Todennäköisyyslaskenta I

Ville Hyvönen, Topias Tolonen[1]

Kesä 2017

[1]Luentomateriaali alun perin Villen käsialaa kesältä 2016, materiaalia muokataan kesän 2017 luentojen mukana ajan tapaa ja luennoitsijan huumorintajua mukaillen

Sisältö

Luku 1

Todennäköisyys

1.1 Klassinen todennäköisyys

Tällä kurssilla käsitellään todennäköisyyttä matemaattisena käsitteenä, jonka tarkoituksena on mallintaa satunnaisuutta reaalimaailman ilmiöissä. Usein puhutaan *satunnaiskokeesta*, jonka mallintaminen on tavoitteena.

Yksinkertaisimmassa tapauksessa satunnaiskokeen mahdollisten tulosten määrä on äärellinen ja ne kaikki ovat yhtä todennäköisiä. Tällöin kyseessä on *äärellinen* ja *symmetrinen* todennäköisyysavaruus.

Esimerkkejä tälläisestä satunnaiskokeesta ovat kolikon tai arpakuution heitto, viiden kortin pokerikäden nostaminen sekoitetusta korttipakasta, lottoarvonta tai ruletin pyöräytys. Tämä historiallisesti varhaisin tapa mallintaa todennäköisyyttä saikin innoituksensa juuri uhkapeleistä (Pascal, circa 1650 innostui asiasta erityisesti).

Tällaisessa tapauksessa, jossa kaikki lopputulokset ovat symmetrisiä, tapahtuman todennäköisyys saadaan jakamalla sille suotuisten lopputulosten määrä kaikkien mahdollisten lopputulosten määrällä. Tätä kutsutaan todennäköisyyden klassiseksi määritelmäksi.

Määritellään ensin perusjoukko, tapahtuma ja klassinen todennäköisyys, ja palataan sen jälkeen tarkemmin edellämainittuihin esimerkkeihin.

Määritelmä 1.1. Olkoon Ω satunnaiskokeen mahdollisten loputulosten joukko. Kutsumme sitä *perusjoukoksi*, sen alkioita $\omega \in \Omega$ *alkeistapauksiksi.*

Määritelmä 1.2. Äärellisen perusjoukon $\Omega = \{\omega_1, \dots, \omega_n\}$ tapauksessa kutsumme kaikkia perusjoukon osajoukkoja $A \subseteq \Omega$ *tapahtumiksi.* Merkitsemme kaikkien tapahtumien joukkoa $\mathcal{F}$:llä. Tässä tapauksessa se on perusjoukon kaikkien osajoukkojen joukko, eli $\mathcal{F} = \mathcal{P}(\Omega)$.

Määritelmä 1.3. (Klassinen todennäköisyys). Olkoon $A \subseteq \Omega$ äärellisen ja symmetrisen perusjoukon Ω tapahtuma. A:n todennäköisyys on

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

missä $|A|$ on A:n alkoiden lukumäärä ja $n := |\Omega|$ on koko perusjoukon alkoiden lukumäärä.

Huomaa, että klassisessa todennäköisyyden mallissa kaikki alkeistapaukset ovat yhtä todennäköisiä, sillä

$$P(\{\omega\}) = \frac{|\{\omega\}|}{n} = \frac{1}{n} \quad \text{kaikille } \omega \in \Omega.$$

Esimerkki 1.4. Heitetään yhtä kolikkoa. Alkeistapaukset ovat $\omega_1 =$ kruuna ja $\omega_2 =$ klaava. Siten perusjoukko

$$\Omega = \{\text{kruuna}, \text{klaava}\},$$

sisältää $n = 2$ alkeistapausta, ja kaikkien tapahtumien joukko on

$$\mathcal{F} = \{\emptyset, \{\text{kruuna}\}, \{\text{klaava}\}, \{\text{kruuna, klaava}\}\}.$$

Näiden tapahtumien todennäköisyydet ovat:

$$\begin{aligned}
&P(\emptyset) = 0 \quad \text{(ei kruunaa eikä klaavaa)} \\
&P(\{\text{kruuna}\}) = \frac{1}{n} = \frac{1}{2} \\
&P(\{\text{klaava}\}) = \frac{1}{n} = \frac{1}{2} \\
&P(\{\text{kruuna, klaava}\}) = P(\Omega) = 1 \quad \text{(kruuna tai klaava)}.
\end{aligned}$$

Huomautus 1.5. Yleensä tapahtuman todennäköisyyttä merkittäessä jätetään kaksista sulkeista toiset pois, esimerkiksi tapahtuman $\{\omega_1, \omega_2\}$ todennäköisyys voidaan kirjoittaa muodossa

$$P(\{\omega_1, \omega_2\}) = P\{\omega_1, \omega_2\}$$

tai muodossa

$$P(\{\omega_1, \omega_2\}) = P(\omega_1, \omega_2).$$

Jatkossa käytämme näistä ensimmäistä muotoa. Siten edellisen esimerkin todennäköisyys tapahtumalle {kruuna, klaava} (sattuu kruuna tai klaava), voidaan kirjoittaa kompaktimmin muodossa

$$P\{\text{kruuna, klaava}\}.$$

Esimerkki 1.6. Yhden kortin nostaminen korttipakasta. Oletetaan, että meillä on (hyvin) sekoitettu 52:n kortin korttipakka. Alkeistapauksia ovat esimerkiksi $\spadesuit 2$, $\clubsuit 7$ ja $\heartsuit J$. Perusjoukko on siis

$$\Omega = \{(x, y) \,|\, x \in \{\clubsuit, \spadesuit, \diamondsuit, \heartsuit\}, y \in \{2, 3, \ldots, 10, J, Q, K, A\}\}.$$

Tapahtumia ovat esimerkiksi B : ”Kortti on musta”:

$$B = \{(x, y) \in \Omega \,|\, x \in \{\spadesuit, \clubsuit\}\}$$

ja C : ”Kortti on ruutu ja arvoltaan pienempi kuin 5”:

$$C = \{(x, y) \in \Omega \,|\, x \in \{\diamondsuit\} \text{ ja } y \in \{2, 3, 4\}\}.$$

Näiden tapahtumien todennäköisyydet saadaan jälleen suoraan klassisesta todennäköisyyden määritelmästä:

$$P(B) = \frac{|B|}{n} = \frac{26}{52} = \frac{1}{2}$$

ja

$$P(C) = \frac{|C|}{n} = \frac{3}{52}.$$

Esimerkki 1.7. Kahden kuusisivuisen nopan heitto. Alkeistapauksia ovat järjestetyt parit

$$(1, 1), (1, 2), \ldots, (6, 5), (6, 6),$$

eli $\Omega = E \times E$, $E = \{1, 2, \ldots, 6\}$. Siten perusjoukossa on $n = 36$ alkiota, jotka voitaisiin edelleen kaikki luetella.

Nyt erilaisia tapahtumia on 2^{36} kappaletta, joten niiden luetteleminen ei onnistu. Tarkastellaan sen sijaan muutamaa esimerkkitapahtumaa. Merkitään A:lla tapahtumaa ”Noppien silmälukujen summa on yli yhdeksän”, jolloin A voidaan kirjoittaa muodossa

$$\begin{aligned} A &= \{(x, y) \in \Omega \,|\, x + y > 9\} \\ &= \{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\}. \end{aligned}$$

Tällöin klassisen todennäköisyyden määritelmän nojalla

$$P(A) = \frac{|A|}{n} = \frac{6}{36} = \frac{1}{6}.$$

Merkitään B:llä tapahtumaa ”molemmat silmäluvut ovat pienempiä kuin kolme”. Tällöin B voidaan esittää muodossa

$$\begin{aligned} B &= \{(x, y) \in \Omega \,|\, x < 3 \text{ ja } y < 3\} \\ &= \{(1, 1), (1, 2), (2, 1), (2, 2)\}. \end{aligned}$$

Jälleen klassisen todennäköisyyden määritelmän nojalla

$$P(B) = \frac{|B|}{n} = \frac{4}{36} = \frac{1}{9}.$$

Huomautus 1.8. Perusjoukon valinta on aina viime kädessä mielivaltainen, ja samaa satunnaiskoetta voidaan mallintaa usealla eri perusjoukolla. Usein, kuten ylläolevissa esimerkeissä, perusjoukon valinta on ilmeinen. Muitakin mahdollisuuksia kuitenkin on: esimerkiksi jos meitä kiinnostaa vain kahden nopan yhteenlaskettu silmäluku, voimme käyttää alkeistapauksina kahden nopan mahdollisten yhteenlaskettujen silmälukujen joukkoa

$$\Omega_1 = \{2, \ldots, 12\}.$$

Tällä perusjoukolla voimme esittää edellisen esimerkin tapahtuman A muodossa

$$A = \{\omega \in \Omega_1 \,|\, \omega > 9\} = \{10, 11, 12\}.$$

Tällöin kuitenkaan emme voi soveltaa klassisen todennäköisyyden määritelmää, koska tilanne ei ole enää symmetrinen (alkeistapaukset eivät ole yhtä todennäköisiä). Tällaisten tilanteiden mallintamiseen palataan myöhemmin.

Tapahtumaa B emme voi kuitenkaan esittää perusjoukon Ω_1 avulla. Perusjoukon tuleekin olla riittävän rikas, jotta kaikki meitä kiinnostavat tapahtumat voidaan esittää sen osajoukkoina.

Klassisen todennäköisyyden käsite ei aina kuitenkaan ole paras mahdollinen. Kuvitellaan tilanne, jossa heitämme kolikkoa niin kauan, kunnes tulee klaava. Mikä on tämän tilanteen perusjoukko? Huomaat, että perusjoukolle pätee $|\Omega| = n = \infty$. Miksi? Klassisen todennäköisyyskäsitteen rajoitukset tulevat nopeasti vastaan, ja siksi tarvitsemme lisää todennäköisyysmalleja. Tällaisia on esimerkiksi frekventistiset ja subjektiiviset todennäköisyydet. Niistä lisää myöhemmissä kappaleissa.

1.2 Kombinatoriikkaa

Kun satunnaiskoe on symmetrinen ja sen arvojoukko on äärellinen, todennäköisyyksien laskeminen on periaatteessa aina yhtä helppoa kuin ylläolevissa esimerkeissä. Kuitenkin kun erilaisia alkeistapauksia on paljon, niitä ei voida enää luetella, vaan tarvitaan laskusääntöjä perusjoukon ja sen osajoukkojen sisältämien alkeistapausten lukumäärän laskemiseksi. Näitä laskusääntöjä kutsutaan kombinatoriikaksi.

1.2.1 Tuloperiaate

Oletetaan, että tehtävä voidaan suorittaa k:ssa peräkkäisessä vaiheessa siten, että ensimmäisessä vaiheessa on n_1 vaihtoehtoa, toisessa vaiheessa on n_2 vaihtoehtoa, ja niin edelleen viimeiseen vaiheeseen, jossa on n_k vaihtoehtoa, asti. Oletetaan myös, että i:nen vaiheen vaihtoehtojen määrä n_i ei riipu (huom. itse vaihtoehdot voivat kyllä riippua) edeltävissä vaiheissa tehdyistä valinnoista. Tällöin tehtävä voidaan suorittaa yhteensä

$$\prod_{i=1}^{k} n_i = n_1 \cdot n_2 \cdot \ldots \cdot n_k$$

eri tavalla.

Esimerkki 1.9. Vakioveikkausrivi. Jokaisessa kohteessa valitaan merkki kolmesta eri vaihtoehdosta (1,X,2), ja kohteita on yhteensä 13. Siten erilaisia mahdollisia rivejä on

$$3^{13} = 1594323$$

kappaletta.

Tuloperiaatteella saadaan myös suoraan laskettua joukon erilaisten osajoukkojen määrä:

Lause 1.10. *$n-$alkioisella joukolla $E = \{x_1, \ldots, x_n\}$ on 2^n erilaista osajoukkoa.*

Todistus. Joukon E mielivaltainen osajoukko voidaan muodostaa n:ssä vaiheessa siten, että ensimmäiseksi valitaan kuuluko x_1 osajoukkoon, seuraavaksi valitaan kuuluko x_2 osajoukkoon, ja niin edelleen aina viimeiseen alkioon x_n asti. Jokaisessa vaiheessa on 2 vaihtoehtoa: joko alkio kuuluu osajoukkoon, tai se ei kuulu siihen. Siten tuloperiaatteen nojalla yhteensä erilaisia vaihtoehtoja on 2^n kappaletta. □

1.2.2 Permutaatiot ja kombinaatiot

Esimerkki 1.11. Tarkastellaan 4-alkioista joukkoa $A = \{a, b, c, d\}$.

- Kuinka monella tavalla joukko voidaan järjestää? Järjestäminen voidaa jakaa neljään vaiheeseen. Jonon ensimmäiseksi alkioksi voidaan valita mikä tahansa joukon alkio, eli erilaisia vaihtoehtoja on 4. Jonon toiseksi alkioksi voidaan valita mikä tahansa paitsi ensimmäiseksi valittu alkio, eli erilaisia vaihtoehtoja on 3. Kolmanneksi alkioksi valitaan toinen jäljelläolevista alkioista, ja viimeiseksi alkioksi vaihtoehtoja on jäljellä enää yksi. Siten tuloperiaatteen nojalla joukko A voidaan järjestää

  $$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$$

  erilaisella tavalla.

- Kuinka monta erilaista 2:n alkion järjestettyä jonoa (eli järjestettyä paria) joukosta A voidaan valita? Kaikki mahdollisuudet voidaan luetella ja laskea:

$$\begin{aligned}&(a,b),(a,c),(a,d)\\&(b,a),(b,c),(b,d)\\&(c,a),(c,b),(c,d)\\&(d,a),(d,b),(d,c),\end{aligned}$$

  eli yhteensä 12 kappaletta. Tämä voidaan laskea myös tuloperiaatteella, sillä 2:n alkion järjestetyn jonon valinta joukosta voidaan suorittaa kahdessa vaiheessa: ensin valitaan jonon ensimmäinen alkio kaikista joukon 4:stä alkiosta, ja sen jälkeen toinen alkio jäljelläolevista kolmesta alkiosta. Siten erilaisia jonoja on yhteensä

$$4 \cdot 3 = 12$$

  kappaletta.

- Kuinka monta erilaista 2:n alkion (järjestämätöntä) osajoukkoa joukosta A voidaan valita? Kaikki vaihtoehdot voidaan jälleen luetella ja laskea:

$$\begin{aligned}&\{a,b\},\{a,c\},\{a,d\}\\&\{b,c\},\{b,d\},\\&\{c,d\},\end{aligned}$$

  eli yhteensä 6 kahden alkion osajoukkoa. Tämä voidaan myös laskea havaitsemalla, että jokainen 2:n alkion joukko voidaan järjestää kahdella eri tavalla; esimerkiksi joukon $\{a,b\}$ mahdolliset järjestykset ovat (a,b) ja (b,a). Siten jokaista kahden alkion osajoukkoa kohti on 2 kahden pituista järjestettyä jonoa. Edellä todettiin, että joukon A kahden pituisia järjestettyjä jonoja on 12 kappaletta, joten kahden alkion osajoukkoja on

$$\frac{12}{2} = 6$$

  kappaletta.

Nimetään seuraavaksi nämä käsitteet, ja todetaan edellisen esimerkin havainnot yleisessä n-alkioisen joukon tapauksessa.

Määritelmä 1.12. Olkoon $E = \{x_1, \dots, x_n\}$ n-alkioinen joukko.

(i) Joukon E alkioiden järjestystä, eli sen alkioista muodostettua (jokainen alkio esiintyy vain kerran) järjestettyä n-alkioista jonoa

$$(x_{i_1}, x_{i_2} \dots, x_{i_n})$$

kutsutaan sen *permutaatioksi.*

(ii) Kaikille $k \in \{1, \ldots, n\}$ joukon E alkioista muodostettua (jokainen alkio esiintyy korkeintaan kerran) järjestettyä k-alkioista jonoa

$$(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$$

kutsutaan sen *k-permutaatioksi* (huom. Tuomisen kirjassa k-variaatioksi).

(iii) Joukon E alkioista muodostettua k-alkioista osajoukkoa (järjestyksellä ei siis ole väliä)

$$\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$$

kutsutaan sen *k-kombinaatioksi.*

Lause 1.13. *Olkoon $E = \{x_1, \ldots, x_n\}$ n-alkioinen joukko.*

(i) Joukolla E on $n!$ permutaatiota.

(ii) Kaikille $k \in \{1, \ldots, n\}$ joukolla E on

$$\frac{n!}{(n-k)!}$$

k-permutaatiota.

(iii) Kaikille $k \in \{1, \ldots, n\}$ joukolla E on

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

k-kombinaatiota.

Todistus.

(i) Mikä tahansa joukon E permutaatio voidaan valita n:ssä vaiheessa siten, että ensimmäiseksi valitaan jonon ensimmäinen alkio, seuraavaksi toinen alkio, ja niin edelleen aina viimeiseen alkioon saakka. Ensimmäiseksi alkioksi voidaan valita mikä tahansa joukon E alkio, joten ensimmäisessä vaiheessa on n vaihtoehtoa. Toiseksi alkioksi voidaan valita mikä tahansa paitsi ensimmäiseksi valittu alkio, joten toisessa vaiheessa on $n-1$ vaihtoehtoa. Näin jatketaan aina viimeiseen vaiheeseen, jossa jäljellä on vain yksi vaihtoehto, saakka. Siten tuloperiaatteen nojalla permutaatio voidaan valita

$$n \cdot (n-1) \cdot \ldots \cdot 1 = n!$$

erilaisella tavalla.

(ii) Vastaavasti kuin (i)-kohta, mutta vaiheita on k kappaletta, jolloin viimeisessä vaiheessa jäljellä on $(n-k+1)$ vaihtoehtoa jonon viimeiseksi alkioksi. Siten tuloperiaatteen nojalla k-permutaatio voidaan valita

$$n \cdot (n-1) \cdot ... \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

erilaisella tavalla.

(iii) Olkoon x joukon E k-kombinaatioiden määrä. Jokainen k-kombinaatio on k-alkioinen joukko, joten kohdan (i) nojalla se voidaan järjestää $k!$:lla eri tavalla. Kaikki joukon k-permutaatiot saadaan järjestämällä sen k-kombinaatiot, joten joukolla E on $x \cdot k!$ k-permutaatiota. Siten kohdan (ii) nojalla saadaan yhtälö, josta voidaan ratkaista k-kombinaatioiden määrä:

$$\begin{aligned} x \cdot k! &= \frac{n!}{(n-k)!} \\ x &= \frac{n!}{k!(n-k)!} = \binom{n}{k}. \end{aligned}$$

□

Kokonaisluvuille $n, k \in \mathbb{N}$ lukuja

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

kutsutaan *binomikertoimiksi*. Kun $k < 0$ tai $k > n$, määritellään

$$\binom{n}{k} = 0.$$

Lause 1.14. *Olkoon $E = \{x_1, \ldots, x_n\}$ n-alkioinen joukko, ja $n_1, \ldots, n_k \in \mathbb{N}$ siten, että $\sum_{i=1}^{k} n_i = n$. Joukko E voidaan jakaa k:n osajoukkoon siten, että kaikille $i \in \{1, \ldots, k\}$ i:nnessä osajoukossa on n_i alkiota*

$$\binom{n}{n_1 n_2 \ldots \; n_k} = \frac{n!}{n_1! n_2! \ldots n_k!}$$

erilaisella tavalla.

Todistus. Osajoukkojen valinta jakautuu luontevasti k:n vaiheeseen:

- Ensimmäiseen osajoukkoon voidaan valita n:stä alkiosta mitkä tahansa n_1, joten lauseen 1.13 kohdan (iii) perusteella erilaisia vaihtoehtoja on $\binom{n}{n_1}$ kappaletta
- Toiseen osajoukkoon valitaan n_2 alkiota niistä alkiosta joita ei valittu ensimmäiseen osajoukkoon, eli vaihtoehtoja on $\binom{n-n_1}{n_2}$ kappaletta.
- Ja niin edelleen aina k:een osajoukkoon asti, jossa vaihtoehtoja on $\binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k}$ kappaletta.

Siten tuloperiaatteen nojalla erilaisia vaihtoehtoja on

$$\begin{aligned}
&\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k} \\
=&\frac{n!}{n_1!(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!(n-n_1-n_2)!}\cdot\cdots\cdot\frac{(n-n_1-n_2-\cdots-n_{k-1})!}{n_k!(n-\sum_{i=1}^k n_i)!} \\
=&\frac{n!}{n_1!n_2!\dots n_k!}=\binom{n}{n_1 n_2\dots\ n_k}
\end{aligned}$$

kappaletta. Tässä käytettiin hyväksi sitä, että peräkäisten osoittajien ja nimittäjien termit kumoavat toisensa tulossa. Viimeiseen termin nimittäjään jää $(n-\sum_{i=1}^k n_i)!$, mutta oletuksen nojalla $\sum_{i=1}^k n_i = n$, joten se on $0! = 1$. □

Esimerkki 1.15. Pokerikäsi. Sekoitetusta korttipakasta nostetaan viisi korttia. Järjestyksellä, jossa kortit on nostettu pakasta, ei ole väliä pelin kannalta. Perusjoukoksi Ω kannattaakin valita kaikki mahdolliset 52:n kortin 5:n alkion osajoukot, eli 5-kombinaatiot. Perusjoukon alkioita $\omega \in \Omega$ ovat esimerkiksi:

$$\{\heartsuit 3, \diamondsuit 3, \heartsuit J, \spadesuit 5, \diamondsuit 9\}$$

ja

$$\{\diamondsuit 5, \spadesuit J, \diamondsuit K, \clubsuit 3, \heartsuit A\}.$$

Näitä korttipakan 5-kombinaatioita on lauseen 1.13 (iii)-kohdan nojalla yhteensä

$$\binom{52}{5} = 2598960$$

kappaletta.

Siten jokaisen viiden kortin yhdistelmän, esimerkiksi herttareedin (herttaväriuora kympistä ässään), todennäköisyys on

$$P\{\heartsuit 10, \heartsuit J, \heartsuit Q, \heartsuit K, \heartsuit A\} = \frac{1}{\binom{52}{5}} = \frac{1}{2598960} \approx 3.85\cdot 10^{-7}.$$

Esimerkki 1.16. Lasketaan todennäköisyys tapahtumalle A : ”nostetaan käsi, jossa on neljä ässää”. Tälläisen 5-kombinaation tulee sisältää ainakin kortit $\heartsuit A, \diamondsuit A, \clubsuit A$ ja $\spadesuit A$. Viimeinen kortti voi olla mikä tahansa jäljelläolevasta 48 kortista, joten tapahtuma A sisältää 48 perusjoukon alkiota (pokerikättä, eli korttipakan 5-kombinaatiota). Siten tapahtuman A todennäköisyys on

$$P(A) = \frac{48}{\binom{52}{5}} \approx 1.85 \cdot 10^{-5}.$$

Esimerkki 1.17. Lasketaan vielä todennäköisyydet tapahtumille B : ”nostetaan käsi, jossa on täsmälleen kolme ässää”, ja C : ” nostetaan käsi, jossa on vähintään kolme ässää”. Tapahtumien B ja C alkioiden määrää laskettaessa tulee olla tarkkana, ettei samaa kättä laske mukaan useampaan kertaan.

Lasketaan ensin käsien, joissa on *täsmälleen* kolme ässää, määrä. Näiden käsien valitseminen voidaan jakaa kahteen osaan: valitaan ensin kolme ässää, ja sen jälkeen kaksi muuta korttia. Kolme ässää voidaan valita neljällä erilaisella tavalla:

$$\{\heartsuit A, \diamondsuit A, \clubsuit A\},$$
$$\{\heartsuit A, \diamondsuit A, \spadesuit A\},$$
$$\{\heartsuit A, \spadesuit A, \clubsuit A\},$$

tai

$$\{\spadesuit A, \diamondsuit A, \clubsuit A\}.$$

Toisessa vaiheessa valitaan 2 korttia muuta korttia. Vaihtoehtoja on 48, koska kolme ässistä on jo käytetty, ja neljättä ei voida valita, koska muuten kädessä olisi *yli* kolme ässää. Lauseen 1.13 kohdan (iii) nojalla 48:sta kortista voidaan valita $\binom{48}{2} = 1128$ kahden kortin joukkoa eli 2-kombinaatiota. Siten tuloperiaatteen nojalla erilaisia käsiä, joissa on täsmälleen 3 ässää, on $4 \cdot 1128 = 4512$ kappaletta, joten todennäköisyys nostaa täsmälleen kolme ässää on

$$P(B) = \frac{4512}{\binom{52}{5}} \approx 0.001736$$

Käsien, joissa on *vähintään* kolme ässää, lukumäärä saadaan laskemalla yhteen käsien, joissa on täsmälleen kolme ässää, ja käsien, joissa on täsmälleen neljä ässää, määrä: erilaisia vaihtoehtoja on yhteensä $4512 + 48 = 4560$ kappaletta. Siten todennäköisyys nostaa vähintään kolme ässää on

$$P(C) = \frac{4560}{\binom{52}{5}} \approx 0.001755.$$

1.2.3 Otanta

Monesti tilastollisessa tutkimuksessa halutaan tutkia jonkin populaation ominaisuuksia. Selvitettäviä kysymyksiä voisivat olla esimerkiksi, kuinka suuri osa suomalaisista kannattaa sotilasliitto NATO:on liittymistä, tai kuinka suuri osa jonkin järven kaloista on alamittaisia. Kaikkien suomalaisten NATO-kannan kysyminen, tai kaikkien järven kalojen kalastaminen on hyvin aikaavievää ja hankalaa, joten yleensä tilastollisessa tutkimuksessa suoritetaan otanta, eli poimitaan satunnaisotos, jonka koko on yleensä paljon koko populaation kokoa pienempi, tutkittavasta populaatiosta. Otanta voidaan suorittaa takaisinpanolla eli palauttaen, jolloin sama yksilö voi päätyä otokseen useamman kerran, tai ilman takaisinpanoa, jolloin sama yksilö voi päätyä otokseen vain kerran.

Esimerkki 1.18. Laatikossa on 8 palloa, joista 5 on valkoista ja $8 - 5 = 3$ mustia. Laatikko voidaan kuvata joukkona

$$E = \{v_1, v_2, v_3, v_4, v_5, m_1, m_2, m_3\},$$

missä v:llä merkitään valkoista, ja m:llä mustaa palloa.

(i) **Otanta ilman takaisinpanoa.** Nostetaan pussista satunnaisesti 4 palloa palauttamatta nostettuja palloja nostojen välillä laatikkoon. Mikä on tapahtuman $A_2 =$ "Nostetaan tasan 2 valkoista palloa todennäköisyys"?

Perusjoukoksi Ω kannattaa valita kaikkien 8 pallon joukon 4 pallon osajoukot, eli 8-kombinaatiot; järjestyksellä, jossa pallot on nostettu, ei ole tutkittavan tapahtuman kannalta merkitystä. Perusjoukon alkioita $\omega \in \Omega$ ovat esimerkiksi

$$\{v_3, v_4, m_1, m_3\}$$

ja

$$\{v_1, v_2, v_4, m_1\}.$$

Näitä on lauseen 1.13 kohdan (iii) nojalla $\binom{8}{4}$ kappaletta. Näistä tapahtumalle A_3 suotuisia alkeistapauksia ovat ne, joissa on tasan 2 valkoista ja $4 - 2 = 2$ mustaa palloa, eli esimerkiksi

$$\{v_3, v_4, m_1, m_3\}$$

ja

$$\{v_1, v_4, m_1, m_2\}.$$

Lauseen 1.13 kohdan (iii) nojalla valkoiset pallot voidaan valita $\binom{5}{2}$ erilaisella, ja mustat pallot $\binom{3}{2}$ erilaisella tavalla. Siten erilaisia mahdollisia yhdistelmiä, joissa on tasan 2 valkoista palloa, on tuloperiaatteen nojalla $\binom{5}{2}\binom{3}{2}$ kappaletta, joten todennäköisyys nostaa tasan 2 valkoista palloa on

$$P(A_2) = \frac{\binom{5}{2}\binom{3}{2}}{\binom{8}{4}} = \frac{10 \cdot 3}{70} = \frac{3}{7} \approx 0.4286.$$

(ii) **Otanta takaisinpanolla.** Otanta takaisinpanolla. Nostetaan laatikosta satunnaisesti 4 palloa palauttaen nostojen välillä nostettu pallo takaisin laatikkoon. Mikä on tapahtuman A_2 = ”Nostetaan tasan 2 valkoista palloa” todennäköisyys?

Nyt perusjoukoksi Ω kannattaa valita kaikki 4:n pallon jonot, jotka on mahdollista muodostaa 8 pallosta siten, että sama pallo voi esiintyä useampaan kertaan. Perusjoukon alkioita $\omega \in \Omega$ ovat esimerkiksi:

$$(m_3, m_3, m_1, m_3)$$

ja

$$(v_3, v_2, v_3, m_2).$$

Tuloperiaatteen nojalla näitä on 8^4 kappaletta.

Lasketaan, kuinka monessa näistä jonoista on täsmälleen 3 valkoista palloa. Selvitetään ensimmäiseksi, kuinka monta erilaista tapaa on järjestää 4 palloa, joista 2 on valkoisia ja 2 mustia, jonoiksi. Tällaisiä jonoja ovat esimerkiksi

$$(m, m, v, v)$$

ja

$$(v, m, v, m).$$

Lauseen 1.13 (iii)-kohdan nojalla näitä erilaisia järjestyksiä on $\binom{4}{2}$ kappaletta. Näiden erilaisten järjestysten valkoiset pallot ja mustat pallot voidaan edelleen valita useammalla eri tavalla. Esimerkiksi järjestyksen (m, m, v, v) toteuttavat

$$(m_2, m_2, v_5, v_3)$$

ja

$$(m_3, m_1, v_1, v_3).$$

Tuloperiaatteen nojalla valkoiset pallot kuhunkin järjestykseen voidaan valita 5^2 erilaisella, ja mustat pallot 3^2 erilaisella tavalla.

Edelleen tuloperiaatteen nojalla erilaisia 4:n pallon jonoja, joissa 2 palloista on valkoisia, on $\binom{4}{2} \cdot 5^2 \cdot 3^2$ kappaletta, joten todennäköisyys nostaa tasan 2 valkoista palloa on

$$P(A_2) = \frac{\binom{4}{2} \cdot 5^2 \cdot 3^2}{8^4} = \frac{1350}{4096} \approx 0.3296.$$

Tarkastellaan kahdentyyppisistä alkoista koostuvaa N-alkioista joukkoa $E = A \cup B$, missä $A = \{a_1, \dots, a_K\}$ ja $B = \{b_1, \dots, b_{N-K}\}$. E voi olla esimerkiksi laatikossa olevien pallojen, joista K on valkoisia ja $N - K$ mustia, joukko. Edellinen esimerkki yleistyy seuraavasti.

Lause 1.19. *Olkoon $n \in \{1, \dots, N\}$ ja $k \in \{0, \dots, n\}$.*

(i) Valitaan yllämainitusta joukosta E n:n kappaleen otos satunnaisesti ilman takaisinpanoa. Todennäköisyys tapahtumalle A_k: "otoksessa on täsmälleen k joukon A alkiota" on

$$P(A_k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

(ii) Valitaan yllämainitusta joukosta E n:n kappaleen otos satunnaisesti takaisinpanolla. Todennäköisyys tapahtumalle A_k: "otoksessa on täsmälleen k joukon A alkiota" on

$$P(A_k) = \binom{n}{k} \frac{K^k (N-K)^{(n-k)}}{N^n}.$$

Todistus. Kuten esimerkissä 1.18. □

Pokerikäden nostamisessa korttipakasta kyse on otannasta ilman takaisinpanoa. Siten esimerkin 1.17 tulos täsmälleen kolmen ässän todennäköisyydelle seuraa suoraan edellisestä lauseesta, kun valitaan $N = 52$, $K = 4$, $n = 5$ ja $k = 3$:

$$P(B) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{K}{n}} = \frac{\binom{4}{3}\binom{48}{2}}{\binom{52}{5}} \approx 0.001736.$$

Jos taas nostetaan korttipakasta 5 korttia siten, että sekoitetaan nostettu kortti pakkaan aina nostojen välillä, kyse on otannasta takaisinpanolla. Siten todennäköisyys nostaa täsmälleen 3 ässää (merkitään tätä tapahtumaa D:llä) saadaan myös suoraan edellisestä lauseesta

$$P(D) = \binom{n}{k} \frac{K^k (N-K)^{(n-k)}}{N^n} = \binom{5}{3} \frac{4^3 \cdot 48^2}{52^5} \approx 0.003878.$$

1.3 Matemaattinen todennäköisyyden käsite

Jos perusjoukko Ω ei ole *symmetrinen*, eli ei ole perusteita olettaa että alkeistapaukset olisivat yhtä todennäköisiä, klassisen todennäköisyyden määritelmää (määritelmä 1.3), ei voida käyttää satunnaiskokeeseen liittyvän epävarmuuden mallintamiseen. Esimerkkejä tällaisesta satunnaiskokeesta ovat esimerkiksi (viiden) tikan heitto, jossa perusjoukoksi voidaan asettaa $\Omega = \{0, 1, 2, \ldots, 10\}^5$, jalkapallo-ottelun lopputulos, jossa luonteva perusjoukko on $\mathbb{N}^2$, tai keihäänheiton tulos, jossa $\Omega = (0, \infty)$.

Kaksi viimeksimainittua esimerkkiä poikkeavat aiemmista tällä kurssilla käsitellyistä esimerkeistä myös siinä suhteessa, että perusjoukko Ω on ääretön; lisäksi viimeiseksi mainitussa esimerkissä perusjoukko on myös ylinumeroituva.

Tarvitaankin määritelmää 1.3 yleisempää todennäköisyyden määritelmää toisaalta epäsymmetristen satunnaiskokeiden, ja toisaalta laajempien perusjoukkojen käsittelemiseen.

Määrittelemme seuraavaksi todennäköisyyden tietyt ominaisuudet, joita kutsumme *todennäköisyyden aksioomiksi*, täyttävänä joukkofunktiona tapahtumien $\mathcal{F}$ joukolta välille $[0, 1]$.

Kun perusjoukko Ω on äärellinen tai korkeintaan numeroituvasti ääretön, niin tapahtumien joukoksi $\mathcal{F}$ voidaan aina valita (voidaan periaatteessa valita myös pienempi joukko) kaikkien perusjoukon osajoukkojen joukko $\mathcal{P}(\Omega)$. Ylinumeroituvan perusjoukon tapauksessa tapahtumien joukolta vaaditaan, että se on ns. σ-algebra. Tämän vuoksi yleiseen, matemaattiseen todennäköisyyden käsitteeseen tarvitsemme $\mathcal{F}$:n olevan σ-algebra.

Määritelmä 1.20. Kokoelma $\mathcal{F}$ perusjoukon Ω osajoukkoja on *σ-algebra*, jos

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Käytännössä σ-algebran ehdot todennäköisyyslaskennan viitekehyksessä tarkoittavat sitä, että tapahtumina voidaan tarkastella vain sellaisia joukkoja, jotka ovat riittävän "siistejä". Seuraava lause helpottaa kokoelman $\mathcal{F}$ käsittelyä:

Lause 1.21. *Olkoon $\mathcal{F}$ σ-algebra. Tällöin*

(i) $\emptyset \in \mathcal{F}$,

(ii) Jos $A, B \in \mathcal{F}$, *niin* $A \cup B, A \cap B, A \setminus B \in \mathcal{F}$,

(iii) $A_i \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} \in \mathcal{F}$.

Todistus. Jätetään harjoitustehtäväksi. Perusidea: kohta (i) seuraa siitä, että $\Omega^c = \emptyset$. Kohta (ii) saadaan soveltamalla σ-algebran määritelmän kohtaa 3, ja edelleen kohta (iii) saadaan ns. De Morganin lauseita soveltamalla. □

Ennen mittateorian tuomaa tulikoneistoa todennäköisyyslaskennalle σ-albegraan liittyvät tulikoneistot eivät ole kovin mielekkäitä, joten jätämme σ-algebraan liittyvät lisätarkastelut myöhemmille kursseille. Määrittelemme nyt todennäköisyyden tai todennäköisyyskuvauksen seuraavasti todennäköisyyden aksioomien (*Kolmogorov: Foundations of the Theory of Probability (1933)*) mukaisesti:

Määritelmä 1.22. Todennäköisyyden aksioomat. Olkoon Ω perusjoukko ja $\mathcal{F}$, σ-algebra, sen tapahtumien joukko . Kuvaus $P : \mathcal{F} \to \mathbb{R}$ on todennäköisyys (lyh. tn), jos

(TN1) $P(A) \geq 0$ kaikille tapahtumille $A \in \mathcal{F}$,

(TN2) $P(\Omega) = 1$,

(TN3) Kaikille tapahtumille $A_1, A_2, \dots \in \mathcal{F}$ pätee

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

kun ne ovat erillisiä, eli $A_i \cap A_j = \emptyset$ kaikille $i \neq j$ (täysadditiivisuus).

Ensimmäinen aksiooma (TN1) vaatii, että kaikkien tapahtumien todennäköisyyden on oltava epänegatiivinen, eli vähintään nolla. Toinen aksiooma (TN2) taas vaatii, että koko perusjoukon Ω todennäköisyyden on oltava 1, eli että jokaisella satunnaiskokeen suorituskerralla joku lopputulos tapahtuu. Kolmas aksiooma (TN3) vaatii, että kaikkien erillisten tapahtumien numeroituvan yhdisteen todennäköisyyden täytyy olla näiden tapahtumien todennäköisyyksien summa. Tätä todennäköisyyden ominaisuutta kutsutaan yleensä täysadditiivisuudeksi.

Määritelmä 1.23. Todennäköisyysavaruus. Kolmikkoa (triplettiä) $(\Omega, \mathcal{F}, P)$, missä Ω on perusjoukko, $\mathcal{F}$ sen σ-algebra, ja $P : \mathcal{F} \to \mathbb{R}$ todennäköisyys, kutsutaan *todennäköisyysavaruudeksi* (lyh. tn-avaruudeksi).

Määritelmä 1.24. Diskreetti todennäköisyysavaruus. Jos perusjoukko Ω on äärellinen, eli $\Omega = \{\omega_1, \dots, \omega_n\}$, tai numeroituvasti ääretön, eli $\Omega = \{\omega_1, \omega_2, \dots\}$, tarkasteltavien tapahtumien joukoksi $\mathcal{F}$ voidaan aina valita perusjoukon kaikkien osajoukkojen joukko $\mathcal{P}(\Omega)$, ja todennäköisyysavaruutta $(\Omega, \mathcal{P}(\Omega), P)$ kutsutaan *diskreetiksi* tn-avaruudeksi.

Koska tapahtumien joukoksi voidaan diskreetissä tapauksessa aina valita kaikkien perusjoukon osajoukkojen joukko, diskreettiä tn-avaruutta merkitään usein lyhyemmin parilla (Ω, P).

Määritelmä 1.25. Äärellinen todennäköisyysavaruus. Diskreettiä todennäköisyysavaruutta (Ω, P) kutsutaan *äärelliseksi* tn-avaruudeksi, jos sen perusjoukko Ω on äärellinen, eli $\Omega = \{\omega_1, \dots, \omega_n\}$ jollekin $n \in \{1, 2, \dots\}$.

Osoitetaan seuraavaksi, että diskreetin perusjoukon tapauksessa kaikkien tapahtumien todennäköisyydet määräytyvät yksikäsitteisesti alkeistapausten todennäköisyyksien summana.

Lause 1.26. *Olkoon Ω joko äärellinen, eli $\Omega = \{\omega_1, \dots, \omega_n\}$, tai ääretön, eli $\Omega = \{\omega_1, \omega_2, \dots\}$, diskreetti perusjoukko, ja vastaavasti luvut $p_1, \dots, p_n$ tai $p_1, p_2, \dots \geq 0$ s.e. $\sum_{i=1}^{\infty} p_i = 1$ [1]. Nyt on olemassa yksi ja vain yksi todennäköisyys, eli aksioomat (TN1)-(TN3) toteuttava joukkofunktio $P : \mathcal{P}(\Omega) \to \mathbb{R}$, jolle pätee*

$$P(\{\omega_i\}) = p_i \quad \textit{kaikille} \quad i \in \{1, 2, \dots\}. \tag{1.27}$$

Todistus. "$\Rightarrow$": Osoitetaan, että on olemassa ainakin yksi todennäköisyys, jolle 1.27 pätee.

Määritellään funktion $P : \mathcal{P}(\Omega) \to \mathbb{R}$ arvoksi jokaiselle joukolle $A \in \mathcal{P}(\Omega)$:

$$P(A) := \sum_{\omega_i \in A} p_i. \tag{1.28}$$

Tästä P:n määritelmästä seuraa suoraan, että P toteuttaa ehdon 1.27. Osoitetaan vielä, että P toteuttaa todennäköisyyden aksioomat:

(TN1) Määritelmän P:lle (1.28) nojalla jokaiselle tapahtumalle $A \in \mathcal{P}(\Omega)$ pätee

$$P(A) = \sum_{\omega_i \in A} p_i \geq 0.$$

Epänegatiivisuus seuraa siitä, että oletuksen nojalla $p_i \geq 0$ kaikille $i \in \{1, 2, \dots\}$.

[1] Äärelliselle perusjoukolle käytetään tässä lauseessa merkintöjen selkeyttämiseksi samaa merkintää $\sum_{i=1}^{\infty} p_i$ kuin äärettömälle perusjoukolle. Tällöin asetetaan $p_i = 0$ kaikille $i > n$.

(TN2) Jälleen määritelmän 1.28 ja oletuksen nojalla perusjoukolle Ω pätee

$$P(\Omega) = \sum_{\omega_i \in \Omega} p_i = \sum_{i=1}^{\infty} p_i = 1.$$

(TN3) Olkoon $A_1, A_2, \dots \in \mathcal{P}(\Omega)$ jono erillisiä tapahtumia. Myös tapahtumien yhdiste $A := \bigcup_{j=1}^{\infty} A_j \in \mathcal{P}(\Omega)$ voidaan esittää numeroituvana yhdisteenä perusjoukon alkioista: $A = \bigcup_{\omega_i \in A} \{\omega_i\}$. Siten määritelmästämme seuraa täysadditiivisuus:

$$\begin{aligned} P\left(\bigcup_{j=1}^{\infty} A_j\right) &= P\left(\bigcup_{j=1}^{\infty} \bigcup_{\omega_i \in A_j} \{\omega_i\}\right) = P\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = P(A) \\ &\overset{(1.28)}{=} \sum_{\omega_i \in A} p_i = \sum_{j=1}^{\infty} \sum_{\omega_i \in A_j} p_i \overset{(1.28)}{=} \sum_{j=1}^{\infty} P(A_j). \end{aligned}$$

Toiseksi viimeisessä yhtäsuuruudessa hyödynnettiin oletusta tapahtumien $A_1, A_2, \dots$ erillisyydestä.

Tällä tavoin voimme siis aina löytää todennäköisyyden P, joka toteuttaa ehdon 1.27. Osoitetaan vielä, että tämä P on yksikäsitteinen, eli jos todennäköisyys P^* toteuttaa ehdon 1.27, on välttämättä $P^* = P$, eli $P^*(A) = P(A)$ kaikille $A \in \mathcal{P}(\Omega)$.

"$\Leftarrow$": Olkoon P^* todennäköisyys, eli aksioomat (TN1)-(TN3) toteuttava joukkofunktio $P^* : \mathcal{P}(\Omega) \to \mathbb{R}$, jolle 1.27 pätee. Osoitetaan, että jokaiselle joukolle $A \in \mathcal{P}(\Omega)$ pätee 1.28, eli että

$$P^*(A) = \sum_{\omega_i \in A} p_i,$$

mistä seuraa $P^* = P$.

Tehdään vastaoletus: on olemassa joukko $A \in \mathcal{P}(\Omega)$ s.e.

$$P^*(A) \neq \sum_{\omega_i \in A} p_i.$$

Vastaoletuksen nojalle erillisille joukoille $\{\omega_i\} \in A$ pätee (Huomaa, että sekä yhdiste $\bigcup_{\omega_i \in A} \{\omega_i\}$, että summa $\sum_{\omega_i \in A} P^*(\{\omega_i\})$ ovat numeroituvia. Jos A on äärellinen, valitaan yhdisteen lopuiksi joukoiksi tyhjiä joukkoja.):

$$P^*\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = P^*(A) \neq \sum_{\omega_i \in A} p_i \overset{1.27}{=} \sum_{\omega_i \in A} P^*(\{\omega_i\}).$$

Mutta tämä on ristiriidassa P^*:n oletetun täysadditiivisuuden kanssa. Siten vastaoletus on väärä, joten on oltava $P^* = P$.

$\square$

Esimerkki 1.29. Symmetrinen tn-avaruus. Olkoon $\Omega = \{\omega_1, \dots, \omega_n\}$ perusjoukko, ja asetetaan

$$P(\{\omega_i\}) = p_i = \frac{1}{n} \quad \text{kaikille} \quad i \in \{1, \dots, n\}.$$

Nyt lauseen 1.26 nojalla (huomaa, että $\sum_{i=1}^{n} p_i = 1$) P määrää todennäköisyyden yksikäsitteisesti kaikille tapahtumille $A \in \mathcal{P}(\Omega)$ siten, että

$$P(A) = \sum_{\omega_i \in A} p_i = \frac{|A|}{n}. \tag{1.30}$$

Tällaista tn-avaruutta (Ω, P), jonka kaikkien alkeistapausten todennäköisyydet ovat samat, kutsutaan *symmetriseksi* todennäköisyysavaruudeksi.

Huomaa, että näin määritelty P on yhtenevä todennäköisyyden klassisen määritelmän (määritelmä 1.3) kanssa. Siten klassisen määritelmän mukainen todennäköisyys on erikoistapaus yleisemmän määritelmän (määritelmä 1.22) mukaisesta todennäköisyydestä.

Kaikki aiemmin tällä kurssilla käsitellyt esimerkit, kuten kolikonheitto, kahden nopan heitto, pokerikäsi ja 4 pallon nostaminen laatikosta, on mallinnettu symmetrisen tn-avaruuden avulla. Esitellään seuraavaksi esimerkit epäsymmetrisestä diskreetistä todennäköisyysavaruudesta äärellisellä ja äärettömällä[2] perusjoukolla.

Esimerkki 1.31. Heitä sikaa. Suositussa Heitä sikaa-pelissä heitetään sian muotoisia arpakuutioita. Pisteitä heitosta saa sen mukaan, mihin asentoon sika jää maahan tullessaan. Pisteet ja frekvenssit [3] yhdelle sialle ovat seuraavat:

	Nimi	Asento	Pisteet	Frekvenssi (%)
ω_1	Siankylki	Kyljellään	0 p.	65.1
ω_2	Sianselkä	Selällään	5 p.	22.5
ω_3	Tavallinen sika	Jaloillaan	5 p.	8.8
ω_4	Siansyltty	Etujalkojen ja kärsän varassa	10 p.	3.0
ω_5	Etupotka	Toisen korvan ja etujalan varassa	15 p.	0.6

[2] Itse asiassa ääretön *diskreetti* tn-avaruus ei voi olla symmetrinen. Tämän voi itse kokeilla osoittaa tn:n aksioomien perusteella.

[3] Simuloitu standardoidulla pinnalla ja automaattisella heittolaitteella (otoskoko $n = 11954$). Lähde: `https://en.wikipedia.org/wiki/Pass_the_Pigs`

Asetetaan P luvuille $p_1 = 0.651, p_2 = 0.225, p_3, = 0.088, p_4 = 0.030$ ja $p_5 = 0.006$ seuraavasti:

$$P(\{\omega_i\}) = p_i \quad \text{kaikille} \quad i \in \{1,2,3,4,5\},$$

Lauseen 1.30 nojalla P on todennäköisyys perusjoukolla $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ (huomaa, että $\sum_{i=1}^5 p_i = 1$). Siten (Ω, P) on äärellinen (ja diskreetti) tn-avaruus.

Lisäksi kaikkien tapahtumien todennäköisyydet saadaan niiden sisältämien alkeistapauksien todennäköisyyksien summana. Esimerkiksi tapahtuman A : ”Heitolla saadaan tasan 5 pistettä” todennäköisyys on

$$P(A) = P(\{\omega_2, \omega_3\}) = p_2 + p_3 = 0.313.$$

Esimerkki 1.32. Kotijoukkueen tekemien maalien määrä jalkapallo-ottelussa. Periaatteessa, vaikkakaan ei käytännössä [4], joukkueen tekemien maalien määrä ei ole rajoitettu, vaan se voi olla mikä tahansa epänegatiivinen kokonaisluku. Perusjoukoksi on tällöin luontevaa asettaa

$$\Omega = \{0, 1, 2 \dots\} = \mathbb{N}.$$

Jos perusjoukko on ääretön, perusjoukon alkioiden todennäköisyyksien valinta siten, että ne summautuvat ykköseksi, on hieman hankalampaa kuin äärellisen perusjoukon tapauksessa, sillä nyt kyse on äärettömästä summasta. Yksi vaihtoehto valita todennäköisyydet on seuraava.

Olkoon $\lambda \in (0, \infty)$. Asetetaan

$$p_i = e^{-\lambda} \frac{\lambda^i}{i!} \quad \text{kaikille} \quad i \in \{0, 1, 2, \dots\}.$$

Nyt eksponenttifunktion sarjakehitelmän $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ nojalla

$$\sum_{i=0}^{\infty} p_i = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1,$$

joten luvut p_i määräävät todennäköisyyden äärettömälle perusjoukolle $\Omega = \{0, 1, 2, \dots\} = \mathbb{N}$. Siten $(\mathbb{N}, P)$, missä

$$P(\{i\}) = e^{-\lambda} \frac{\lambda^i}{i!} \quad \text{kaikille} \quad i \in \mathbb{N}$$

[4] Suurin voittomarginaali ammattilaistason jalkapallo-ottelussa on 12. syyskuuta 1885, kun Arbroath voitti Bon Accordin 36-0 Skotlannin cupin ottelussa. Tosin lokakuussa 2002 AS Adema voitti SO l'Emyrne:n 149-0 Madagaskarin mestaruudesta pelatussa neljän joukkueen turnauksessa. Tätä ei kuitenkaan lasketa, sillä kaikki maalit olivat omia: SO l'Emyrne hävisi ottelun tahallaan protestiksi ottelusarjan aiempien ottelujen tuomarityöskentelystä. Lähde: `https://en.wikipedia.org/wiki/Arbroath_36-0_Bon_Accord`, `https://en.wikipedia.org/wiki/AS_Adema_149-0_SO_l%27Emyrne`

on diskreetti tn-avaruus. Näin määriteltyä todennäköisyyttä kutsutaan Poisson-jakaumaksi (tästä lisää myöhemmin satunnaismuuttujien yhteydessä).

Jos jonkin asian sattumisen todennäköisyys jollain aikavälillä on vakio, ja näiden sattumisten ajankohdat ovat toisistaan riippumatonta, näiden tapahtumien määrää tietyllä aikavälillä voidaan mallintaa Poisson-jakaumalla. Onnekkaasti osoittautuu, että esimerkiksi joukkueen jalkapallo-ottelussa tekemien maalien määrä noudattaa melko tarkkaan Poisson-jakaumaa. Seuraavaan taulukkoon on merkitty kotijoukkueen ottelussa tekemien maalien määrät ja frekvenssit kausilla La Ligassa (Espanjan pääsarja) 2010-2015 (yhteensä 2116 ottelua).

Maalit	Ottelut	Frekvenssi (%)	Poisson-jakaumasta ($\lambda = 1.631$) laskettu tn
0	466	22.02	0.1957
1	668	31.57	0.3192
2	498	23.53	0.2604
3	262	12.38	0.1416
4	147	6.95	0.0577
5	50	2.36	0.0188
6	15	0.71	0.0051
7	7	0.33	0.0012
8	2	0.09	0.0002
9	1	0.05	0.0000

Kotijoukkue on siis jäänyt 466 ottelussa nollaan maaliin, 668 tehnyt 1 maalia, ja niin edelleen. Lisäksi taulukkoon on merkitty Poisson-jakaumasta parametrin $\lambda = 1.631$ arvolla (kotijoukkueen maalimäärien keskiarvo aineistossa, tilastollisen mallin sovittamiseen aineistoon perehdytään tarkemmin kurssilla Tilastollinen päättely I) lasketut todennäköisyydet. Vaikka aineistossamme suurin maalimäärä on 9, Poisson-jakauman avulla saadaan positiiviset todennäköisyydet kaikille perusjoukon alkioille: esimerkiksi $P(\{10\}) = 7.20 \cdot 10^{-6}$.

Seuraavassa lauseessa osoitetaan joitain todennäköisyyslaskennan keskeisiä laskusääntöjä, jotka ovat johdettavissa suoraan todennäköisyyden aksioomista.

Lause 1.33. *Olkoon* $(\Omega, \mathcal{F}, P)$ *todennäköisyysavaruus ja* A *ja* B *sen tapahtumia. Tällöin*

(i) $P(\emptyset) = 0$.

(ii) (Äärellinen additiivisuus) Jos tapahtumat $A_1, \dots, A_n$ *ovat erillisiä, eli* $A_i \cap A_j = \emptyset$ *kaikille* $i \neq j$*, niin* $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

(iii) $P(A^C) = 1 - P(A)$. *kaikille* $A_1, \dots, A_n \in \mathcal{F}$.

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

(v) (Monotonisuus) Jos $A \subseteq B$*, niin* $P(A) \leq P(B)$.

(vi) (Yhteenlaskukaava) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Todistus. Todistetaan kohdat (i)-(iii), ja jätetään loput harjoitustehtäväksi.

(i) Koska tyhjät joukot ovat erillisiä, todennäköisyyden additiivisuuden (TN3) nojalla tapahtumille $\emptyset, \emptyset, \dots$ pätee

$$P(\emptyset) = P(\bigcup_{i=1}^{\infty} \emptyset) = P(\emptyset) + P(\emptyset) + \cdots = \sum_{i=1}^{\infty} P(\emptyset)$$

Mutta ainoa reaaliluku, joka toteuttaa tämän yhtälön on 0, joten on oltava $P(\emptyset) = 0$.

(ii) Kun $i > n$, valitaan $A_i = \emptyset$. Nyt väite seuraa täysadditiivisuudesta (TN3) erillisille tapahtumille $A_1, A_2, \dots$ ja kohdasta (i).

(iii) Joukon komplementin määritelmän nojalla $\Omega = A \cup A^C$ ja $A \cap A^C = \emptyset$. Siten todennäköisyyden toisen aksiooman (TN2) ja äärellisen additiivisuuden (osoitettiin edellisessä kohdassa) nojalla

$$1 = P(\Omega) = P(A \cup A^C) = P(A) + P(A^C),$$

mistä väite seuraa.

□

Seuraavassa taulukossa vielä muutamia todennäköisyyslaskennan kannalta keskeisiä joukko-opin merkintöjä tulkittuna todennäköisyyslaskennan termeinä.

Joukko-opin merkintä	Tn-laskennan tulkinta
Ω	perusjoukko / varma tapahtuma
$\omega \in \Omega$	alkeistapaus / satunnaiskokeen tulosmahdollisuus
$\emptyset$	mahdoton tapahtuma
$\mathcal{F}$	kaikkien tapahtumien joukko
A	A tapahtuu
A^C	A ei tapahdu
$A \cup B$	A tai B tapahtuu
$A \cap B$	A ja B tapahtuvat
$\bigcup_{j=1}^{\infty} A_j$	ainakin yksi tapahtumista $A_1, A_2, \ldots$ tapahtuu
$\bigcap_{j=1}^{\infty} A_j$	kaikki tapahtumat $A_1, A_2, \ldots$ tapahtuvat
$A \setminus B$	A tapahtuu, mutta B ei tapahdu
$A \subseteq B$	jos A tapahtuu, niin myös B tapahtuu
$A \cap B = \emptyset$	A ja B eivät voi tapahtua yhtä aikaa

1.4 Ehdollinen todennäköisyys

Määritelmä 1.34. Olkoot A ja B tapahtumia siten että $P(B) \geq 0$. Tapahtuman A *todennäköisyys ehdolla* B on

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

Voidaan myös osoittaa, että ehdollinen todennäköisyys kiinteän tapahtuman B, jolle $P(B) \geq 0$, suhteen toteuttaa todennäköisyyden aksioomat (TN1)-(TN3), eli että se on todennäköisyys (sen ensimmäisen argumentin A suhteen). Tämä mahdollistaa kaikkien todennäköisyyden laskusääntöjen soveltamisen myös ehdolliselle todennäköisyydelle.

Esimerkki 1.35. Kahden nopan heitto. Valitaan perusjoukoksi jälleen $\Omega = \{1, \ldots, 6\}^2$, ja todennäköisyydeksi

$$P(\{w\}) = \frac{1}{|\Omega|} = \frac{1}{36} \quad \text{kaikille} \quad \omega \in \Omega,$$

jolloin (Ω, P) on symmetrinen tn-avaruus.

(i) Mikä on tn tapahtumalle A: ”Silmälukujen summa on vähintään 8”?

Laskemalla tapahtumalle A suotuisat alkeistapaukset saadaan

$$P(A) = \frac{15}{36}$$

(ii) Mikä on tapahtuman A todennäköisyys ehdolla B : ”Ensimmäisen nopan silmäluku on kuusi”?

Yhdistelmät, joilla kumpikin tapahtuma toteutuu, ovat $(6,2), (6,3), (6,4), (6,5)$ ja $(6,6)$, joten

$$P(A \cap B) = \frac{5}{36}.$$

Siten ehdollisen todennäköisyyden kaavasta saadaan

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{5}{36}}{\frac{1}{6}} = \frac{5}{6}.$$

Tieto siitä, että ensimmäisen nopan silmäluku on kuusi, muutti siis huomattavasti arviotamme tapahtuman A todennäköisyydestä.

Kirjoittamalla ehdollisen todennäköisyyden määritelmä tulomuodossa saadaan kertolaskusääntö, jota kutsutaan myös todennäköisyyden ketjusäännöksi.

Lause 1.36. *Kertolaskusääntö. Olkoot A ja B tapahtumia ja $P(B) \geq 0$. Nyt*

$$P(A \cap B) = P(A|B)P(B).$$

Todistus. Suoraan ehdollisen tn:n määritelmästä saadaan:

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

□

Kertolaskulaskusääntö yleistyy myös suoraan useammalle tapahtumalle.

Lause 1.37. *Yleinen ketjusääntö. Olkoot $A_1, \ldots, A_n$ tapahtumia siten, että $P(A_1 \cap \cdots \cap A_{n-1}) \geq 0$. Nyt*

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cap \cdots \cap P(A_n|A_1 \cap \cdots \cap A_{n-1}) \\ &= P(A_1) \prod_{i=2}^{n} P(A_i| \cap_{j=1}^{i-1} A_j). \end{aligned}$$

Todistus. Induktiolla kertolaskusäännöstä. □

1.5 Tapahtumien riippumattomuus

Määritelmä 1.38. Kahden tapahtuman riippumattomuus. Tapahtumia A ja B kutsutaan *riippumattomiksi*, jos

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

Tällöin merkitään $A \perp\!\!\!\perp B$.

Tapahtumien riippumattomuus voidaan yhtäpitävästi esittää myös ehdollisen todennäköisyyden avulla, kuten seuraavasta lauseesta nähdään. Tapahtuma A on siis riippumaton tapahtumasta B, jos ehdollistaminen B:llä ei muuta sen todennäköisyyttä; toisin sanoen tieto siitä, että B on tapahtunut, ei muuta arviotamme A:n todennäköisyydestä.

Lause 1.39. *Olkoot tapahtumat A ja B siten että $P(B) \geq 0$. Nyt $A \perp\!\!\!\perp B$ jos ja vain jos $P(A|B) = P(A)$.*

Todistus. "$\Rightarrow$": Ehdollisen tn:n määritelmästä saadaan tapahtumien A ja B riippumattomuuden nojalla:

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

"$\Leftarrow$": Kertolaskusäännön ja oletuksen nojalla:

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B).$$

□

Riippumattomuuden määritelmä yleistyy myös useammalle kuin kahdelle tapahtumalle. Tällöin vaaditaan, että kaikkien mahdollisten tarkasteltavien tapahtumien keskinäisten leikkausten todennäköisyydet saadaan niiden todennäköisyyksien tulona.

Määritelmä 1.40. Tapahtumia $A_1, \ldots, A_n$ kutsutaan riippumattomiksi, jos

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

kaikille indeksijoukoille $I \subseteq \{1, \ldots, n\}$. Tällöin merkitään $A_1, \ldots, A_n \perp\!\!\!\perp$.

Esimerkiksi kolmelle tapahtumalla A, B ja C riippumattomuus tarkoittaa, että tapahtumat ovat pareittain riippumattomia, eli että

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

ja lisäksi, että

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Esimerkki 1.41. Esimerkki riippumattomista tapahtumista ja tapahtumista, jotka eivät ole riippumattomia.

(i) Heitetään kahta noppaa, jolloin perusjoukko on jälleen $\Omega = \{1, 2, \ldots, 6\}^2$. Tapahtumien A : "Ensimmäisen nopan silmäluku on 6" ja tapahtuman B : "toisen nopan silmäluku on kuusi" todennäköisyydet ovat $P(A) = \frac{1}{6}$ ja $P(B) = \frac{1}{6}$. Toisaalta vain alkeistapaus $(6, 6)$ toteuttaa kummankin näistä tapahtumista, joten

$$P(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(A)P(B).$$

Tapahtumat A ja B ovat siis riippumattomia.

(ii) Nostetaan kaksi korttia pakasta palauttamatta ensimmäiseksi nostettua korttia välillä takaisin pakkaan. Perusjoukoksi voidaan valita korttipakan 2-permutaatiot, eli kahden alkion järjestetyt jonot. Sekä tapahtuman A : "Ensimmäiseksi nostettu kortti on ässä", että tapahtuman B : "Toiseksi nostettu kortti on ässä" tn on perusjoukon symmetrian nojalla

$$P(A) = \frac{4 \cdot 51}{52 \cdot 51} = \frac{1}{13} = \frac{51 \cdot 4}{52 \cdot 51} = P(B).$$

Mahdollisia yhdistelmiä, joissa sekä ensimmäinen, että toinen nostettu kortti ovat ässiä, on $4 \cdot 3$, joten tapahtumien A ja B leikkauksen todennäköisyys on

$$P(A \cap B) = \frac{4 \cdot 3}{52 \cdot 51} = \frac{1}{221} \neq \frac{1}{169} = P(A)P(B),$$

joten tapahtumat A ja B eivät ole riippumattomia.

Nähdään myös, että todennäköisyys nostaa toisella kortilla ässä, jos ensimmäinen nostettu kortti oli ässä on

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{221}}{\frac{1}{13}} = \frac{1}{17} \neq \frac{1}{13} = P(B).$$

1.6 Kokonaistodennäköisyys ja Bayesin kaava

Esimerkki 1.42. Tarkastellaan satunnaiskoetta, jossa ensin valitaan satunnaisesti yksi kolmesta laatikosta, joissa on valkoisia ja mustia palloja seuraavan taulukon mukaisesti.

Laatikko	Valkoiset	Mustat
1	1	1
2	2	1
3	3	1

Valitaan sen jälkeen ensimmäisessä vaiheessa valitusta laatikosta satunnaisesti yksi pallo. Mikä on tapahtuman A : ”valittu pallo on valkoinen” todennäköisyys?

Tätä satunnaiskoetta ei voi mallintaa symmetristen alkeistapauksien avulla, sillä eri pallojen todennäköisyydet poikkeavat toisistaan. Monissa tilanteissa tapahtumien ehdolliset todennäköisyydet on helpompi päätellä kuin ehdottomat todennäköisyydet. Niin tässäkin: kun merkitsemme B_i : ”valitaan i:s laatikko” kaikille $i \in \{1, 2, 3\}$, tapahtuman A ehdolliset todennäköisyydet ehdolla B_i ovat:

$$P(A|B_1) = \frac{1}{2}, \quad P(A|B_2) = \frac{2}{3}, \quad P(A|B_3) = \frac{3}{4}.$$

Koska laatikko valitaan satunnaisesti, $P(B_i) = \frac{1}{3}$ kaikille $i \in 1, 2, 3$.

Valkoinen pallo voidaan nostaa kolmesta eri laatikosta, joten tapahtuma A voidaan jakaa erillisiin toisensa poisulkeviin osiin $A \cap B_1$, $A \cap B_2$ ja $A \cap B_3$. Näiden todennäköisyydet saadaan kertolaskusäännöstä, ja koska kyseiset tapahtumat ovat erillisiä, A:n todennäköisyys saadaan todennäköisyyden additiviisuuden nojalla:

$$\begin{aligned} P(A) &= P\left(\bigcup_{i=1}^{3}(A \cap B_i)\right) = \sum_{i=1}^{3} P(A \cap B_i) \\ &= \sum_{i=1}^{3} P(B_i)P(A|B_i) = \left(\frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{3}{4}\right) = \frac{23}{36}. \end{aligned}$$

Osoittautuu, että vastaavaa päättelyä voidaan soveltaa yleisemminkin: yleistä tapausta kutsutaan *kokonaistodennäköisyyden kaavaksi*. Määritellään ensin tarkemmin, mitä tarkoittaa joukon, eli tapahtuman *ositus*.

Määritelmä 1.43. Ositus. Joukot $B_1, \ldots, B_n$ osittavat joukon X, eli muodostavat joukon X *osituksen*, jos ne toteuttavat seuraavat ehdot:

(i) Ne ovat erillisiä:

$$B_i \cap B_j \quad \text{kaikille} \quad i \neq j.$$

(ii) Niiden yhdiste on koko tarkasteltava joukko

$$\bigcup_{i=1}^{n} B_i = X.$$

Todennäköisyyslaskennassa tärkeä erityistapaus on $X = \Omega$, jolloin sanotaan, että joukot $B_1, \ldots, B_n$ ovat perusjoukon ositus.

Lause 1.44. *Kokonaistodennäköisyyden kaava. Olkoon* $(\Omega, \mathcal{F}, P)$ *tn-avaruus,* $A \in \mathcal{F}$ *tapahtuma, ja* $B_1, \dots, B_n \in \mathcal{F}$ *perusjoukon* Ω *ositus. Nyt*

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

Todistus. Joukot $(A \cap B_1), \dots, (A \cap B_n)$ muodostavat joukon A osituksen, joten todennäköisyyden additiivisuuden ja kertolaskusäännön nojalla

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

□

Esimerkki 1.45. Jatkoa esimerkkiin 1.42. Oletetaan, että koe suoritetaan verhon takana, eli emme näe mistä laatikosta pallo on nostettu, vaan pelkästään nostetun pallon värin. Pallo on valkoinen, eli havaitsemme tapahtuman A. Mikä on nyt tapahtuman B_1 : "Valittiin ensimmäinen laatikko" todennäköisyys ehdolla A, eli mikä on todennäköisyys sille, että pallo nostettiin laatikosta 1, kun tiedetään että nostettu pallo on valkoinen?

Ehdollisen todennäköisyyden määritelmää ja kertolaskusääntöä soveltamalla saadaan:

$$P(B_1|A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(B_1)P(A|B_1)}{P(A)} = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{23}{36}} = \frac{6}{23}.$$

Vastaavasti voidaan laskea todennäköisyydet toiselle ja kolmannelle laatikolle:

$$P(B_2|A) = \frac{\frac{1}{3} \cdot \frac{2}{3}}{\frac{23}{36}} = \frac{8}{23}$$
$$P(B_3|A) = \frac{\frac{1}{3} \cdot \frac{3}{4}}{\frac{23}{36}} = \frac{9}{23}.$$

Tämä päättely on esimerkki todennäköisyyden keskeisen laskusäännön, Bayesin kaavan käytöstä. Bayesin kaavaa voidaan soveltaa tilanteessa, jossa olemme havainneet tapahtuman A, ja tiedämme tapahtuman A ehdolliset todennäköisyydet jonkin perusjoukon osituksen suhteen:

$$P(A|B_i) \quad \text{kaikille} \quad i \in \{1, \dots, n\}.$$

Tällöin Bayesin kaavan avulla saadaan laskettua tapahtumien $B_1, \dots, B_n$ ehdolliset todennäköisyydet ehdolla A:

$$P(B_i|A) \quad \text{kaikille} \quad i \in \{1, \dots, n\}.$$

Lause 1.46. *Bayesin kaava. Olkoon A tapahtuma, jolle $P(A) \geq 0$, ja $B_1, \ldots, B_n$ perusjoukon ositus. Jokaisen tapahtuman B_i ehdollinen todennäköisyys ehdolla A saadaan Bayesin kaavasta:*

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

Todistus. Käyttämällä ehdollisen todennäköisyyden määritelmää ja kertolaskusääntöä saadaan

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

Jälkimmäinen muotoilu saadaan soveltamalla kokonaistodennäköisyyden kaavaa nimittäjään. □

Tarkastellaan vielä klassista Bayesin kaavan sovellusta.

Esimerkki 1.47. Oletetaan, että harvinainen sairaus esiintyy 0.1 prosentilla väestöstä, ja että meillä on erittäin tarkka testi sen toteamiseksi. Testin sensitiivisyys, eli positiivisen testituloksen tn jos henkilölle, jolla on tutkittava sairaus, on 0.95. Testin spesifisyys, eli positiivisen testituloksen tn terveelle henkilölle on 0.01. Mikä on todennäköisyys, että testattavalla (satunnaisesti väestöstä valitulla) henkilöllä on tutkittava sairaus, jos testin tulos positiivinen?

Merkitään tapahtumia S : ”henkilöllä on tutkittava sairaus”, T : ”henkilön testitulos on positiivinen”. Tehtävänannosta saadaan seuraavat todennäköisyydet:

$$\begin{aligned} P(S) &= 0.001 \\ P(S^C) &= 0.999 \\ P(T|S) &= 0.95 \\ P(T|S^C) &= 0.01. \end{aligned}$$

Tapahtumat S ja S^C osittavat perusjoukon, joten todennäköisyys, että henkilö on sairas ehdolla että testitulos on positiivinen, saadaan Bayesin kaavasta:

$$\begin{aligned} P(S|T) &= \frac{P(S)P(T|S)}{P(S)P(T|S) + P(S^C)P(T|S^C)} \\ &= \frac{0.001 \cdot 0.95}{0.001 \cdot 0.95 + 0.999 \cdot 0.01} \approx 0.087. \end{aligned}$$

Todennäköisyys, että positiivisen testituloksen saanut henkilö on oikeasti sairas, saatiin siis tutkimalla, mikä on odotettu oikeiden positiivisten (positiivinen testitulos sairaalle) osuus kaikista positiivisista testituloksista, eli yhteenlasketuista oikeista positiivista ja vääristä positiivista (positiivinen testitulos terveelle) testituloksista.

Sairaiden pieni odotettu osuus positiivisen testituloksen saaneista selittyy tietenkin sairaiden pienellä osuudella väestöstä: testin tarkkuudesta huolimatta on huomattavasti todennäköisempää, että satunnaisesti väestöstä valittu testihenkilö on terve **ja** saa positiivisen testituloksen, kuin että henkilö on sairas **ja** saa positiivisen testituloksen.

1.7 Toistokokeet

Joskus on mielekästä tarkastella tilannetta, jossa haluamme toistaa tiettyä satunnaiskoetta n kertaa. Rajoitamme tilanteen siten, että tarkastelemme vain sellaisten satunnaiskokeiden toistoa, jotka ovat riippumattomia toisistaan. Lisäksi kiinnitämme yksinkertaisuuden vuoksi huomiota yksittäisessä kokeessa vain siihen, sattuuko joku tietty tapahtuma A vai ei. Jos merkitsemme $P(A) = p$, $P(A^c) = 1 - p = q$, ja edelleen

$$B_k = \text{"}A \text{ tapahtuu vain ja ainoastaan } k \text{ kertaa"},$$

voimme laskea todennäköisyyden $P(B_k)$.

Esimerkki 1.48. Tarkastellaan tilannetta, jossa haluamme toistaa yhden nopan heittoa 5 kertaa ja olemme kiinnostuineita silmäluvun 6 ilmentymistä. Nyt $P(A) =$"kutonen", ja $B_3 =$"viidestä heitosta saamme 3 kutosta". Lisäksi $p = 1/6$ ja $q = 5/6$. 5-kertaista toistokoetta kuvailevan todennäköisyysavaruuden alkeistapauksiksi valitsemme kaikki jonot $(\omega_1, \omega_2, \ldots, \omega_5)$, missä $\omega_i = \{A, A^c\}$ kaikilla $i \in 1, 2, 3, 4, 5$.

Koska erilliset toistot ovat toisistaan riippumattomia, sellaisen jonon todennäköisyys, missä saamme kutosia 3 kappaletta ja ei-kutosia 2 kappaletta, on

$$p^3q^2 = (\frac{1}{6})^3(\frac{5}{6})^2.$$

kpl. Tämän lisäksi tällaisia jonoja (eli mahdollisia tapahtumajärjestyksiä, missä saamme 3 kutosta) on yhteensä $\binom{5}{3}$ kappaletta. Edelleen tuloperiaatteen nojalla

$$P(B_3) = \binom{5}{3}(\frac{1}{6})^3(\frac{5}{6})^2 \approx 0.032.$$

Yleisesti toistokoe ja siihen liittyvä todennäköisyys $P(B_k)$ käyttäytyy seuraavasti:

Lause 1.49. *Toistokoe. Merkitään $P(A) = p$. Tällöin n-kertaisessa toistokokeessa tapahtuman $B_k =$ "A tapahtuu tasan k kertaa" todennäköisyys on*

$$P(B_k) = \binom{n}{k} p^k q^{n-k},$$

missä $k = 0, 1, \ldots, n$ ja $q = 1 - p$.

Todistus. Todistus kuten esimerkissä 1.48. □

Esimerkki 1.50. Otanta takaisinpanolla -tyylinen tilanne voidaan tulkita n-kertaiseksi toistokokeeksi, missä $A =$”satunnaisesti valittu pallo on valkoinen”, ja edelleen $p = K/N$ lauseen 1.19 merkinnöin.

Luku 2

Satunnaismuuttujat

Satunnaismuuttujaksi (lyh. sm, engl. *random variable*) kutsutaan satunnaiskokeeseen liittyvää funktiota, joka liittää jokaiseen satunnaiskokeen mahdolliseen tulokseen, eli alkeistapaukseen, jonkin reaaliluvun.

Määritelmä 2.1. Satunnaismuuttuja. Olkoon $(\Omega, \mathcal{F}, P)$ tn-avaruus. Funktiota [1] $X : \Omega \to \mathbb{R}$ kutsutaan *satunnaismuuttujaksi.*

Joskus käytämme lyhennettyä merkintää $X(\omega) = X$.

Määritelmä 2.2. Diskreetti satunnaismuuttuja. Tn-avaruuden $(\Omega, \mathcal{F}, P)$ satunnaismuuttujaa $X : \Omega \to \mathbb{R}$ kutsutaan *diskreetiksi satunnaismuuttujaksi*, jos sen arvojoukko $X(\Omega)$ on numeroituva, eli äärellinen tai numeroituvasti ääretön.

Esimerkki 2.3. Tarkastellaan kolmea yksinkertaista esimerkkiä diskreetistä satunnaismuuttujasta.

(i) Yhden nopan heitto. Määritellään perusjoukolle $\Omega = \{1, 2, 3, 4, 5, 6\}$ satunnaismuuttuja $X : \Omega \to \mathbb{R},\ X(\omega) = \omega$. Satunnaismuuttuja X siis kuvaa heitetyn nopan silmälukua. Esimerkiksi joukko $\{X(\omega) = X = 3\}$ kuvaa tapahtumaa, jossa heitetty noppa saa silmäluvun kolme.

Tämän satunnaismuuttujan arvojoukko on $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$, joten kyseessä on diskreetti satunnaismuuttuja.

(ii) Kahden nopan heitto. Määritellään perusjoukolle $\Omega = \{(a, b) : a, b \in \{1, 2, 3, 4, 5, 6\}\}$ satunnaismuuttuja $Y(\omega) = Y(a, b) = a + b$. Satunnaismuuttuja Y kuvaa nyt heitettyjen noppien silmälukujen summaa.

[1] Tarkemmin ottaen funktiolta vaaditaan, että se on Borel-mitallinen, eli että jokaisen riittävän Borel-joukon, eli riittävän säännölisen joukon, kuten välien ja välien numeroituvien yhdisteiden ja leikkausten, alkukuva on perusjoukon tapahtuma, eli että $X^{-1}(B) \in \mathcal{F}$ kaikille riittävän "siisteille"$B \subseteq \mathbb{R}$.

Satunnaismuuttujan Y arvojoukko on $Y(\Omega) = \{2, 3, \ldots, 12\}$, joten kyseessä on myös diskreetti satunnaismuuttuja.

(iii) Heitä sikaa (vrt. esimerkki 1.31). Olkoon perusjoukko $\Omega = \{\omega_1, \ldots, \omega_5\}$ seuraavilla todennäköisyyksillä.

	Nimi	Asento	Pisteet	$P\{\omega_i\}$
ω_1	Siankylki	Kyljellään	0 p.	0.651
ω_2	Sianselkä	Selällään	5 p.	0.225
ω_3	Tavallinen sika	Jaloillaan	5 p.	0.088
ω_4	Siansyltty	Etujalkojen ja kärsän varassa	10 p.	0.030
ω_5	Etupotka	Toisen korvan ja etujalan varassa	15 p.	0.006

Määritellään satunnaismuuttuja $Z : \Omega \to \mathbb{R}$ kunkin perusjoukon alkion, eli sian laskeutumisasennon, tuottamaksi pistemääräksi; esimerkiksi $Z(\omega_1) = 0$. Satunnaismuuttujan Z arvojoukko on $Z(\Omega) = \{0, 5, 10, 15\}$, eli kyseessä on myös diskreetti satunnaismuuttuja.

Tapahtumaa, että satunnaismuuttuja X saa arvoja joukosta $B \in \mathbb{R}$ merkitään seuraavasti:

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}.$$

Tällöin todennäköisyyttä, että satunnaismuuttuja saa arvoja joukosta B voidaan merkitä jollain seuraavista tavoista:

$$P(X \in B) = P\{X \in B\} = P(\{X \in B\}) = P(\{\omega \in \Omega : X(\omega) \in B\}).$$

Jälleen (vrt huom. 1.5) toiset sisäkkäisistä sulkeista yleensä tiputetaan pois merkinnöistä. Tällä kurssilla sovelletaan jatkossa merkintää $P\{X \in B\}$.

Samalla tavoin kuin todennäköisyys on joukkofunktio, joka on määritelty perusjoukon osajoukoille eli tapahtumille, satunnaismuuttujan jakauma on joukkofunktio, joka on määritelty satunnaismuuttujan arvojoukon osajoukoille.

Määritelmä 2.4. Satunnaismuuttujan jakauma. Olkoon $X : \Omega \to \mathbb{R}$ satunnaismuuttuja. Satunnaismuuttujan X *jakauma* on joukkofunktio

$$P_X := P\{X \in B\},$$

joka on määritelty kaikille satunnaismuuttujan X arvojoukon riittävän siisteille[2] osajoukoille $B \subseteq X(\Omega)$.

Huomautus 2.5. Voidaan osoittaa, että satunnaismuuttujalle X sen jakauma P_X toteuttaa todennäköisyyden aksioomat (TN1-TN3), ja siten määrittää todennäköisyyden satunnaismuuttujan arvojoukossa. Tämän takia monissa käytännön sovelluksissa alkuperäistä todennäköisyysavaruutta ja perusjoukkoa ei tarvitse erikseen määritellä, vaan puhutaan yleensä pelkästään satunnaismuuttujasta ja sen jakaumasta. Voidaankin ajatella, että perusjoukko on satunnaismuuttujan arvojoukko $X(\Omega)$, tapahtumat ovat sen osajoukkoja ja että todennäköisyys on satunnaismuuttujan jakauma P_X.

Määritelmä 2.6. Satunnaismuuttujan $X : \Omega \to \mathbb{R}$ *pistetodennäköisyysfunktio* (lyh. ptnf) on funktio $f : \mathbb{R} \to \mathbb{R}$, jolle

$$f(x) = P\{X = x\} \quad \text{kaikille} \quad x \in \mathbb{R}.$$

Pistetodennäköisyysfunktio on periaattessa määritelty kaikille reaaliluvuille, mutta sen arvo on nolla satunnaismuuttujan X arvojoukon ulkopuolella:

$$f(x) = 0 \quad \text{kaikille} \quad x \notin X(\Omega).$$

Useampia satunnaismuuttujia käsiteltäessä satunnaismuuttujaa, jonka ptnf on kyseessä, merkitään usein sekaannusten välttämiseksi alaindeksillä:

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

Huomautus 2.7. Pistetodennäköisyysfunktio siis kertoo sen tapahtuman todennäköisyyden, että satunnaismuuttujan arvo on x.

$$f(x) = P\{X = x\} = P(\{\omega \in \Omega : X(\omega) = x\}).$$

Kun $(\Omega, \mathcal{F}, P)$ on alkuperäinen tn-avaruus, niin $\{X = x\} \in \mathcal{F}$, ja yllä oleva P on alkuperäisen tn-avaruuden todennäköisyys. Ptnf siis antaa satunnaismuuttujan arvojen todennäköisyydet alkuperäisen perusjoukon tapahtumien todennäköisyyksinä.

Diskreetin satunnaismuuttujan, kuten aiemmin diskreetin tn-avaruuden (vrt. lause 1.26), tapauksessa tilanne yksinkertaistuu siten, että arvojoukon pisteiden todennäköisyydet määrittävät satunnaismuuttujan jakauman yksikäsitteisesti, ja tapahtumien todennäköisyydet saadaan yksittäisten arvojen todennäköisyyksien summana:

[2] Jälleen tarkemmin reaalilukujen Borel-joukoille.

Lause 2.8. *Olkoon X diskreetti satunnaismuuttuja ja f sen ptnf. Nyt f määrää satunnaismuuttujan X jakauman kaavalla*

$$P\{X \in B\} = \sum_{x \in B} f(x).$$

Todistus. Harjoitustehtävä. Kokonaistodennäköisyyden kaavasta käyttämällä perusjoukolle $\Omega = \{x_1, x_2, \dots\}$ ositusta

$$\{X \notin X(\Omega)\}, \{X = x_1\}, \{X = x_2\}, \dots$$

$\square$

Esimerkki 2.9. Määritellään edellisen esimerkin satunnaismuuttujien pistetodennäköisyysfunktiot.

(i) Yhden nopan tapaus. Oletetaan symmetrinen perusjoukko, eli että kaikkien perusjoukon alkoiden tn on

$$P\{\omega\} = \frac{1}{6}.$$

Nyt $\{X = x\} = \{\omega \in \Omega : \omega = x\} = \{x\}$, joten myös

$$f_X(x) = P\{X = x\} = P\{x\} = \frac{1}{6} \quad \text{kaikille} \quad x \in X(\Omega).$$

Lisäksi $f(x) = 0$ kaikille $x \notin X(\Omega) = \{1, 2, \dots, 6\}$; monesti tästä ei erikseen mainita.

(ii) Kahden nopan tapaus. Oletetaan jälleen symmetrinen perusjoukko, eli että kaikkien perusjoukon alkoiden tn on

$$P\{\omega\} = \frac{1}{36}.$$

Nyt pistetodennäköisyydet saadaan laskemalla satunnaismuuttujan arvoille suotuisat alkeistapaukset:

y	Suotuisat alkeistapaukset	$f_Y(y)$
2	$(1,1)$	$1/36$
3	$(1,2),(2,1)$	$2/36$
4	$(1,3),(2,2),(3,1)$	$3/36$
5	$(1,4),(2,3),(3,2),(4,1)$	$4/36$
6	$(1,5),(2,4),(3,3),(4,2),(5,1)$	$5/36$
7	$(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)$	$6/36$
8	$(2,6),(3,5),(4,4),(5,3),(6,2)$	$5/36$
9	$(3,6),(4,5),(5,4),(6,3)$	$4/36$
10	$(4,6),(5,5),(6,4)$	$3/36$
11	$(5,6),(6,5)$	$2/36$
12	$(6,6)$	$1/36$

(iii) Heitä sikaa -tapaus. Satunnaismuuttujan pistetodennäköisyydet saadaan jälleen laskemalla yhteen kunkin satunnaismuuttujan arvon tuottavien perusjoukon alkioiden todennäköisyydet:

z	Suotuisat alkeistapaukset	Asento	$f_Z(z)$
0	ω_1	Siankylki	0.651
5	ω_2, ω_3	Sianselkä, Tavallinen sika	0.313
10	ω_4	Siansyltty	0.030
15	ω_5	Etupotka	0.006

Edellisen esimerkin tapauksessa huomattiin, että kaikkien tarkasteltujen satunnaismuuttujien ptnf:t summautuivat ykköseksi. Tämä pätee myös yleisesti:

Lause 2.10. *Olkoon $X : \Omega \to \mathbb{R}$ diskreetti satunnaismuuttuja. Sen ptnf:lle f pätee:*

(i) $f(x) \geq 0$ kaikille $x \in \mathbb{R}$

(ii) jos $f(x > 0)$, $x \in X(\Omega) = \{x_1, x_2 \ldots\}$, eli x kuuluu X:n arvojoukkoon.

(iii) $\sum_{i=1}^{\infty} x_i = 1$.

Voidaan osoittaa, että nämä ehdot pätevät myös toiseen suuntaan, eli että jokainen ei-negatiivinen funktio $f : \mathbb{R} \to \mathbb{R}$, joka saa positiivisia arvoja vain äärellisessä tai numeroituvasti äärettömässä joukossa ja summautuu kaikkien indeksiensä yli ykkökseksi, on jonkin diskreetin satunnaismuuttujan pistetodennäköisyysfunktio.

Määritelmä 2.11. *Diskreetti kertymäfunktio.* Diskreetin satunnaismuuttujan X kertymäfunktio (lyh. kf) on funktio $F : \mathcal{R} \to \mathcal{R}$, jolle pätee

$$F(x) = P\{X \leq x\} \sum_{k \leq x} P\{X = k\} = \sum_{k \leq x} f(k),$$

missä $k \in \{0, 1, 2, \ldots, x\}$.

Diskreeteille jakaumille laskettu kertymäfunktio ei ole aivan yhtä hyödyllinen kuin jatkuville satunnaismuuttujille määritelty ystävänsä, sillä sen laskeminen symbolisesti on yleensä huomattavan työlästä summattavien termien lisääntyessä. Ennen tietokoneiden aikaa tämä saattoi olla iso ongelma, ja diskreettien satunnaismuuttujien kertymäfunktioiden arvoja approksimoitiin muilla jakaumilla. Tästä lisää viimeisessä luvussa.

2.1 Esimerkkejä diskreeteistä jakaumista

2.1.1 Diskreetti tasajakauma

Määritelmä 2.12. Satunnaismuuttujalla X on *diskreetti tasajakauma* joukossa $\{1, \ldots, N\}$, jos sen ptnf on

$$f(x) = \frac{1}{n} \quad \text{kaikille} \quad x \in \{1, \ldots, N\}.$$

Esimerkkien 2.3 ja 2.9 kohdan (i) satunnaismuuttuja X noudattaa diskreettiä tasajakaumaa joukossa $\{1, \ldots, 6\}$.

2.1.2 Binomijakauma

Määritelmä 2.13. Satunnaismuuttuja X noudattaa *binomijakaumaa* parametreilla $n \in \{1, 2, \ldots\}$ ja $p \in [0, 1]$, jos sen ptnf on

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{kaikille} \quad k \in \{0, 1, \ldots, n\}.$$

Tällöin merkitään $X \sim \mathrm{Bin}(n, p)$.

Binomijakautuneella satunnaismuuttujalla voidaan mallintaa *toistokoetta*, mitä käsittelimme kappaleessa 1.7.

Jos esimerkin 1.18 kohdassa (ii) nostettujen valkoisten pallojen määrää neljän pallon otoksessa kuvataan satunnaismuuttujalla X, niin X noudattaa binomijakaumaa parametreilla $n = 4$ ja $p = \frac{5}{8}$.

Samoin lauseen 1.19 kohdassa (ii) nostettujen valkoisten pallojen määrää n:stä nostetusta pallosta voidaan kuvata satunnaismuuttujalla X, joka noudattaa binomijakaumaa parametreilla n ja $p = \frac{K}{N}$, missä K on korissa olevien valkoisten pallojen, ja N kaikkien korissa olevien pallojen määrä. Helpolla laskulla nähdään, että lauseen kaava todellakin on binomijakautuneen satunnaismuuttujan pistetodennäköisyysfunktio.

2.1.3 Hypergeometrinen jakauma

Määritelmä 2.14. Satunnaismuuttuja X noudattaa *hypergeometrista jakaumaa* parametreilla $N \in \{1, 2, \dots\}$ ja $n, K \in \{1, \dots, N\}$, jos sen pistetodennäköisyysfunktio on

$$f(k) = P\{X = k\} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \quad \text{kaikille} \quad k \in \{0, 1, \dots, n\}.$$

Tällöin käytetään merkintää

$$X \sim \mathrm{Hyperg}(N, K, n).$$

Hypergeometrisella jakaumalla voidaan mallintaa otantaa ilman takaisinpanoa. Jos esimerkin 1.18 kohdassa (ii) nostettujen valkoisten pallojen määrää otoksessa kuvataan satunnaismuuttujalla X, se noudattaa hypergeometrista jakaumaa parametreilla $N = 8$, $K = 5$, ja $n = 4$. Samoin lauseen 1.19 kohdassa (ii) nostettujen valkoisten pallojen määrää voidaan kuvata hypergeometrisella jakaumalla.

Huomautus 2.15. Huomaa, että koska binomikertoimen $\binom{x}{y}$ arvo on määritelmän mukaan 0, jos $y > x$, hypergeometrista jakaumaa noudattavan satunnaismuuttujan arvojoukkoon kuuluvat vain sellaiset $k \in \{0, 1, \dots, n\}$, joille $k \leq K$ ja $n - k \leq N - K$. Korista ei voida nostaa enempää valkoisia tai mustia palloja kuin niitä siellä on, jos palloja ei palauteta nostojen välillä!

2.1.4 Geometrinen jakauma

Määritelmä 2.16. Satunnaismuuttuja X noudattaa *geometrista jakaumaa* parametrilla $p \in (0, 1)$, jos sen ptnf on

$$f(k) = P\{X = k\} = p(1-p)^k \quad \text{kaikille} \quad k \in \mathbb{N}.$$

Tällöin merkitään

$$X \sim \mathrm{Geom}(p).$$

Toistetaan koetta, jonka onnistumistodennäköisyys on p, kunnes saadaan ensimmäinen onnistuminen. Tällöin sm X, joka kuvaa "turhien" koekertojen määrää ennen ensimmäistä onnistumista, noudattaa geometrista jakaumaa parametrilla p. Geometrisen jakauman tiheysfunktion takaa löytyvää intuitiota voi hahmotella ajattelemalla riippumattomia toistoja satunnaiskokeessa, jossa epäonnistumisia sattuu k kappaletta, ja tämän jälkeen viimein onnistuminen todennäköisyydellä p.

Ensimmäiseen onnistumiseen voi periaatteessa tarvita kuinka monta toistokertaa tahansa, joten geometrista jakaumaa noudattavan satunnaismuuttujan arvojoukkoon kuuluvat kaikki luonnolliset luvut.

Esimerkki 2.17. Heitetään arpakuutiota, kunnes saadaan ensimmäinen kutonen. Merkitään satunnaismuuttujalla X heittojen määrää, joka tarvitaan ensimmäisen kutosen saamiseksi (heittoa, jolla saadaan ensimmäinen kutonen, ei lasketa näihin). Satunnaismuuttuja X noudattaa geometrista jakaumaa parametrilla $p = \frac{1}{6}$. Satunnaismuuttujan X ptnf on myös helppo päätellä suoraan:

- $X = 0$, jos heti ensimmäisellä heitolla saadaan kutonen, eli $P\{X = 0\} = \frac{1}{6} = p(1-p)^0$.
- $X = 1$, jos ensimmäisellä heitolla ei saada kutosta, mutta toisella saadaan, eli $P\{X = 1\} = \frac{5}{6} \cdot \frac{1}{6} = p(1-p)^1$
- $X = 2$, jos kahdella ensimmäisellä heitolla ei saada kutosta, mutta kolmannella saadaan, eli $P\{X = 2\} = \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} = p(1-p)^2$ ja niin edelleen.

Huomautus 2.18. Voidaan helposti tarkastaa, että myös geometrista jakaumaa noudattavan satunnaismuuttujan ptnf toteuttaa lauseen 2.10 ehdot. Erityisesti geometrisen sarjan (sarja suppenee, koska $1 - p < 1$) summan kaavaa soveltamalla nähdään, että pistetodennäköisyydet summautuvat ykköseksi:

$$\sum_{k=0}^{\infty} p(1-p)^k = p\frac{1}{1-(1-p)} = 1.$$

2.1.5 Poisson-jakauma

Määritelmä 2.19. Satunnaismuuttuja X noudattaa *Poisson-jakaumaa* parametrilla $\lambda > 0$, jos sen ptnf on

$$f(k) = P\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{kaikille} \quad k \in \mathbb{N}.$$

Tällöin merkitään

$$X \sim \text{Poisson}(\lambda).$$

Poisson-jakautuneella satunnaismuuttujalla voidaan kuvata joidenkin tapahtumien määrää aika- tai tilayksikköä kohti, jos tapahtumat oletetaan toisistaan riippumattomiksi ja niiden odotettu määrä aika- tai tilayksikköä kohti vakioksi. Poisson-jakautuneella satunnaismuuttujalla voidaan kuvata esimerkiksi:

- Joukkueen tekemien maalien määrää jalkapallo-ottelussa.
- Liikenneonnettomuuksien määrää tiekilometria kohti.
- Onkijan tunnissa saamien kalojen määrää.

Esimerkki 2.20. Kuvataan satunnaismuuttujalla X kotijoukkueen tekemien maalien määrä La Ligassa (Espanjan jalkapallon pääsarja) kausilla 2010-2015 (yhteensä 2116 ottelua). Oletetaan, että $X \sim \text{Poisson}(1.631)$[3]

Oheiseen taulukkoon on kirjattu kotijoukkeen maalimäärien havaitut määrät ja osuudet peleissä (esimerkiksi 466:ssä pelissä kotijoukkue on tehnyt 0 maalia, joten osuus on $466/2116 \approx 0.2202$) ja Poisson-jakauman pistetodennäköisyysfunktion (parametrilla $\lambda = 1.631$) arvot. Huomataan, että havaitut frekvenssit ovat hyvin lähellä Poisson-jakaumasta laskettuja, joten oletus siitä, että maalimäärät noudattavat Poisson-jakaumaa, vaikuttaa järkevältä.[4]

Maalit	Ottelut	Osuus	$f_X(x)$
0	466	0.2202	0.1957
1	668	0.3157	0.3192
2	498	0.2353	0.2604
3	262	0.1238	0.1416
4	147	0.0695	0.0577
5	50	0.0236	0.0188
6	15	0.0071	0.0051
7	7	0.0033	0.0012
8	2	0.0009	0.0002
9	1	0.0005	0.0000
≥ 10	0	0.0000	0.0000

[3]Käytetty parametrin arvo $\lambda = 1.631$ on aineistosta laskettu maalien keskiarvo ottelua kohti. Jakauman parametrien arvojen päättelyyn aineiston perusteella, eli ns. tilastolliseen päättelyyn tutustutaan tarkemmin kurssilla Tilastollinen päättely I.

[4]Tosin suuria maalimääriä on hieman enemmän kuin oletuksen Poisson-jakautuneisuudesta perusteella pitäisi olla. Tämä selittynee sillä, että sarjan joukkueet ovat eritasoisia, joten oletus siitä, että kotijoukkueen maalien syntymistodennäköisyys olisi vakio ottelua kohti, ei täyty. Maalimäärien tarkempaan mallintamiseen tulisikin ottaa huomioon ainakin pelaavien joukkueiden tasoero. Tällöin kysymyksessä olisi Poisson-regressio; tähän palataan kurssilla Yleistetyt lineaariset mallit.

Tarkka todennäköisyys sille, että kotijoukkue tekee vähintään kymmenen maalia ottelussa, saadaan laskettua komplementtitapahtuman kautta:

$$P\{X \geq 10\} = 1 - P\{X \leq 9\} = 1 - \sum_{i=0}^{9} P\{X = i\} \approx 8.42 \cdot 10^{-6}.$$

Tämä todennäköisyys pyöristyy nollaan taulukossa, jossa käytetään neljän desimaalin tarkkuutta.

Vertaa esimerkkiin 1.32, jossa pistetodennäköisyydet asetettiin suoraan alkeistapauksille. Todennäköisyydet ovat samat, mutta satunnaismuuttujan käyttäminen yksinkertaistaa merkintöjä ja laskuja. Myöhemmillä kursseilla osoittautuu, että monille (tarkemmin exponenttiperheen) jakaumille hyvä parametrin estimaatti tulee havaintojen otoskeskiarvosta.

2.2 Jatkuvat satunnaismuuttujat

Jos satunnaismuuttujan arvojoukko on ylinumeroituva, esimerkiksi koko reaalilukujen joukko $\mathbb{R}$ tai väli $[0,1]$, pistetodennäköisyyksiä ei voida enää käyttää satunnaismuuttujan jakauman määrittämiseen (itse asiassa osoittautuu, että jatkuvan jakauman tapauksessa kaikkien yksittäisten pisteiden todennäköisyydet ovat nollia).

Pistetodennäköisyysfunktiota vastaava käsite jatkuvasti jakautuneelle satunnaismuuttujalle on tiheysfunktio. Summaamisen sijaan satunnaismuuttujan arvojoukon osajoukkojen todennäköisyydet saadaan integroimalla tiheysfunktiota. Määritellään ensin täsmällisesti jatkuvan jakauman ja tiheysfunktion käsitteet.

Määritelmä 2.21. Satunnaismuuttujalla X on *jatkuva jakauma tiheysfunktiolla* (lyh. tf) f, jos

$$P\{a \leq X \leq b\} = \int_a^b f(x)\,\mathrm{d}x$$

kaikille $a, b \in \mathbb{R}, a < b$.

Jatkuvan jakauman tapauksessa siis kaikkien suljettujen välien todennäköisyydet $P\{X \in [a,b]\}$ saadaan integroimalla tiheysfunktiota tämän välin yli. Kuten pistetodennäköisyysfunktio, myös tiheysfunktio on kaikkialla epänegatiivinen, ja sen integaali koko reaaliakselin yli on yksi. Samoin myös jokainen nämä ehdot täyttävä funktio on jonkin jatkuvan satunnaismuuttujan tiheysfunktio:

Lause 2.22. *Funktio $f : \mathbb{R} \to \mathbb{R}$ on jonkin jatkuvan satunnaismuuttujan tiheysfunktio, jos ja vain jos*

(i) $f(x) \geq 0$ *kaikille* $x \in \mathbb{R}$,

(ii) f on integroituva koko reaaliakselilla ja

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.$$

Todistus. Harjoitustehtävä. □

Jatkuvia jakaumia käsitellään monesti myös kertymäfunktion avulla. Kertymäfunktio on määritelty jokaiselle reaaliluvulle $x \in \mathbb{R}$, ja nimensä mukaisesti kertoo kuinka suuri osa satunnaismuuttujan "todennäköisyysmassasta" on arvon x alapuolella:

Määritelmä 2.23. Satunnaismuuttujan X *kertymäfunktio* (lyh. kf) on funktio $F : \mathbb{R} \to \mathbb{R}$, jolle

$$F(x) = P\{X \leq x\} \quad \text{kaikille} \quad x \in \mathbb{R}.$$

Huomautus 2.24. Koska diskreetille satunnaismuuttujalle X joukkojen todennäköisyydet saadaan niiden alkioiden pistetodennäköisyyksien summana, myös sen kertymäfunktio saadaan ptnf:n f summana:

$$F(X) = P\{X \leq x\} = \sum_{k \leq x} P\{X = k\} = \sum_{k \leq x} f(k).$$

Koska diskreetin sm:n arvojoukko on numeroituva, myös kyseinen summa on numeroituva.

Samalla tavoin jatkuvalle jakaumalle kertymäfunktio saadaan integroimalla tiheysfunktiota (tai vastaavasti tiheysfunktio saadaan derivoimalla kertymäfunktiota):

Lause 2.25. *Jatkuvasti jakautuneelle satunnaismuuttujalle X (tiheysfunktiolla f) :*

(i) Kertymäfunktio F saadaan integroimalla tiheysfunktiota:

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \quad \textit{kaikille} \quad x \in \mathbb{R}.$$

(ii) Toisaalta tiheysfunktiolle f pätee

$$f(x) = F'(x)$$

kaikissa f:n jatkuvuuspisteissä.

Seuraava lause osoittaa, että jatkuvalle jakaumalle kaikki pistetodennäköisyydet tosiaan ovat nollia. Tämän takia ei ole merkitystä, otetaanko välien päätepisteet mukaan tarkasteltaessa välien todennäköisyyksiä; ts. ei ole merkitystä, ovatko tarkasteltavat välit avoimia, puoliavoimia, vai suljettuja.

Lause 2.26. *Jatkuvalle satunnaismuuttujalle X (tiheysfunktiolla f) pätee:*

(i) $P\{X = x\} = 0$ *kaikille* $x \in \mathbb{R}$,

(ii) $P\{a < X < b\} = P\{a < X \le b\} = P\{a \le X < b\} = P\{a \le X \le b\}$ *kaikille* $a < b$.

Todistus. Ensimmäinen kohta seuraa jatkuvan satunnaismuuttujan määritelmästä, kun asetetaan $a = x$, ja annetaan b:n lähestyä x:ää oikealta. Toinen kohta seuraa todennäköisyyden additiivisuudesta ja ensimmäisestä kohdasta. □

Huomautus 2.27. Kaikille $a < b$ tapahtuma, että jatkuva satunnaismuuttuja X on pienempi tai yhtä suuri kuin b, voidaan esittää yhdisteenä erillisistä tapahtumista:

$$\{X \le b\} = \{X < a\} \cup \{a \le X \le b\}.$$

Siten todennäköisyyden additiivisuuden ja lauseen 2.26 nojalla

$$\begin{aligned} P\{a \le X \le b\} &= P\{X \le b\} - P\{X < a\} \\ &= P\{X \le b\} - P\{X \le a\} \\ &= F(b) - F(a). \end{aligned}$$

Välien todennäköisyydet voidaan siis ilmaista erotuksena kertymäfunktion arvoista välin päätepisteissä.

2.3 Esimerkkejä jatkuvista jakaumista

2.3.1 Jatkuva tasajakauma

Määritelmä 2.28. Satunnaismuuttuja X noudattaa välin (a, b) *tasajakaumaa*, jos sen tf on

$$\frac{1}{b-a} \quad \text{kaikille} \quad x \in (a, b),$$

ja $f(x) = 0$, kun $x \le a$ tai $x \ge b$. Tällöin merkitään $X \sim \mathrm{Tas}(a, b)$.

Kertymäfunktio saadaan integroimalla tiheysfunktiota. Kun $x \in (a, b)$:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{a}^{x} \frac{1}{b-a}\,\mathrm{d}t = \frac{x-a}{b-a}.$$

Siten tasajakautuneen satunnaismuuttujan kertymäfunktio on

$$F(x) = \begin{cases} 0, & \text{kun} \quad x \le a \\ \frac{x-a}{b-a}, & \text{kun} \quad a < x < b \\ 1, & \text{kun} \quad x \ge b. \end{cases}$$

Esimerkki 2.29. Generoidaan satunnaisluku väliltä $(0, 1000)$. Millä todennäköisyydellä luku on väliltä $[500, 600]$?

Kyseessä on satunnaisluku, jolloin kaikki välin luvut ovat periaatteessa yhtä todennäköisiä. Tilannetta voidaan siis kuvata satunnaismuuttujalla $X \sim \mathrm{Tas}(0, 1000)$. Kysytty todennäköisyys saadaan erotuksena tasajakauman kertymäfunktion arvoista välin päätepisteissä:

$$P\{500 \leq X \leq 600\} = F(600) - F(500) = \frac{600}{1000} - \frac{500}{1000} = \frac{1}{10}.$$

2.3.2 Eksponenttijakauma

Määritelmä 2.30. Satunnaismuuttuja X noudattaa *eksponenttijakaumaa* parametrilla $\lambda > 0$, jos sillä on jatkuva jakauma tiheysfunktiolla

$$f(x) = \lambda e^{-\lambda x} \quad \text{kaikille} \quad x > 0.$$

Tällöin merkitään $X \sim \mathrm{Exp}(\lambda)$.

Eksponenttijakauman kertymäfunktio saadaan integroimalla tiheysfunktiota:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{0}^{x} \lambda e^{-\lambda t}\,\mathrm{d}t = - \Big/_{0}^{x} e^{-\lambda t} = 1 - e^{-\lambda x},$$

kun $x > 0$ ja 0 muuten.

Jos tapahtumien esiintymistodennäköisyys aikaväliä tai tilayksikköä kohti on vakio, ja ne esiintyvät toisistaan riippumatta, näiden tapahtumien esiintymisten välistä aikaa tai etäisyyttä kuvaava satunnaismuuttuja noudattaa eksponenttijakaumaa. Jos tapahtumien määrä aikayksikköä kohti noudattaa Poisson-jakaumaa, niin niiden välinen aika noudattaa eksponenttijakaumaa. Tällaista prosessia, jonka tuottamien satunnaisten tapahtumien määrä aikayksikköä kohti noudattaa Poisson-jakaumaa, ja niiden väliaika eksponenttijakaumaa, kutsutaan Poisson-prosessiksi.

Eksponenttijakautuneella satunnaismuuttujalla voidaan siis kuvata odotusaikaa tai välimatkaa toisistaan riippumattomien tapahtumien, joiden sattumistodennäköisyys on vakio, välillä, kuten:

- Odotusaikaa seuraavaan maaliin jalkapallo-ottelussa.
- Liikenneonnettomuuksien tapahtumispaikkojen välejä tieosuudella.
- Onkijan odotusaikaa seuraavaan kalaan.

Esimerkki 2.31. Oletetaan, että hehkulampun kestoaika tunneissa noudattaa eksponenttijakaumaa parametrilla $\lambda = 0.001$. Millä todennäköisyydellä hehkulamppu palaa yli 2000 tuntia?

Merkitään lampun palamisaikaa satunnaismuuttujalla X, jolloin $X \sim \mathrm{Exp}(0.001)$. Kysytyn tapahtuman tn saadaan laskettua kertymäfunktion avulla komplementtitapahtuman todennäköisyyden kautta:

$$P\{X > 2000\} = 1 - P\{X \leq 2000\} = 1 - F(2000) = e^{-2} \approx 0.135.$$

Toinen mahdollisuus laskea todennäköisyys on integroida suoraan tiheysfunktiota:

$$P\{X > 2000\} = \int_{2000}^{\infty} \lambda e^{-\lambda t}\,\mathrm{d}t = - \Big/_{2000}^{\infty} e^{-\lambda t} = e^{-2000\lambda} = e^{-2}.$$

2.3.3 Normaalijakauma

Määritelmä 2.32. Jatkuva satunnaismuuttuja X noudattaa *normaalijakaumaa* parametreilla $\mu \in \mathbb{R}$ ja $\sigma^2 > 0$, jos sen tiheysfunktio on

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Tällöin merkitään $X \sim N(\mu, \sigma^2)$.

Normaalijakauman parametrit kertovat jakauman tiheysfunktion sijainnin ja leveyden. Näitä parametreja sanotaan usein *odotusarvoparametriksi* ja *varianssiparametriksi*.

Jos satunnaismuuttuja X noudattaa jakaumaa $N(0, 1)$, eli normaalijakaumaa parametreilla $\mu = 0$ ja $\sigma^2 = 1$, sanotaan että X noudattaa *standardinormaalijakaumaa*. Tällöin X:n tiheysfunktio yksinkertaistuu muotoon:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Huomautus 2.33. Normaalijakautunut satunnaismuuttuja $X \sim N(\mu, \sigma^2)$ voidaan aina muuntaa standardinormaalijakaumaa noudattavaksi satunnaismuuttujaksi vähentämällä siitä odotusarvoparametri μ ja jakamalla se keskihajonnalla σ. Satunnaismuuttujalle

$$Z = \frac{X - \mu}{\sigma}$$

siis pätee $Z \sim N(0, 1)$. Tätä muunnosta kutsutaan standardoinniksi.

Normaalijakauman keskeinen rooli todennäköisyyslaskennassa johtuu siitä, että riippumattomien satunnaismuuttujien summan (ja siten myös keskiarvon) jakauma lähestyy normaalijakaumaa otoskoon kasvaessa. Tähän keskeisenä raja-arvolauseena tunnettuun tulokseen palataan kurssin loppupuolella.

2.4 Satunnaismuuttujan odotusarvo

2.4.1 Diskreetin satunnaismuuttujan odotusarvo

Määritelmä 2.34. Olkoon X diskreetti satunnaismuuttuja, jonka ptnf on f, ja arvojoukko $X(\Omega) = \{x_1, x_2, \dots\}$. Satunnaismuuttujan X *odotusarvo* on

$$E(X) = \sum_{i=1}^{\infty} x_i f(x_i),$$

jos kyseinen sarja suppenee itseisesti, eli jos

$$\sum_{i=1}^{\infty} |x_i| f(x_i) < \infty.$$

Mikäli sarja ei suppene itseisesti, sanotaan että X:llä ei ole odotusarvoa. Jos satunnaismuuttujan $X(\Omega)$ arvojoukko on äärellinen, niin myös kyseinen summa on aina äärellinen, ja siten odotusarvo on aina olemassa.

Määritelmä 2.35. Oletetaan, että satunnaismuuttuja X kuvaa yhden nopan heiton silmälukua, jolloin sen ptnf on

$$f(k) = P\{X = k\} = \frac{1}{6} \quad \text{kaikille} \quad k \in \{1, 2, \dots, 6\}.$$

Tällöin X:n odotusarvoksi saadaan

$$E(X) = \sum_{k=1}^{6} x_k f(x_k) = \sum_{k=1}^{6} k \cdot \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = 3.5.$$

Satunnaismuuttujan odotusarvoa voi ajatella sen jakauman painopisteenä: Jos massattoman sauvan pisteisiin $\{x_1, x_2, \dots\}$ asetetaan pistetodennäköisyysfunktion arvoja näissä pisteissä vastaavat painot $\{f(x_1), f(x_2), \dots\}$, niin sauva pysyy tasapainossa, jos sitä tuetaan odotusarvon kohdalta.

Toinen tapa havainnollistaa odotusarvoa on uhkapelin keskimääräisenä voittosummana kierrosta kohti: jos peliä jatketaan monta kierrosta, keskimääräinen voitto kierrosta kohti lähestyy kierroksen voittosummaa kuvaavan satunnaismuuttujan odotusarvoa.

Esimerkki 2.36. Tarkastellaan peliä, jossa pelaaja heittää kahta noppaa, ja voittaa silmälukujen summan verran rahaa. Kannattaako tästä pelistä maksaa seitsemän euroa kierrokselta?

Merkitään silmälukujen summaa, joka kuvaa myös kierroksella voitettua rahamäärää, satunnaismuuttujalla X. Arvojoukko on $\{2, 3, \dots, 12\}$, ja ptnf laskettiin esimerkissä 2.9. Lasketaan X:n odotusarvo:

$$\begin{aligned} E(X) &= \sum_{i=2}^{12} x_i f(x_i) = \frac{2\cdot 1 + 3\cdot 2 + 4\cdot 3 + 5\cdot 4 + 6\cdot 5 + 7\cdot 6 + 8\cdot 5 + 9\cdot 4 + 10\cdot 3 + 11\cdot 2 + 12\cdot 1}{36} \\ &= \frac{252}{36} = 7. \end{aligned}$$

Voiton odotusarvo on tasan 7, joten seitsemän euron panoksella peli on *reilu*: keskimäärin kumpikaan pelaajasta ja pelin tarjoajasta ei jää voitolle.

Esimerkki 2.37. Poisson-jakaumaa parametrilla λ noudattavan satunnaismuuttujan X odotusarvo saadaan laskettua eksponenttifunktion sarjakehitelmän

$$e^x = \sum_{i=0}^{\infty} \frac{x^n}{n!}$$

avulla:

$$E(X) = \sum_{k=1}^{\infty} kf(k) = 0 + \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{(k-1)}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Esimerkki 2.38. Pietarin paradoksi. Tarkastellaan seuraavaa peliä: heitetään kolikkoa, kunnes saadaan ensimmäinen kruuna. Pelaaja voittaa 2^k dukaattia, missä k on heittojen määrä. Kuinka paljon tästä pelistä kannattaa maksaa?

Merkitään pelistä saatavaa voittoa sm:llä X, jonka arvojoukko on $\{2, 4, 8, \dots\}$. Koska tn saada ensimmäinen kruuna k:nnella heitolla on $\frac{1}{2^k}$, X:n pistetodennäköisyysfunktio on

$$f(x) = P\{X = x\} = \frac{1}{x} \quad \text{kaikille} \quad x \in \{2, 4, 8 \dots\}.$$

Voiton odotusarvoksi saadaan:

$$E(X) = \sum_{k=1}^{\infty} x_k f(x_k) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \sum_{i=1}^{\infty} 1 = \infty.$$

Siten pelistä kannattaa periaatteessa maksaa mielivaltaisen paljon!

Tämä perustuu siihen, että pelaajan panos on rajoitettu, mutta kasinon (= pelin tarjoajan) maksama voittosumma on periaatteessa rajoittamaton. ”Paradoksi” ratkeaa, kun oletetaan, että kasinon maksamalla voittosummalla on jokin yläraja.

2.4.2 Jatkuvan satunnaismuuttujan odotusarvo

Määritelmä 2.39. Olkoon X jatkuva satunnaismuuttuja tiheysfunktiolla f. Sen *odotusarvo* on

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x,$$

mikäli kyseinen integraali suppenee itseisesti, eli

$$\int_{-\infty}^{\infty} |x| f(x)\,\mathrm{d}x < \infty.$$

Jos integraali ei suppene itseisesti, sanotaan että X:llä ei ole odotusarvoa.

Esimerkki 2.40. Olkoon X tasajakaumaa $\mathrm{Tas}(a, b)$ noudattava satunnaismuuttuja. Tällöin X:n odotusarvo on

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_{a}^{b} x \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{2(b-a)} \Big/_{a}^{b} x^2 = \frac{1}{2}(a+b).$$

Välin (a, b) tasajakaumaa noudattavan sm:n odotusarvo sijaitsee siis aina tämän välin puolessavälissä. Tämä sopii hyvin yhteen oletusarvon massakeskipistetulkinnan kanssa: tasajakaumassa todennäköisyysmassa on jakautunut tasaisesti koko välille, joten painopiste on välin puolessa välissä.

Esimerkki 2.41. Satunnaismuuttuja X, noudattattaa (standardi-) *Cauchy-jakaumaa*, jos sen tiheysfunktio

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2} \quad \text{kaikille} \quad x \in \mathbb{R}.$$

Tällöin merkitään $X \sim \mathrm{Cauchy}(0, 1)$. Cauchy-jakaumaa käytetään yleisesti esimerkkinä jakaumasta, jolla ei ole äärellistä odotusarvoa eikä varianssia. Odotusarvolle tämä voidaan todeta integroimalla. Jakauman symmetrisyyden nojalla:

$$\int_{-\infty}^{\infty} |x| f(x)\,\mathrm{d}x = \frac{1}{\pi} \int_{0}^{\infty} \frac{2x}{1+x^2}\,\mathrm{d}x = \Big/_{0}^{\infty} \ln(1+x^2) = \infty,$$

Esimerkki 2.42. Oletetaan, että hehkulampun kestoaikaa tunteina kuvaava satunnaismuuttuja X noudattaa eksponenttijakaumaa parametrilla $\lambda = 0.001$, eli $X \sim \mathrm{Exp}(0.001)$. Mikä on lampun kestoajan odotusarvo?

Lasketaan ensin eksponenttijakaumaa noudattavan sm:n odotusarvo yleisessä tapauksessa osittaisintegroinnilla:

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_{0}^{\infty} x\lambda e^{-\lambda x}\,\mathrm{d}x \\
&= \Big/_{\!\!0}^{\infty} x(-e^{-\lambda x}) - \int_{0}^{\infty} 1\cdot(-e^{-\lambda x})\,\mathrm{d}x \\
&= 0 - \frac{1}{\lambda} \Big/_{\!\!0}^{\infty} e^{-\lambda x} = \frac{1}{\lambda}.
\end{aligned}
$$

Siten hehkulampun kestoiän odotusarvo on $E(X) = \frac{1}{\lambda} = 1000$ tuntia.

Huomautus 2.43. Satunnaismuuttujan jakauman ei välttämättä tarvitse olla diskreetti tai jatkuva, vaan satunnaismuuttuja voi noudattaa myös ns. *sekatyypin* jakaumaa. Esimerkki tällaisesta satunnaismuuttujasta on hehkulampun kestoikä, jos oletetaan että kymmenesosa lampuista on viallisia, eli niiden kestoikä on nolla, ja loppujen lamppujen kestoikä noudattaa eksponenttijakaumaa parametrilla 0.001. Tämän jakauman kertymäfunktio on

$$
F(x) = \begin{cases} 0, & \text{kun} \quad x < 0 \\ \frac{1}{10}, & \text{kun} \quad x = 0 \\ \frac{1}{10} + \frac{9}{10}(1 - e^{-0.001x}), & \text{kun} \quad x > 0. \end{cases}
$$

Yleisessä tapauksessa satunnaismuuttujan odotusarvo määritellään Riemannin integraalia yleisempänä Lebesguen integraalina. Odotusarvolla on seuraavat ominaisuudet, joita voi kokeilla itse osoittaa diskreetin tai jatkuvan satunnaismuuttujan tapauksessa.

Lause 2.44. *Satunnaismuuttujille X ja Y, joiden odotusarvot ovat olemassa, pätee:*

(i) (Positiivisuus) Jos $X \geq 0$, niin $E(X) \geq 0$.

(ii) (Lineaarisuus) $E(aX + bY) = aE(X) + bE(Y)$ kaikille $a, b \in \mathbb{R}$.

(iii) Jos $X \leq Y$, niin $E(X) \leq E(Y)$.

(iv) Jos $X \geq 0$ ja $E(X) = 0$, niin sm X on vakio arvolla 0, eli $P\{X = 0\} = 1$.

(v) Jos X:llä ja Y:llä on sama jakauma, niin $E(X) = E(Y)$.

2.5 Satunnaismuuttujien riippumattomuus

Intuitiivisesti satunnaismuuttujien riippumattomuus tarkoittaa, että niiden arvot eivät vaikuta toisiinsa: jos tiedämme toisen satunnaismuuttujan arvon, se ei vaikuta arvioomme toisen satunnaismuuttujan arvosta. Satunnaismuuttujien riippumattomuus määritellään tapahtumien riippumattomuuden kautta:

Määritelmä 2.45. Satunnaismuuttujat X ja Y ovat *riippumattomia*, jos

$$P(\{X \in A\} \cap \{Y \in B\}) = P\{X \in A\}P\{Y \in B\},$$

eli jos

$$\{X \in A\} \perp\!\!\!\perp \{Y \in B\}$$

kaikille (riittävän säännöllisille) joukoille $A, B \subseteq \mathbb{R}$. Tällöin merkitään

$$X \perp\!\!\!\perp Y.$$

Huomautus 2.46. Jos $X \perp\!\!\!\perp Y$, ja $P\{Y \in B\} > 0$, niin joukkojen ehdolliset todennäköisyydet palautuvat ehdottomiin todennäköisyyksiin (vrt. lause 1.39):

$$\begin{aligned}
P(\{X \in A\}|\{Y \in B\}) &= \frac{P(\{X \in A\} \cap P\{Y \in B\})}{P\{Y \in B\}} \\
&= \frac{P\{X \in A\}P\{Y \in B\}}{P\{Y \in B\}} \\
&= P\{X \in A\}.
\end{aligned}$$

Jos satunnaismuuttujat ovat keskenään riippumattomia, niin myös niiden muunnokset ovat keskenään riippumattomia:

Lause 2.47. *Jos $X \perp\!\!\!\perp Y$, niin kaikille (riittävän säännöllisille) funktioille $g, h : \mathbb{R} \to \mathbb{R}$:*

$$g(X) \perp\!\!\!\perp h(Y).$$

Riippumattomuus yleistyy useammalle satunnaismuuttujalle seuraavasti:

Määritelmä 2.48. Joukko satunnaismuuttujia $X_1, \ldots, X_n$ on riippumaton, jos

$$P\left(\bigcap_{i=1}^{n} \{X \in B_i\}\right) = \prod_{i=1}^{n} P\{X_i \in B_i\}$$

kaikille (riittävän säännöllisille) joukoille $B_1, \ldots, B_n$. Tällöin merkitään

$$X_1, \ldots, X_n \perp\!\!\!\perp .$$

2.6 Varianssi ja keskihajonta

Varianssi ja keskihajonta kuvaavat satunnaismuuttujan jakauman *hajontaa*, eli sitä kuinka paljon satunnaismuuttujan arvot keskimäärin poikkeavat sen odotusarvosta. Varianssi on neliöidyn poikkeman odotusarvosta odotusarvo, ja keskihajonta taas varianssin neliöjuuri.

Määritelmä 2.49. Olkoon X satunnaismuuttuja, jonka odotusarvo $E(X)$ on olemassa ja äärellinen. X:n *varianssi* on

$$\mathrm{Var}(X) = E\left[(X - E(X))^2\right],$$

mikäli kyseinen odotusarvo on olemassa. Jos $E(X)$ on olemassa, mutta $E\left[(X - E(X))^2\right] = \infty$, sanotaan, että X:n varianssi on ääretön. Varianssille voidaan käyttää myös merkintöjä $D^2(X)$ tai σ^2 (tai σ_X^2, jos käsitellään useampia satunnaismuuttujia ja halutaan täsmentää, että kyseessä on nimenomaan X:n varianssi).

Määritelmä 2.50. Satunnaismuuttujan X varianssin neliöjuurta

$$\sqrt{\mathrm{Var}(X)}$$

kutsutaan X:n *keskihajonnaksi*. Keskihajonnalle voidaan käyttää myös merkintöjä $D(X)$, σ tai σ_X.

Huomautus 2.51. Satunnaismuuttujan neliön odotusarvoa $E(X^2)$ kutsutaan sen toiseksi momentiksi. Jatkuvalle satunnaismuuttujalle X, jonka tf on f se voidaan laskea seuraavana integraalina:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x,$$

ja diskreetille satunnaismuuttujalle vastaavana summana. Varianssin kaava voidaan esittää toisen momentin avulla, mikä voi helpottaa laskuja:

Lause 2.52. *Jos satunnaismuuttujan toinen momentti $E(X^2)$ on olemassa, myös sen varianssi on äärellinen, ja saadaan kaavasta*

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2.$$

Todistus. Odotusarvon olemassaolo seuraa aina toisen momentin olemassaolosta, sillä

$$|x| \leq 1 + x^2 \quad \text{kaikille} \quad x \in \mathbb{R}.$$

Nyt väite seuraa odotusarvon lineaarisuudesta (Lauseen 2.44 toinen kohta):

$$\mathrm{Var}(X) = E\left[(X - E(X))^2\right] = E\left[X^2 - 2XE(X) + E(X)^2\right] = E(X^2) - [E(X)]^2.$$

$\square$

Esimerkki 2.53. Olkoon X välin (a, b) jatkuvaa tasajakaumaa noudattava satunnaismuuttuja, eli $X \sim \text{Tas}(a, b)$, jolloin X:n odotusarvo on $E(X) = \frac{1}{2}(a + b)$. Lasketaan ensin X:n toinen momentti:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_a^b \frac{x^2}{b-a}\,\mathrm{d}x = \frac{1}{3(b-a)} \Big/_{a}^{b} x^3 = \frac{b^3 - a^3}{3(b-a)}.$$

Varianssi saadaan nyt laskettua lauseen 2.52 nojalla:

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{b^3 - a^3}{3(b-a)} - \left[\frac{1}{2}(a+b)\right]^2 = \frac{1}{12}(b-a)^2.$$

Laskemalla varianssi suoraan määritelmästä oltaisiin päädytty hyvin samankaltaisiin laskuihin.

Satunnaismuuttujan lineaarisen muunnoksen varianssi saadaan seuraavalla kaavalla.

Lause 2.54. *Jos satunnaismuuttujan X varianssi on olemassa ja $a, b \in \mathbb{R}$, niin*

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Todistus.

Odotusarvon lineaarisuuden nojalla $E(aX + b) = aE(X) + b$, joten

$$\text{Var}(aX + b) = E([aX + b - aE(X) - b]^2) = a^2 E([X - E(X)]^2) = a^2 \text{Var}(X).$$

□

Vakion lisääminen satunnaismuuttujaan ei siis muuta jakauman varianssia, mutta vakiolla kertomisen vaikutus on neliöllinen. Tämä on intuitiivista siinä mielessä, että varianssi mittaa nimenomaan keskimääräistä *neliöityä* poikkeamaa odotusarvosta.

2.7 Kovarianssi ja korrelaatio

Satunnaismuuttujien välistä lineaarista riippuvuutta kuvataan niiden yhteisvaihtelulla, eli kovarianssilla.

Määritelmä 2.55. Olkoot X ja Y satunnaismuuttujia, joilla on odotusarvot $E(X)$ ja $E(Y)$. Niiden välinen *kovarianssi* on

$$\text{Cov}(X, Y) = E([X - E(X)][Y - E(Y)]),$$

mikäli kyseinen odotusarvo on olemassa. Osoittautuu, että myös kovarianssi olemassaololle riittävä ehto on, että X:n ja Y:n toiset momentit ovat äärellisiä.

Suoraan määritelmästä nähdään, että varianssi on erityistapaus kovarianssista: se on satunnaismuuttujan kovarianssi itsensä kanssa:

$$\mathrm{Var}(X) = \mathrm{Cov}(X, X).$$

Kuten varianssi, myös kovarianssi voidaan esittää muodossa, joka monesti yksinkertaistaa laskuja:

Lause 2.56. *Olkoot X ja Y satunnaismuuttujia, joiden toiset momentit ovat äärellisiä. Tällöin*

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Todistus. Harjoitustehtävä. Kannattaa rohkeasti aloittaa kertomalla odotusarvon sisällä oleva tulo auki. □

Yleensä kovarianssin sijaan satunnaismuuttujien välisen lineaarisen riippuvuuden kuvaamiseen käytetään normalisoitua kovarianssia, eli korrelaatiota, joka saadaan jakamalla tarkasteltavien muuttujien kovarianssi niiden keskihajonnoilla.

Määritelmä 2.57. Olkoot X ja Y satunnaismuuttujia, joiden odotusarvot ja toiset momentit ovat äärellisiä. Niiden välinen *korrelaatio* on

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Korrelaatio sijaitsee aina välillä $[-1, 1]$. Korrelaation arvo 1 kuvaa täydellistä lineaarista riippuvuutta, ja -1 täydellistä negatiivista lineaarista riippuvuutta. Korrelaation arvo 0 taas tarkoittaa, ettei tarkasteltavien muuttujien välillä ole *lineaarista* riippuvuutta.

Osoittautuu, että riippumattomien satunnaismuuttujien välinen kovarianssi, ja siten myös korrelaatio, todellakin on 0. Tätä varten tarvitaan seuraavaa aputulosta, jonka mukaan riippumattomien satunnaismuuttujien odotusarvo saadaan niiden odotusarvojen tulona.

Lause 2.58. *Jos satunnaismuuttujien X ja Y odotusarvot ovat olemassa, ja $X \perp\!\!\!\perp Y$, niin*

$$E(XY) = E(X)E(Y).$$

Lause 2.59. *Jos X ja Y ovat riippumattomia, eli $X \perp\!\!\!\perp Y$, niin*

$$\mathrm{Cov}(X, Y) = 0,$$

ja siten myös

$$\mathrm{Corr}(X, Y) = 0.$$

Todistus.

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0.$$

□

Huomautus 2.60. Jos $\mathrm{Cov}(X, Y) = 0$, sanotaan, että X ja Y ovat *korreloimattomia*. Edellisen lauseen nojalla riippumattomuudesta seuraa aina korreloimattomuus. Mutta korreloimattomuudesta ei kuitenkaan välttämättä seuraa riippumattomus. Tämä johtuu siitä, että kovarianssi ja korrelaatio mittaavat pelkästään muuttujien välistä *lineaarista* riippuvuutta. Jos riippuvuus on epälineaarista, kovarianssi ei välttämättä tavoita sitä.

Standardinormaalijakauman tiheysfunktio on parillinen, joten $E(X) = E(X^3) = 0$. Tästä seuraa, että standardinormaalijakautuneelle muuttujalle $X \sim N(0, 1)$ ja sen neliölle

$$\mathrm{Cov}(X, X^2) = E(X \cdot X^2) - E(X)E(X^2) = 0 - 0 = 0,$$

mutta tietenkään X ja X^2 eivät ole riippumattomia.

Satunnaismuuttujien summan varianssi saadaan laskettua niiden kovarianssin avulla. Satunnaismuuttujien erotukselle ja lineaarikombinaatioille voidaan laskea kaava vastaavalla tavalla. Erityisesti, jos satunnaismuuttujat ovat riippumattomia, niiden summan varianssi saadaan niiden varianssien summana.

Lause 2.61. *Jos satunnaismuuttujien X ja Y varianssit ovat äärellisiä, niin*

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y).$$

Erityisesti, jos $X \perp\!\!\!\perp Y$, niin

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Todistus. Todistus saadaan laskemalla neliöt auki ja soveltamalla odotusarvon lineaarisuutta:

$$\begin{aligned}\mathrm{Var}(X + Y) &= E[(X + Y)^2] - E(X + Y)^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y).\end{aligned}$$

Väitteen toinen osa seuraa lauseesta 2.59, jonka mukaan riippumattomille satunnaismuuttujille $\mathrm{Cov}(X, Y) = 0$. □

Riippumattomien satunnaismuuttujien varianssin summan kaava yleistyy myös useammalle satunnaismuuttujalle.

Lause 2.62. *Jos satunnaismuuttujat* $X_1, \dots, X_n$ *ovat riippumattomia, eli* $X_1, \dots, X_n \perp\!\!\!\perp$*, ja niiden varianssit ovat äärellisiä, niin*

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

2.8 Epäyhtälöitä

Seuraavaksi osoitamme kaksi todennäköisyyslaskennan keskeisintä epäyhtälöä: Markovin ja Tšebyševin epäyhtälöt[5]. Näiden avulla voidaan laskea ylärajoja satunnaismuuttujien ”häntätodennäköisyyksille”, eli todennäköisyyksille, että satunnaismuuttujien arvot poikkeavat satunnaismuuttujan odotusarvosta vähintään jonkin vakion verran, vaikkei tiedettäisi satunnaismuuttujan pistetodennäköisyys- tai tiheysfunktiota.

Markovin epäyhtälössä oletetaan pelkästään, että epänegatiivisen satunnaismuuttujan odotusarvo on tunnettu: tämän perusteella todennäköisyys, että X:n arvo on vähintään a, on pienempää kuin X:n odotusarvo jaettuna a:lla.

Lause 2.63. *Markovin epäyhtälö. Jos* $X \geq 0$ *ja* X*:n odotusarvo on olemassa, niin*

$$P\{X \geq a\} \leq \frac{E(X)}{a} \quad \textit{kaikille} \quad a > 0.$$

Todistus. Määritellään satunnaismuuttuja

$$Y = \mathbf{1}_{[a,\infty)}(X) = \begin{cases} 1, & \text{kun} \quad X \geq a \\ 0, & \text{muuten.} \end{cases}$$

Koska aina $X \geq Y$, lauseen 2.44 kohdan (iii) ja odotusarvon lineaarisuuden nojalla

$$E(X) \geq E(Y) = aE(\mathbf{1}_{[a,\infty)}(X)) = a(0 \cdot P\{X < a\} + 1 \cdot P\{X \geq a\}) = aP\{X \geq a\}.$$

Nyt väite seuraa jakamalla epäyhtälön kumpikin puoli a:lla. □

Jos $\mu = E(X) > 0$, asettamalla $a = k\mu$ Markovin epäyhtälö voidaan myös kirjoittaa muodossa

$$P\{X \geq k\mu\} \leq \frac{1}{k} \quad \text{kaikille} \quad k > 0.$$

[5]Nimitykset vaihtelevat ulkomaisessa kirjallisuudessa: monesti myös Markovin epäyhtälöä kutsutaan Tšebyševin epäyhtälöksi.

Markovin epäyhtälön mukaan siis todennäköisyys, että X:n arvo on vähintään k kertaa sen odotusarvo, on pienempää kuin $1/k$. Tämä muoto näyttää myös selkeämmin yhteyden Tšebyševin epäyhtälöön. Näitä kutsutaankin joskus ensimmäisen (Markov) ja toisen momentin (Tšebyšev) epäyhtälöiksi.

Tšebyševin epäyhtälöä voidaan soveltaa kaikille satunnaismuuttujille, joiden odotusarvo ja varianssi tunnetaan. Tšebyševin epäyhtälön mukaan todennäköisyys, että satunnaismuuttujan arvot ovat poikkeavat vähintään k:n keskihajonnan verran satunnaismuuttujan keskiarvosta, on pienempää kuin $1/k^2$.

Lause 2.64. *Tšebyševin epäyhtälö. Jos X on satunnaismuuttuja, jonka odotusarvo $\mu = E(X)$ ja varianssi $\sigma^2 = \mathrm{Var}(X) > 0$ ovat olemassa, niin*

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2} \quad \textit{kaikille} \quad k > 0.$$

Todistus. Sovelletaan Markovin epäyhtälöä satunnaismuuttujalle $|X-\mu|^2$ ja vakiolle $a = k^2\sigma^2$, sekä varianssin määritelmää:

$$P\{|X-\mu| \geq k\sigma\} = P\{|X-\mu|^2 \geq k^2\sigma^2\} \leq \frac{E(|X-\mu|^2)}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}.$$

□

Esimerkki 2.65. Tarkastellaan jälleen (katso esimerkit 2.31 ja 2.42) hehkulamppuja, joiden kestoaika X noudattaa eksponenttijakaumaa, ja jonka odotusarvo on 1000 tuntia, eli $X \sim \mathrm{Exp}(0.001)$.

Koska $X \geq 0$, Markovin epäyhtälöstä saadaan ylärajat todennäköisyydelle, että lamppu palaa vähintään 3000 tuntia:

$$P\{X \geq 3000\} = P\{X \geq 3\mu\} \leq \frac{1}{3}.$$

Eksponenttijakauman varianssiksi saadaan osittaisintegroinnilla $1/\lambda^2$; siten X:n keskihajonta on $\sigma = 1/\lambda = 1000$. Kun tunnetaan varianssi, Tšebyševin epäyhtälöstä saadaan tiukempi yläraja:

$$P\{X \geq 3000\} \leq P\{|X-\mu| \geq 2\sigma\} \leq \frac{1}{4}.$$

Tässä tapauksessa tietenkin tiedämme X:n kertymäfunktion F, joten voimme laskea kyseiselle todennäköisyydelle tarkan arvon:

$$P\{X \geq 3000\} = 1 - F(3000) = e^{-0.001 \cdot 3000} = e^{-3} \approx 0.05.$$

Seuraavaan taulukkoon on vielä laskettu Markovin ja Tšebyševin epäyhtälöstä lasketut ylärajat, sekä tarkat arvot yllä tarkastellun X:n häntätodennäköisyyksille.

x	Markov	Tšebyšev	$P\{X \geq x\}$
3000	0.33	0.25	0.050
4000	0.25	0.11	0.018
5000	0.20	0.0625	0.007
10000	0.10	0.012	$4.54 \cdot 10^{-5}$

Huomataan toisaalta, että epäyhtälöiden avulla saadut ylärajat ovat melko karkeita, ja toisaalta, että suurille poikkeamille Tšebyševin epäyhtälöstä saadaan huomattavasti tiukempi yläraja. Tämä selittyy sillä, että Tšebyševin epäyhtälössä käytössä on enemmän informaatiota: odotusarvon lisäksi myös varianssi.

Markovin ja Tšebyševin epäyhtälöitä tarvitaan monien todennäköisyyslaskennan keskeisten tulosten todistuksissa. Tällä kurssilla näistä käsitellään suurten lukujen laki ja keskeinen raja-arvolause.

Luku 3

Raja-arvolauseita

Tässä luvussa esiteltävät suurten lukujen laki ja keskeinen raja-arvolause käsittelevät riippumattoman ja samoin jakautuneen[1] satunnaismuuttujajonon $X_1, \ldots, X_n$ keskiarvon (tai vastaavasti summan) käyttäytymistä, kun satunnaiskoetta toistetaan rajatta. Osoittautuu, että keskiarvojen jono lähestyy satunnaismuuttujien yhteistä odotusarvoa, ja niiden jakauma lähestyy normaalijakaumaa, jonka odotusarvo on myös satunnaismuuttujien yhteinen odotusarvo.

Nämä raja-arvotulokset ovat keskeisiä tilastollisen päättelyn teoriassa. Estimaattoreille ei monesti pystytä laskemaan tarkkoja jakaumia, vaan sovelletaan keskeistä raja-arvolausetta, jonka mukaan suurella otoskoolla riippumattomien ja samoin jakautuneiden satunnaismuuttujien summat ja keskiarvot ovat likimain normaalisti jakautuneita: tämän perusteella estimaattorin jakaumaa voidaan approksimoida normaalijakaumalla.

3.1 Suurten lukujen laki

Aiemmin käytetty odotusarvon tulkinta uhkapelin keskimääräisenä voittona kierrosta kohti, kun peliä pelataan monta kierrosta, perustuu yhteen todennäköisyyslaskennan keskeisistä tuloksista, nimittäin suurten lukujen lakiin. Sen mukaan samoin riippumattomien satunnaismuuttujien $X_1, \ldots X_n$, joilla on sama odotusarvo, keskiarvo lähenee niiden odotusarvoa, kun n kasvaa rajatta, eli satunnaiskoetta toistetaan n kertaa, ja $n \to \infty$. Matemaattisesti ilmaistuna keskiarvojen jonon raja-arvo (toistojen lukumäärän funktiona) on satunnaismuuttujien X_i yhteinen odotusarvo μ.

Suurten lukujen laki on myös monien tieteellisten koeasetelmien taustalla: jos otos-

[1] Suurten lukujen laista ja keskeisestä raja-arvolauseesta on olemassa useita eri versioita erilaisilla oletuksilla. Esittelemme tässä klassiset versiot näistä lauseista; vastaavat tulokset voi kuitenkin osoittaa myös näitä löyhemmillä oletuksilla.

koko on riittävän suuri, niin sen nojalla voidaan olettaa, että keskimääräinen onnistumisten osuus kokeessa on lähellä todellista onnistumistodennäköisyyttä. Esimerkiksi jos testataan, parantaako uusi lääke sairauden, voidaan olettaa että parantuneiden potilaiden osuus otoksesta on lähellä todellista parantumistodennäköisyyttä, jos otoskoko on suuri.

Määritellään ensin, mitä tarkoitamme suppenemisella eli konvergenssilla satunnaismuuttujajonon, joka siis on jono funkioita perusjoukolta reaalilukujen joukolle, tapauksessa. Satunnaismuuttujien konvergenssi voidaan määritellä usealla eri tavalla; heikon suurten lukujen lain yhteydessä käytämme ns. stokastista konvergenssia.

Määritelmä 3.1. Stokastinen konvergenssi. Satunnaismuuttujajono $X_1, X_2, \ldots$ *suppenee stokastisesti* kohti satunnaismuuttujaa X, jos

$$\lim_{n\to\infty} P\{|X_n - X| \geq \epsilon\} = 0 \quad \text{kaikille} \quad \epsilon > 0.$$

Tällöin merkitään

$$X_n \xrightarrow{p} X.$$

Huomaa, että vaikka stokastisen konvergenssin yleisessä muotoilussa X voi olla satunnaismuuttuja, seuraavassa tuloksessa keskiarvojen jono suppenee kohti vakiota $\mu = E(X_i)$.

Nyt voimme muotoilla heikon suurten lukujen lain (weak law of large numbers: WLLN) ja todistaa sen Tšebyševin epäyhtälön avulla.

Lause 3.2. *Heikko suurten lukujen laki (SLL). Olkoon $X_1, X_2, \ldots$ jono riippumattomia satunnaismuuttujia, joilla on sama odotusarvo $\mu = E(X_i)$ ja varianssi $\sigma^2 = \mathrm{Var}(X_i) \leq \infty$. Nyt satunnaismuuttujien keskiarvojen*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

jono $\bar{X}_1, \bar{X}_2, \ldots$ suppenee stokastisesti kohti satunnaismuuttujien yhteistä odotusarvoa:

$$\bar{X}_n \xrightarrow{p} \mu,$$

eli

$$\lim_{n\to\infty} P\{|\bar{X}_n - \mu| \geq \epsilon\} = 0 \quad \textit{kaikille} \quad \epsilon > 0.$$

Todistus. Satunnaismuuttujien riippumattomuuden nojalla niiden keskiarvon $\bar{X}_n$ varianssi on

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

ja siten keskihajonta on

$$\sqrt{\mathrm{Var}(\bar{X}_n)} = \frac{\sigma}{\sqrt{n}}$$

kaikille $n \in \{1, 2, \dots\}$.

Nyt kaikille $\epsilon > 0$ saadaan Tšebyševin epäyhtälöstä käyttäen arvoa $k = \sqrt{n}\epsilon/\sigma$:

$$P\{|\bar{X}_n - \mu| \geq \epsilon\} = P\{|\bar{X}_n - \mu| \geq k\frac{\sigma}{\sqrt{n}}\} \leq \frac{1}{k^2} = \frac{\sigma^2}{n\epsilon^2} \to 0, \text{ kun } n \to \infty.$$

□

Huomautus 3.3. Jos satunnaismuuttujat $X_1, X_2, \dots$ ovat riippumattomia ja samoin jakautuneita, tulos pätee stokastista suppenemista voimakkaammalle konvergenssin muodolle, *melkein varmalle suppenemiselle*,[2] myös ilman oletusta varianssin olemassaolosta. Tätä *vahvaksi suurten lukujen laiksi* kutsuttua versiota ei käsitellä tällä kurssilla.

3.2 Keskeinen raja-arvolause

Keskeisen raja-arvolauseen mukaan samoin jakautuneen ja riippumattoman satunnaismuuttujajonon keskiarvon (tai vastaavasti summan) jakauma suppenee kohti tiettyä normaalijakaumaa. Määrittelemme ensin täsmällisesti kertymäfunktioiden avulla, mitä tarkoittaa satunnaismuuttujajonon suppeneminen jakaumamielessä.

Määritelmä 3.4. Jakaumasuppeneminen. Olkoon $X_1, X_2, \dots$ jono satunnaismuuttujia, joiden kertymäfunktiot ovat $F_1, F_2, \dots,$ ja X satunnaismuuttuja, jonka kertymäfunktio on G. Jono (X_n) *suppenee jakaumaltaan* kohti X:ää, mikäli niiden kertymäfunktioiden jono suppenee pisteittäin kohti X:än kertymäfunktiota, eli

$$\lim_{n\to\infty} F_n(x) = G(x)$$

kaikissa G:n jatkuvuuspisteissä x. Tällöin merkitään

$$X_n \xrightarrow{d} X.$$

[2]Melkein varma suppeneminen tarkoittaa, että jono $X_1, X_2, \dots$ konvergoi kohti satunnaismuuttujaa X todennäköisyydellä yksi, eli kaikkialla muualla paitsi mahdollisesti perusjoukon nollamittaisessa osajoukossa:

$$P\{\lim_{n\to\infty} X_n = X\} = 1.$$

Tällöin merkitään (a.s. = almost surely):

$$X_n \xrightarrow{a.s.} X.$$

Seuraavaksi esitämme klassisen version keskeisestä raja-arvolauseesta (central limit theorem : CLT). Todistuksessa tarvitaan satunnaismuuttujan karakteristisia funktioita, joita ei olla käsitelty tällä kurssilla; todistus löytyy Tuomisen kirjasta (s. 118).

Lause 3.5. *Keskeinen raja-arvolause. Olkoon $X_1, X_2, \ldots$ jono riippumattomia samoin jakautuneita satunnaismuuttujia, joiden yhteinen odotusarvo on $\mu = E(X_i)$ ja varianssi $\sigma^2 = \mathrm{Var}(X_i) < \infty$.*

Tällöin standardoitujen keskiarvojen jono suppenee jakaumaltaan kohti standardinormaalijakaumaa:

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0, 1),$$

tai yhtäpitävästi

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

3.2.1 Normaaliapproksimaatio

Keskeistä raja-arvolausetta voidaan käyttää otoskeskiarvon jakauman approksimoimiseen, sillä jos n on riittävän suuri, riippumattomien ja samoin jakautuneiden satunnaismuuttujien $X_1, \ldots, X_n$ standardoitu ($\bar{X}_n$:stä vähennetään sen odotusarvo μ, ja tulos jaetaan $\bar{X}_n$:n keskihajonnalla $\frac{\sigma}{\sqrt{n}}$) keskiarvo

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma$$

noudattaa likimain standardinormaalijakaumaa, mitä voidaan merkitä

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \overset{d}{\approx} N(0, 1).$$

Tällöin normaalijakauman ominaisuuksien (katso huomautus 2.33) nojalla keskiarvon $\bar{X}_n$ jakaumaa voidaan approksimoida normaalijakaumalla $N(\mu, \frac{\sigma^2}{n})$, eli

$$\bar{X}_n \overset{d}{\approx} N(\mu, \frac{\sigma^2}{n}),$$

missä $\mu = E(X_i)$ on satunnaismuuttujien yhteinen odotusarvo, ja $\sigma^2 = \mathrm{Var}(X_i)$ niiden yhteinen varianssi.

Keskeisen raja-arvolauseen mukaan myös riippumattomien ja samoin jakautuneiden satunnaismuuttujien $X_1, \ldots, X_n$ summan $\sum_{i=1}^{n} X_i$ jakaumaa voidaan approksimoida seuraavalla normaalijakaumalla:

$$\sum_{i=1}^{n} X_i \overset{d}{\approx} N(n\mu, n\sigma^2).$$

Approksimaation tarkkuus riippuu n:n lisäksi myös satunnaismuuttujien $X_1, \ldots, X_n$ jakaumasta. Jos se muistuttaa normaalijakaumaa, eli on yksihuippuinen ja suhteellisen symmetrinen, approksimaatio voi olla suhteellisen tarkka jo muutamalla kymmenellä havainnolla. Jos taas jakauma on hyvin vino, kuten esimerkiksi eksponenttijakauma, tarkkaan approksimaatioon vaaditaan huomattavasti suurempaa otoskokoa. Approksimaation tarkkuus riippuu myös siitä, mitä kohtaa jakaumasta tarkastellaan: normaaliapproksimaatio on usein tarkempi jakauman keskellä, ja epätarkempi sen hännissä.

Monien diskreettien jakaumien, kuten binomi- ja Poisson-jakaumien, kertymäfunktiot voidaan ilmaista pelkästään summina. Ennen tietokoneaikaa nämä olivat hankalia laskea, joten monesti turvauduttiin normaaliapproksimaatioon, jossa pystyttiin hyödyntämään standardinormaalijakauman kertymäfunktion ($\Phi(z) = F_Z(z) = P\{Z < z\}$ standardinormaalijakaumaa noudattavalle satunnaismuuttujalle $Z \sim N(0, 1)$) taulukoituja arvoja. Diskreettejä jakaumia approksimoitaessa ns. jatkuvuuskorjaus monesti parantaa approksimaation tarkkuutta.

Huomautus 3.6. Jatkuvuuskorjaus. Kun kokonaislukuarvoisen satunnaismuuttujan, eli satunnaismuuttujan jonka arvojoukko kuuluu $\mathbb{Z}$:aan, X jakaumaa approksimoidaan normaalijakaumalla (tai jollain muulla jatkuvalla jakaumalla), voidaan tarkastella jokaiselle satunnaismuuttujan arvolle $k \in \mathbb{Z}$ todennäköisyyttä

$$P\{k - \frac{1}{2} \leq X \leq k + \frac{1}{2}\}$$

k:n pistetodennäköisyyden

$$P\{X = k\}$$

sijaan, sillä kokonaislukuarvoiselle satunnaismuuttujalle

$$\{X = k\} = \{k - \frac{1}{2} \leq X \leq k + \frac{1}{2}\}.$$

Esimerkiksi jos X on yhden nopan heiton silmäluku, niin

$$\{X = 2\} = \{1.5 \leq X \leq 2.5\},$$

sillä nopan silmäluku voi saada arvoja vain joukosta $\{1, 2, \ldots, 6\}$.

Tämä on myös erittäin järkevää, sillä normaalijakautuneelle (kuten kaikille muillekin jatkuvaa jakaumaa noudattaville) satunnaismuuttujalle Y kaikki pistetodennäköisyydet ovat nollia:

$$P\{Y = k\} = 0.$$

Koska kaikkien joukkojen todennäköisyydet saadaan diskreeteille satunnaismuuttujille joukon pisteiden pistetodennäköisyyksien summina, niin jatkuvuuskorjausta käytettäessä väleistä tulee $\frac{1}{2}$:n verran pidempiä kummastakin päästä:

$$P\{k_1 \le X \le k_2\} = P\{k_1 - \frac{1}{2} \le X \le k_2 + \frac{1}{2}\} \quad \text{kaikille} \quad k_1, k_2 \in \mathbb{Z}, k_1 \le k_2.$$

Esimerkiksi jos X on yhden nopan heiton silmäluku, niin

$$P\{2 \le X \le 4\} = P\{2.5 \le X \le 4.5\}.$$

Samoin vain toiselta puolelta rajatuille väleille

$$P\{X \le k\} = P\{X \le k + \frac{1}{2}\} \quad \text{kaikille} \quad k \in \mathbb{Z},$$

ja

$$P\{X \ge k\} = P\{X \ge k - \frac{1}{2}\} \quad \text{kaikille} \quad k \in \mathbb{Z}.$$

Pienellä otoskoolla jatkuvuuskorjaus voi parantaa normaaliapproksimaation tarkkuutta huomattavasti.

Määritelmä 3.7. Satunnaismuuttuja X noudattaa *Bernoullin jakaumaa* parametrilla $p \in (0, 1)$, jos sen ptnf on

$$f(k) = P\{X = k\} = p^k(1-p)^{1-k} \quad \text{kaikille} \quad k \in \{0, 1\}.$$

Tällöin merkitään $X \sim \text{Bernoulli}(p)$.

Bernoullin jakaumaa noudattavan satunnaismuuttujan odotusarvo on $E(X) = p$, ja varianssi on $\text{Var}(X) = p(1-p)$.

Bernoullin jakaumalla voidaan kuvata satunnaiskoetta, jossa on kaksi vaihtoehtoa, onnistuminen ($X = 1$) ja epäonnistuminen ($X = 0$), ja onnistumistodennäköisyys on p. Esimerkiksi kolikonheitossa kruunaa voidaan merkitä ykkösellä ja klaavaa nollalla, jolloin tulos noudattaa Bernoullin jakaumaa parametrilla $\frac{1}{2}$, tai nopanheitossa kutonen voidaan määritellä onnistumiseksi ja muu tulos epäonnistumiseksi, jolloin tulos noudattaa Bernoullin jakaumaa parametrilla $\frac{1}{6}$.

Bernoullin jakaumalla voidaan siis mallintaa toistokokeen yksittäisiä toistoja; n:n riippumattoman samaa Bernoullin jakaumaa parametrilla p noudattavan satunnaismuuttujan summa noudattaakin binomijakaumaa $\text{Bin}(n, p)$.

Esimerkki 3.8. Heitetään noppaa sata kertaa. Mikä on todennäköisyys, että saadaan vähintään 15, mutta enintään 20 kutosta?

Kyseessä on toistokoe, jossa satunnaismuuttujat $X_1, \dots, X_{100}$, voidaan määritellä seuraavasti

$$X_i = \begin{cases} 1, & \text{jos } i\text{:nnellä heitolla saadaan kutonen} \\ 0, & \text{muuten,} \end{cases}$$

ovat riippumattomia ja noudattavat samaa Bernoullin jakaumaa: $X_i \sim \text{Bernoulli}(\frac{1}{6})$ kaikille $i \in \{1, 2, \dots, n\}$. Tällöin niiden summalle $X = \sum_{i=1}^{100} X_i$ saadaan tarkka jakauma: se noudattaa binomijakaumaa otoskoolla $n = 100$, ja onnistumistodennäköisyydellä $p = \frac{1}{6}$.

Lasketaan kuitenkin likiarvo kysytylle todennäköisyydelle käyttäen normaaliapproksimaatiota. Odotusarvon lineaarisuuden nojalla summan odotusarvoksi saadaan

$$E(X) = E\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} E(X_i) = np$$

ja varianssiksi saadaan riippumattomuuden nojalla

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} \text{Var}(X_i) = np(1-p).$$

Keskeisen raja-arvolauseen nojalla standardoitu keskiarvo

$$Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}} = \frac{X - np}{\sqrt{np(1-p)}}$$

noudattaa likimain standardinormaalijakaumaa[3] $N(0,1)$ suurella otoskoolla n, joten (jatkuvuuskorjausta käyttäen) kysytyksi todennäköisyydeksi saadaan:

$$\begin{aligned} P\{15 \le X \le 20\} &= P\{14.5 \le X \le 20.5\} \\ &= P\left\{\frac{14.5 - np}{\sqrt{np(1-p)}} \le Z \le \frac{20.5 - np}{\sqrt{np(1-p)}}\right\} \\ &= \Phi\left(\frac{20.5 - 100 \cdot \frac{1}{6}}{\sqrt{100 \cdot \frac{1}{6}(1 - \frac{1}{6})}}\right) - \Phi\left(\frac{14.5 - 100 \cdot \frac{1}{6}}{\sqrt{100 \cdot \frac{1}{6}(1 - \frac{1}{6})}}\right) \\ &\approx \Phi(1.03) - \Phi(-0.58) \\ &\approx 0.848 - 0.281 \\ &= 0.567. \end{aligned}$$

[3] Standardinormaalijakauman kertymäfunktion Φ arvot voi lukea esimerkiksi Tuomisen kirjan toiseksi viimeisen sivun taulukosta, tai laskea tilasto-ohjelmisto `R`:llä esimerkiksi komennolla `pnorm(1.03)`.

Tarkaksi arvoksi saadaan binomijakauman kertymäfunktiosta F[4]

$$P\{15 \leq X \leq 20\} = P\{X \leq 20\} - P\{X \leq 14\} = F(20) - F(14) \approx 0.561,$$

eli normaaliapproksimaatio on tässä tapauksessa erittäin tarkka.

Kiitos tarkkaavaisuudesta ja läsnäolosta kurssilla! Jos huomaatte luentomateriaaleissa virheitä tai haluatte antaa luennoista/laskareista/kurssista palautetta, niin laittakaa huomiot näistä osoitteeseen topias.tolonen@helsinki.fi.

[4] R:ssä komennolla `pbinom(20,100,1/6) - pbinom(14,100,1/6)`.

Kirjallisuutta

[1] Pekka Tuominen: Todennäköisyyslaskenta I. 10. muuttumaton painos, Limes ry, 2010.

[2] Petri Koistinen: Todennäköisyyslaskenta. Luentomoniste, Helsingin yliopisto, 2013.

Liite A

Jakaumia

Diskreettejä jakaumia

Jakauma	Arvojoukko	Parametrit	Ptnf $f_X(k)$	$E(X)$	$\mathrm{Var}(X)$
Bernoulli(p)	$\{0,1\}$	$p \in (0,1)$	$p^k(1-p)^{1-k}$	p	$p(1-p)$
Bin(n,p)	$\{0,1,\ldots,n\}$	$n \in \mathbb{N}^+, p \in (0,1)$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Geom(p)	$\{0,1,2,\ldots\}$	$p \in (0,1)$	$p(1-p)^k$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	$\{0,1,2,\ldots\}$	$\lambda > 0$	$e^{-\lambda} \cdot \frac{\lambda^k}{k!}$	λ	λ
Hyperg(N,K,n)	$\{0,1,\ldots,n\}$	$N, K, n \in \mathbb{N}^+$, $n \leq N,\ K \leq N$	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$n \cdot \frac{K}{N}$	$n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$

Jatkuvia jakaumia

Jakauma	Arvojoukko	Parametrit	Tf $f_X(x)$	$E(X)$	$\mathrm{Var}(X)$
Tas(a,b)	(a,b)	$a,b \in \mathbb{R},\ a < b$	$\frac{1}{b-a}$	$\frac{1}{2}(a+b)$	$\frac{(b-a)^2}{12}$
Exp(λ)	$(0,\infty)$	$\lambda > 0$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$N(\mu,\sigma^2)$	$\mathbb{R}$	$\mu \in \mathbb{R}, \sigma^2 > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2