenta ja kombinatoriikka: Yhteenveto

Todennäköisyyden ominaisuuksia

- Epänegatiivisuus $\quad P(A) \geq 0, \quad A \subset \Omega.$
- Monotonisuus $\quad P(A) \leq P(B), \quad \text{kun } A \subset B \subset \Omega.$
- Additiivisuus $\quad P(A) = \sum_{i=1}^{n} P(A_i), \quad$ jos $A_1,\ A_2,\ \ldots A_n$ on A:n jako.
- Komplementti $\quad P(A^c) = 1 - P(A).$
- Yhteenlaskulause $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B).$

Symmetriaan perustuva todennäköisyys

$$p(\omega_i) = \frac{1}{n}, \qquad \text{kaikilla } \omega_i \in \Omega = \{\omega_1, \omega_2, \ldots, \omega_n\};$$

$$P(A) = \sum_{\omega_i \in A} p(\omega_i) = \frac{|A|}{n} = \frac{\text{'suotuisat'}}{\text{'kaikki'}}.$$

Kombinatoriikkaa

- Järjestettyjen n-otosten lukumäärä, kun perusjoukon koko on N:

  1) valinta paluttaen: $\quad N^n,$
  2) valinta palauttamatta: $\quad N^{(n)} = N(N-1)(N-2)\cdots(N-n+1), \ 0 \leq n \leq N,$
     $N^{(N)} = N!.$

- Otokset (palauttamatta) eli n-kombinaatiot

$$\binom{N}{n} = \frac{N^{(n)}}{n!} = \frac{N!}{n!\,(N-n)!}.$$

- Multinomikerroin

$$\binom{n}{n_1\ n_2\ \dots\ n_k} = \frac{n!}{n_1!\,n_2!\cdots n_k!}.$$

- Binomilause

$$(1+t)^n = \sum_{r=0}^{n} \binom{n}{r} t^r,$$

kaikilla $t \in \mathbb{R}$ ja positiivisilla kokonaisluvuilla n.

Satunnaismuuttuja

- Satunnaismuuttuja X on kuvaus $X : \Omega \to \mathbb{R}$.
- X:n arvojoukko S: $X(\omega) \in S \subset \mathbb{R}$.
- Jos S on numeroituva, niin X on diskreetti satunnaismuuttuja.
- Jos X ja Y ovat satunnaismuuttujia, niin

$$aX, \qquad X+Y, \qquad X-Y, \qquad XY \qquad \text{ja} \qquad \frac{X}{Y}, \quad Y \neq 0$$

ovat satunnaismuuttujia, missä a on reaalivakio.

- Diskreetin satunnaismuuttujan X jakauma

$$P(X \in A) = \sum_{x \in A} P(X = x) \qquad \text{kaikilla } A \subset S.$$

- X:n kertymäfunktio F: $F(x) = P(X \leq x)$.
- Jos X:n todennäköisyysfunktio f on diskreetti, niin $f(x) = P(X = x)$.
- Hypergeometrinen jakauman todennäköisyysfunktio

$$f(x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}, \qquad x \leq a \quad \text{ja} \quad n - x \leq b.$$

- Binomijakauman todennäköisyysfunktio

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \dots, n.$$

Kolmogorovin aksioomat

Olkoon $\mathcal{F}$ jokin Ω:n osajoukkojen muodostama σ-algebra. Kuvaus $P\colon \mathcal{F} \to \mathbb{R}$ määrittelee todennäköisyysmitan, jos

1. $0 \le P(A) \le 1 \quad$ kaikilla $A \in \mathcal{F}$.
2. $P(\emptyset) = 0 \quad$ ja $\quad P(\Omega) = 1$.
3. Jos tapahtumat $A_i \in \mathcal{F}$ $(i = 1, 2, \ldots)$ ovat parittain erilliset, niin

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Harjoituksia

1. Laske seuraavat lausekkeet:

(a) $6^{(3)}$, $0^{(5)}$, $5^{(0)}$, $7!$, $\binom{7}{-3}$, $\binom{-7}{3}$.

(b) $\binom{10}{7\ 3}$, $\binom{14}{2\ 3\ 5\ 4}$, $\binom{-1.5}{4}$.

2. Olkoon n positiivinen kokonaisluku ja $0 \le p \le 1$. Osoita, että

(a) $\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1$;
(Vihje: Merkitse $(1-p) + p = (1-p)(1 + \frac{p}{1-p})$ ja käytä binomilausetta.)

(b) $\sum_{x=0}^{n} x\binom{n}{x} p^x (1-p)^{n-x} = np$.
(Vihje: Käytä hyväksi tulosta $x\binom{n}{x} = n\binom{n-1}{x-1}$ ja binomilausetta.)

3. (a) Valitaan satunnaisesti 2 lukua luvuista 1, 2, ..., 39 palauttamatta. Millä todennäköisyydellä saadaan peräkkäiset luvut?

(b) Valitaan 7 lukua luvuista 1, 2, ..., 39 palauttamatta (Lotto). Millä todennäköisyydellä saadaan peräkkäiset luvut?

(c) Valitaan 2 lukua luvuista 1, 2, ..., n palauttamatta. Millä todennäköisyydellä saadaan peräkkäiset luvut?

4. Kahdestatoista verinäytteestä 4 oli positiivisia ja 8 negatiivisia. Sekaannuksen takia näytteet unohtuivat merkitsemättä, joten ne oli analysoitava uudestaan yksitellen (satunnaisessa järjestyksessä).

(a) Millä todennäköisyydellä tarvitaan vain 4 analyysia (4 ensimmäistä positiivisia)?

(b) Millä todennäköisyydellä tarvitaan täsmälleen 5 analyysia?

(c) Millä todennäköisyydellä positiiviset tulokset saadaan peräkkäin?

5. Erääseen lääketieteelliseen hoitokokeeseen osallistui 15 miestä ja 20 naista. Kymmenen satunnaiseti valittua potilasta sai tutkittavaa uutta hoitoa (hoitoryhmä) ja loput kuuluivat vertailuryhmään. Mikä on todennäköisyys, että hoitoryhmään tulee

(a) ainakin yksi kumpaakin sukupuolta?

(b) ainakin kolme kumpaakin sukupuolta?

6. (a) Valitaan 30 kännykän tuote-erästä 4 satunnaisesti palauttamatta tarkastukseen. Jos tuote-erässä on 3 viallista, niin millä todennäköisyydellä otoksessa on

i. täsmälleen 2 viallista?

ii. ainakin 2 viallista?

(b) Olkoon 30:n kännykän tuote-erässä d viallista. Tuote-erästä tarkastetaan n:n kännykän otos. Erä lähetetään myyntiin, jos otoksessa ei ole yhtään viallista, muutoin erä palautetaan. Halutaan, että 5 viallista sisältävät tuote-erät palautetaan todennäköisyydellä $p \geq 0.95$. Kuinka suuri otoskoko silloin tarvitaan?

7. (a) Sijoitetaan 22 palloa satunnaisesti 120 laatikkoon. Mikä on todennäköisyys, että yhdessäkään laatikossa ei ole enempää kuin yksi pallo?

(b) Eräässä 120 päivän jaksossa kaapattiin 22 liikennekonetta. Mikä on todennäköisyys, että samana päivänä kaapataan ainakin 2 konetta, jos eri kaappaukset ajoittuvat täysin satunnaisesti ja toisistaan riippumatta.

8. Valitaan satunnaisesti palauttaen 3 palloa laatikosta, jossa on 3 punaista, 4 keltaista ja 5 sinistä palloa. Laske todennäköisyys, että

(a) pallot ovat samanvärisiä;

(b) pallot ovat erivärisiä.

Laske vastaavat todennäköisyydet, kun otanta on palauttamatta.

9. Eräässä 10000 vaalikelpoisen asukkaan kaupungissa tehtiin juuri ennen vaalia mielipidekysely valitsemalla 100 henkilön otos vaalikelpoisten populaatiosta. Ehdokkaat olivat A ja B. Vaalin tuloksen perusteella tiedetään, että A:n kannatus oli 45 % ja B:n kannatus 55 %. Mikä on todennäköisyys, että kyselyssä

(a) 51 henkilöä kannattaa A:ta?

(b) yli puolet kannattaa A:ta?

(c) Kuinka suuri otos on tehtävä, jotta otoksessa olisi B:n kannattajia enemmän kuin A:n kannattajia vähintään todennäköisyydellä 0.9?

10. Oletetaan, että neliönmuotoinen maa-alue on jaettu kolmeen pinta-alaltaan yhtä suureen kaistaleeseen A, B ja C. Lisäksi oletetaan, että kaistaleiden yksikköhinnat ovat toisiinsa suhteessa $1 : 2 : 3$. Jokaisen (mitallisen) osa-alueen M suhteellinen hinta verrattuna koko maa-alueen hintaan saadaan kaavalla

$$V(M) = \frac{P(M \cap A) + 2\,P(M \cap B) + 3\,P(M \cap C)}{2},$$

missä $P(M) = \frac{|M|}{|\Omega|}$, $|M|$ on M:n pinta-ala ja $|\Omega|$ on koko maa-alueen pinta-ala. Osoita, että $V(M)$ on todennäköisyys(mitta) (Määritelmä 4.1).

11. Olkoon otosavaruus Ω äärellinen. Osoita, että Määritelmän 4.1 aksioomeista seuraavat Määritelmän 1.1 mukaiset todennäköisyysfunktion ominaisuudet.

12. Monivalintatehtävässä on 6 väittämää, joista jokaiseen on vastattava tosi (T) tai epätosi (E). Vastaus on oikein tai väärin ja oikeasta vastauksesta saa 1 pistettä ja väärästä -1 pistettä. Oletetaan, että Mr RW (Random Walker) vastaa väittämiin täysin satunnaisesti (Heittää esimerkiksi lanttia).

(a) Millä todennäköisyydellä RW saa negatiivisen pistemäärän?

(b) Mikä on RW:n pistemäärän todennäköisyysjakauma?

(c) Jos kolmantena vaihtoehtona on mahdollisuus vastata ”en tiedä” (N), niin millä todennäköisyydellä RW saa negatiivisen pistemäärän?

13. Mitä voit sanoa tapahtumasta A, joka on riippumaton itsensä kanssa? Miten luonnehdit tapahtumia A ja B, jotka ovat toisensa poissulkevat ja riippumattomat?

14. Todista Yhteenlaskulause 4.4 osoittamalla ensin, että

$$(A \cup B) - A = B - (A \cap B).$$

15. Oletetaan, että 550 omenan laatikossa on 2 % pilaantuneita.

(a) Millä todennäköisyydellä 25 omenan satunnaisotoksessa (otanta palauttamatta) on 2 pilaantunutta?

(b) Millä todennäköisyydellä 25 omenan satunnaisotoksessa on korkeintaan 2 pilaantunutta?

(c) Halutaan, että ainakin 2 % pilaantuneita sisältävät laatikot hylätään todennäköisyydellä $p \geq 0.95$. Kuinka suureksi otoskoko on valittava?

16. Ovessa on kaksi (erilaista) lukkoa ja avaimet ovat niiden kuuden joukossa, joita kannat aina mukanasi. Olet kiireessä pudottanut yhden näistä kuudesta jonnekin.

(a) Mikä on todennäköisyys, että vielä saat oven auki avaimillasi?

(b) Millä todennäköisyydellä saat oven auki heti kahdella ensiksi kokeilemallasi avaimella (Oletetaan, että avaimet näyttävät täysin samanlaisilta.)

17. (a) Kuinka monta kokonaislukuarvoista ratkaisua yhtälöllä $x_1 + x_2 = 5$ on, kun ratkaisujen tulee olla epänegatiivisia?

(b) Investoit 20 tuhatta euroa 4:ään mahdolliseen kohteeseen. Jokaisen sijoituksen tulee olla 100 euron monikerta. Montako investointistrategiaa on, jos koko summa on sijoitettava?

(c) Montako strategiaa on silloin, jos koko summaa ei tarvitse investoida?

18. Tietokoneessa on n prosessoria ja r työtä jaetaan prosessoreille satunnaisesti. Eri prosessoreille tulevien töiden lukumäärät ovat r_1, r_2, $\ldots$, r_n, $r_i \geq 0$, $i = 1, 2, \ldots, n$ ja

$$r_1 + r_2 + \cdots + r_n = r. \tag{2.9.1}$$

(a) Mikä on erilaisten varausjakaumien (Yhtälön (4.9.1) ratkaisujen lukumäärä)?

(b) Mikä on todennäköisyys, että tietyllä prosessorilla on k, $0 \leq k \leq r$ työtä?

(c) Oletetaan, että annetut n lukua r_1, r_2, $\ldots$, r_n toteuttavat yhtälön (4.9.1) ja kaikki mahdolliset n^r töiden sijoittelua prosessoreille ovat yhtä mahdollisia. Mikä on todennäköisyys saada varausluvut r_1, r_2, $\ldots$, r_n?