

Todennäköisyyslaskenta

Petri Koistinen
Matematiikan ja tilastotieteen laitos
Helsingin yliopisto

28. elokuuta 2012

Esipuhe

Tämä opetusmoniste on tarkoitettu Helsingin yliopiston matematiikan ja tilastotieteen laitoksella pidettävälle 10 opintopisteen laajuiselle aineopintotasoiselle todennäköisyyslaskennan kurssille. Kurssilla käsitellään sellaisia puolia todennäköisyyslaskennasta, joita jokaisen tilastotieteilijän (tai muun todennäköisyyslaskennan soveltajan) tulisi tuntea. Stokastisiin prosesseihin liittyvät käsitteet ovat kuitenkin rajattu kurssin ulkopuolelle. Lukijalla oletetaan ennestään olevan jonkin verran esitietoja aiheesta, jotka todennäköisesti ovat peräisin Tuomisen teokseen *Todennäköisyyslaskenta I* [10] perustuvalta kurssilta. Tämän takia kurssin alkuvaiheessa tiettyjä perusasioita vain kerrataan nopeasti. Kokemus on osoittanut, että joitakin monisteen kohtia joudutaan jättämään väliin ajan puutteen takia.

Lähestymistapa pyrkii olemaan käytännönläheinen: tavoitteena on opetella satunnaismuuttujiin ja todennäköisyysjakaumiin liittyvää laskentoa pikemmin kuin käsitellä todennäköisyyslaskentaa matemaattisena strukturina. Eräs tämän kurssin tärkeimpiä tavoitteita on oppia käsittelemään kaksi- ja useampiulotteisia jakaumia. Tällaiset tiedot ja taidot ovat välttämättömiä esim. sellaiselle opiskelijalle, joka tahtoo ymmärtää tilastotieteen menetelmien perusteita ja lukea alan kirjallisuutta.

Matemaattisesti aukoton esitys vaatisi mittateoriaa ja Lebesguen integraalia, mutta näitä työkaluja ei tällä kurssilla käytetä. Valtaosan todennäköisyyslaskennan sovelluksista pystyy ymmärtämään ilman mitta- ja integrointiteoriaa (mutta tiettyjä sovelluksia ei). Tietyissä kohdissa luentomonisteessa kuitenkin esitetään yhteyksiä mittateoreettisen todennäköisyyslaskennan käsitteisiin. Nämä huomautukset on tarkoitettu niille opiskelijoille, jotka ovat kiinnostuneet matemaattisesta täsmällisyydestä; muut opiskelijat voivat sivuuttaa ne kaikessa rauhassa.

Kaikki tässä monisteessa esitettävät asiat löytyvät myös monesta teoreettisen tilastotieteen ja todennäköisyyslaskennan oppikirjasta, jollaisia ovat esimerkiksi teokset [5, 1, 12, 8, 2].

Todennäköisyyslaskennan matemaattisesti täsmällistä mittateoreettista muotoilua voi opiskella paitsi todennäköisyysteorian kurssilla myös lukuisista teoksista, kuten esimerkiksi [6, 4, 11, 7, 3]. Esimerkiksi Schervish [9] näyttää, kuinka tilastotieteen teoria voidaan esittää matemaattisesti täsmällisesti todennäköisyysteoriaan pohjautuen.

Kirjallisuutta

- [1] S. F. Arnold. *Mathematical Statistics*. Prentice-Hall, Inc., 1990.
- [2] Robert B. Ash. *Basic Probability Theory*. Dover, 2008.
- [3] Krishna B. Athreya and Soumendra N. Lahiri. *Measure Theory and Probability Theory*. Springer Texts in Statistics. Springer, 2006.
- [4] P. Billingsley. *Probability and Measure*. John Wiley & Sons, Inc., 2nd edition, 1986.
- [5] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, 2nd edition, 2002.

- [6] G. Elfving and P. Tuominen. *Todennäköisyyslaskenta II*. Limes ry, Helsinki, 1985.
- [7] Jean Jacod and Philip Protter. *Probability Essentials*. Springer, 2nd edition, 2002.
- [8] Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, 7th edition, 2000.
- [9] Mark J. Schervish. *Theory of Statistics*. Springer series in statistics. Springer-Verlag, 1995.
- [10] P. Tuominen. *Todennäköisyyslaskenta I*. Limes ry, Helsinki, 1993.
- [11] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.
- [12] David Williams. *Weighing the Odds: A Course in Probability and Statistics*. Cambridge University Press, 2001.

Sisältö

1	Tapahtumat ja niiden todennäköisyydet	1
1.1	Johdanto	1
1.2	Joukko-oppia	3
1.3	Matemaattinen todennäköisyyden käsite	5
1.4	Kombinatoriikkaa	7
1.5	Multinomikertoimet	10
1.6	Ehdollinen todennäköisyys	11
1.7	Tapahtumien riippumattomuus	12
1.8	Kokonaistodennäköisyys ja Bayesin kaava	14
1.9	Monotoninen jatkuvuus	15
2	Satunnaismuuttujat	17
2.1	Satunnaismuuttuja ja sen jakauma	17
2.2	Kertymäfunktio	19
2.3	Diskreetti jakauma	22
2.4	Esimerkkejä diskreeteistä jakaumista	23
2.5	Jatkuva jakauma	24
2.6	Esimerkkejä jatkuvista jakaumista	28
2.7	Satunnaismuuttujan muunnos	29
2.8	Kvantiilifunktio ja sen käyttö simuloinnissa	31
2.9	Satunnaismuuttujien riippumattomuus	33
2.10	Muunnetun satunnaismuuttujan tiheys	36
3	Odotusarvo	41
3.1	Diskreetin satunnaismuuttujan odotusarvo	41
3.2	Jatkuvasti jakautuneen satunnaismuuttujan odotusarvo	45
3.3	Odotusarvon ominaisuuksia	46
3.4	Muunnoksen odotusarvo	47
3.5	Momentit	48
3.6	Varianssi, keskihajonta ja kovarianssi	49
3.7	Momenttiemäfunktio ja kumulanttiemäfunktio	52
4	Tärkeitä jakaumia	55
4.1	Diskreettejä jakaumia	55
4.1.1	Binomijakauma	55
4.1.2	Hypergeometrinen jakauma	56
4.1.3	Geometrinen jakauma	56
4.1.4	Negatiivinen binomijakauma	57

4.1.5	Poissonin jakauma	59
4.2	Gamma- ja beetafunktio	60
4.3	Jatkuvia jakaumia	62
4.3.1	Siirto ja skaalaus	62
4.3.2	Tasajakauma	62
4.3.3	Eksponenttijakauma	63
4.3.4	Gammajakauma	63
4.3.5	Beetajakauma	64
4.3.6	Normaalijakauma	65
4.3.7	Normaalijakaumasta johdettuja jakaumia	66
5	Epäyhtälöitä	67
5.1	Markovin ja Tšebyševin epäyhtälöt	67
5.2	Konveksit funktiot ja Jensenin epäyhtälö	69
5.3	Hölderin epäyhtälö	71
5.4	Cauchyn-Schwarzin epäyhtälö, kovarianssi ja korrelaatio	72
5.5	Momenttiemäfunktiota koskevia epäyhtälöitä	73
6	Kaksiulotteiset jakaumat	74
6.1	Kaksiulotteinen satunnaisvektori ja sen jakauma	74
6.2	Diskreetti kaksiulotteinen jakauma	75
6.3	Trinomijakauma	78
6.4	Jatkuva kaksiulotteinen jakauma	79
6.5	Tasajakauma alueessa	83
6.6	Riippumattomuus	85
6.7	Satunnaisvektorin muunnoksen odotusarvo	86
6.8	Kovarianssi ja muita yhteismomenteja	87
6.9	Paras lineaarinen ennuste	89
6.10	Odotusarvovektori ja kovarianssimatriisi	90
6.11	Tiheysfunktion muuntokaava	92
6.12	t -jakauman ominaisuuksia	96
7	Ehdollinen jakauma	98
7.1	Ehdolliset jakaumat	98
7.2	Kertolaskusääntö eli ketjusääntö	100
7.3	Diskreetin ja jatkuvan muuttujan yhteisjakauma	102
7.4	Ehdollinen odotusarvo	103
7.5	Yhteisjakauman määrittely kertolaskukaavan avulla	106
8	Moniulotteiset jakaumat	110
8.1	Satunnaisvektori	110
8.2	Odotusarvovektori ja kovarianssimatriisi	112
8.3	Ehdolliset jakaumat, kertolaskusääntö ja ehdollinen odotusarvo	115
8.4	Ehdollinen riippumattomuus	117
8.5	Tilastollisia malleja	118
8.6	Multinomijakauma	121
8.7	Tiheysfunktion muuntokaava	122
8.8	Satunnaisvektorin momenttiemäfunktio	124

9	Moniulotteinen normaalijakauma	127
9.1	Standardinormaalijakauma $N_n(\mathbf{0}, \mathbf{I})$	127
9.2	Yleinen multinormaalijakauma	128
9.3	Affinin muunnoksen jakauma	129
9.4	Tiheysfunktio	130
9.5	Tiheysfunktion tasa-arvopinnat	131
9.6	Korreloimattomuus ja riippumattomuus	133
9.7	Ehdolliset jakaumat	134
9.8	Kaksiulotteinen normaalijakauma	135
9.9	Normaalijakauman otoskeskiarvon ja otosvarianssin yhteisjakauma	136
10	Raja-arvolauseita ja approksimaatioita	140
10.1	Suurten lukujen laki	140
10.2	Jakaumasuppeneminen	141
10.3	Suppenemistuloksia	142
10.4	Keskeinen raja-arvolause	144
10.5	Normaaliapproksimaatio	145
10.6	Deltamenetelmä	146

Luku 1

Tapahtumat ja niiden todennäköisyydet

1.1 Johdanto

Tilastotiedettä, ja tätä kautta todennäköisyyslaskentaa sovelletaan kaikilla tieteiden ja useilla elinkeinoelämän aloilla. Jotkin ilmiöt ovat luonteeltaan deterministisiä, mikä tarkoittaa sitä, että niiden kehitys voidaan ennustaa (esim. jonkin luonnonlain avulla), kun tunnetaan ilmiöön liittyvät alkuehdot ja muut relevantit tekijät. Toisaalta joidenkin ilmiöiden lopputulosta ei osata varmasti ennustaa käytettävän tiedon perusteella. Lopputulokseen liittyvää epävarmuutta voidaan yrittää käsitellä stokastisen mallin (todennäköisyysmallin, tilastollisen mallin) avulla. Todennäköisyyslaskenta on stokastisiin malleihin liittyvää matematiikkaa.

Tarkastellaan kahta erilaista ilmiötä, joiden lopputuloksia voidaan kuvailla todennäköisyyksien avulla:

- a) nopanheitto,
- b) jonkin tietyn vaalin tulokset.

Nopanheiton yhteydessä tavallisesti sovelletaan todennäköisyyden ns. *klassista tulkintaa*, jonka mukaan eri silmäluvut ovat yhtä todennäköisiä nopan fyysikaalisen *symmetrian* nojalla. Yleisesti, jos mahdollisten lopputulosten joukko Ω on äärellinen, ja $A \subset \Omega$ on kiinnostuksen kohteena olevan tapahtuman lopputulosten joukko ja todennäköisyydet määritellään symmetrian perusteella, niin tapauksen A todennäköisyydeksi valitaan A :lle suotuisten tapauksien lukumäärä jaettuna kaikkien mahdollisten tapauksien lukumäärällä, eli

$$P(A) = \frac{\#A}{\#\Omega} = \frac{A\text{:n alkioden lkm}}{\Omega\text{:n alkioden lkm}}$$

Esimerkiksi nopanheiton tapauksessa mahdolliset lopputulokset (silmäluvut) ovat $\Omega = \{1, 2, 3, 4, 5, 6\}$, ja silmäluvun 6 todennäköisyydeksi saadaan ylläolevalla määritelmällä (valitsemalla $A = \{6\}$)

$$P(\text{“kuutonen”}) = \frac{1}{6}.$$

Nopanheittoa on mahdollista toistaa useita kertoja. Yksittäisen nopanheiton lopputulosta on mahdotonta ennustaa, mutta pitkissä toistosarjoissa eri silmälukujen suhteelliset esiintymisfrekvenssit käyttäytyvät säännöllisesti. Nopanheitossa toistot ovat *riippumattomia* siinä mielessä, että edeltävien heittojen tulokset eivät vaikuta seuraavien heittojen tuloksiin. Klassisen tulkinnan sijaan nopanheittoon voidaan myös soveltaa todennäköisyyden *frekvenssitulkintaa*, jonka mukaan tietyn tapahtuman tai lopputuloksen A (esim. “yhdessä nopan heitossa saadaan silmäluku kuusi”) todennäköisyys on tämän tapahtuman esiintymisten suhteellisen frekvenssin raja-arvo, kun riippumattomien toistojen lukumäärä kasvaa rajatta. Ts. tapahtuman A todennäköisyys voidaan yrittää määrittellä kaavalla

$$P(A) = \lim_{N \rightarrow \infty} \frac{N(A)}{N}, \quad (1.1)$$

jossa $N(A)$ on niiden kokeiden frekvenssi (eli lukumäärä), joissa saatiin lopputulos A , ja N on toistojen lukumäärä. Suure $N(A)/N$ on tapahtuman A suhteellinen esiintymisfrekvenssi. Tämän lähestymistavan teoreettinen oikeutus perustuu todennäköisyyslaskennan ns. suurten lukujen lakiin, joka on eräs todennäköisyyslaskennan kuuluisimmista tuloksista. Tämän takia todennäköisyyttä ei voida määrittellä frekvenssitulkinnan avulla. Sitä paitsi tällaista raja-arvoa ei voida koskaan saada reaali maailman kokeilujen avulla tarkasti selville.

Frekvenssitulkinta ei kelpaa todennäköisyyden määritelmäksi, mutta sitä kannatta kuitenkin yrittää mieleissään soveltaa todennäköisyyslaskennan käsitteiden ja erilaisten todennäköisyysmallien ominaisuuksien ymmärtämiseksi. Erityisen hyödyllistä on yrittää simuloida todennäköisyysmallin käyttäytymistä tietokoneella. Tällöin puhtaan stokastisesta simuloinnista tai Monte Carlo -menetelmästä. Tietyn tapahtuman todennäköisyyttä voidaan tällöin arvioida toistamalla simulointia useita kertoja ja laskemalla ko. tapahtuman esiintymien suhteellinen frekvenssi.

Toisin kuin nopanheitto, yksittäiset vaalit ovat ainutkertainen tapahtuma, joten tässä yhteydessä ei voida puhua kokeen toistamisesta. Silti monet tahot ovat valmiita arvioimaan vaalien lopputulosta ennen kuin se on tiedossa. Esimerkiksi vaalituloksista on tavallisesti mahdollista lyödä vetoa jossakin vedonlyöntitoimistossa. Vedonlyöntitoimisto asettaa kertoimensa arvioimalla kyseessä olevien tapahtumien todennäköisyyksiä. Todennäköisyydelle on ainutkertaisen tapahtumien yhteydessä annettava toisenlainen, ns. *subjektiivinen* eli henkilökohtainen tulkinta. Tällöin todennäköisyys kuvaa henkilön (tai muun tahon) uskomuksen astetta väitteen totuuteen hänen käytettävissä olevan tiedon pohjalta. Eri tahot voivat tällöin saada erilaisia numeerisia vastauksia saman tapahtuman todennäköisyydelle. Saman henkilön todennäköisyydet tietylle tapahtumalle voivat olla eri aikoina erilaisia, jos henkilön käytössä oleva informaatio muuttuu.

Joskus tarvitaan vielä edellisistä poikkeavia tapoja tulkita todennäköisyyden käsitettä. Esimerkiksi kvanttimekaniikan yhteydessä tietävästi edelleen kiistellään siitä, miten kvanttimaailman ilmiöiden todennäköisyydet pitäisi tulkita. Lisäksi kvanttimekaniikan todennäköisyyslaskenta poikkeaa joiltain osin siitä, mitä tällä kurssilla käsitellään.

1.2 Joukko-oppia

Aluksi kertaamme joukko-opin merkintöjä, sillä todennäköisyyslaskenta perustuu joukko-oppiin.

Määritelmä 1.1. Olkoon Ω tarkasteltavan satunnaiskokeen kaikkien mahdollisten lopputulosten joukko. Kutsumme joukkoa Ω *perusjoukoksi*, ja sen alkioita $\omega \in \Omega$ *alkeistapauksiksi* (engl. *elementary event*). *Tapahtumiksi* (engl. *event*) kutsumme perusjoukon Ω osajoukkoja.

Huomautus. Usein tilastotieteessä perusjoukolle käytetään nimitystä otosavaruus (engl. *sample space*) sekä esim. merkintää S .

Olkoot nyt A ja B perusjoukon Ω osajoukkoja. Seuraavat merkinnät lienevät lukijalle tuttuja:

- Merkintä $x \in A$ tarkoittaa, että x on joukon A alkio, eli että x kuuluu joukkoon A . Merkintä $x \notin A$ tarkoittaa, että x ei ole joukon A alkio, eli että x ei kuulu joukkoon A .
- A^c tai \bar{A} tarkoittaa joukon A komplementtia, eli niitä $x \in \Omega$, jotka eivät ole A :n alkioita.
- $A \subset B$ tarkoittaa, että joukko A on joukon B osajoukko eli että jokainen A :n alkio on myös B :n alkio; tämä asia voidaan merkitä myös $B \supset A$.
- $A = B$ tarkoittaa, että joukot A ja B ovat samoja, eli että niillä on samat alkiot (ts. sitä, että $A \subset B$ ja $B \subset A$).
- $A \cup B = \{x : x \in A \text{ tai } x \in B\}$ on joukkojen A ja B yhdiste.
- $A \cap B = \{x : x \in A \text{ ja } x \in B\}$ on joukkojen A ja B leikkaus.
- $A \setminus B = A \cap B^c = \{x : x \in A \text{ ja } x \notin B\}$ on joukkojen A ja B erotus.
- \emptyset on tyhjä joukko.

Joukko-operaatiota ja niiden ominaisuuksia kannattaa havainnollistaa Vennin diagrammien avulla.

Joukko-operaatioilla on lukuisia tärkeitä algebrallisia ominaisuuksia. Yhdisteellä ja leikkauksella on mm. seuraavat ominaisuudet.

- Kommutatiivisuus (eli vaihdantalait): kaikilla A ja B pätee

$$A \cup B = B \cup A, \quad A \cap B = B \cap A.$$

- Assosiatiivisuus (eli liitälait): kaikilla A , B ja C pätee

$$\begin{aligned}(A \cup B) \cup C &= A \cup (B \cup C) = A \cup B \cup C \\ (A \cap B) \cap C &= A \cap (B \cap C) = A \cap B \cap C.\end{aligned}$$

- Distributiivisuus (eli osittelulait): kaikilla A , B ja C pätee

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

Osittelulain muistamista auttaa vastaavan säännön palauttaminen mieleen reaalityöiden laskutoimituksille: $(a + b)c = ac + bc$. Joukko-opissa kumpi tahansa operaatioista \cup tai \cap voi olla yhteenlaskun roolissa, ja toiselle jää kertolaskun rooli.

De Morganin lait yhdistävät yhdisteen, leikkauksen ja komplementoinnin,

$$(A \cup B)^c = A^c \cap B^c \quad \text{ja} \quad (A \cap B)^c = A^c \cup B^c \quad \text{kaikille } A \text{ ja } B.$$

De Morganin lait yleistyvät myös usemman kuin kahden joukon yli lasketulle yhdisteelle ja leikkaukselle.

Yhdisteen ja leikkauksen voi laskea myös äärettömän monen joukon kokoelmalle. Jos A_1, A_2, \dots ovat perusjoukon osajoukkoja, niin niiden yhdistettä merkitään

$$\bigcup_{j=1}^{\infty} A_j = \{x : x \in A_j \text{ jollakin } j\}$$

ja leikkausta merkitään

$$\bigcap_{j=1}^{\infty} A_j = \{x : x \in A_j \text{ kaikilla } j\}.$$

Todennäköisyyslaskennan yhteydessä joukko-operaatioille voidaan antaa havainnollisempia tulkintoja. Taulukko 1.1 antaa tästä esimerkkejä.

Taulukko 1.1 Eräitä joukko-opin merkintöjen tulkintoja.

Kaava	Tulkinta
Ω	varma tapahtuma
\emptyset	mahdoton tapahtuma
$x \in A$	x on A :lle suotuisa alkeistapaus
A	A sattuu
A^c	A ei satu
$A \cup B$	A tai B (tai molemmat) sattuu
$A \cap B$	sekä A ja B sattuvat
$A \cap B = \emptyset$	A ja B ovat erillisiä eli toisensa poissulkevia
$A \subset B$	jos A sattuu, niin myös B sattuu
$A \setminus B$	A sattuu, mutta B ei satu
$\bigcup_{j=1}^{\infty} A_j$	ainakin yksi tapahtumista $A_j, j \geq 1$ sattuu
$\bigcap_{j=1}^{\infty} A_j$	kaikki tapahtumat $A_j, j \geq 1$ sattuvat

Esimerkki 1.1. Perusjoukko eräissä sovelluksissa. Reaalimaailman satunnaisilmiön mallintamiseksi perusjoukko voidaan valita muuten vapaasti, mutta sen pitää olla riittävän rikas, jotta kaikki kiinnostuksen kohteena olevat tapahtumat voidaan esittää sen osajoukkoina. Tietty sovellus voidaan käsitellä käyttämällä erilaisia valintoja.

- Nopanheitossa tulos on jokin silmäluvuista 1, 2, ..., 6. Perusjoukoksi voidaan valita

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

- Kahden nopan heitossa voidaan alkeistapaukset esittää kokonaislukupareilla (i, j) , jossa ensimmäinen koordinaatti i kertoo ensimmäisen nopan tuloksen ja toinen koordinaatti j toisen nopan tuloksen. Siis voidaan valita

$$\Omega = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}.$$

Tämä valinta kelpaisi myös yhden nopan heiton kuvailuun; meidän tarvitsisi vain jättää laskuissa jälkimmäinen koordinaatti huomioimatta.

- Matemaattisissa ohjelmistoissa on yleensä ns. satunnaislukugeneraattori, joka poimii satunnaisesti reaaliluvun väliltä $(0, 1)$. Tällöin luonteva perusjoukko on kyseinen lukuväli, $\Omega = (0, 1)$. Toisin kuin edellisissä sovelluksissa, tällä kertaa perusjoukko ei ole äärellinen, vaan on ääretön ja sisältää ylinumeroituvan monta alkioita.
- Eräs tärkeä stokastisten prosessien tyyppi on sellainen, jonka jokainen realisaatio on välillä $(0, \infty)$ määritelty jatkuva funktio. Tällöin luonteva perusjoukko on kaikkien välillä $(0, \infty)$ määriteltyjen jatkuvien funktioiden muodostama joukko. Tällaisen perusjoukon käsittely on huomattavasti monimutkaisempaa kuin minkään aikaisemmin esitetyn perusjoukon käsittely, mutta tähän asiaan emme tällä kurssilla puutu.

△

1.3 Matemaattinen todennäköisyyden käsite

Johdantoluvussa tapasimme erilaisia todennäköisyyden tulkintoja. Matemaattinen todennäköisyyden käsite ei ota kantaa siihen, miten sitä tulkitaan eli miten sitä sovelletaan reaali maailman ilmiöiden kuvailuun. Sen sijaan ajatuksena on määritellä aksiomien avulla, minkälaisia ominaisuuksia todennäköisyydellä on. Nämä todennäköisyyden aksiomat esitti A. N. Kolmogorov 1930-luvulla

Todennäköisyys(mitta) (engl. *probability measure*) P liittyy perusjoukon Ω tapahtumiin $A \subset \Omega$ reaaliluvun $P(A)$, jota kutsutaan tapahtuman A todennäköisyydeksi. Todennäköisyydeltä vaaditaan seuraavat ominaisuudet.

Määritelmä 1.2. P on todennäköisyys (lyh. tn) eli todennäköisyysmitta (lyh. tn-mitta), mikäli se toteuttaa seuraavat kolme ominaisuutta.

- (1) $P(A) \geq 0$ kaikille tapahtumille $A \subset \Omega$,
- (2) $P(\emptyset) = 0$ ja $P(\Omega) = 1$,
- (3) (täysadditiivisuus) $P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$, kun A_1, A_2, \dots ovat erillisiä tapahtumia.

Se seikka, että tapahtumat A_1, A_2, \dots ovat *erillisiä* (engl. *disjoint*) (ks. aksioma (3)) tarkoittaa sitä, että

$$A_i \cap A_j = \emptyset \quad \text{kun} \quad i \neq j.$$

Tämä voidaan ilmaista myös sanomalla, että kyseiset tapahtumat ovat *toisensa poissulkevia* (engl. *mutually exclusive*).

Jos A ja B ovat erillisiä tapahtumia (ts. $A \cap B = \emptyset$), niin täysadditiivisuuden aksiomasta seuraa (valinnoilla $A_1 = A$, $A_2 = B$ ja $A_j = \emptyset$, kun $j \geq 3$) se, että

$$P(A \cup B) = P(A) + P(B) \quad \text{kun} \quad A \cap B = \emptyset. \quad (1.2)$$

Tätä ominaisuutta kutsutaan n -mitan (äärelliseksi) *additiivisuudeksi*. Pysähdymme nyt pohtimaan todennäköisyyden frekvenssitulkinnan valossa, miksi (äärellinen) additiivisuus on luonnollinen todennäköisyyden ominaisuus.

Olkoot A ja B tietyn satunnaiskokeen erillisiä tapahtumia. Oletetaan, että kyseistä satunnaiskokeita toistetaan riippumattomasti N kertaa. (Emme tässä vaiheessa määrittele, mitä tällainen riippumaton toistaminen oikeastaan tarkoittaa, vaan pohdinta pitää yrittää ymmärtää intuitioon nojautuen.) Sovitaan, että $N(E)$ tarkoittaa niiden kokeiden lukumäärää, joissa tapahtuma E sattuu. Koska A ja B ovat erillisiä, niin $N(A \cup B) = N(A) + N(B)$, sillä kussakin toistossa korkeintaan yksi tapahtumista A tai B voi sattua. Tämän takia

$$\frac{N(A \cup B)}{N} = \frac{N(A)}{N} + \frac{N(B)}{N},$$

joten additiivisuus pätee suhteellisille frekvensseille. Rajankäynnin jälkeen additiivisuuden pitää päteä myös todennäköisyyksille.

Sen sijaan täysadditiivisuus (aksioma (3)) ei ole intuitiivisesti selvä ominaisuus. Se oletetaan sen takia, että tällä tavalla menetellen saadaan matemaattisesti kaunis teoria, jonka puitteissa voidaan formuloida ja todistaa äärettömiä pitkiä tapahtumajonoja koskevia lauseita (kuten esim. vahva suurten lukujen lause).

Todennäköisyyden aksiomeista seuraa yksinkertaisilla laskuilla monia sen ominaisuuksia, kuten esimerkiksi seuraavassa lauseessa todetut ominaisuudet. Näistä osa jätetään harjoitustehtäviksi. Ideana todistuksissa on jakaa sopivasti valittu tapahtuma erillisiin tapahtumiin, ja sitten soveltaa additiivisuutta sekä muita todennäköisyyden ominaisuuksia.

Lause 1.1. *Todennäköisyydellä on seuraavat ominaisuudet.*

- a) *(Äärellinen additiivisuus)* Jos $A \cap B = \emptyset$, niin $P(A \cup B) = P(A) + P(B)$.
- b) $P(A^c) = 1 - P(A)$ kaikille A .
- c) $0 \leq P(A) \leq 1$ kaikille A .
- d) *(n -mitan monotonisuus)* Jos $A \subset B$, niin $P(A) \leq P(B)$.
- e) $P(A \setminus B) = P(A) - P(A \cap B)$ kaikille A ja B .
- f) *(Yhteenlaskukaava)* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ kaikille A ja B .
- g) *(Boolen epäyhtälö)* $P(A \cup B) \leq P(A) + P(B)$ kaikille A ja B .
- h) *(Bonferronin epäyhtälö)* $P(A \cap B) \geq P(A) + P(B) - 1$ kaikille A ja B .

Todistus. Äärellinen additiivisuus (a-kohta) todistettiin kaavassa (1.2). Todistetaan esimerkin vuoksi kohdat b) ja c). Jos A on tapahtuma, niin $\Omega = A \cup A^c$, jossa osat ovat erillisiä. Additiivisuuden ja aksioman (2) nojalla

$$1 = P(A) + P(A^c),$$

joten b-kohta on todistettu. Koska A^c on tapahtuma, niin $P(A^c) \geq 0$ aksiooman (1) nojalla, joten

$$P(A) = 1 - P(A^c) \leq 1,$$

mikä yhdessä aksiooman (1) kanssa todistaa c-kohdan. \square

Huomautus. $0 \leq P(A) \leq 1$ kaikille tapahtumille A . Jos joskus saat jonkin tapahtuman todennäköisyydeksi luvun, joka ei ole välillä $[0, 1]$, olet tehnyt las-kuvirheen. Tällaisista virheistä sakotetaan pisteitä tämän kurssin kokeissa ja tenteissä!

Yhdessä perusjoukkoa Ω , sen tapahtumia eli sitä Ω :n osajoukkojen kokoelmaa, jolla P on määritelty, sekä Ω -mittaa P kutsutaan *todennäköisyysavaruudeksi* (tai todennäköisyyskentäksi).

Täydentävä huomautus. Jos perusjoukko Ω on äärellinen tai korkeintaan numeroituvasti ääretön, niin tavallisesti sallitaan tapahtumaksi mikä tahansa perusjoukon osajoukko. Jos kuitenkin Ω on suurempi kuin numeroituvasti ääretön, niin todennäköisyyden aksioomia ei saada toteutumaan kaikille sen osajoukoille, vaan joukkofunktion P määrittelyjoukkoa pitää rajoittaa. Kaikkien tapahtumien joukolla eli todennäköisyyden P määrittelyjoukolla on se ominaisuus, että se on ns. σ -algebra. Tästä seuraa se, että jos lähdetään liikkeelle korkeintaan numeroituvasta määrästä tapahtumia, ja sovelletaan niihin korkeintaan numeroituva määrä joukko-operaatiota, niin tulokseksi saadaan tapahtuma. Tällä kurssilla sivuutamme σ -algebroidiin liittyvät tarkastelut, ja **oletamme, että todennäköisyys on määritelty kaikille meitä kiinnostaville perusjoukon osajoukoille.**

1.4 Kombinatoriikkaa

Lantinheiton, nopanheiton, korttipelien, loton ja muiden tällaisten uhkapelin stokastisena mallina pidetään — mikäli ei kerrota tarkempia tietoja — klassista todennäköisyyden tulkintaa, jonka mukaan

$$P(A) = \frac{\#A}{\#\Omega}. \quad (1.3)$$

Tapahtuman A todennäköisyys on tällöin tapahtumalle suotuisten alkeistapauksien lukumäärän (ts. joukon A alkioden lukumäärän) ja kaikkien mahdollisten alkeistapauksien lukumäärän (ts. perusjoukon Ω alkioden lukumäärän) osamäärä. Tällainen valinta on mahdollista tietenkin vain silloin, kun perusjoukko on äärellinen. Alkeistapaukset $\omega \in \Omega$ ovat tällöin siinä mielessä symmetrisiä, että niillä on kaikilla sama todennäköisyys, eli

$$P(\{\omega\}) = \frac{1}{\#\Omega}, \quad \text{kaikilla } \omega \in \Omega.$$

Tällaiseen malliin viitataan usein termillä *klassinen todennäköisyys*. Vaihtoehtoisesti voidaan puhua *tasaisesta todennäköisyysmallista* tai esim. *symmetrisestä todennäköisyysavaruudesta*. Kun käsitellään klassista todennäköisyyttä (eli symmetristä todennäköisyysavaruutta), niin stokastinen malli on määritelty heti, kun perusjoukko on valittu. Perusjoukko Ω pitää tietenkin valita järkevästi,

jotta sen alkiot ja reaali maailman symmetriat vastaavat toisiaan. Jotta osaisimme laskea todennäköisyyksiä, meidän pitää osata laskea eräitä kombinatorisia laskuja, jotka lienevät lukijalle pääosin jo ennestään tuttuja.

Tarkastellaan n -alkioista joukkoa E ,

$$E = \{e_1, e_2, \dots, e_n\}.$$

Kuinka monella eri tavalla voidaan poimia k alkioita n -alkoisesta joukosta E ?

Saamme tähän kysymyksen erilaisia vastauksia sen mukaan, valitaanko alkiot

- *takaisinpanolla* (eli palauttaen, engl. *with replacement*), jolloin sama alkio voidaan valita useaan kertaan, ja k voi olla suurempi kuin n ;
- *ilman takaisinpanoa* (eli takaisinpanotta eli palauttamatta, engl. *without replacement*), jolloin valitaan $k \leq n$ eri alkioita.

Termi “takaisinpano” viittaa tilanteeseen, jossa esim. arpalippuja poimitaan arvontauurnasta umpimähkään yksi kerrallaan. Otannassa takaisinpanolla kukin arpalippu palautetaan, eli pannaan takaisin uurnaun ennen seuraavan arvan poimintaa. Otannassa ilman takaisinpanoa arpalippua ei laiteta takaisin ennen seuraavan arpalipun poimintaa.

Lisäksi pitää kertoa, onko poimintajärjestyksellä väliä vai ei.

- Jos *poimintajärjestyksellä on väliä*, niin valinta esitetään (järjestettynä) jonona.
- Jos *poimintajärjestyksellä ei ole väliä*, niin sellaiset kaksi valintaa ovat ekvivalentteja, joissa on poimittu samat alkiot (ja kukin alkio on molemmissa valinnoissa poimittu yhtä monta kertaa).

Pysytymme laskemaan lukumäärät eri tilanteissa ns. tuloperiaatteen avulla.

Määritelmä 1.3 (Tuloperiaate). Tarkastellaan tehtävää, joka voidaan jakaa k :hon osatehtävään. Ensimmäinen osatehtävä voidaan suorittaa n_1 :llä eri tavalla. Kun ensimmäinen osatehtävä on suoritettu, toinen osatehtävä voidaan suorittaa n_2 :lla eri tavalla (riippumatta siitä, mikä valinta tehtiin ensimmäisessä vaiheessa) jne. ja viimein, kun $k - 1$ ensimmäistä osatehtävää on suoritettu, k :s osatehtävä voidaan suorittaa n_k eri tavalla (riippumatta siitä, mitkä valinnat tehtiin aikaisemmissa vaiheissa). Mikäli koko tehtävän ratkaisu saadaan koottua yksikäsitteisellä tavalla osatehtävien ratkaisuksista, niin tällöin koko tehtävä voidaan suorittaa

$$n_1 n_2 \cdots n_k = \prod_{j=1}^k n_j$$

eri tavalla.

Poiminta takaisinpanolla, kun poimintajärjestyksellä on väliä: n -alkioisesta joukosta E voidaan muodostaa

$$n^k$$

erilaista k -alkioista jonoa (x_1, \dots, x_k) takaisinpanolla. Tämä lukumäärä on sama kuin joukon E k -kertaisen karteesisen tulon

$$E \times \dots \times E \quad (k \text{ tekijää})$$

kardinaliteetti. Kunkin koordinaatin x_i arvolle on n eri mahdollisuutta riippumatta muiden koordinaattien arvoista. Näiden jonojen muodostama joukko ja vastaava tasainen todennäköisyysmalli on luonteva malli esim. k :lle lantin tai nopan heitolle: lantinheitossa joukolla E on $n = 2$ alkioita, ja nopanheitossa $n = 6$.

Poiminta ilman takaisinpanoa, kun poimintajärjestyksellä on väliä: n -alkioisesta joukosta E voidaan muodostaa

$$n(n-1) \cdots (n-k+1).$$

erilaista k -jonoa ($k \leq n$), kun kaikkien jonon alkioiden pitää olla erisuuria. Tällaisiin jonoihin viitataan termillä E :n k -permutaatio (tai termillä k -variaatio, kuten Tuomisen kirjassa). Ensimmäinen jonon alkio voidaan valita n erilaisella tavalla. Tämän jälkeen toinen alkio voidaan valita $(n-1)$:llä eri tavalla, sillä toisen alkion pitää olla eri suuri kuin ensimmäisen. Tekemällä tällä tavoin peräkkäin k valintaa, joissa aina edelliset valinnat on suljettu pois päädytään yllä olevaan kaavaan.

Muistathan n -kertoman $n!$ määritelmän kokonaisluvulle $n \geq 0$:

$$n! = n \cdot (n-1) \cdots 2 \cdot 1, \quad \text{kun } n \geq 1, \text{ ja } \quad 0! = 1. \quad (1.4)$$

Kertoma $n!$ ilmaisee n -alkoisen joukon n -permutaatioiden, eli lyhyesti permutaatioiden lukumäärän; n alkioita voidaan siis järjestää $n!$ erilaisella tavalla. Kertoman avulla kirjoitettuna n -alkoisen joukon k -permutaatioita on

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

Poiminta ilman takaisinpanoa, kun poimintajärjestyksellä ei ole väliä: kun n -alkioisesta joukosta E valitaan k alkioita ilman takaisinpanoa, niin tämä on sama asia kuin että valitaan joukon E k -alkioinen osajoukko. Näitä k -osajoukkoja kutsutaan myös joukon E k -kombinaatioiksi. Niiden lukumäärän ilmaisee binomikerroin n yli k :n,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \begin{cases} \frac{n(n-1) \cdots (n-k+1)}{k!}, & \text{kun } 0 \leq k \leq n, \\ 0, & \text{muuten.} \end{cases} \quad (1.5)$$

Tämä johtuu siitä, että kukin tällainen osajoukko voidaan järjestää $k!$ eri tavalla k -jonoksi, jolloin saadaan generoitua kertaalleen kaikki joukon E k -permutaatiot.

Poiminta takaisinpanolla, kun poimintajärjestyksellä ei ole väliä: sivuutetaan tällä kurssilla.

Lukija muistaa varmaankin, että binomikertoimet esiintyvät ns. *binomikaavassa*, jonka mukaan

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}, \quad (1.6)$$

kun a ja b ovat reaalilukuja ja $n \geq 0$ on kokonaisluku. Binomikaava seuraa binomikertoimien kombinatorisesta luonnehdinnasta sekä reaalilukujen laskutoimitusten ominaisuuksista.

1.5 Multinomikertoimet

Eräs tulkinta binomikertoimelle $\binom{n}{k}$ on, että se kertoo, kuinka monella tavalla n -alkioinen joukko voidaan osittaa kahteen osajoukkoon siten, että ensimmäiseen osaan tulee k alkioita ja toiseen osaan $n - k$ alkioita.

Kuinka monta vaihtoehtoa on, jos n -alkioinen joukko ositetaan kolmeen osajoukkoon siten, että ensimmäiseen osaan tulee k_1 alkioita, toiseen osaan k_2 alkioita ja kolmanteen osaan k_3 alkioita? Jotta kysymys olisi mielekäs, pitää tietenkin olettaa, että kukin $0 \leq k_i \leq n$, ja että

$$k_1 + k_2 + k_3 = n.$$

Samme vastattua kysymykseen helposti tuloperiatteen avulla. Voimme valita ensimmäisen osajoukon alkiot $\binom{n}{k_1}$ tavalla, ja sen jälkeen toisen osan alkiot $\binom{n-k_1}{k_2}$ tavalla, jonka jälkeen kaikki jäljelle jääneet alkiot kuuluvat kolmanteen osaan. Näin ollen eri mahdollisuuksia on yhteensä

$$\binom{n}{k_1} \binom{n-k_1}{k_2} = \frac{n!}{k_1!(n-k_1)!} \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} = \frac{n!}{k_1!k_2!k_3!}.$$

Tälle luvulle käytetään seuraavia merkintöjä,

$$\binom{n}{k_1 k_2 k_3} = \binom{n}{k_1, k_2, k_3} = \frac{n!}{k_1!k_2!k_3!},$$

eli toisinaan alakerrassa olevat luvut erotetaan toisistaan selvyden vuoksi pilkuilla, ja toisinaan taas pilkkuja ei käytetä. Tälle luvulle käytetään nimitystä trinomikerroin (tai multinomikerroin).

Yleistetään edellinen kysymys m :lle osalle. Olkoon annettuna luvut k_1, \dots, k_m siten, että kukin $0 \leq k_i \leq n$, ja

$$k_1 + \dots + k_m = n.$$

Miten monella eri tavalla n -alkioinen joukko voidaan osittaa m osaan A_1, \dots, A_m siten, että kussakin osassa A_i on k_i alkioita? Vastaavalla tavalla päättelämällä kuin edellä tapauksessa $m = 3$ nähdään, että eri mahdollisuuksia on

$$\frac{n!}{k_1!k_2! \dots k_m!} = \binom{n}{k_1, k_2, \dots, k_m} \quad (1.7)$$

kappaletta. Näitä lukuja kutsutaan multinomikertoimiksi. Binomikerroin on tietenkin multinomikertoimen erikoistapaus, sillä

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

Multinomikertoimet esiintyvät ns. *multinomikaavassa*, jonka mukaan

$$(a_1 + a_2 + \dots + a_m)^n = \sum \binom{n}{k_1, k_2, \dots, k_m} a_1^{k_1} a_2^{k_2} \dots a_m^{k_m}. \quad (1.8)$$

Tässä kaavassa summataan kaikkien ei-negatiivisten kokonaislukujen k_1, k_2, \dots, k_m , joiden summa on n . Binomikaava (1.6) on multinomikaavan erikoistapaus.

1.6 Ehdollinen todennäköisyys

Olkoot A ja B tapahtumia, ja $P(B) > 0$. Tapahtuman A ehdollinen todennäköisyys ehdolla B määritellään kaavalla

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \quad \text{jossa } P(B) > 0. \quad (1.9)$$

Tämä on tapahtuman A todennäköisyys, kun tiedetään, että B on sattunut.

“Alkuperäiset” eli ei-ehdolliset todennäköisyydet $P(A)$ voidaan ymmärtää ehdollisina todennäköisyyksinä ehdolla Ω ,

$$P(A | \Omega) = \frac{P(A \cap \Omega)}{P(\Omega)} = \frac{P(A)}{1}.$$

On lisäksi mahdollista tarkistaa, että ehdollinen todennäköisyys $P(A | B)$ on argumentin A funktiona todennäköisyys, eli että se toteuttaa todennäköisyyden kolme aksiomaa. Tästä seuraa, että **ehdolliselle todennäköisyydelle saadaan käyttää samoja laskusääntöjä (pystyviivan vasemmalla puolella esiintyvän argumentin suhteen), kuin tavalliselle todennäköisyydelle.**

Ehdollisen todennäköisyyden kaavan frekvenssitulkintaa varten tarkastellaan N -kertaista toistokoetta, jossa voi sattua A tai B tai molemmat. Tapahtuman A suhteellinen esiintymisfrekvenssi niissä kokeissa, joissa on sattunut tapahtuma B on

$$\frac{N(A \cap B)}{N(B)} = \frac{N(A \cap B)/N}{N(B)/N}.$$

Frekvenssitulkinnan mukaan jälkimmäisessä muodossa osoittaja lähestyy todennäköisyyttä $P(A \cap B)$ ja nimittäjä todennäköisyyttä $P(B)$, kun N kasvaa rajatta. Siis ehdollisen todennäköisyyden $P(A | B)$ määritelmä osamääränä $P(A \cap B)/P(B)$ on järjestyksessä.

Ehdollisen todennäköisyyden määritelmä (1.9) voidaan kirjoittaa myös tulomuodossa,

$$P(A \cap B) = P(A | B) P(B). \quad (1.10)$$

Tämä erittäin tärkeä kaava on nimeltään todennäköisyyksien *kertolaskusääntö* eli *kertolaskukaava*. Siitä käytetään myös nimeä todennäköisyysslaskennan *ketjusääntö*.

Mikäli $P(A) > 0$ ja $P(B) > 0$, on

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

josta saadaan ratkaistua toinen ehdollisista todennäköisyyksistä, nimittäin

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

Kertolaskusääntö (1.10) voidaan yleistää myös usemmalle tapahtumalle. Tarkastellaan esimerkiksi kolmea tapahtumaa A , B ja C , ja oletetaan, että $P(A \cap B) > 0$. Tällöin myös $P(A) > 0$, ja

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

Jos tapahtumia on n kappaletta A_1, \dots, A_n ja $P(A_1 \cap \dots \cap A_{n-1}) > 0$, niin niiden leikkauksen todennäköisyys saadaan kertolaskusäännöllä

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) \\ = P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \cdots P(A_n | A_1 \cap \dots \cap A_{n-1}). \end{aligned} \quad (1.11)$$

Esimerkki 1.2. Perusteellisesti sekoitetusta 52 kortin pakasta jaetaan kaksi korttia. Tarkastellaan tapahtumia

$$A_1 = \{\text{“1. kortti on ässä”}\}, \quad A_2 = \{\text{“2. kortti on ässä”}\}.$$

Laske todennäköisyydet $P(A_1)$, $P(A_2)$, $P(A_1 \cap A_2)$ ja $P(A_2 | A_1)$.

Ratkaisu 1. Valitaan perusjoukoksi 52 kortin 2-permutaatiot, joita on $52 \cdot 51$ kappaletta. Tuloperiaatteen nojalla

$$\#A_1 = \#A_2 = 4 \cdot 51, \quad \#(A_1 \cap A_2) = 4 \cdot 3.$$

Siis

$$P(A_1) = P(A_2) = \frac{4 \cdot 51}{52 \cdot 51} = \frac{1}{13}, \quad P(A_1 \cap A_2) = \frac{4 \cdot 3}{52 \cdot 51},$$

ja näistä saadaan

$$P(A_2 | A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{4 \cdot 3}{4 \cdot 51} = \frac{1}{17}.$$

Ratkaisu 2. Ajatellaan tilannetta kortti kerrallaan. Todennäköisyys, että ensimmäinen kortti on ässä on tietenkin $4/52$, koska ässiä on pakassa neljä. Kun tarkastellaan vain yhden kortin arvoa kerrallaan, on samantekevää, momentenko se jaetaan, joten $4/52$ on myös tn sille, että toinen kortti on ässä. Siis

$$P(A_1) = P(A_2) = \frac{4}{52} = \frac{1}{13}.$$

Kun ensimmäinen kortti on jaettu ja osoittautunut ässäksi, niin jäljellä on perusteellisesti sekoitettu 51 kortin pakka, joka sisältää 3 ässää. Tämän takia

$$P(A_2 | A_1) = \frac{3}{51}.$$

Kertolaskusäännön nojalla

$$P(A_1 \cap A_2) = P(A_1) P(A_2 | A_1) = \frac{4}{52} \frac{3}{51}$$

△

1.7 Tapahtumien riippumattomuus

Määritelmä 1.4 (Kahden tapahtuman riippumattomuus). Tapahtumat A ja B ovat riippumattomia, jos

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

Tämä asia voidaan merkitä $A \perp B$.

Huomautus. Tarkista, että ymmärrät, mitä eroa on seuraavilla käsitteillä.

- a) A ja B ovat erillisiä tapahtumia,
 b) A ja B ovat riippumattomia tapahtumia.

Jos A ja B ovat riippumattomia, niin tapauksen $A \cap B$ (sekä A että B sattuvat) todennäköisyys saadaan kertomalla keskenään kyseisten tapauksien ei-ehdolliset todennäköisyydet. (Yleisemmässä tapauksessa tarvitaan kertolaskusääntöä.)

Jos $P(B) > 0$ ja $A \perp B$, niin

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

Siis mikäli A ja B ovat riippumattomia, niin tieto B :n sattumisesta ei muuta A :n todennäköisyyttä. Toisaalta, jos $P(B) > 0$ ja $P(A | B) = P(A)$, niin $A \perp B$, sillä tällöin

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

Yhteenveto edellisistä tarkasteluista: jos $P(B) > 0$, on

$$A \perp B \iff P(A | B) = P(A). \quad (1.12)$$

Jos $A \perp B$, niin helpoilla laskuilla saadaan tarkistettua, että myös

$$A^c \perp B, \quad A \perp B^c, \quad A^c \perp B^c.$$

Esimerkki 1.3. Kahden nopan heitto. Perusjoukko on $\Omega = E \times E$, jossa $E = \{1, 2, 3, 4, 5, 6\}$. Sen alkeistapaukset ovat symmetrisiä, ja alkeistapauksessa $(x, y) \in E \times E$ ensimmäinen koordinaatti x kertoo ensimmäisen nopan silmäluvun ja toinen koordinaatti y toisen nopan silmäluvun. Olkoot $1 \leq i, j \leq 6$ annettuja kokonaislukuja, ja tarkastellaan tapahtumia

$$A = \{\text{“nopan yksi tulos on } i\text{”}\}, \quad B = \{\text{“nopan kaksi tulos on } j\text{”}\}$$

Tällöin

$$P(A) = \frac{6}{36}, \quad P(B) = \frac{6}{36},$$

ja

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

joten $A \perp B$. △

Määritelmä 1.5 (Useamman tapahtuman riippumattomuus). Tapahtumat A_1, \dots, A_n ovat riippumattomia, mikä asia voidaan merkitä $A_1, \dots, A_n \perp$, jos kaikilla $k \geq 2$ on voimassa

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$$

aina kun i_1, \dots, i_k ovat keskenään erisuuria kokonaislukuja, jotka kaikki ovat välillä $1 \leq i_j \leq n$.

Esimerkiksi kolme tapahtumaa A , B ja C ovat riippumattomia, jos kaikki seuraavaat neljä ehtoa pätevät,

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

Tässä neljäs ehto ei seuraa kolmesta ensimmäisestä, eli tapahtumien A , B ja C parittaisesta riippumattomuudesta, mikä voidaan osoittaa yksinkertaisella esimerkillä.

Jos tapahtumat A_1, \dots, A_n ovat riippumattomia, ja kukin joukoista B_i on joko A_i tai sen komplementti, niin voidaan osoittaa, että tapahtumat B_1, \dots, B_n ovat myös riippumattomia.

Riippumattomuuden käsite voidaan laajentaa koskemaan myös äärettömän montaa tapahtumaa A_1, A_2, \dots .

Määritelmä 1.6. Tapahtumat A_1, A_2, \dots ovat riippumattomia, jos kaikilla $k \geq 2$ pätee

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$$

aina kun i_1, \dots, i_k ovat keskenään erisuuria positiivisia kokonaislukuja.

Määritelmä 1.7 (Ehdollinen riippumattomuus). Olkoon C tapahtuma, jolle $P(C) > 0$. Tapahtumat A ja B ovat ehdollisesti riippumattomia ehdolla C , jos

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

Tämä voidaan merkitä $(A \perp B) | C$.

Toisin sanoen kaksi tapahtumaa ovat ehdollisesti riippumattomia ehdolla C , jos ne ovat riippumattomia ehdollisen todennäköisyyden $P(\cdot | C)$ mielessä. Ehdollisen riippumattomuuden käsite voidaan tällä periaatteella tietenkään yleistää useammalle kuin kahdelle tapahtumalle.

1.8 Kokonaistodennäköisyys ja Bayesin kaava

Määritelmä 1.8. Tapahtumat B_1, \dots, B_n muodostavat perusjoukon Ω osituksen, jos

- ne ovat erillisiä: $B_i \cap B_j = \emptyset$, kun $i \neq j$
- ne peittävät (eli tyhjentävät) Ω :n, eli $\Omega = B_1 \cup \dots \cup B_n$.

Olkoon A tapahtuma ja B_1, \dots, B_n perusjoukon Ω ositus. Joukko A voidaan osittaa leikkamalla se kullakin joukoista B_i , eli

$$A = (A \cap B_1) \cup \dots \cup (A \cap B_n),$$

jossa osat $(A \cap B_1), \dots, (A \cap B_n)$ ovat erillisiä, joten (additiivisuus ja kertolaskusääntö)

$$P(A) = \sum_{i=1}^n P(B_i)P(A | B_i). \quad (1.13)$$

Tämä on *kokonaistodennäköisyyden kaava*.

Olkoon A tapahtuma siten, että $P(A) > 0$, ja olkoon B_1, \dots, B_n perusjoukon Ω ositus. Jos satunnaiskokeessa on sattunut tapahtuma A , niin tapahtuman B_i (ehdollinen) tn on

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

Kun tässä $P(A)$ vielä esitetään kokonaistodennäköisyyden kaavalla, saadaan *Bayesin kaava*

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)}, \quad 1 \leq i \leq n. \quad (1.14)$$

1.9 Monotoninen jatkuvuus

Tarkastellaan jonoa erillisiä tapahtumia A_1, A_2, \dots . Määritellään kaikilla $n \geq 1$ tapahtuma B_n yhdisteenä n ensimmäisestä A_i -tapahtumasta,

$$B_n = \bigcup_{i=1}^n A_i.$$

Tapahtumat B_1, B_2, \dots muodostavat nyt kasvavan jonon, ts.

$$B_1 \subset B_2 \subset \dots$$

Lisäksi

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{n=1}^{\infty} B_n.$$

Kehitetään seuraavaksi esitys ylläolevan numeroituvasti äärettömän yhdisteen tn:lle.

Todennäköisyyden täysadditiivisuuden ja tapahtumien A_1, A_2, \dots erillisyyden nojalla on voimassa

$$P(B_n) = \sum_{i=1}^n P(A_i) \xrightarrow{n \rightarrow \infty} \sum_{i=1}^{\infty} P(A_i) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{m=1}^{\infty} B_m\right).$$

Toisaalta, jos lähdetään liikkeelle kasvavasta jonosta tapahtuma B_1, B_2, \dots , jossa siis

$$B_1 \subset B_2 \subset \dots,$$

niin on helppoa konstruoida erilliset tapahtumat A_1, A_2, \dots siten, että kullakin n joukko B_n on yhdiste n ensimmäisestä joukosta A_i . Meidän tarvitsee vain valita

$$A_1 = B_1, \quad A_2 = B_2 \setminus A_1, \quad A_3 = B_3 \setminus (A_1 \cup A_2)$$

ja niin edelleen, ts. asetetaan $A_i = B_i \setminus (\cup_{j=1}^{i-1} A_j)$, kun $i \geq 2$. Edellisen päättelyn nojalla on voimassa

$$P\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{n \rightarrow \infty} P(B_n).$$

Komplementteihin siirtymällä nähdään että samanlainen raja-arvotulos pätee, kun tarkastellaan laskevaa jonoa tapahtumia B_1, B_2, \dots , jossa siis

$$B_1 \supset B_2 \supset \dots$$

Tällöin on voimassa

$$P\left(\bigcap_{i=1}^{\infty} B_i\right) = \lim_{n \rightarrow \infty} P(B_n).$$

Kirjataan nämä huomiot lauseeksi.

Lause 1.2 (Todennäköisyyden monotoninen jatkuvuus). *Olkoon B_1, B_2, \dots jono tapahtumia.*

(a) *Jos jono on kasvava, eli $B_1 \subset B_2 \subset B_3 \subset \dots$, niin*

$$P\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{n \rightarrow \infty} P(B_n).$$

(b) *Jos jono on laskeva, eli $B_1 \supset B_2 \supset B_3 \supset \dots$, niin*

$$P\left(\bigcap_{i=1}^{\infty} B_i\right) = \lim_{n \rightarrow \infty} P(B_n).$$

Luku 2

Satunnaismuuttujat

2.1 Satunnaismuuttuja ja sen jakauma

Satunnaismuuttuja (lyhenne sm, engl. *random variable*, *rv*; myös *variate*) on satunnaiskokeeseen liittyvä numeerinen muuttuja, jonka arvo määräytyy kokeen lopputuloksesta. Satunnaismuuttuja on siis perusjoukolla määritelty reaaliarvoinen funktio.

Määritelmä 2.1. Olkoon Ω perusjoukko. Kuvaus $X : \Omega \rightarrow \mathbb{R}$ on *satunnaismuuttuja*.

Huomautus. Tarkasti ottaen edellinen määritelmä ei ole riittävä. Lisäksi pitäisi vaatia, että jokaisen riittävän säännöllisen joukon (kuten esim. jokaisen välin) $B \subset \mathbb{R}$ alkukuva kuvauksessa X pitää olla tapahtuma, eli sellainen Ω :n osajoukko, jonka tn on määritelty. Tämä vaatimus ilmaistaan sanomalla, että X on Borelin funktio (eli Borel-mitallinen funktio). Tästä lähtien sivuutamme tällaiset mitallisuustarkastelut.

Esimerkki 2.1. Esimerkkejä satunnaismuuttujista.

- Silmäluku nopanheitossa. Jos alkeistapaus $\omega \in \{1, \dots, 6\}$ kertoo nopan silmäluvun, niin

$$X(\omega) = \omega$$

on tuo silmäluku.

- Silmälukujen summa kahdessa nopanheitossa. Jos $E = \{1, \dots, 6\}$, perusjoukko $\Omega = E \times E$, ja alkeistapaukselle $(i, j) \in E \times E$ ensimmäinen koordinaatti i kertoo nopan yksi ja toinen koordinaatti j nopan kaksi silmäluvun, niin niiden summan ilmoittaa

$$X(i, j) = i + j.$$

△

Määritelmä 2.2 (Tapahtuman indikaattori). Olkoon $A \subset \Omega$ tapahtuma. Sen indikaattori (eli ilmaisoin eli osoitinmuuttuja) 1_A on satunnaismuuttuja, jonka määrittelee lauseke

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

Huomautus. Tapahtuman A indikaattorille käytetään yleisesti myös merkintää I_A . Jos tapahtuma A esitetään monimutkaisella lausekkeella, niin usein käytetään edellisten sijasta merkintää $1(A)$ tai $I(A)$. Jos on lisäksi tarpeen merkitä argumentti ω näkyviin (mutta harvoin on), niin voidaan käyttää merkintöjä $1(A)(\omega)$ tai $I(A)(\omega)$.

Usein samalla perusjoukolla on määritelty useampi kuin yksi satunnaismuuttuja, esim. X ja Y . Tällöin voidaan tarkastella myös *satunnaismuuttujien välisiä laskutoimituksia*, kuten aX (vakiolle $a \in \mathbb{R}$), $X + Y$, XY jne. Myös ne ovat satunnaismuuttujia. Esim. $X + Y$ tarkoittaa funktioiden yhteenlaskua, ts. $X + Y$ on se funktio $\Omega \rightarrow \mathbb{R}$, jolle

$$(X + Y)(\omega) = X(\omega) + Y(\omega).$$

Toisin sanoen satunnaismuuttujien (ja muiden funktioiden) laskutoimitukset tulkitaan pisteittäin. Jos Y ei saa arvoa nolla, niin myös X/Y on sm.

Jos X on sm, niin voidaan kysyä, millä todennäköisyydellä se saa arvon joukosta B , jossa $B \subset \mathbb{R}$. Kyseistä tapahtumaa voidaan merkitä

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

mutta sitä merkitään tavallisesti lyhyemmin seuraavasti,

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

Sen todennäköisyyttä voidaan merkitä jollakin seuraavista tavoista,

$$P(X \in B) = P\{X \in B\} = P(\{X \in B\}) = P(\{\omega \in \Omega : X(\omega) \in B\}). \quad (2.1)$$

Tavallisesti käytetään lyhyitä merkintöjä, joissa ei esiinny perusjoukkoa Ω eikä sisäkkäisiä sulkuja.

Erityisesti voidaan olla kiinnostuneita seuraavista valinnoista joukolle B . Huomaa, minkälaisia merkintöjä niiden yhteydessä käytetään.

- B on yksiö $\{x\}$, jossa $x \in \mathbb{R}$. Tällöin kysytään pistetodennäköisyyttä

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

- B on väli kuten esim. suljettu väli $[a, b]$, jossa $a < b$,

$$P(a \leq X \leq b) = P(\{\omega \in \Omega : a \leq X(\omega) \leq b\}).$$

Vastaavasti määritellään $P(a < X \leq b)$, $P(a \leq X < b)$ ja $P(a < X < b)$.

- Voidaan myös tarkastella tapausta $B = (-\infty, x]$. Tällöin merkitään

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

Määritelmä 2.3 (Satunnaismuuttujan jakauma). Jos X on satunnaismuuttuja, niin sen *jakauma* on

$$P(X \in B)$$

ymmärrettyinä argumentin $B \subset \mathbb{R}$ funktiona.

Huomautus. Itse asiassa tn $P(X \in B)$ on määritelty vain silloin, kun $B \subset \mathbb{R}$ on ns. Borelin joukko, jollaisia ovat esim. kaikki välit ja kaikki joukot, jotka voidaan muodostaa numeroituvasta määrästä välejä soveltamalla numeroituva määrä joukko-operaatiota. Tästä lähtien jätämme tämän seikan vaille huomiota. (Oletamme, että kaavoissa tarkastelemme vain sellaisia $A \subset \Omega$, jotka ovat tapahtumia; vain sellaisia $B \subset \mathbb{R}$, jotka ovat Borelin joukkoja; vain sellaisia funktiota $X : \Omega \rightarrow \mathbb{R}$, jotka ovat Borelin funktiota eli satunnaismuuttujia; vain sellaisia funktioita $f : \mathbb{R} \rightarrow \mathbb{R}$ jotka ovat Borelin funktiota. Näitä seikkoja ei tarvitse pelästyä; muunlaisia joukkoja $B \subset \mathbb{R}$ tai funktioita $f : \mathbb{R} \rightarrow \mathbb{R}$ on huomattavan vaikea konstruoida.)

On mahdollista osoittaa, että sm:n X jakauma

$$P_X(B) = P(X \in B), \quad B \subset \mathbb{R}$$

toteuttaa todennäköisyyden aksioomat, joten jakauma on perusjoukossa \mathbb{R} määritelty tn-mitta. Käyttöön otetut lyhyet merkinnät johtavat siihen, että alkuperäinen perusjoukko Ω häviää tavallisesti kokonaan merkinnöistä. Tilanne voidaan mieltää myös siten, että perusjoukkona onkin \mathbb{R} ja että tapahtumat ovat sen osajoukkoja.

Jakauma on reaaliakselin osajoukkojen kokoelmalla määritelty funktio, joten se on hyvin abstrakti olio. Seuraavaksi tarkastelemme konkreettisempia välineitä, joiden avulla voimme kuvailla ja hallita jakaumia: kertymäfunktio, diskreetin jakauman pistetodennäköisyysfunktio ja jatkuvan jakauman tiheysfunktio.

2.2 Kertymäfunktio

Määritelmä 2.4. Jos X on satunnaismuuttuja, niin funktio $F : \mathbb{R} \rightarrow \mathbb{R}$

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

on X :n kertymäfunktio (lyhenne kf).

Huomautus. Englannin kielellä jakauma on (*probability distribution* (joskus *law*), ja kertymäfunktio on *distribution function* tai *cumulative distribution function, cdf*.

Esimerkki 2.2. Tapahtuman indikaattorin kf. Olkoon A tapahtuma, ja olkoon $X = 1_A$, eli X on tapahtuman A indikaattori. Tällöin X :n kf on

$$F(x) = \begin{cases} 0, & \text{jos } x < 0, \\ 1 - P(A), & \text{jos } 0 \leq x < 1, \\ 1, & \text{jos } x \geq 1. \end{cases}$$

Huomaa, että kf:lla voi olla hyppyjä ja että se voi olla vakio jollakin välillä. \triangle

Palautetaan mieleen ennen seuraavaa määritelmää, että monotonisella funktiolla $G : \mathbb{R} \rightarrow \mathbb{R}$ on jokaisessa pisteessä $x \in \mathbb{R}$ olemassa sekä vasemmanpuoleinen että oikeanpuoleinen raja-arvo. Lisäksi sillä on raja-arvo pisteissä $-\infty$ ja ∞ . Näitä raja-arvoja merkitään seuraavasti

$$\begin{aligned} G(x-) &= \lim_{y \rightarrow x-} G(y), & G(x+) &= \lim_{y \rightarrow x+} G(y) \\ G(-\infty) &= \lim_{y \rightarrow -\infty} G(y), & G(\infty) &= \lim_{y \rightarrow \infty} G(y). \end{aligned}$$

Lause 2.1 (Kertymäfunktion ominaisuudet). *Satunnaismuuttujan X kertymäfunktiolla $F : \mathbb{R} \rightarrow \mathbb{R}$ on seuraavat ominaisuudet.*

(a) F on kasvava funktio.

(b) F on oikealta jatkuva funktio, eli

$$F(x+) = F(x), \quad \text{kaikilla } x \in \mathbb{R}.$$

(c) $F(-\infty) = 0$ ja $F(\infty) = 1$.

Kääntäen, jos funktiolla $F : \mathbb{R} \rightarrow \mathbb{R}$ on nämä ominaisuudet, niin on olemassa sm X siten, että F on X :n kf.

Todistus. Jos F on X :n kf ja $x < y$, niin $\{X \leq x\} \subset \{X \leq y\}$, joten

$$F(x) = P(X \leq x) \leq P(X \leq y) = F(y),$$

mikä todistaa ominaisuuden (a).

Koska F on kasvava funktio, niin raja-arvo $F(x+)$ on olemassa missä tahansa pisteessä $x \in \mathbb{R}$. Koska oikeanpuoleinen raja-arvo on olemassa, niin se voidaan laskea minkä tahansa pistettä x oikealta lähestyvän jonon avulla, esim. seuraavasti,

$$\begin{aligned} F(x+) &= \lim_{n \rightarrow \infty} F(x + \frac{1}{n}) = \lim_{n \rightarrow \infty} P(X \leq x + \frac{1}{n}) = P\left(\bigcap_{n=1}^{\infty} \{X \leq x + \frac{1}{n}\}\right) \\ &= P(X \leq x) = F(x). \end{aligned}$$

Päätely perustui siihen, että tapahtumat $\{X \leq x + \frac{1}{n}\}$ muodostavat laskevan jonon, kun $n = 1, 2, \dots$. Sen takia voitiin käyttää tn-mitan monotonista jatkuvuutta (ks. lause 1.2).

Samaan tapaan c-kohta saadaan perusteltua monotonisen jatkuvuuden avulla, ts.

$$F(-\infty) = \lim_{n \rightarrow \infty} F(-n) = P\left(\bigcap_n \{X \leq -n\}\right) = P(\emptyset) = 0.$$

$$F(\infty) = \lim_{n \rightarrow \infty} F(n) = P\left(\bigcup_n \{X \leq n\}\right) = P(\Omega) = 1.$$

Sivuutamme perustelun sille, miksi ominaisuuksista (a), (b) ja (c) seuraa se, että F on jonkin satunnaismuuttujan kf. \square

Lasketaan tn $P(a < X \leq b)$, jossa $a < b$. Koska

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\},$$

jossa osat ovat erillisiä, on

$$P(a < X \leq b) = F(b) - F(a).$$

Seuraavan lauseen jälkeen pystymme toistamaan vastaavan laskun myös väleille (a, b) , $[a, b)$ ja $[a, b]$.

Lause 2.2. Jos F on $sm:n$ X kf, niin kaikilla x

$$P(X < x) = F(x-).$$

Todistus. Todistus onnistuu helposti käyttämällä tn-mitan monotonista jatkuvuutta. \square

Huomaa, että kertymäfunktioilla voi olla hyppyjä. Koska

$$\{X \leq x\} = \{X < x\} \cup \{X = x\},$$

jossa osat ovat erillisiä, niin

$$P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x-).$$

Toisin sanoen pistetodennäköisyys $P(X = x)$ on yhtä suuri kuin kertymäfunktion hyppy pisteessä x .

Merkintäsopimus (satunnaismuuttuja voidaan ilmoittaa funktion alaindeksillä). Usein tarkastellaan useampaa kuin yhtä satunnaismuuttujaa, vaikkapa X ja Y . Tällöin käytetään sellaista merkintätapaa, jossa satunnaismuuttuja kirjoitetaan alaindeksiksi: esim. F_X on $X:n$ kf ja F_Y on $Y:n$ kf. Tätä merkintätapaa voidaan käyttää muidenkin jakauman esitysten kuin kertymäfunktioiden kohdalla.

Määritelmä 2.5. Satunnaismuuttujilla X ja Y on *sama jakauma* eli ne ovat *samoin jakautuneet*, jos

$$P(X \in B) = P(Y \in B), \quad \text{kaikilla } B \subset \mathbb{R}.$$

Tämä seikka voidaan ilmaista merkinnällä

$$X \stackrel{d}{=} Y.$$

Seuraava lause toteaa, että kertymäfunktioit ja jakaumat vastaavat toisiaan yksikäsitteisesti. Tämän takia voidaan puhua tietyn jakauman kertymäfunktioista, eikä aina tarvitse puhua $sm:n$ kertymäfunktioista.

Lause 2.3 (Kertymäfunktio määrää jakauman yksikäsitteisesti). *Seuraavat asiat ovat yhtäpitäviä.*

(a) X ja Y ovat samoin jakautuneet.

(b) $F_X = F_Y$ (ts. $X:n$ ja $Y:n$ kertymäfunktioit ovat samat).

Todistus. (a) \Rightarrow (b): Koska jakaumat ovat samat, on kaikilla $t \in \mathbb{R}$ voimassa

$$F_X(t) = P(X \leq t) = P(Y \leq t) = F_Y(t),$$

joten kertymäfunktioit ovat samat.

Implikaation (b) \Rightarrow (a) todistaminen sivuutetaan, sillä se ei onnistu ilman mittateoriaa. \square

2.3 Diskreetti jakauma

Määritelmä 2.6. Satunnaismuuttujan X jakauma on *diskreetti*, jos sen arvojoukko $\{x_1, x_2, \dots\}$ on äärellinen tai korkeintaan numeroituvasti ääretön.

Se, että sm:n X jakauma on diskreetti voidaan vaihtoehtoisesti ilmaista myös jollakin seuraavista tavoista

- X on diskreetti sm,
- sm X on diskreetisti jakautunut
- X :llä on diskreetti jakauma.

Joskus määritelmää laajennetaan siten, että vaaditaan vain, että X saa todennäköisyydellä yksi arvon korkeintaan numeroituvasti äärettömästä joukosta S . Tällöin diskreetti sm X voi saada arvoja joukon S ulkopuolelta, mutta tämän poikkeustapahtuman todennäköisyys on nolla. Yksinkertaisuuden vuoksi kutsumme myös tässä tapauksessa X :n arvojoukoksi jotakin sellaista korkeintaan numeroituvaa joukkoa S , jolle $P(X \in S) = 1$.

Määritelmä 2.7. Satunnaismuuttujan X pistetodennäköisyysfunktio (lyh. ptnf) on funktio $f : \mathbb{R} \rightarrow \mathbb{R}$, jolle

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

Huomautus. Pistetodennäköisyysfunktio $x \mapsto P(X = x)$ (engl. *probability (mass) function, pmf*) on periaattessa määritelty koko reaaliakselilla kaikenlaisille satunnaismuuttujille X . Tämä käsite on kuitenkin mielenkiintoinen vain diskreeteille satunnaismuuttujille.

Merkintäsopimuksia:

- Tyypillisesti pistetodennäköisyydet ilmoitetaan vain niille x_i , jotka kuuluvat diskreetin sm:n määritelmässä esiintyvään, korkeintaan numeroituvasti äärettömään joukkoon, joka voi esimerkiksi olla jokin kokonaislukujen osajoukko. Tällöin asiayhteydestä pitää ymmärtää, että muilla $x \in \mathbb{R}$ on $P(X = x) = 0$.
- Diskreetin satunnaismuuttujan ptnf:lle käytetään muuten samaa symbolia kuin sen kf:lle, mutta ptnf:n symboli kirjoitetaan pienellä kirjaimella (esim. f , g tai f_X) ja kf:n symboli vastaavalla isolla kirjaimella (esim. F , G tai F_X).

Lause 2.4 (Ptnf:n ominaisuuksia). *Olkoon X diskreetti sm, f sen ptnf, ja olkoon S sen arvojoukko. Tällöin*

(a) $0 \leq f(x) \leq 1$ kaikilla x , ja $f(x) = 0$, kun $x \notin S$.

(b) f määrää X :n jakauman kaavalla

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

Todistus. Olkoon $S = \{x_1, x_2, \dots\}$. Kohta (a) on ilmeinen. Kohdan (b) pohjustamiseksi pitää huomauttaa, että siinä lasketaan yhteen korkeintaan numeroituva määrä ei-negatiivisia termejä $f(x_i)$, jossa $x_i \in S \cap B$, minkä takia kyseinen summamerkintä on mielekäs. Kohdan (b) kaava seuraa kokonaistodennäköisyyden kaavasta (tai mahdollisesti sen versiosta numeroituvasti äärettömällä ositukselle), kun perusjoukolle Ω käytetään ositusta

$$\{X \in S^c\}, \{X = x_1\}, \{X = x_2\}, \dots$$

□

Lause 2.5. *Olkoon $f : \mathbb{R} \rightarrow \mathbb{R}$ ei-negatiivinen funktio, joka saa positiivisia arvoja korkeintaan numeroituvassa \mathbb{R} :n osajoukossa siten, että*

$$\sum_x f(x) = 1$$

Tällöin se on jonkin diskreetin satunnaismuuttujan ptnf.

Todistus. Ideana on se, että ensin konstruoidaan pistetodennäköisyyksistä vastaava kertymäfunktioehdokas, ja sitten tarkistetaan, että sillä on kertymäfunktion ominaisuudet. □

2.4 Esimerkkejä diskreeteistä jakaumista

Diskreetti tasajakauma. Satunnaismuuttujalla X on diskreetti tasajakauma joukossa $\{1, \dots, N\}$, mikäli sen ptnf on

$$f(x) = \frac{1}{N}, \quad \text{kun } x = 1, 2, \dots, N.$$

Esimerkiksi, jos X on silmäluku nopanheitossa, niin X :llä on diskreetti tasajakauma joukossa $1, \dots, 6$.

Bernoullin jakauma. Olkoon A on tapahtuma, ja merkitään $p = P(A)$, jolloin $0 \leq p \leq 1$. Tällöin indikaattorilla $X = 1_A$ on Bernoullin jakauma parametrilla p (eli X noudattaa Bernoullin jakaumaa parametrilla p). Tämä asia voidaan merkitä

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

X :n arvojoukko on $\{0, 1\}$, ja sen ptnf on

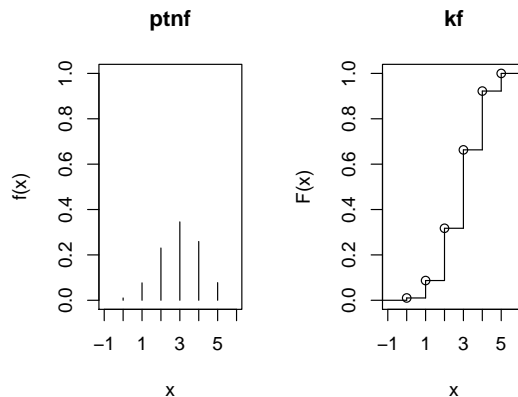
$$f(x) = f(x | p) = \begin{cases} 1 - p, & \text{kun } x = 0, \\ p, & \text{kun } x = 1, \\ 0, & \text{muuten.} \end{cases}$$

Tämä ptnf voidaan kirjoittaa lyhyemmin

$$f(x | p) = p^x(1 - p)^{1-x}, \quad \text{kun } x = 0, 1,$$

kun sovelletaan edellä mainittua merkintäsopimusta, jonka mukaan ptnf:n arvot tarvitsee kertoa vain kyseisen satunnaismuuttujan arvojoukossa. Edellä käytetään tulkintaa $0^0 = 1$, mikäli $p = 0$ tai $p = 1$.

Tavallisesti ajatellaan niin, että kun sattuu $A = \{X = 1\}$, niin satunnaiskoe onnistuu ja muussa tapauksessa epäonnistuu. Tällöin p on yhtä kuin kokeen onnistumistodennäköisyys.

Kuva 2.1 Binomijakauman $\text{Bin}(n, p)$ ptnf ja kf, kun $n = 5$ ja $p = 0.6$.

Huomautus. Useimmat mielenkiintoiset jakaumat riippuvat yhden tai useamman parametrin arvosta. Pitääksemme lukua parametrin arvosta, merkitsemme sen tarvittaessa pystyviivan oikealle puolelle. Tätä merkintää voidaan käyttää kertymäfunktioille, pistetodennäköisyysfunktioille ynnä muille funktioille. Jos taas sekaantumisen vaaraa ei ole, parametri voidaan jättää pois merkinnöistä.

Binomijakauma. Olkoon $n \geq 1$ kokonaisluku ja $0 \leq p \leq 1$. Sm X noudattaa binomijakaumaa parametreillä n ja p , jos sen arvojoukko on $\{0, 1, \dots, n\}$ ja ptnf on

$$f(x) = f(x | n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{kun } x = 0, 1, \dots, n.$$

Tämä voidaan merkitä

$$X \sim \text{Bin}(n, p).$$

Binomijakaumaa $\text{Bin}(n, p)$ noudattava sm syntyy toistokokeessa, jossa toistetaan riippumattomasti n kertaa sellaista (Bernoullin) koetta, jossa yhdessä kokeessa onnistumistodennäköisyys on p . Jos X on onnistumisten lukumäärä n toistossa, niin $X \sim \text{Bin}(n, p)$. Tämä tulos perustellaan esimerkissä 2.6. Bernoullin jakauma Bernoulli(p) on tietenkin sama kuin binomijakauma $\text{Bin}(1, p)$.

2.5 Jatkuva jakauma

Määritelmä 2.8. Satunnaismuuttujalla X on jatkuva jakauma tiheysfunktioilla f (lyh. tf), jos f on ei-negatiivinen ja

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

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

Tiheysfunktio on englanniksi (*probability density function*, *pdf*). Toisinaan puhutaan tiheysfunktion sijasta lyhyesti tiheydestä.

Se, että sm:lla X on jatkuva jakauma voidaan vaihtoehtoisesti ilmaista myös jollakin seuraavista tavoista.

- X :n jakauma on jatkuva,
- X on jatkuvasti jakautunut,
- X on jatkuva sm.

Sanontapa “ X on jatkuva sm” on yleisesti käytössä, vaikka se on harhaanjohdettava. Asiaa tuntematon henkilö voi nimittäin sekoittaa sen siihen aivan erisisältöiseen (ja tässä yhteydessä hyvin vähäisesti kiinnostavaan) väitteeseen, että funktio $X : \Omega \rightarrow \mathbb{R}$ olisi jatkuva.

Merkintäsopimus:

- Jatkuvan jakauman tiheysfunktiolle käytetään muuten samaa symbolia kuin sen kf:lle, mutta tiheysfunktion symboli kirjoitetaan pienellä kirjaimella (esim. f , g tai f_X) ja kf:n symboli vastaavalla isolla kirjaimella (esim. F , G tai F_X).
- Joskus tiheysfunktio häviää jonkin joukon $B \subset \mathbb{R}$ ulkopuolella. Tällöin tiheysfunktion kaava voidaan ilmoittaa vain joukossa B , ja asiayhteydestä pitää ymmärtää, että $f(x) = 0$, kun $x \notin B$.

Jatkuvan jakauman määritelmässä esiintyvä integraali tarkoittaa itseasiassa Lebesguen integraalia, mutta sovelluksissa esiintyvät tiheysfunktiot ovat vähintään paloittain jatkuvia funktioita, ja niiden integraalit voidaan käytännössä laskea Riemannin integraalien avulla. Todennäköisyyslaskennassa integrointialue ei välttämättä ole äärellinen väli. Näiden asioiden takia integrointialue joudutaan usein ensin pilkkomaan paloihin, minkä jälkeen tiheysfunktion integraali saadaan palautettua summaksi (mahdollisesti epäoleellisia) Riemannin integraaleja.

Jakauman tiheysfunktio ei ole yksikäsitteinen, mutta melkein yksikäsitteinen. Jos f on tietyn jakauman tf ja g on yhtä kuin f muualla paitsi äärellisessä määrässä pisteitä, niin

$$\int_a^b f(x) dx = \int_a^b g(x) dx$$

kaikilla $a < b$. Tämän takia myös g on kyseisen jakauman tf. Täsmällinen tulos on seuraava. Ei-negatiiviset funktiot f ja g ovat saman jakauman tiheysfunktioita, jos ja vain jos ne yhtyvät melkein kaikkialla, eli joukko jossa $f \neq g$ on nollamittainen (mikä käsite määritellään reaalianalyysin kursseilla).

Jos X :n jakauma on jatkuva ja sen tf on f , niin

$$P(X = x) = P(x \leq X \leq x) = \int_x^x f(u) du = 0$$

kaikilla x , joten pistetodennäköisyys on aina nolla. Tämän takia ptmf ei ole hyödyllinen käsite jatkuville jakaumille.

Koska kaikki pistetodennäköisyydet ovat nollia, niin

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(x) dx,$$

kun $a < b$. Rajankäynnin jälkeen huomataan, että puoliäärettömien välien todennäköisyydet saadaan myöskin tf:a integroimalla. Kaikille x pätee

$$P(X \leq x) = P(X < x) = \int_{-\infty}^x f(u) du, \quad P(X > x) = P(X \geq x) = \int_x^{\infty} f(u) du.$$

Itse asiassa pätee vielä yleisempi tulos

$$P(X \in B) = \int_B f(x) dx, \quad \text{kaikille } B \subset \mathbb{R}, \quad (2.2)$$

jossa integraalin alaindeksi B tarkoittaa joukon B yli laskettua integraalia, ts.

$$\int_B f(x) dx = \int_{-\infty}^{\infty} 1_B(x) f(x) dx.$$

Ominaisuus (2.2) oltaisiin voitu alunperin valita tiheysfunktion (ja jatkuvan jakauman) määritelmäksi.

Erityisesti

$$1 = P(\Omega) = P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x) dx,$$

joten f :n integraali koko reaaliakselin yli on yksi.

Lause 2.6 (Kertymäfunktion ja tiheysfunktion yhteys). *Olkoon X :llä jatkuva jakauma.*

(a) *Jos X :n tiheysfunktio on f , niin sen kertymäfunktio F on*

$$F(x) = \int_{-\infty}^x f(u) du.$$

Jos x on tiheysfunktion f jatkuvuus piste, niin $F'(x) = f(x)$.

(b) *Jos X :n kertymäfunktio on F , niin F on derivoituva melkein kaikkialla, ja X :n tiheysfunktiksi voidaan valita F' .*

Todistus. Kertymäfunktion a-kohdassa annettu lauseke tuli jo perusteltua. Kertymäfunktion derivoituvuus f :n jatkuvuus pisteissä seuraa analyysin peruslauseesta. Kohdan (b) väite (tässä yleisyydessä) seuraa mitta- ja integrointiteoriasta. \square

Huomautus. Kun edellä sanotaan, että jatkuvan jakauman tiheysfunktiksi voidaan valita $f = F'$, niin tällä tarkoitetaan sitä, että tiheysfunktiksi valitaan jokin funktio f , joka yhtyy funktion F derivaattaan niissä pisteissä, joissa derivaatta on määritelty, ja muissa pisteissä f saadaan määritellä mielivaltaisesti.

Lause 2.7 (Riittävä ehto sille, että annettu kf on jatkuvan jakauman kf). *Olkoon F sellainen kertymäfunktio, joka on jatkuva koko reaaliakselilla ja joka on jatkuvasti derivoituva kaikkialla muualla, paitsi äärellisessä määrässä pisteitä. Tällöin F on jatkuvan jakauman kertymäfunktio, ja jakauman tiheysfunktiksi voidaan valita sen derivaatta $f = F'$.*

Todistus. (Hahmotelma) Jos $[a, b]$ on sellainen äärellinen väli, joka ei sisällä yhtään pistettä, jossa F ei ole jatkuvasti derivoituva, niin analyysin peruslauseen mukaan

$$F(b) - F(a) = \int_a^b f(u) du,$$

jossa f on kf:n F derivaatta. Tämä yhtälö todistetaan seuraavaksi mielivaltaisille $a \leq b$ (äärettömät mukaan lukien) ottamalla raja-arvoja sekä käyttämällä niitä lauseen oletuksia, joiden mukaan $F(-\infty) = 0$, $F(+\infty) = 1$ ja F :llä ei ole hyppyä myöskään niissä pisteissä, joissa se ei ole jatkuvasti derivoituva. \square

Täydentäviä huomautuksia.

- Jos X :llä on jatkuva jakauma, niin sen kf on jatkuva. Sen sijaan tf:n ei tarvitse olla jatkuva eikä edes paloittain jatkuva funktio, vaikka tilastollisissa sovelluksissa esiintyvät tiheysfunktioit ovat paloittain jatkuvia. Tiheysfunktion ei myöskään tarvitse olla rajoitettu funktio.
- Täydellinen karakterisointi niille kertymäfunktioille, jotka ovat jonkin jatkuvan jakauman kertymäfunktioita on se, että kf F on jatkuvan jakauman kf täsmälleen silloin, kuin F on *absoluuttisesti jatkuva*, mutta tämän asian määrittäminen jätetään reaalianalyysin kurssien huoleksi.
- Siitä, että jakauman kf on jatkuva ei vielä seuraa, että jakauma olisi jatkuva. Tätä varten kf:n pitää siis olla absoluuttisesti jatkuva. Täsmällisempi nimitys jatkuvalla jakaumalla olisi *absoluuttisesti jatkuva jakauma*, mutta tätä nimitystä käytetään harvoin.

Lause 2.8. Jos $f : \mathbb{R} \rightarrow \mathbb{R}$ on ei-negatiivinen, ja sen integraali koko reaaliakselin yli on yksi, niin se on jonkin satunnaismuuttujan tiheysfunktio.

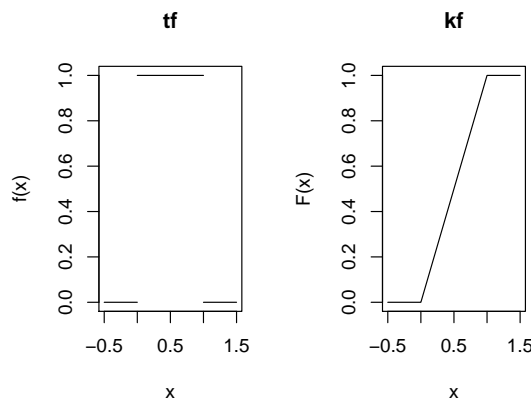
Todistus. Ideana on konstruoida kertymäfunktioehdokas integroimalla funktiota f , ja tarkistaa, että konstruoidulla funktiolla on kertymäfunktion ominaisuudet. \square

Huomautus. Jos $g \geq 0$ on sellainen reaaliakselilla määritelty funktio, jonka integraali $c = \int_{-\infty}^{\infty} g(x) dx$ on äärellinen, niin tällöin $f = g/c$ on tiheysfunktio. Tällä tavalla jokainen *normalisoimaton tiheysfunktio* g määrää (edellisen lauseen perusteella) yksikäsitteisellä tavalla jatkuvan jakauman, jonka tiheysfunktio on tällä tavalla konstruoitu f .

Esimerkki 2.3. Tiheysfunktion frekvenssitulkinta. Olkoon sm:lla X tiheysfunktio f . Oletetaan, että meillä on käytettävissä suuri aineisto x_1, x_2, \dots, x_N , jota voidaan pitää N riippumattoman sm:n X :n jakaumaa noudattavan sm:n havaittuina arvoina. Toisin sanoen

$$x_1 = X_1(\omega^{\text{act}}), x_2 = X_2(\omega^{\text{act}}), \dots, x_N = X_N(\omega^{\text{act}}),$$

jossa ω^{act} on aktualisoitunut alkeistapaus, kullakin X_i on sama jakauma kuin X :llä ja lisäksi X_1, \dots, X_N ovat riippumattomia (mikä määritellään myöhemmin). Kuinka tällaisessa tilanteessa voitaisiin arvioida tiheysfunktion arvoa pisteessä u ? Oletetaan, että u on f :n jatkuvuus piste.

Kuva 2.2 Tasajakauman $U(0, 1)$ tiheys- ja kertymäfunktio.

Tällaisessa tilanteessa voitaisiin menetellä seuraavasti. Valitaan jokin pieni luku $h > 0$ ja lasketaan, kuinka moni otospisteistä x_i osuu h :n pituiselle välille

$$\left[u - \frac{h}{2}, u + \frac{h}{2}\right].$$

Olkoon tämä lukumäärä $N_{u,h}$. Tiheysfunktion määritelmän nojalla ja todennäköisyyden frekvenssitulkinnan nojalla

$$P\left(u - \frac{h}{2} \leq X \leq u + \frac{h}{2}\right) = \int_{u-h/2}^{u+h/2} f(x) dx \approx \frac{N_{u,h}}{N}.$$

Toisaalta, integraalilaskennan väliarvolauseen nojalla

$$\int_{u-h/2}^{u+h/2} f(x) dx \approx h f(u).$$

Kun nämä arviot yhdistetään, saadaan arvio

$$f(u) \approx \frac{N_{u,h}}{hN}.$$

Tietenkin tässä voitaisiin käyttää myös muita h :n pituisia välejä, jotka sisältävät pisteen u kuin tätä, jonka keskipiste on u . Sivuutamme nyt sen ongelman, kuinka arvo h kannattaisi valita, jotta arviosta saataisiin mahdollisimman tarkka.

Tällaista ns. parametrifontia tiheysfunktion estimointia varten löytyy kirjallisuudesta monia erilaisia menetelmiä, joista eräät ovat perusajatukseltaan todellakin näin yksinkertaisia. \triangle

2.6 Esimerkkejä jatkuvista jakaumista

Tasajakauma (engl. *uniform distribution*). Olkoon $a, b \in \mathbb{R}$ ja $a < b$. Satunnaismuuttujalla X on välin (a, b) tasajakauma, $X \sim U(a, b)$, jos sillä on

jatkuva jakauma tiheysfunktiona f , jossa

$$f(x) = f(x | a, b) = \begin{cases} \frac{1}{b-a}, & \text{kun } a < x < b, \\ 0, & \text{muuten.} \end{cases}$$

Huomaa, että tämä tiheysfunktio on vain paloittain jatkuva; se on epäjatkua pisteissä $x = a$ ja $x = b$ ja jatkuva muualla. Aiemmin mainitun merkintäsopimuksen mukaan tämä tiheysfunktio voidaan ilmaista myös kaavalla

$$f(x) = \frac{1}{b-a}, \quad \text{kun } a < x < b.$$

Tällöin pitää ymmärtää asiayhteydestä, että $f(x) = 0$, kun $x \leq a$ tai $x \geq b$.

Tasajakauman kertymäfunktio on

$$F(x) = \int_{-\infty}^x f(u) \, du = \begin{cases} 0, & \text{kun } x \leq a, \\ \frac{x-a}{b-a}, & \text{kun } a < x < b, \\ 1, & \text{kun } x \geq b. \end{cases}$$

Eksponttijakauma (engl. *exponential distribution*). Satunnaismuuttujalla X on eksponenttijakauma parametrilla $\lambda > 0$, $X \sim \text{Exp}(\lambda)$, jos X :llä on tf

$$f(x) = f(x | \lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{kun } x > 0, \\ 0, & \text{kun } x \leq 0. \end{cases}$$

Tiheysfunktio voidaan ilmoittaa lyhyemmin kaavalla

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

Tarkistetaan varmuuden vuoksi, että tämä funktio f todellakin on tiheysfunktio. Ensiksi $f \geq 0$, ja toiseksi

$$\int_{-\infty}^{\infty} f(x) \, dx = \int_0^{\infty} \lambda e^{-\lambda x} \, dx = \left|_0^{\infty} (-e^{-\lambda x}) \right| = 1,$$

joten f on tf.

Kun $x > 0$, niin eksponenttijakauman kf:lla on arvo

$$F(x) = \left|_0^x (-e^{-\lambda u}) \right| = 1 - e^{-\lambda x},$$

ja $F(x) = 0$, kun $x \leq 0$.

2.7 Satunnaismuuttujan muunnos

Jos X on sm, ja $g : \mathbb{R} \rightarrow \mathbb{R}$ on funktio, niin myös $Y = g(X)$ on sm. Merkintä $Y = g(X)$ tarkoittaa sitä, että

$$Y(\omega) = g(X(\omega)), \quad \omega \in \Omega,$$

ts. Y on yhdistetty funktio $g \circ X$. Mitä voimme sanoa Y :n jakaumasta? Diskreetille jakaumalle muunnoksen jakauma on helppo karakterisoida, mutta jatkuvan tapauksen käsittely on monimutkaisempaa.

Jos $g : \mathbb{R} \rightarrow \mathbb{R}$ on funktio ja $A \subset \mathbb{R}$, niin merkintä $g^{-1}(A)$ tarkoittaa joukon A alkukuvaa kuvauksessa g , eli

$$g^{-1}(A) = \{x \in \mathbb{R} : g(x) \in A\}.$$

Huomaa, että

$$g(x) \in A \iff x \in g^{-1}(A).$$

Lause 2.9. *Olkoon X diskreetti sm ptnf:lla f_X , ja olkoon $g : \mathbb{R} \rightarrow \mathbb{R}$ funktio. Tällöin sm $Y = g(X)$ diskreetti, ja sen ptnf f_Y on*

$$f_Y(y) = \sum_{x \in g^{-1}(\{y\})} f_X(x).$$

Todistus. Koska X :n arvojoukko on korkeintaan numeroituvasti ääretön, voi Y saada korkeintaan numeroituvasti äärettömän määrän eri arvoja. Siksi Y on diskreetti. Mille tahansa $y \in \mathbb{R}$ on

$$\begin{aligned} P(Y = y) &= P(g(X) = y) = P(g(X) \in \{y\}) = P(X \in g^{-1}(\{y\})) \\ &= \sum_{x \in g^{-1}(\{y\})} f_X(x), \end{aligned}$$

jossa viimeinen vaihe seuraa täysadditiivisuudesta. □

Huomautus. Jos yhtälöllä $y = g(x)$ on yksikäsitteinen ratkaisu $x = h(y)$ kaikilla y , jotka kuuluvat sm:n Y arvojoukkoon S_Y , niin

$$f_Y(y) = f_X(h(y)), \quad y \in S_Y. \quad (2.3)$$

Jos taas ratkaisuja on useita, niin jokainen erisuuri ratkaisu antaa ptnf:n arvoon pisteessä y yhden tällaisen termin lisää.

Jos X :llä on jatkuva jakauma, niin sen muunnoksen $Y = g(X)$ jakauma voi funktion g luonteesta riippuen joko diskreetti, jatkuva tai ei kumpaakaan näistä. Esimerkiksi, jos g on paloittain vakio funktio, niin $Y = g(X)$ on diskreetti satunnaismuuttuja. Myöhemmin tarkastelemme tapausta, jossa satunnaismuuttujalla $Y = g(X)$ on jatkuva jakauma. On myös helppo konstruoida esimerkkejä, joissa muunnoksella $Y = g(X)$ ei ole diskreetti eikä jatkuva jakauma, vaan ns. sekatyypin jakauma.

Esimerkki 2.4. Sensuroinnin kautta syntyvä sekatyypin jakauma. Olkoon X :llä välin $[0, 4]$ tasajakauma $U(0, 4)$. Määritellään, että $Y = g(X)$, jossa

$$g(x) = \min(\max(1, x), 2).$$

Tällöin voidaan sanoa, että Y saadaan satunnaismuuttujasta X sensuroimalla (tai rajaamalla) sitä vasemmalta pisteessä 1 ja oikealta pisteessä 2. Helpolla laskulla nähdään, että Y :n kf voidaan esittää konveksina kombinaationa tietyn diskreetin jakauman kf:sta sekä tietyn jatkuvan jakauman kf:sta, nimittäin

$$F_Y = \frac{3}{4} F_d + \frac{1}{4} F_c.$$

Tässä F_d on seuraava diskreetin jakauman kf,

$$F_d(x) = \begin{cases} 0 & \text{kun } x < 1, \\ \frac{1}{3} & \text{kun } 1 \leq x < 2, \\ \frac{2}{3} & \text{kun } x \geq 2, \end{cases}$$

ja F_c on välin $[1, 2]$ tasajakauman kf. (Konvekssi kombinaatio on sellainen lineaarikombinaatio, jossa painot [edellä $\frac{3}{4}$ ja $\frac{1}{4}$] ovat ei-negatiivisia ja niiden summa on yksi.) Tällöin sanotaan, että Y :llä on sekatyypin jakauma.

Sensuroituja jakaumia esiintyy käytännössä esim. elinaika-analyysissa. \triangle

2.8 Kvantiilifunktio ja sen käyttö simuloinnissa

Tarkastellaan nyt sellaista jatkuvaa kertymäfunktiota F , joka on aidosti kasvava välillä (a, b) , ja lisäksi $F(a) = 0$ ja $F(b) = 1$. Sallimme tapaukset $a = -\infty$ tai $b = \infty$ tai molemmat. Tällöin F :llä on olemassa käänteisfunktio $F^{-1} : (0, 1) \rightarrow (a, b)$. (Tarkemmin sanoen F :n rajoittumalla joukkon (a, b) on käänteisfunktio F^{-1} .) Tässä yhteydessä merkintä F^{-1} ei siis tarkoita funktiota $1/F$, vaan kertymäfunktion käänteisfunktiota. Se toteuttaa identiteetit

$$F^{-1}(F(x)) = x, \quad \text{kaikilla } a < x < b,$$

ja

$$F(F^{-1}(u)) = u, \quad \text{kaikilla } 0 < u < 1.$$

Tällaista kertymäfunktion käänteisfunktiota F^{-1} kutsutaan kyseisen jakauman *kvantiilifunktioksi* (tai *fraktiilifunktioksi*).

Kvantiilifunktion tärkeys johtuu siitä, että osuus $0 < u < 1$ jakauman todennäköisyysmassasta jää pisteen $F^{-1}(u)$ vasemmalle puolelle. Toisin sanoen, jos sm:lla X on kf F , niin

$$P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u, \quad \text{mielivaltaiselle } 0 < u < 1.$$

Olkoon X sm, jonka kf:lla F on edellä mainitut ominaisuudet. Tarkastellaan satunnaismuuttujaa $F(X)$, joka on yhdistetty funktio

$$\omega \mapsto F(X(\omega)).$$

Palautetaan mieleen se seikka, että koska F on aidosti kasvava funktio, niin seurauksena myös F^{-1} on aidosti kasvava funktio. Jos $0 < u < 1$, niin

$$\begin{aligned} P[F(X) \leq u] &= P[F^{-1}(F(X)) \leq F^{-1}(u)] = P[X \leq F^{-1}(u)] \\ &= F(F^{-1}(u)) = u. \end{aligned}$$

Tämä tarkoittaa sitä, että satunnaismuuttujalla $F(X)$ on tuttu jakauma,

$$F(X) \sim U(0, 1). \quad (2.4)$$

Satunnaismuuttujaan $F(X)$, ja tulokseen $F(X) \sim U(0, 1)$ viitataan toisinaan nimellä *kertymäfunktio muunnos* (engl. *probability integral transform*). Mainitaan ilman perusteluja, että tulos (2.4) pätee myös silloin, kun F on mielivaltainen

jatkuva kertymäfunktio — jos F :llä on ainakin yksi hyppy, tulos ei tietenkään pidä paikkaansa.

Edellisen tuloksen voi kääntää. Olkoon $U \sim U(0, 1)$, olkoon F kf, jolla on tämän jakson alussa mainitut ominaisuudet, ja tarkastellaan satunnaismuuttujaa $Z = F^{-1}(U)$. Jos $a < x < b$, niin

$$\begin{aligned} P(Z \leq x) &= P[F^{-1}(U) \leq x] = P[F(F^{-1}(U)) \leq F(x)] \\ &= P[U \leq F(x)] = F(x) \end{aligned}$$

Siis

$$U \sim U(0, 1) \quad \Rightarrow \quad \text{sm:n } F^{-1}(U) \text{ kertymäfunktio on } F.$$

Tällöin sanotaan, että sm $Z = F^{-1}(U)$ on saatu *käänteisfunktio menetelmällä* (engl. (esim.) *inverse transform*).

Matemaattisissa ohjelmistoissa on yleensä satunnaislukugeneraattori, jonka tuottamia arvoja voidaan pitää riippumattomina jakaumaa $U(0, 1)$ noudattavien satunnaismuuttujien arvoina. Jos jakauman kvantiilifunktiolla on yksinkertainen lauseke, niin käänteisfunktio menetelmä antaa kätevän keinon simuloida kyseistä jakaumaa.

Esimerkki 2.5. Jakauman $\text{Exp}(1)$ simulointi käänteisfunktio menetelmällä. Jakauman kf on

$$F(x) = \begin{cases} 1 - e^{-x}, & \text{jos } x \geq 0 \\ 0, & \text{muuten,} \end{cases}$$

Kvantiilifunktion arvo $F^{-1}(u)$ pisteessä $0 < u < 1$ saadaan ratkaisemalla x yhtälöstä

$$F(x) = u \quad \Leftrightarrow \quad x = -\ln(1 - u).$$

Tämän takia jakaumaa $\text{Exp}(1)$ voidaan simuloida asettamalla $X = -\ln(1 - U)$, kun U on ensin generoitu jakaumasta $U(0, 1)$. \triangle

Määritelmä 2.9 (Jakauman kvantiili). Olkoon X satunnaismuuttuja, jonka kertymäfunktio F on tämän jakson alussa vaaditut ominaisuudet. Tällöin lukua

$$x_p = F^{-1}(p) \quad \Leftrightarrow \quad F(x_p) = p$$

kutsutaan X :n jakauman *p-kvantiiliksi* (eli *p*-fraktiiliksi).

Jakauman tietyille kvantiileille käytetään erikoisnimityksiä:

- mediaani on $\frac{1}{2}$ -kvantiili,
- alakvantiili on $\frac{1}{4}$ -kvantiili,
- yläkvantiili on $\frac{3}{4}$ -kvantiili.

Johdimme edellä käänteisfunktio menetelmän asettamalla F :lle oletuksia. Itse asiassa nämä oletukset ovat tarpeettomia, ja menetelmä yleistyy mille tahansa kertymäfunktioille ts. lauseen 2.1 kolme ominaisuutta (a), (b) ja (c) toteuttavalle funktioille. Tätä varten kvantiilifunktion määritelmää pitää ensin yleistää.

Määritelmä 2.10 (Kertymäfunktion yleistetty käänteisfunktio, kvantiilifunktio). Olkoon F mielivaltainen kf. Sen *yleistetty käänteisfunktio* määritellään kaavalla

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 < u < 1. \quad (2.5)$$

Sanomme, että yleistetty käänteisfunktio F^{-1} on kertymäfunktioita F vastaava *kvantiilifunktio*, tai että F^{-1} on kyseisen *jakauman kvantiilifunktio*.

Edellä merkintä $\inf B$ tarkoittaa joukon $B \subset \mathbb{R}$ suurinta alarajaa. Koska F on kasvava ja oikealta jatkuva, tässä määritelmässä esiintyvä joukko $\{x : F(x) \geq u\}$ on muotoa $[t, \infty)$ oleva rajoittamaton väli, jonka suurin alaraja on välin vasen päätepiste t .

Lause 2.10. *Jos F on mielivaltainen kf, ja $U \sim U(0,1)$, niin satunnaismuuttujalla $F^{-1}(U)$ on jakauma, jonka kertymäfunktio on F .*

Todistus. Lauseen erikoistapaus todistettiin jo. Yleisen tapauksen todistus si-
vuutetaan. \square

Edellä määrittelimme kvantiilin ainoastaan tämän jakson alussa käsitellyssä tilanteessa. Yleisemmässä tapauksessa kvantiili määritellään siten, että mikä tahansa luku x_p , joka toteuttaa ominaisuudet

$$P(X \leq x_p) \geq p \quad \text{ja} \quad P(X < x_p) \leq p$$

on X :n jakauman p -kvantiili. On helppo nähdä, että $F^{-1}(p)$ (jossa F^{-1} on nyt X :n kf:n yleistetty käänteisfunktio) on jakauman p -kvantileista pienin.

2.9 Satunnaismuuttujien riippumattomuus

Samalla tn-avaruudella voi tietenkin olla määriteltynä useampi kuin yksi sm. Määrittelemme myöhemmin usean sm:n yhteisjakauman, ja tarkastelemme systemaattisesti, miten sitä käsitellään. Tässä vaiheessa tyydymme tarkastelemaan, milloin sm:t ovat riippumattomia.

Määritelmä 2.11 (Kahden sm:n riippumattomuus). Satunnaismuuttujat X ja Y ovat riippumattomia, jos

$$P(X \in A \text{ ja } Y \in B) = P(X \in A) P(Y \in B), \quad \text{kaikilla } A, B \subset \mathbb{R}.$$

Tämä voidaan merkitä $X \perp Y$.

Merkintäsopimus (pilkku tapahtumien välissä tarkoittaa niiden leikkausta): Kun lasketaan tapahtumien leikkauksen todennäköisyyttä, niin tapahtumien väliin voidaan symbolin \cap sijasta merkitä pilkku, siis seuraavaan tapaan

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

Lause 2.11. *Seuraavat seikat ovat yhtäpitäviä,*

(a) $X \perp Y$

(b) $P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$ kaikilla $x, y \in \mathbb{R}$.

Todistus. **(a) \Rightarrow (b):** Valitaan $A = (-\infty, x]$ ja $B = (-\infty, y]$ ja sovelletaan riippumattomuuden määritelmää.

Implikaation $(b) \Rightarrow (a)$ todistaminen sivuutetaan, sillä todistus vaatisi mitateoriaa. \square

Huomautus. Lauseketta $P(X \leq x, Y \leq y)$ ymmärrettynä argumenttien x ja y funktiona kutsutaan satunnaismuuttujaparin (X, Y) (yhteis)kertymäfunktioiksi. Sille käytetään tavallisesti merkintää $F_{X,Y}(x, y)$. Jos sm:n X kf on F_X ja sm:n Y kf on F_Y , niin edellisessä lauseessa todetaan, että

$$X \perp\!\!\!\perp Y \iff F_{X,Y}(x, y) = F_X(x) F_Y(y) \quad \forall x, y.$$

Palaamme näihin asioihin myöhemmin.

Lause 2.12. *Jos X ja Y ovat diskreettejä satunnaismuuttujia, niin seuraavat seikat ovat yhtäpitäviä,*

(a) $X \perp\!\!\!\perp Y$

(b) $P(X = x, Y = y) = P(X = x) P(Y = y)$ kaikilla $x, y \in \mathbb{R}$.

Todistus. Implikaatio $(a) \Rightarrow (b)$ seuraa valitsemalla riippumattomuuden määritelmässä $A = \{x\}$ ja $B = \{y\}$.

Todistetaan implikaatio $(b) \Rightarrow (a)$ lähtemällä liikkeelle oletuksesta, että ominaisuus (b) on voimassa. Olkoot $A, B \subset \mathbb{R}$ mielivaltaisia. Olkoot sm X :n mahdolliset eri arvot x_1, x_2, \dots ja sm Y :n mahdolliset eri arvot y_1, y_2, \dots . Tällöin

$$\begin{aligned} P(X \in A, Y \in B) &= P\left(\bigcup_{\substack{x_i \in A \\ y_j \in B}} \{X = x_i, Y = y_j\}\right) \\ &= \sum_{x_i \in A, y_j \in B} P(X = x_i, Y = y_j) \quad (\text{täysadditiivisuus}) \\ &= \sum_{x_i \in A, y_j \in B} P(X = x_i) P(Y = y_j) \quad (\text{oletus}) \\ &= \sum_{x_i \in A} P(X = x_i) \sum_{y_j \in B} P(Y = y_j) \\ &= P(X \in A) P(Y \in B), \end{aligned}$$

joten X ja Y ovat riippumattomia. \square

Huomautus. Lauseketta $P(X = x, Y = y)$ ymmärrettynä argumenttien x ja y funktiona kutsutaan diskreetin satunnaismuuttujaparin (X, Y) (yhteis)pistetodennäköisyysfunktioiksi. Sille käytetään tavallisesti merkintää $f_{X,Y}(x, y)$. Jos sm:n X ptnf on f_X ja sm:n Y ptnf on f_Y , niin edellisessä lauseessa todetaan, että

$$X \perp\!\!\!\perp Y \iff f_{X,Y}(x, y) = f_X(x) f_Y(y) \quad \text{kaikilla } x, y.$$

Huomautus. Myöhemmin (luvuissa 6 ja 7) todistetaan, että jatkuvan yhteisjakauman tapauksessa satunnaismuuttujat ovat riippumattomia silloin ja vain silloin, kun niiden yhteistiheysfunktio voidaan esittää samaan tapaan reunajakaumien tiheysfunktioiden tulona.

Jos A ja B ovat tapahtumia, niin edellistä lausetta soveltamalla nähdään helposti, että tapahtumat ovat riippumattomia silloin ja vain silloin kuin niiden indikaattorit ovat riippumattomia sm:ia, eli

$$A \perp B \iff 1_A \perp 1_B.$$

Seuraavaksi näytetään, että riippumattomien satunnaisfunktioiden muunnokset ovat riippumattomia.

Lause 2.13. Jos $X \perp Y$, ja $g : \mathbb{R} \rightarrow \mathbb{R}$ sekä $h : \mathbb{R} \rightarrow \mathbb{R}$ ovat funktioita, niin myös $g(X) \perp h(Y)$.

Todistus. Olkoot $A, B \subset \mathbb{R}$ mielivaltaisia. Tällöin

$$\{g(X) \in A, h(Y) \in B\} = \{X \in g^{-1}(A), Y \in h^{-1}(B)\}.$$

Tässä $g^{-1}(A)$ on joukon A alkukuva kuvauksessa g , ja $h^{-1}(B)$ on joukon B alkukuva kuvauksessa h . Koska $X \perp Y$, on

$$\begin{aligned} P(g(X) \in A, h(Y) \in B) &= P(X \in g^{-1}(A), Y \in h^{-1}(B)) \\ &= P(X \in g^{-1}(A)) P(Y \in h^{-1}(B)) \\ &= P(g(X) \in A) P(h(Y) \in B). \end{aligned} \quad \square$$

Määritelmä 2.12 (Usean sm:n riippumattomuus). Satunnaismuuttujat X_1, \dots, X_n ovat riippumattomia, jos

$$P\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = \prod_{i=1}^n P(X_i \in B_i), \quad \text{kaikilla } B_1, \dots, B_n \subset \mathbb{R}.$$

Tämä voidaan merkitä $X_1, \dots, X_n \perp$.

Esimerkki 2.6. Binomijakauman synty toistetussa Bernoullin kokeessa. Olkoot Y_1, \dots, Y_n riippumattomia Bernoullin jakaumaa noudattavia satunnaismuuttujia, joilla on yhteinen onnistumistn $0 \leq p \leq 1$. Ts. kullakin niistä on pntf:na

$$f(y) = p^y (1-p)^{1-y}, \quad \text{kun } y = 0, 1.$$

Jos $Y_i = 1$ niin sanotaan, että onnistuttiin toistossa i , ja muuten sanotaan, että epäonnistuttiin toistossa i . Olkoon X onnistumisten lukumäärä n toistossa, eli

$$X = Y_1 + \dots + Y_n = \sum_{i=1}^n Y_i.$$

Mikä on X :n jakauma?

Jos (y_1, \dots, y_n) on jokin jono nollia ja ykkösiä, niin

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} = p^s (1-p)^{n-s},$$

jossa $s = \sum_{i=1}^n y_i$, eli s on ykkösten ja $n-s$ on nollien lukumäärä jonossa (y_1, \dots, y_n) .

Satunnaismuuttuja X saa arvon $0 \leq x \leq n$ jos ja vain jos satunnaismuuttujat Y_1, \dots, Y_n saavat sellaiset arvot y_1, \dots, y_n , että jonossa (y_1, \dots, y_n) on x ykköstä (ja $n - x$ nollaa). Tällaisia jonoja on $\binom{n}{x}$ kappaletta, ja niistä kunkin tn on $p^x (1 - p)^{n-x}$. Todennäköisyyden additiivisuuden nojalla

$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad \text{kun } x = 0, 1, \dots, n.$$

Ts. $X \sim \text{Bin}(n, p)$, eli X noudattaa binomijakaumaa otoskokoparametrilla n ja onnistumistodennäköisyydellä p . \triangle

Huomautus. Kun määrittelimme useamman tapahtuman riippumattomuuden, jouduimme tarkastelemaan kaikkia jonoja i_1, \dots, i_k jossa indeksit ovat keskenään erisuuria. Satunnaismuuttujien riippumattomuuden määritelmässä tätä ei tarvita. Tämä johtuu siitä, että voimme poistaa tarkastelusta sellaiset sm:t

$$X_j, \quad \text{joilla } j \notin \{i_1, \dots, i_k\}$$

valitsemalla niiden kohdalla $B_j = \mathbb{R}$.

Jos A_1, \dots, A_n ovat tapahtumia, niin niiden indikaattorit $1_{A_1}, \dots, 1_{A_n}$ ovat satunnaismuuttujia. Seuraavat kaksi seikkaa ovat yhtäpitäviä.

- Tapahtumat A_1, \dots, A_n ovat riippumattomia,
- satunnaismuuttujat $1_{A_1}, \dots, 1_{A_n}$ ovat riippumattomia.

Numeroituvasti äärettömän monen satunnaismuuttujan X_1, X_2, \dots riippumattomuus määritellään seuraavasti.

Määritelmä 2.13 (Äärettömän monen sm:n riippumattomuus). Satunnaismuuttujat X_1, X_2, \dots ovat riippumattomia, jos satunnaismuuttujat X_1, \dots, X_n ovat riippumattomia kaikilla $n \geq 2$.

2.10 Muunnetun satunnaismuuttujan tiheys

Tarkastelemaan tapausta, jossa X :llä on jatkuva jakauma, ja $g : \mathbb{R} \rightarrow \mathbb{R}$. Satunnaismuuttujalla $Y = g(X)$ ei tällöin välttämättä ole jatkuva jakauma, kuten olemme edellä nähneet.

Jos kuitenkin g on riittävän sileä ja aidosti monotoninen, tai jos sen määrittelyjoukko \mathbb{R} voidaan osittaa paloihin, joissa g on riittävän sileä ja aidosti monotoninen, niin osoittautuu, että tällöin Y :llä on jatkuva jakauma. Jakauman tiheysfunktion lauseke on helpointa johtaa laskemalla ensin sen kertymäfunktio ja sitten kertymäfunktion derivaatta. Tämän lisäksi pitää jollakin tavalla varmistaa, että Y :n jakauma todellakin on jatkuva.

Esimerkki 2.7. Olkoon X :llä jatkuva jakauma, jolla on jatkuva tiheysfunktio f_X . Tarkastellaan satunnaismuuttujaa $Y = e^X$. Selvästi $Y > 0$, joten $F_Y(y) = 0$, kun $y < 0$. Kun $y > 0$, on

$$F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \ln(y)) = F_X(\ln(y)),$$

joten arvaamme, että tiheysfunktioksi kelpaa

$$f_Y(y) = F_Y'(y) = f_X(\ln(y))/y, \quad \text{kun } y > 0,$$

ja $f_Y(y) = 0$, kun $y \leq 0$.

Tarkistetaan, että tn $P(a \leq Y \leq b)$ saadaan laskettua integroimalla tätä funktiota välin (a, b) yli, kun $0 < a < b$. Tällöin ollaan tarkistettu suoraan määritelmän perusteella, että Y :llä on jatkuva jakauma. Tarkistus on helppoa, sillä

$$\begin{aligned} P(a \leq Y \leq b) &= P(a \leq e^X \leq b) \\ &= P(\ln a \leq X \leq \ln b) \\ &= \int_{\ln a}^{\ln b} f_X(x) dx && (X\text{:n tf on } f_X) \\ &= \int_a^b f_X(\ln(y)) \frac{1}{y} dy && (\text{Sij. } y = e^x \Leftrightarrow x = \ln(y)) \end{aligned}$$

Vaihtoehtoinen tapa tarkistaa sm:n Y jakauman jatkuvuus olisi vedota lauseeseen 2.7. \triangle

Tämän esimerkin päättely on helppo yleistää seuraavassa lauseessa väitettyyn muotoon. Kerrataan kuitenkin ensin yksi käsite.

Määritelmä 2.14 (Diffeomorfismi). Kuvaus $g : A \rightarrow B$, jossa A ja B ovat d -ulotteisen avaruuden \mathbb{R}^d avoimia osajoukkoja, on diffeomorfismi, jos

- (a) g on bijektio joukkojen A ja B välillä
- (b) sekä g että sen käänteisfunktio $g^{-1} : B \rightarrow A$ ovat jatkuvasti derivoituvia.

Lause 2.14 (Muuttujanvaihtokaava tiheysfunktioille). *Olkoon sm:lla X jatkuva jakauma tf:lla f_X . Olkoon $g : A \rightarrow B$ diffeomorfismi, jossa $A, B \subset \mathbb{R}$ ovat avoimia välejä, ja olkoon*

$$P(X \in A) = 1.$$

Määritellään $h(y) = g^{-1}(y)$, kun $y \in B$. Tällöin sm:lla $Y = g(X)$ on jatkuva jakauma tiheysfunktioilla

$$f_Y(y) = \begin{cases} f_X(h(y)) |h'(y)|, & \text{kun } y \in B \\ 0, & \text{muuten.} \end{cases} \quad (2.6)$$

Huomautus. Tulos (2.6) voidaan tietenkin ilmaista lyhyemmin kaavalla

$$f_Y(y) = f_X(h(y)) |h'(y)|, \quad \text{kun } y \in B.$$

Vertaa jatkuvan jakauman tapauksen tulosta (2.6) vastaavan diskreetin tapauksen tulokseen (2.3). Vain jatkuvan jakauman tapauksessa tarvitaan käänteiskuvauksen derivaatan itseisarvoa.

Todistus. Koska g on bijektio avoimelta väliltä avoimelle välille, se on joko aidosti kasvava tai aidosti vähenevä funktio. Oletetaan ensin, että g on aidosti kasvava. Tällöin myös sen käänteisfunktio h on aidosti kasvava. Jos $a < b$ ja $a, b \in B$, niin

$$P(a \leq Y \leq b) = P(a \leq g(X) \leq b) = P(h(a) \leq X \leq h(b)) = \int_{h(a)}^{h(b)} f_X(x) dx.$$

Tekemällä integraalissa muuttujanvaihdos

$$y = g(x) \iff x = h(y)$$

saadaan

$$P(a \leq Y \leq b) = \int_a^b f_X(h(y)) h'(y) dy.$$

(Tässä kohdassa sovelletaan muuttujanvaihtoa Lebesguen integraaliin; lauseen oletuksilla käsiteltävät funktiot eivät välttämättä olisi Riemann-integroituvia.) Koska tämä pätee kaikille väleille $[a, b] \in B$ ja koska $P(Y \in B) = 1$, ja koska $h' > 0$, niin väite on todistettu siinä tapauksessa, että g on aidosti kasvava.

Jos taas g on aidosti vähenevä, niin $h = g^{-1}$ on myös aidosti vähenevä, joten siis $h' < 0$. Edellinen lasku pätee pienten muutosten jälkeen: epäyhtälön suunta vaihtuu, mikä vaihtaa integraalin merkin, mutta tämä kompensoidaan ottamalla h :n derivaatasta itseisarvo. \square

Tilastotieteilijä joutuu tuon tuosta laskemaan satunnaismuuttujan muunnoksen tiheysfunktion, eikä silloin aina ole kirjallisuutta käsillä. Jos g on diffeomorfismi, niin kaavan muistamista helpottaa seuraava muistisääntö. Yhtälön

$$f_X(x) |dx| = f_Y(y) |dy| \tag{2.7}$$

pitää säilyä bijektiivisessä muuttujanvaihdossa

$$y = g(x) \iff x = h(y).$$

Kun tiedosta (2.7) ratkaistaan $f_Y(y)$, saadaan

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = f_X(h(y)) |h'(y)|,$$

mutta tietenkin tämä kaava pätee vain g :n kuvajoukossa B . Yksinkertaisissa tapauksissa muunnoksen tf on tällä tavalla menetellen mahdollista johtaa jopa päässälaskulla.

Huomautus. Kun muunnoksen $Y = g(X)$ tiheysfunktio ilmoitetaan, on tärkeää paitsi kertoa kaava $f_X(h(y)) |h'(y)|$ myös ilmoittaa selkeästi kaavan pätevyysalue, eli se joukko B , jossa kyseinen kaava pätee.

Yhtälöstä (2.7) voidaan ratkaista $f_Y(y)$ myös toisella tavalla, nimittäin seuraavasti,

$$f_Y(y) = f_X(x) \frac{1}{\left| \frac{dy}{dx} \right|} = \frac{f_X(x)}{|g'(x)|} = \frac{f_X(h(y))}{|g'(h(y))|}.$$

Myös tämä kaava pitää paikkansa joukossa B , sillä kaava

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}$$

kertoo oikean yhteyden funktion ja sen käänteisfunktion derivaattojen välillä, ja aidosta monotonisuudesta seuraa se, että $|g'| > 0$ joukossa A ja $|h'| > 0$ joukossa B .

Joskus tiheysfunktion muuttujanvaihtokaavaa tarvitaan myös sellaisessa tilanteessa, jossa g ei ole monotoninen, mutta jossa sen määrittelyjoukko voidaan pilkkoa väleiksi, joille rajoitettuna g on aidosti monotoninen. Tällaiset tilanteet saadaan hoidettua samaan tapaan kuin seuraavassa esimerkissä.

Esimerkki 2.8. Olkoon X :llä jatkuva jakauma, jolla on jatkuva tiheysfunktio f_X . Tarkastellaan satunnaismuuttujaa $Y = X^2$, ja lasketaan sen kf. Selvästi $Y \geq 0$, joten $F_Y(y) = 0$, kun $y < 0$. Kun $y > 0$, on

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Derivaatan laskemisen jälkeen arvaamme, että Y :n tf on

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}, \quad \text{kun } y > 0,$$

ja $f_Y(y) = 0$, kun $y \leq 0$.

Tarkistetaan, että tn $P(a \leq Y \leq b)$ todella saadaan laskettua integroimalla tätä tiheysfunktioita välin (a, b) yli, kun $0 < a < b$. Lasku sujuu seuraavasti.

$$\begin{aligned} P(a \leq Y \leq b) &= P(a \leq X^2 \leq b) = P(\sqrt{a} \leq |X| \leq \sqrt{b}) \\ &= P(-\sqrt{b} \leq X \leq -\sqrt{a}) + P(\sqrt{a} \leq X \leq \sqrt{b}) \\ &= \int_{-\sqrt{b}}^{-\sqrt{a}} f_X(x) dx + \int_{\sqrt{a}}^{\sqrt{b}} f_X(x) dx \end{aligned}$$

Nyt ensimmäisessä integraalissa tehdään muuttujanvaihto

$$y = x^2 \quad \Longleftrightarrow \quad x = -\sqrt{y} \quad (\text{jossa } x < 0 \text{ ja } y > 0)$$

ja toisessa

$$y = x^2 \quad \Longleftrightarrow \quad x = \sqrt{y} \quad (\text{jossa } x, y > 0).$$

Tällöin saadaan

$$P(a \leq Y \leq b) = \int_a^b \left(f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} \right) dy$$

△

Yleistämällä tämän esimerkin päättely päädytään seuraavaan tulokseen.

Lause 2.15. *Olkoon sm :lla X jatkuva jakauma tiheysfunktioilla f_X ja olkoon $Y = g(X)$. Oletetaan, että \mathbb{R} voidaan osittaa joukkoihin A_0 ja A_1, A_2, \dots siten, että*

- $P(X \in A_0) = 0$,
- *kukin A_i on avoin väli, kun $i \geq 1$ ja välejä $A_i, i \geq 1$ on äärellinen tai numeroituvasti ääretön määrä*
- *kuvaus $g|_{A_i}$, eli g rajoitettuna joukkoon A_i , on diffeomorfisimi avoimelta väliltä A_i sen kuvajoukolle B_i kaikilla $i \geq 1$*

Koska kukin kuvaus $g|_{A_i}$ on kääntövä kun $i \geq 1$, on sillä käänteisfunktio, jota merkitään

$$h_i = (g|_{A_i})^{-1}, \quad h_i : B_i \rightarrow A_i.$$

Tällöin Y :llä on jatkuva jakauma, jonka tiheysfunktio on

$$\sum_{i \geq 1} 1_{B_i}(y) f_X(h_i(y)) |h_i'(y)|.$$

Tässä termi

$$1_{B_i}(y) f_X(h_i(y)) |h'_i(y)|$$

tulkitaan nolllaksi, jos $y \notin B_i$ (jolloin indikaattorin jälkeinen termi ei ole hyvin määritelty).

Todistus. Ideana on pilkkoa tapahtuma $\{a \leq Y \leq b\}$ erillisiin tapahtumiin

$$\{a \leq g(X) \leq b, X \in A_i\}, \quad i \geq 0.$$

Kun $i = 0$, tämän tapahtuman tn on nolla. Muilla i tapahtuman todennäköisyys esitetään integraalina, jossa tehdään muuttujanvaihdos. \square

Huomautus. Jos edellisen lauseen oletukset ovat voimassa, niin muunnoksen $Y = g(X)$ tiheysfunktio saadaan laskea myös johtamalla ensin sm:n Y kertymäfunktio F_Y ja sitten derivoimalla tämä funktio. Tämän lauseen avulla voidaan käsitellä monenlaisia muunnoksia, kuten vaikkapa $Y = |X|$ tai $Y = \sin(X)$, kun sm:lla X on jatkuva jakauma. Tiheysfunktion $f_Y(y)$ kaavaan saadaan kutakin yhtälön $y = g(x)$ ratkaisua kohti yksi termi, joka on samaa muotoa (2.6) kuin diffeomorfinen muunnoksen tapauksessa.

Esimerkki 2.9. Esimerkissä 2.8 käsiteltiin tapausta $Y = X^2$. Siinä yhteydessä tehtiin implisiittisesti valinnat

$$A_0 = \{0\}, \quad A_1 = (-\infty, 0), \quad A_2 = (\infty, 0),$$

jolloin

$$B_1 = (0, \infty), \quad B_2 = (0, \infty)$$

ja kuvaukset h_1 ja h_2 ovat

$$\begin{aligned} h_1(y) &= -\sqrt{y}, & y > 0 \\ h_2(y) &= \sqrt{y}, & y > 0. \end{aligned}$$

\triangle

Luku 3

Odotusarvo

Seuraavaksi kerrataan, miten odotusarvo määritellään diskreetissä ja jatkuvassa tapauksessa. Odotusarvolle käytetään englanninkielisessä kirjallisuudessa lukuisia nimityksiä, kuten seuraavia: *mean (value)*, *(mathematical) expectation*, *expected value*.

3.1 Diskreetin satunnaismuuttujan odotusarvo

Määritelmä 3.1 (Diskreetin sm:n odotusarvo). Jos X on diskreetti sm, jonka arvojoukko on $S = \{x_1, x_2, \dots\}$ ja ptnf on f , niin sen odotusarvo on reaaliluku

$$EX = E(X) = \sum_i x_i f(x_i) = \sum_x x f(x)$$

mikäli tämä sarja on itseisesti summautuva, eli mikäli $\sum_i |x_i| f(x_i) < \infty$. Muussa tapauksessa sanotaan että X :llä *ei ole odotusarvoa*. Se seikka, että X :llä on odotusarvo voidaan ilmaista myös sanomalla, että X on *integroituva* sm.

Edellä kysymys sarjan itseisestä summautumisesta on relevantti tietenkin vain silloin, kun summattavia arvoja on äärettömän monta. Palautetaan mieleen analyysistä se seikka, että jos sarja summautuu itseisesti, niin sille saadaan sama summa kaikilla summausjärjestyksillä. (Jos summa suppenee kohti jotakin reaalilukua, mutta ei suppene itseisesti, niin erilaisilla summausjärjestyksillä sarjalle voidaan saada erilaisia summia.) Itseisen suppenemisen takia edellisessä määritelmässä ei tarvitse kiinnittää tiettyä summausjärjестystä. Toinen tapaus, jossa sarjan summausjärjестyksellä ei ole väliä on se, jossa kaikki summan termit ovat ei-negatiivisia; tällöin sarjan summa voi olla jokin reaaliluku tai ∞ .

Jos satunnaismuuttujilla X ja Y on sama diskreetti jakauma, niin niillä on sama odotusarvo. Sen takia diskreetin jakauman odotusarvo voidaan määritellä lukuna EX , jossa X on mikä tahansa kyseistä jakaumaa noudattava sm.

Odotusarvolla on *fysikaalinen tulkinta* massajakauman painopisteenä. Jos massattoman sauvan (=reaaliakseli) pisteisiin x_i asetetaan massat p_i , niin sauva pysyy tasapainossa, mikäli sitä tuetaan alta päin pisteestä EX . Tämä odotusarvon tulkinta säilyy myös muunlaisille jakaumille.

Historiallisesti edellä esitettyyn määritelmään on päädytty odotusarvon *frekvenssitulkinnan* avulla. Ajatellaan diskreettiä satunnaismuuttujaa X , jonka

mahdolliset arvot ovat $\{x_i\}$. Konkreettisuuden vuoksi olkoon X sinun voittamasi rahasumma tietyn uhkapelin yhdessä toistossa. Jos toistat peliä riippumattomasti N kertaa (jossa N on hyvin suuri luku), niin voitat summan x_i osapuulleen $NP(X = x_i)$ kertaa, joten keskimääräinen voittonsi yhtä peliä kohti on

$$\frac{1}{N} \sum_i x_i NP(X = x_i) = \sum_i x_i f(x_i) = EX.$$

Tällainen odotusarvon frekvenssitulkinta perustuu ns. vahvaan suurten lukujen lakiin, joka on yksi todennäköisyysteorian kuuluisimmista tuloksista. Itse asiassa todennäköisyysteoriassa todistetaan, että tällainen riippumattomien ja samoin jakautuneitten satunnaismuuttujien keskiarvo suppenee (otoskoon N kasvaessa rajatta) kohti odotusarvoa jos ja vain jos odotusarvo on olemassa, eli diskreetissa tapauksessa silloin, kun $\sum_i |x_i| f(x_i) < \infty$.

Esimerkki 3.1. Vakion odotusarvo. Vakiota $a \in \mathbb{R}$ voidaan pitää diskreettina satunnaismuuttujana, joka saa aina arvon a . Vastaavaa jakaumaa sanotaan *degeneroituneeksi jakaumaksi*. Nyt

$$Ea = a \cdot 1 = a,$$

joten vakion odotusarvo on kyseinen vakio. \triangle

Esimerkki 3.2. Tapahtuman A indikaattorin odotusarvo (eli Bernoullin jakauman odotusarvo). Jos $p = P(A)$ ja $X = 1_A$, niin $X \sim \text{Bernoulli}(p)$, ja

$$EX = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Tapahtuman indikaattorin odotusarvo on sama kuin kyseisen tapahtuman todennäköisyys. \triangle

Esimerkki 3.3. Binomijakauman odotusarvo. Olkoon $X \sim \text{Bin}(n, p)$, ja asetetaan $q = 1 - p$. Odotusarvo EX saadaan laskettua suoraan määritelmän nojalla seuraavasti.

$$\begin{aligned} EX &= \sum_{x=0}^n xf(x) = \sum_{x=1}^n xf(x) = \sum_{x=1}^n x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\ &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} = np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-1-k} \\ &= np(p+q)^{n-1} = np \end{aligned}$$

Laskussa sovellettiin binomikaavaa (1.6). Laskemme binomijakauman odotusarvon myöhemmin helpommalla tavalla, jossa käytämme hyväksi odotusarvon lineaarisuutta. \triangle

Esimerkki 3.4. Diskreetti jakauma, jolla ei ole odotusarvoa. Tunnetusti sarja $\sum_{x=1}^{\infty} 1/x^2$ suppenee, ja sarja $\sum_{x=1}^{\infty} 1/x$ hajaantuu. Olkoon $c = \sum_{x=1}^{\infty} 1/x^2$, ja määritellään sm X siten, että sen arvojoukko on kokonaisluvut poislukien nolla, ja ptnf on

$$f(x) = P(X = x) = \frac{1}{2c} \frac{1}{x^2}, \quad x = \pm 1, \pm 2, \dots$$

Tällöin X :llä ei ole odotusarvoa. \triangle

Jos diskreetin sm:n X arvojoukko on $\{x_1, x_2, \dots\}$, niin X voidaan esittää erillisten tapahtumien $A_i = \{X = x_i\}$ indikaattorien avulla summana

$$X = \sum_{i \geq 1} x_i 1_{A_i}.$$

Mikäli X :llä on odotusarvo, niin se saadaan summasta

$$EX = \sum_{i \geq 1} x_i P(A_i).$$

Joskus on kätevää käyttää jotakin muuta perusjoukon Ω ositusta B_1, B_2, \dots . Jos X on vakio kussakin osituksen palassa B_i , niin seuraava lause totetaa, että edellinen kaava pitää yhä paikkansa.

Lause 3.1. *Jos B_1, B_2, \dots on perusjoukon (äärellinen tai numeroituvasti ääretön) ositus, ja X on integroitava sm, joka saa vakioarvon y_j kullakin palalla B_j , eli*

$$X = \sum_{j \geq 1} y_j 1_{B_j},$$

niin

$$EX = \sum_{j \geq 1} y_j P(B_j).$$

Todistus. Selvästi X on diskreetti sm. Olkoot sen eri arvot x_1, x_2, \dots , ja määritellään

$$A_i = \{X = x_i\}, \quad i \geq 1.$$

Kun kukin A_i ositetaan tapahtumilla B_j , niin nähdään, että

$$\begin{aligned} EX &= \sum_{i \geq 1} x_i P(A_i) \\ &= \sum_{i \geq 1} x_i \sum_{j \geq 1} P(A_i \cap B_j) = \sum_{i \geq 1} \sum_{j \geq 1} x_i P(A_i \cap B_j) \end{aligned}$$

Mikäli $A_i \cap B_j \neq \emptyset$, niin edellä $x_i = y_j$, sillä X saa arvon x_i joukossa A_i ja arvon y_j joukossa B_j . Muussa tapauksessa summattava termi on nolla. Kaikissa tapauksissa

$$x_i P(A_i \cap B_j) = y_j P(A_i \cap B_j), \quad \text{kaikilla } i, j \geq 1.$$

Näin ollen

$$\begin{aligned} EX &= \sum_{i \geq 1} \sum_{j \geq 1} y_j P(A_i \cap B_j) \\ &= \sum_{j \geq 1} y_j \sum_{i \geq 1} P(A_i \cap B_j) = \sum_{j \geq 1} y_j P(B_j), \end{aligned}$$

sillä tapahtumat A_1, A_2, \dots osittavat kunkin tapahtuman B_j . □

Lause 3.2. *Olkoot X ja Y diskreettejä satunnaismuuttujia, joiden odotusarvot ovat olemassa. Odotusarvolla on seuraavat ominaisuudet.*

- (a) (Positiivisuus) Jos $X \geq 0$, niin $EX \geq 0$.
- (b) Jos $X \geq 0$ ja $EX = 0$, niin X on vakio nolla, ts. $P(X = 0) = 1$.
- (c) Jos $X \leq Y$, niin $EX \leq EY$.
- (d) (Vakion odotusarvo) Jos $a \in \mathbb{R}$, niin $Ea = a$.
- (e) (Lineaarisuus) Jos $a, b \in \mathbb{R}$, niin $E(aX + bY) = aEX + bEY$.
- (f) Jos $X \perp Y$, niin satunnaismuuttujalla XY on odotusarvo, ja se saadaan kaavalla

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

- (g) EX riippuu vain X :n jakaumasta: jos X :llä ja Y :llä on sama jakauma, niin $EX = EY$.

Todistus. Kohdat (a), (b), ja (g) ovat ilmeisiä. Vakion odotusarvo (kohta (d)) on jo laskettu.

Todistetaan lineaarisuus (kohta (e)) sekä riippumattomien satunnaismuuttujien tulon odotusarvo koskeva väite (kohta (f)). Olkoon X :n eri arvot x_1, x_2, \dots ja Y :n eri arvot y_1, y_2, \dots . Määritellään

$$A_i = \{X = x_i\}, \quad B_j = \{Y = y_j\}.$$

Tällöin joukot $A_i \cap B_j, i, j \geq 1$ osittavat perusjoukon Ω , ja X :llä on arvo x_i ja Y :llä arvo y_j joukossa $A_i \cap B_j$, mikäli kyseinen joukko on epätyhjä. Muussa tapauksessa tämän joukon tn on nolla.

Jos $a, b \in \mathbb{R}$ ovat vakioita, niin lauseen 3.1 perusteella

$$\begin{aligned} E(aX + bY) &= \sum_{i,j} (ax_i + by_j)P(A_i \cap B_j) \\ &= a \sum_{i,j} x_i P(A_i \cap B_j) + b \sum_{i,j} y_j P(A_i \cap B_j) \\ &= a \sum_i x_i P(A_i) + b \sum_j y_j P(B_j) = aEX + bEY. \end{aligned}$$

Tämä todistaa lineaarisuuden. Jos $X \perp Y$, ja X ja Y ovat integroituvia, niin

$$\begin{aligned} E(XY) &= \sum_{i,j} x_i y_j P(A_i \cap B_j) \\ &= \sum_{i,j} x_i y_j P(A_i)P(B_j) = \sum_{i,j} x_i P(A_i) y_j P(B_j) \\ &= \left[\sum_i x_i P(A_i) \right] \left[\sum_j y_j P(B_j) \right] = E(X)E(Y) \end{aligned}$$

Tämä todistaa kohdan (f).

Se seikka, että odotusarvo säilyttää järjestyksen (kohta (c)) seuraa lineaarisuudesta ja odotusarvon positiivisuudesta seuraavasti. Jos $X \leq Y$, niin $Y - X \geq 0$, joten

$$0 \leq E(Y - X) = EY - EX. \quad \square$$

Esimerkki 3.5. Binomijakauman odotusarvo uudestaan. Jos $X \sim \text{Bin}(n, p)$, niin se voidaan esittää summana

$$X = \sum_{i=1}^n Y_i,$$

jossa $Y_i \sim \text{Bernoulli}(p)$ ja Y_i :t ovat riippumattomia. Odotusarvon lineaarisuuden nojalla

$$EX = E \left[\sum_{i=1}^n Y_i \right] = \sum_{i=1}^n EY_i = np.$$

Tässä laskussa ei tarvittu satunnaismuuttujien Y_i riippumattomuutta. \triangle

3.2 Jatkuvasti jakautuneen satunnaismuuttujan odotusarvo

Määritelmä 3.2 (Odotusarvo jatkuvassa tapauksessa). Jos X on jatkuvasti jakautunut sm, ja sen tiheysfunktio on f , niin sen odotusarvo on luku

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

mikäli kyseinen integraali suppenee itseisesti, eli mikäli $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$. Muussa tapauksessa sanotaan että X :llä ei ole odotusarvoa. Se seikka, että X :llä on odotusarvo voidaan ilmaista myös sanomalla, että X on *integroituva* sm.

Esimerkki 3.6. Tasajakauman odotusarvo. Jos $a < b$ ja $X \sim U(a, b)$, niin helppo lasku osoittaa, että $EX = \frac{1}{2}(a + b)$. \triangle

Esimerkki 3.7. Cauchyn jakauma esimerkkinä jatkuvasta jakaumasta, jolla ei ole odotusarvoa. Sm X noudattaa (perusmuotoista) Cauchyn jakaumaa, jos sillä on tiheys

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

Tämä on tf, sillä $f \geq 0$ ja

$$\int_{-\infty}^{\infty} f(x) dx = \frac{2}{\pi} \int_0^{\infty} \frac{1}{1 + x^2} dx = \frac{2}{\pi} \Big|_0^{\infty} \arctan x = \frac{2}{\pi} \frac{\pi}{2} = 1$$

X :llä ei ole odotusarvoa, sillä

$$\int_{-\infty}^{\infty} |x| f(x) dx = \frac{1}{\pi} \int_0^{\infty} \frac{2x}{1 + x^2} dx = \frac{1}{\pi} \lim_{M \rightarrow \infty} \Big|_0^M \ln(1 + x^2) = \infty.$$

\triangle

3.3 Odotusarvon ominaisuuksia

Diskreettien ja jatkuvien jakaumien lisäksi on olemassa myös muunlaisia jakaumia. Yleisessä tapauksessa odotusarvo EX määritellään perusjoukon Ω yli laskettuna abstraktina Lebesguen integraalina

$$EX = \int X dP = \int_{\Omega} X(\omega) P(d\omega)$$

mikäli integrandi on itseisesti integroituva. Odotusarvo saadaan laskettua myös hieman konkreettisemmin, reaaliakselin yli laskettuna Lebesguen-Stieltjesin integraalina

$$\int_{\mathbb{R}} x dF(x),$$

jossa F on X :n kertymäfunktio. Vaikka emme puutu näihin määritelmiin tämän tarkemmin, käsittelemme seuraavaksi kuitenkin hieman sitä, miten merkiltään rajoittamattoman satunnaismuuttujan odotusarvo määritellään.

Integrointiteoriassa määritellään ensin ei-negatiivisen sm:n $Y \geq 0$ integraali EY tietyn rajankäynnin kautta. Tulokseksi saadaan joko jokin ei-negatiivinen reaaliluku (jolloin sanotaan, että Y on integroituva) tai ∞ . Tämän jälkeen merkiltään rajoittamattoman sm:n X odotusarvo määritellään jakamalla se ensin *positiiviseen osaan* $X^+ \geq 0$ sekä *negatiiviseen osaan* $X^- \geq 0$ seuraavasti

$$X^+ = \max(X, 0), \quad X^- = \max(-X, 0). \quad (3.1)$$

Tällöin

$$X = X^+ - X^-, \quad \text{ja} \quad |X| = X^+ + X^-. \quad (3.2)$$

Mikäli sekä EX^+ ja EX^- ovat äärellisiä, niin EX määritellään kaavalla

$$EX = EX^+ - EX^-. \quad (3.3)$$

Tällöin sanotaan, että X on *integroituva*.

Joissakin tarkasteluissa (mutta harvoin) odotusarvolle sallitaan reaalilukujen lisäksi arvot $+\infty$ tai $-\infty$, ts. odotusarvo saa olla laajennettu reaaliluku. Jos vain toinen integraaleista EX^+ ja EX^- on äärellinen, niin EX voidaan määritellä käyttämällä kaavaa $EX = EX^+ - EX^-$ sekä laskusääntöjä

$$\infty - a = \infty, \quad a - \infty = -\infty, \quad a \in \mathbb{R}.$$

Sen sijaan lausekkeelle $\infty - \infty$ ei voida määritellä arvoa. Mikäli X :n odotusarvo on määriteltävissä tähän tapaan, sanomme, että odotusarvo on *olemassa laajennettuna reaalilukuna* tai että se on *olemassa laajennetussa mielessä* tai että X on *kvasi-integroituva*.

Edellä nähtiin esimerkit sekä diskreetistä että jatkuvasti jakautuneesta satunnaismuuttujasta, joilla kummallakaan ei ole olemassa odotusarvoa edes tässä laajennetussa mielessä. Tämän kurssin puitteissa olemme pääasiassa kiinnostuneita vain äärellisistä odotusarvoista, ja sanomme muussa tapauksissa, että odotusarvo ei ole olemassa.

Yleisessä tapauksessa odotusarvolla on edelleen ne ominaisuudet, jotka toditimme diskreetissä tapauksessa. Seuraavaa lausetta emme todista, sillä se vaatisi mitta- ja integrointiteoriaa. Lauseen ominaisuuksia saadaan käyttää ikään kuin ne olisivat aksioomeja.

Lause 3.3. Olkoot X ja Y integroituvia satunnaismuuttujia (siis: odotusarvot $EX, EY \in \mathbb{R}$). Odotusarvolla on seuraavat ominaisuudet.

- (a) (Positiivisuus) Jos $X \geq 0$, niin $EX \geq 0$.
- (b) Jos $X \geq 0$ ja $EX = 0$, niin X on vakio nolla, ts. $P(X = 0) = 1$.
- (c) Jos $X \leq Y$, niin $EX \leq EY$.
- (d) (Vakion odotusarvo) Jos $a \in \mathbb{R}$, niin $Ea = a$.
- (e) (Lineaarisuus) Jos $a, b \in \mathbb{R}$, niin $E(aX + bY) = aEX + bEY$.
- (f) Jos $X \perp Y$, niin satunnaismuuttujalla XY on odotusarvo, ja se saadaan kaavalla
- $$E(XY) = E(X)E(Y).$$
- (g) EX riippuu vain X :n jakaumasta: jos X :llä ja Y :llä on sama jakauma, niin $EX = EY$.

Lause 3.4. X on integroitava eli sen odotusarvo on olemassa jos ja vain jos $|X|$ on integroitava. Tällöin

$$|EX| \leq E|X|$$

Todistus. Jos X on integroitava, niin $EX^+ \in \mathbb{R}$ ja $EX^- \in \mathbb{R}$, joten yhtälön (3.2) sekä lineaarisuuden nojalla

$$E|X| = EX^+ + EX^- \in \mathbb{R}.$$

Toisaalta, jos $E|X| < \infty$, niin välttämättä

$$0 \leq EX^+ < \infty, \quad \text{ja} \quad 0 \leq EX^- < \infty,$$

joten $EX = EX^+ - EX^- \in \mathbb{R}$. Jos X on integroitava, niin kolmioepäyhtälön nojalla

$$|EX| = |EX^+ - EX^-| \leq EX^+ + EX^- = E|X|. \quad \square$$

3.4 Muunnoksen odotusarvo

Jos X on sm, ja $g : \mathbb{R} \rightarrow \mathbb{R}$ on funktio, niin yhdistetty funktio $Y = g(X) = g \circ X$ on myös satunnaismuuttuja. Yksi tapa laskea Y :n odotusarvo olisi ensin johtaa sm:n Y jakauma. Tämä onnistuu periaatteessa helposti diskreetissä tapauksessa. Jatkuvasa tapauksessa johto onnistuu ainakin, jos g on riittävän säännöllinen. Mikäli Y :n jakauma on joko diskreetti tai jatkuva, voitaisiin sitten soveltaa tuttuja kaavoja. Toisen tavan tarjoaa seuraava lause.

Lause 3.5 (Satunnaismuuttujan muunnoksen odotusarvo). *Olkoon X sm, $g : \mathbb{R} \rightarrow \mathbb{R}$ ja $Y = g(X)$.*

- (a) Jos X :llä on diskreetti jakauma ptnf:lla f_X , niin

$$EY = Eg(X) = \sum_x g(x) f_X(x),$$

mikäli summa suppenee itseisesti, eli mikäli $\sum_x |g(x)| f_X(x) < \infty$

(b) Jos X :llä on jatkuva jakauma f_X :lla, niin

$$EY = Eg(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

mikäli integraali suppenee itseisesti, eli mikäli $\int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty$.

Todistus. Esitetään todistus vain diskreetissä tapauksessa, jolloin myös Y on diskreetti sm. Olkoot X :n eri arvot x_1, x_2, \dots , ja määritellään

$$A_i = \{X = x_i\}.$$

Tällöin $Y(\omega)$ saa vakioarvon $y_i = g(x_i)$, kun $\omega \in A_i$. Lauseen 3.1 perusteella

$$EY = \sum_{i \geq 1} y_i P(A_i) = \sum_{i \geq 1} g(x_i) f_X(x_i). \quad \square$$

Huomautus. Aloittelijat luulevat usein virheellisesti, että $Eg(X)$ on yhtäsuuri kuin $g(EX)$. Tämä pitää paikkansa affiinille funktiolle $g(x) = a + bx$, mutta ei yleisesti.

Esimerkki 3.8. Odotusarvo sensuroidulle jakaumalla. Esimerkissä 2.4 käsitelimme jakaumaa, joka saadaan muunnoksella $Y = g(X)$, kun $X \sim U(0, 4)$ ja $g(x) = \min(\max(1, x), 2)$. Osoittautui, että sm:n Y jakauma ei ole jatkuva eikä diskreetti vaan ns. sekatyypin jakauma, joten Y :n odotusarvon laskuun suoraan sen jakaumasta tarvittaisiin Lebesguen integraalia. Koska Y on muunnos jatkuvasti jakautuneesta satunnaismuuttujasta X , niin EY :n saa kuitenkin laskettua suoraviivaisesti lauseen 3.5 avulla, sillä

$$\begin{aligned} EY = Eg(X) &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \\ &= \int_0^1 \frac{1}{4} dx + \int_1^2 \frac{x}{4} dx + \int_2^4 \frac{2}{4} dx = \frac{13}{8}. \end{aligned}$$

△

Lause 3.6. (Riippumattomien sm:ien muunnoksien tulon odotusarvo.) Olkoot satunnaismuuttujat $X \perp Y$, ja $g: \mathbb{R} \rightarrow \mathbb{R}$ ja $h: \mathbb{R} \rightarrow \mathbb{R}$ sellaisia funktioita, että $g(X)$ ja $h(Y)$ ovat integroituvia sm:ia. Tällöin

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Todistus. Koska $g(X) \perp h(Y)$ lauseen 2.13 nojalla, tämä lause seuraa suoraan lauseen 3.3 f-kohdasta. □

3.5 Momentit

Määritelmä 3.3. Satunnaismuuttujan X

- k :s origomomentti eli k :s momentti μ'_k on

$$\mu'_k = EX^k, \quad \text{kun } k = 1, 2, \dots$$

- k :s keskusmomentti (engl. *central moment*) μ_k on

$$\mu_k = E[(X - EX)^k], \quad \text{kun } k = 1, 2, \dots$$

- kertaluvun r absoluuttinen momentti (engl. *absolute moment*) on

$$E|X|^r, \quad r > 0$$

mikäli kyseinen odotusarvo on olemassa. Muuten sanotaan, että kyseinen momentti ei ole olemassa.

Huomautuksia.

- Odotusarvo EX on 1. momentti μ'_1 , ja 1. keskusmomentti $\mu_1 = 0$.
- Keskusmomentti μ_k voidaan esittää origomomenttien $\mu'_j, j \leq k$ avulla kehittämällä lauseke $(X - EX)^k$ binomikaavalla. Kun merkitään $\mu = EX$, niin esim.

$$\mu_2 = E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - \mu^2 = \mu'_2 - (\mu'_1)^2$$

Tässä esiintyvä toinen keskusmomentti μ_2 on sovelluksissa niin tärkeä, että sille käytetään erityistä nimitystä: se on nimeltään *varianssi*.

- Jos kertaluvun $s > 0$ absoluuttinen momentti on olemassa, $E|X|^s < \infty$, niin myöhemmin osoitetaan, että kaikki sitä pienemmän kertaluvun absoluuttiset momentit

$$E|X|^r, \quad 0 < r < s$$

ovat myös olemassa. Erityisesti myös kaikki origomomentit sekä keskusmomentit kertalukua $1 \leq k \leq r$ ovat olemassa.

- Sovelluksissa tärkeimpiä ovat ensimmäinen ja toinen momentti. Korkeammista momenteista esim. kolmas keskusmomentti kuvaa jakauman vinoutta ja neljäs keskusmomentti sen huipukkuutta. Tätä korkeampia momenteja käytetään harvoin.

3.6 Varianssi, keskihajonta ja kovarianssi

Määritelmä 3.4 (Varianssi ja keskihajonta). Satunnaismuuttujan X *varianssi* var X on sen 2. keskusmomentti, eli

$$\text{var } X = \text{var}(X) = E[(X - EX)^2],$$

mikäli kyseinen odotusarvo on olemassa. Tällöin X :n *keskihajonta* (engl. *standard deviation*) on varianssin var X (positiivinen) neliöjuuri. Usein X :n varianssia merkitään σ^2 (tai σ_X^2). Tällöin varianssin positiivinen neliöjuuri $\sigma \geq 0$ (tai $\sigma_X \geq 0$) on X :n keskihajonta.

- Varianssille var X käytetään yleisesti myös merkintää $D^2(X)$.
- Edellä nähtiin, että mikäli varianssi on äärellinen, niin var $X = EX^2 - (EX)^2$. (On myös olemassa jakaumia, joille odotusarvo on olemassa, mutta varianssi ja toinen momentti ovat äärettömiä.)

Lause 3.7 (Varianssin ominaisuuksia).

(i) Mikäli $\text{var } X$ on olemassa, niin $\text{var } X \geq 0$. Jos $\text{var } X = 0$, niin X on vakio.

(ii) Jos $a, b \in \mathbb{R}$ ja $\text{var } X$ on olemassa, niin $\text{var}(aX + b) = a^2 \text{var } X$.

Todistus. (i): Merkitään $\mu = EX$. Koska $(X - \mu)^2 \geq 0$, niin $\text{var } X \geq 0$, mikäli se on reaalityyppinen olemassa.

Jos $\text{var } X = 0$, niin tällöin

$$P[X = \mu] = P[(X - \mu)^2 = 0] = 1,$$

joten X saa arvon μ todennäköisyydellä yksi.

(ii): $E(aX + b) = aEX + b$, joten

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

□

Toisinaan on tarpeen osata laskea kahden satunnaismuuttujan summan tai erotuksen varianssi. Tämä onnistuu niiden kovarianssin avulla.

Määritelmä 3.5. Satunnaismuuttujien X ja Y (välinen) *kovarianssi* $\text{cov}(X, Y)$ on odotusarvo

$$\text{cov}(X, Y) = E[(X - EX)(Y - EY)],$$

mikäli kyseinen odotusarvo on olemassa.

Tietenkin

- $\text{cov}(Y, X) = \text{cov}(X, Y)$,
- $\text{var } X = \text{cov}(X, X)$,

mikäli kyseiset suureet ovat äärellisiä. Kun kerrotaan sulut auki, ja lasketaan odotusarvo termi termiltä, saadaan

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - EX)(Y - EY)] \\ &= E[XY - X(EY) - (EX)Y + (EX)(EY)] \\ &= E(XY) + (-1 - 1 + 1)(EX)(EY), \end{aligned}$$

joten kovarianssi voidaan laskea myös kaavalla

$$\text{cov}(X, Y) = E(XY) - (EX)(EY). \quad (3.4)$$

Lause 3.8 (Summan varianssi). *Olko X ja Y satunnaismuuttujia, joilla on äärelliset toiset momentit. Tällöin*

(i) $\text{var}(X + Y) = \text{var } X + \text{var } Y + 2 \text{cov}(X, Y)$ ja
 $\text{var}(X - Y) = \text{var } X + \text{var } Y - 2 \text{cov}(X, Y)$.

(ii) Jos $X \perp Y$, niin $\text{cov}(X, Y) = 0$ ja

$$\text{var}(X + Y) = \text{var}(X - Y) = \text{var } X + \text{var } Y.$$

Todistus. (i): $E(X+Y) = EX + EY$, joten odotusarvon lineaarisuuden nojalla

$$\begin{aligned}\text{var}(X+Y) &= E[(X-EX+Y-EY)^2] \\ &= E[(X-EX)^2 + 2(X-EX)(Y-EY) + (Y-EY)^2] \\ &= \text{var } X + \text{var } Y + 2 \text{cov}(X, Y).\end{aligned}$$

Se, että äärellisistä toisista momenteista seuraa, että tässä kaikkien termien odotusarvot ovat äärellisiä seuraa luvun 5 tuloksista. Erotuksen varianssia koskeva tulos osoitetaan samalla menetelmällä, eli kertomalla binomin neliö auki ja laskemalla odotusarvo termi termiltä.

(ii): Jos $X \perp Y$, niin

$$\text{cov}(X, Y) = E[(X-EX)(Y-EY)] = E(X-EX)E(Y-EY) = 0 \cdot 0 = 0,$$

joten ensimmäisen kohdan mukaan $\text{var}(X \pm Y) = \text{var } X + \text{var } Y$. \square

Esimerkki 3.9. Huomaa seuraava lauseen 3.7 erikoistapaus,

$$\text{var}(-X) = (-1)^2 \text{var}(X) = \text{var}(X).$$

Summan $X + X$ varianssi voidaan laskea joko lauseen 3.7 avulla,

$$\text{var}(X + X) = \text{var}(2X) = 2^2 \text{var } X = 4 \text{var } X,$$

tai lauseen 3.8 avulla,

$$\text{var}(X + X) = \text{var } X + \text{var } X + 2 \text{cov}(X, X) = 4 \text{var } X.$$

\triangle

Huomaa, että summan $X + Y$ varianssi on yhtä kuin muuttujien varianssien summa täsmälleen silloin, kun $\text{cov}(X, Y) = 0$. Tässä tapauksessa sanotaan, että X ja Y eivät korreloi. Edellä nähtiin, että riippumattomuudesta seuraa korreloimattomuus.

Esimerkki 3.10. Korreloimattomuudesta ei seuraa riippumattomuus. Olkoon sm:lla X jatkuva jakauma tf:lla f , ja olkoon f parillinen funktio. Oletetaan lisäksi, että EX^4 on äärellinen. Tällöin $EX = E(X^3) = 0$, joten kaavan (3.4) avulla nähdään helposti, että

$$\text{cov}(X, X^2) = E(X \cdot X^2) - (EX)(EX^2) = 0.$$

Siis X ja X^2 eivät korreloi, mutta ne eivät tietenkään ole riippumattomia satunnaismuuttujia. \triangle

Lauseen 3.8 kohta (ii) yleistyy helposti myös äärellisen monen riippumattoman satunnaismuuttujan summan varianssille (HT): jos X_1, \dots, X_n ovat riippumattomia ja niillä on kaikilla sama varianssi, niin tällöin

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i)$$

Esimerkki 3.11. Binomijakauman varianssi. Jos $X \sim \text{Bin}(n, p)$, niin se voidaan esittää summana $\sum_{i=1}^n Y_i$, jossa $Y_i \sim \text{Bernoulli}(p)$ ja Y_i :t ovat riippumattomia. Nyt $EY_i = p$, joten

$$\text{var } Y_i = E[(Y_i - p)^2] = (0 - p)^2(1 - p) + (1 - p)^2p = p(1 - p).$$

Riippumattomuuden nojalla $\text{var } X = \sum_{i=1}^n \text{var } Y_i = np(1 - p)$. \triangle

3.7 Momenttiemäfunktio ja kumulanttiemäfunktio

Määritelmä 3.6 (Momenttiemäfunktio). Satunnaismuuttujan X (jakauman) momenttiemäfunktio eli momenttigeneroivafunktio eli momentit generoiva funktio (engl. *moment generating function, mgf*) on funktio

$$M(t) = M_X(t) = Ee^{tX},$$

joka on (reaalifunktiona) olemassa niissä pisteissä t , joissa kyseinen odotusarvo on (äärellisenä) olemassa.

Jos seuraavassa laskussa derivoinnin ja odotusarvon järjestyksen vaihto pysytään jotenkin perustelevaan, niin helposti nähdään, että

$$M'_X(t) = \frac{d}{dt} M_X(t) = \frac{d}{dt} Ee^{tX} = E \frac{\partial}{\partial t} e^{tX} = EXe^{tX},$$

joten $EX = M'_X(0)$. Laskemalla toinen derivaatta nähdään, että

$$M''_X(t) = \frac{d}{dt} EXe^{tX} = E \frac{\partial}{\partial t} Xe^{tX} = EX^2e^{tX},$$

joten $EX^2 = M''_X(0)$. Samalla tavalla edeten nähdään, että aina $EX^k = M^{(k)}(0)$. Osoittautuu, että laskun läpiviennille riittävä ehto on se, että momenttiemäfunktio on määritelty jossakin origon ympäristössä.

Lause 3.9. Jos momenttiemäfunktio $M_X(t)$ on reaalisenä olemassa jossakin origon ympäristössä $|t| < h$, jossa $h > 0$, niin $M_X(t)$ on äärettömän monta kertaa derivoitua origossa, X :n kaikki momentit ovat olemassa, ja M_X voidaan esittää suppenevana Taylorin sarjana

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} EX^k, \quad |t| < h.$$

Erityisesti $EX^k = M_X^{(k)}(0)$, $k \geq 1$.

Todistus. Sarjaesityksen todistus löytyy esim. Billingsleyn teoksesta *Probability and Measure*. Momenttien esitys derivaattojen avulla seuraa derivoimalla Taylorin sarjaa. \square

Määritelmä 3.7 (Kumulanttiemäfunktio). Jos momenttiemäfunktio $M_X(t)$ on määritelty jossakin origon ympäristössä, niin X :n (jakauman) kumulanttiemäfunktio $K(t) = K_X(t)$ (engl. *cumulant generating function, cgf*) määritellään kyseisessä ympäristössä momenttiemäfunktion logaritmina,

$$K(t) = K_X(t) = \ln M_X(t).$$

Odotusarvo ja varianssi saadaan laskettua derivoimalla kumulanttiemäfunktiota origossa, sillä

$$K'_X(0) = \frac{M'_X(0)}{M_X(0)} = \frac{EX}{1} = EX,$$

ja

$$K''_X(0) = \frac{M''_X(0)M_X(0) - M'_X(0)M'_X(0)}{M_X(0)^2} = EX^2 - (EX)^2 = \text{var } X.$$

Tämä tarjoaa näppärän tavan johtaa odotusarvo ja varianssi joillekin jakaumille, mutta toisille jakaumille sattaa olla kätevämpää johtaa odotusarvo ja varianssi laskemalla jakauman ensimmäinen ja toinen momentti.

Origossa laskettuja kumulanttiefunktion derivaattoja kutsutaan X :n (jakauman) *kumulanteiksi*: k :s kumulantti on $K_X^{(k)}(0)$. Ensimmäinen kumulantti on saman kuin odotusarvo ja toinen kumulantti on sama kuin varianssi.

Esimerkki 3.12. Binomijakauman odotusarvo ja varianssi kumulanttiefunktion avulla. Jos $X \sim \text{Bin}(n, p)$, niin binomikaavan avulla nähdään helposti, että

$$M_X(t) = (pe^t + 1 - p)^n,$$

joten

$$K_X(t) = n \ln(pe^t + 1 - p).$$

Laskemalla ensimmäinen ja toinen derivaatta origossa nähdään, että

$$EX = K_X'(0) = np, \quad \text{var } X = K_X''(0) = np(1 - p).$$

△

Lause 3.10 (Momenttiefunktion ominaisuuksia). *Seuraavassa kohtien (a) ja (b) kaavat pätevät niissä pisteissä, joissa kyseiset funktiot ovat määritellyjä. (Ne pätevät muissa pisteissä muodossa $\infty = \infty$.)*

(a) Jos a, b ovat vakioita, niin $M_{aX+b}(t) = e^{bt} M_X(at)$.

(b) Jos $X \perp Y$, niin $M_{X+Y}(t) = M_X(t) M_Y(t)$.

Todistus. Nämä ominaisuudet todistetaan yksinkertaisilla laskuilla, sillä

$$M_{aX+b}(t) = E \exp(t(aX + b)) = e^{bt} E \exp(atX) = e^{bt} M_X(at),$$

ja kun $X \perp Y$, niin

$$M_{X+Y}(t) = E \exp(t(X + Y)) = E[\exp(tX) \exp(tY)] = Ee^{tX} Ee^{tY},$$

missä käytimme hyväksi lausetta 3.6. □

Momenttiefunktiota ja kumulanttiefunktiota käytetään useissa asymp-toottisissa tarkasteluissa (esim. suurten poikkeamien teoriassa sekä satulapistekehitemässä). Näiden funktioiden käyttöön liittyy kuitenkin hankaluuksia. Jos jokin X :n momentti ei ole olemassa, niin momenttiefunktio on kyllä määritelty origossa, mutta ei missään origon ympäristössä. Lisäksi on olemassa jakaumia (esim. log-normaalin jakauma), joilla on kaikkien kertalukujen momentit, mutta joiden momenttiefunktio ei ole olemassa missään origon ympäristössä. Tällaisilta ongelmilta vältytään, mikäli momenttiefunktion sijasta käytetään jakauman karakteristista funktiota, sillä karakteristinen funktio on kaikkialla määritelty.

Määritelmä 3.8. Satunnaismuuttujan X karakteristinen funktio (engl. *characteristic function*) on koko reaaliakselilla määritelty kompleksiarvoinen funktio

$$\phi(t) = \phi_X(t) = Ee^{itX}, \quad \text{jossa } i = \sqrt{-1}.$$

Tässä kompleksinen eksponentti saadaan laskettua kaavalla

$$e^{itX} = \cos(tX) + i \sin(tX),$$

ja sen odotusarvo määritellään laskemalla yhteen reaali- ja imaginääriosan odotusarvot,

$$Ee^{itX} = E \cos(tX) + i E \sin(tX).$$

Karakteristinen funktio $\phi_X(t)$ on määritelty kaikilla X ja kaikilla t sen takia, että se on rajoitetun funktion odotusarvo.

Lause 3.11. *Jos momenttiemäfunktio $M_X(t)$ on määritelty jossakin origon ympäristössä, niin X :n karakteristinen funktio on*

$$\phi_X(t) = M_X(it).$$

Huomautus. Tässä $M_X(it)$ tarkoittaa sitä, että Taylorin sarjaan sijoitetaan argumentin t sijalle it . Tämä on luvallista, koska kyseessä on suppeneva potenssisarja. Jos funktiolle $M_X(t)$ saadaan johdettua lauseke, niin tulos saadaan laskettua myös sijoittamalla ko. lausekkeessa argumentin t sijalle it .

Lause 3.12 (Momenttiemäfunktio määrää jakauman, karakteristinen funktio määrää jakauman).

- (a) *Jos X :n ja Y :n momenttiemäfunktiot ovat molemmat olemassa jossakin origon ympäristössä, ja ovat siellä yhtäsuuret, niin X :llä ja Y :llä on sama jakauma.*
- (b) *Jos X :n ja Y :n karakteristiset funktiot ovat samat, niin X :llä ja Y :llä on sama jakauma.*

Todistus. Ks. esim. Billingsleyn teos. □

Luku 4

Tärkeitä jakaumia

Tässä luvussa tarkastellaan tiettyjä sovelluksissa usein esiintyviä jakaumia. Näistä jakaumista löytyy lisätietoja ja kuvaajia Wikipediasta, ks.

http://en.wikipedia.org/wiki/List_of_probability_distributions

Varoitus. Useille näistä jakaumista käytetään kirjallisuudessa erilaisia parametreinteja. Kussakin lähteessä käytetty parametointi paljastuu tavallisesti tutkimalla lähemmin siinä käytettyjä kaavoja. Eri lähteet käyttävät jakaumille eri tunnuksia.

4.1 Diskreettejä jakaumia

Suurin osa sovelluksissa käytettävistä diskreeteistä satunnaismuuttujista on kokonaislukuarvoisia. Annamme ptfn:n f lausekkeen vain niissä pisteissä, joissa $f(x)$ voi olla nollaa suurempi.

4.1.1 Binomijakauma

Synty. Onnistumisten lkm n -kertaisessa toistetussa Bernoullin kokeessa, kun onnistumisten on p .

Tunnus $X \sim \text{Bin}(n, p)$, jossa $n \geq 1$ kokonaisluku (otoskokoparametri) ja $0 \leq p \leq 1$ (onnistumisten; todennäköisyysparametri).

Ptnf

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

Binomikaavan mukaan

$$1 = (p + (1-p))^n = \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x}.$$

Koska lisäksi jokainen termi on ei-negatiivinen, muodostuu niistä ptnf kokonaisluvuilla $x = 0, 1, \dots, n$.

Odotusarvo, varianssi, momenttiemäfunktio

$$EX = np, \quad \text{var } X = np(1-p), \quad M(t) = (pe^t + 1 - p)^n.$$

Yhteyksiä muihin jakaumiin. $\text{Bin}(1, p)$ on Bernoulli(p).

Yhteenlaskuominaisuus. Jos $X \sim \text{Bin}(m, p)$ ja $Y \sim \text{Bin}(n, p)$ (sama onnistumistn) ja $X \perp Y$, niin

$$X + Y \sim \text{Bin}(n + m, p).$$

Esimerkiksi, jos laatikossa on N palloa, joista $0 \leq K \leq N$ on valkoista ja loput mustia, ja laatikosta poimitaan umpimähkään n palloa takaisinpanolla, niin nostettujen valkoisten pallojen lukumäärällä X on jakauma $\text{Bin}(n, K/N)$, jonka odotusarvo ja varianssi ovat

$$EX = n \frac{K}{N}, \quad \text{var } X = n \frac{K}{N} \left(1 - \frac{K}{N}\right).$$

4.1.2 Hypergeometrinen jakauma

Synty. Laatikossa on N palloa, joista $0 \leq K \leq N$ on valkoista ja loput ovat mustia, ja laatikosta nostetaan umpimähkäisesti $n \leq N$ palloa ilman takaisinpanoa. Jos X on sm, joka kertoo, kuinka moni nostetuista palloista on valkoinen, niin X :n **ptnf** on

$$f(x) = f(x | N, K, n) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \dots, n.$$

Tämä on hypergeometrinen jakauma parametreilla N , K ja n . Edellä käytettiin binomikertoimille $\binom{m}{k}$ sopimusta (1.5), jonka mukaan $\binom{m}{k} = 0$ ellei $0 \leq k \leq m$. Hypergeometrisen jakauman pistetodennäköisyysfunktio poikkeaa nolasta vain, kun $0 \leq x \leq \min(n, K)$ ja $n - x \leq N - K$.

Hypergeometrisen jakauman ptnf:n saa johdettua suoraan kombinatoriikan tuloperiaatteella. Nyt symmetrisiä alkeistapauksia ovat N pallon n -osajoukot, joiden lukumäärä on $\binom{N}{n}$. Sellaisen osajoukon valinta, jossa on k valkoista ja $n-k$ mustaa palloa, voidaan ajatella tehtävän kahdessa vaiheessa. Ensin valitaan k :n valkoisen pallon osajoukko K valkoisesta pallosta (eri mahdollisuuksia on $\binom{K}{k}$), ja sen jälkeen valitaan $n-k$ mustan pallon osajoukko $N-K$ mustasta pallosta (eri mahdollisuuksia on $\binom{N-K}{n-k}$). Jakauman ptnf saadaan jakamalla suotuisten alkeistapauksien lukumäärä kaikkien alkeistapausten lukumäärällä.

Hypergeometrisen jakauman odotusarvo ja varianssi ovat

$$EX = n \frac{K}{N}, \quad \text{var } X = n \frac{K}{N} \left(1 - \frac{K}{N}\right) \frac{N-n}{N-1}.$$

Nämä tulokset johdetaan esim. Tuomisen kirjassa. Hypergeometrisen jakauman käsittely on hankalaa, minkä takia suurten populaatioiden tapauksessa (N suuri ja $n \ll N$) sitä mielellään approksimoidaan vastaavalla binomijakaumalla $\text{Bin}(n, K/N)$, joka syntyy otannasta takaisinpanolla.

4.1.3 Geometrinen jakauma

Synty. Toistetaan riippumattomasti Bernoullin koetta, jossa onnistumistn on $0 < p < 1$. Olkoon

$$X = \text{“epäonnistumisten lkm ennen ensimmäistä onnistumista”}.$$

Tällöin X :llä on geometrinen jakauma parametrilla p , eli $X \sim \text{Geom}(p)$.

Ptnf on

$$f(x) = f(x | p) = P(X = x) = p(1-p)^x, \quad x = 0, 1, 2, \dots$$

Tämä johtuu siitä, että $X = x$ jos ja vain jos koetta vastaavat riippumattomat Bernoullin muuttujat Y_1, \dots, Y_x, Y_{x+1} saavat arvot

$$Y_1 = 0, Y_2 = 0, \dots, Y_x = 0 \quad \text{ja} \quad Y_{x+1} = 1.$$

Riippumattomuuden nojalla tämän tapahtuman tn on $p(1-p)^x$.

Se että kyseessä on todella ptnf nähdään *geometrisen sarjan* summasta,

$$\sum_{j=0}^{\infty} r^j = 1 + r + r^2 + \dots = \frac{1}{1-r}, \quad |r| < 1,$$

jonka seurauksena

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

Koska lisäksi jokainen termeistä $f(x) = p(1-p)^x$ on positiivinen, muodostuu niistä ptnf ei-negatiivisilla kokonaisluvuilla.

Momenttiemäfunktio saadaan helpolla laskulla,

$$M(t) = \frac{p}{1-(1-p)e^t}, \quad t < -\ln(1-p).$$

Tästä johdetaan helposti jakauman odotusarvo ja varianssi,

$$EX = \frac{1-p}{p}, \quad \text{var } X = \frac{1-p}{p^2}.$$

Huomautus. Useissa lähteissä geometrinen jakauma määritellään edellisesti poikkeavasti niin, että se on satunnaismuuttujan Y jakauma, jossa

Y = "sen toiston järjestysnumero, jolla onnistutaan ensimmäisen kerran".

Tässä $Y = X + 1$.

4.1.4 Negatiivinen binomijakauma

Synty. Toistetaan riippumattomasti Bernoullin koetta, jossa onnistumistn on $0 < p < 1$. Olkoon $r > 0$ kokonaisluku, ja tarkastellaan sm:a

X = "epäonnistumisten lkm, ennenkuin onnistutaan r :nnen kerran".

Tällöin

$$\{X = x\} = \{r+x-1 \text{ ensimmäisessä toistossa epäonnistutaan } x \text{ kertaa} \\ \text{ja onnistutaan } r-1 \text{ kertaa}\} \cap \{\text{toistossa } r+x \text{ onnistutaan}\}.$$

Koska toistot ovat riippumattomia, on

$$f(x) = f(x | r, p) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, 2, \dots$$

Tässä binomikertoimet ilmaistaan usein gammafunktion avulla, joka yleistää kertoman positiivisille reaaliluvuille kaavan $r! = \Gamma(r + 1)$ kautta. Tällöin parametrille $r > 0$ voidaan sallia myös muut kuin kokonaislukuarvot, ja ptnf voidaan kirjoittaa muotoon

$$f(x) = f(x | r, p) = \frac{\Gamma(r+x)}{\Gamma(r)x!} p^r (1-p)^x, \quad x = 0, 1, 2, \dots$$

Jos $r > 0$ ei ole kokonaisluku, jakaumalla ei kuitenkaan enää ole toistokokeeseen liittyvää tulkintaa

Huomautus: useissa lähteissä (esim. Tuomisen TN I) negatiivinen binomijakauma määritellään siten, että se on sm:n Y jakauma, jossa

$Y =$ "sen toiston järjestysnumero, jolla onnistutaan r :nnen kerran".

Tällöin $Y = X + r$.

Mikä negatiivisessa binomijakaumassa sitten on negatiivista? Tätä varten meidän pitää ensin tehdä alustavia tarkasteluja. Binomikertoimet määritellään seuraavalla tavalla, kun ylempi indeksi ei ole välttämättä kokonaisluku, mutta alempi indeksi on,

$$\binom{r}{k} = \begin{cases} \frac{r(r-1)\cdots(r-k+1)}{k!}, & k \geq 0 \text{ kokonaisluku,} \\ 0, & k < 0 \text{ kokonaisluku.} \end{cases} \quad (4.1)$$

Tätä merkintää tarvitaan esim. *binomisarjassa*, jonka mukaan kaikilla $r \in \mathbb{R}$

$$(1+z)^r = \sum_{j=0}^{\infty} \binom{r}{j} z^j, \quad |z| < 1. \quad (4.2)$$

Binomikaava on binomisarjan erikoistapaus: jos $n \geq 0$ on kokonaisluku ja kokonaisluku $j \geq n$, niin $\binom{n}{j} = 0$, joten

$$(1+z)^n = \sum_{j=0}^{\infty} \binom{n}{j} z^j = \sum_{j=0}^n \binom{n}{j} z^j.$$

Olkoon $r > 0$ ja olkoon $j \geq 0$ kokonaisluku. Tällöin

$$\begin{aligned} \binom{-r}{j} &= \frac{(-r)(-r-1)\cdots(-r-j+1)}{j!} \\ &= (-1)^j \frac{(r+j-1)\cdots(r+1)r}{j!} = (-1)^j \binom{r+j-1}{j}. \end{aligned}$$

NegBin(r, p)-jakauman ptnf voidaan siis kirjoittaa vielä uuteen, binomijakaumaa muistuttavaan muotoon, kun merkitään $q = 1 - p$,

$$f(x) = \binom{-r}{x} p^r (-q)^x, \quad x = 0, 1, 2, \dots$$

Binomisarjan avulla on nyt helppo näyttää, että kyseessä on todellakin ptnf, sillä

$$\sum_{x=0}^{\infty} f(x) = p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-q)^x = p^r (1-q)^{-r} = 1.$$

Ts. negatiivisen binomijakauman ptnf saadaan kehittämällä lausekkeessa $p^r(1-q)^{-r}$ binomin $1-q$ negatiivinen potenssi $(1-q)^{-r}$ binomisarjalla.

Jakauman **momenttiemäfunktio** saadaan laskettua helposti binomisarjan avulla,

$$M(t) = \left(\frac{p}{1 - (1-p)e^t} \right)^r, \quad t < -\ln(1-p).$$

Odotusarvo ja varianssi löytyvät helposti derivoimalla,

$$EX = \frac{r(1-p)}{p}, \quad \text{var } X = \frac{r(1-p)}{p^2}.$$

Yhteyksiä muihin jakaumiin. NegBin($1, p$) on Geom(p). Negatiiviselle binomijakaumalla käytetään myös muita parametreja. Usein parametreiksi otetaan jakauman odotusarvo μ sekä r (dispersioparametri). Tällöin $p = r/(\mu + r)$ ja jakauman varianssi on $\mu + \mu^2/r$. Jos tässä parametroinnissa annetaan $r \rightarrow \infty$, niin rajalla saadaan Poi(μ) (ts. pistetodennäköisyydet suppenevat kohti tämän Poissonin jakauman pistetodennäköisyyksiä). Tämän takia negatiivista binomijakaumaa käytetään mallina sellaisissa yhteyksissä, joissa tavallisesti tahdottaisiin käyttää Poissonin jakaumaa, mutta joissa aineiston empiirinen varianssi näyttää selvästi suuremmalta kuin Poissonin jakauman varianssi (*negative binomial as an overdispersed Poisson distribution*).

Huomautus. Kun $r > 0$ on kokonaisluku, niin NegBin(r, p)-jakaumaa kutsutaan toisinaan Pascalin jakaumaksi. Jos $r > 0$ on kokonaisluku, niin joissakin lähteissä sm:n $Y = X + r$ (eikä sm:n X) jakaumaa kutsutaan negatiiviseksi binomijakaumaksi.

4.1.5 Poissonin jakauma

Tunnus $X \sim \text{Poi}(\theta)$, jossa $\theta \geq 0$.

Ptnf on

$$f(x) = f(x | \theta) = e^{-\theta} \frac{\theta^x}{x!}, \quad x = 0, 1, 2, \dots$$

Tämä on ptnf sillä perusteella, että kaikki arvot ovat ei-negatiivisia, ja eksponenttifunktion sarjakehitelmän mukaan

$$e^u = \exp(u) = \sum_{j=0}^{\infty} \frac{u^j}{j!}, \quad \forall u \in \mathbb{R}.$$

Mikäli $u > 0$, niin sarjan kaikki termit ovat positiivisia, joten luvut $e^{-u} u^j/j!$ muodostavat tällöin pistetodennäköisyysfunktion, kun $u = 0, 1, 2, \dots$

Momenttiemäfunktio saadaan laskettua eksponenttifunktion sarjakehitelmän avulla,

$$M(t) = \sum_{x=0}^{\infty} e^{xt} e^{-\theta} \frac{\theta^x}{x!} = \exp(\theta(e^t - 1)).$$

Odotusarvo ja varianssi saadaan tästä helposti derivoimalla,

$$EX = \text{var } X = \theta.$$

Yhteenlaskuominaisuus. Jos $X \sim \text{Poi}(\theta_1)$ ja $Y \sim \text{Poi}(\theta_2)$, ja $X \perp Y$, niin

$$M_{X+Y}(t) = M_X(t) M_Y(t) = \exp((\theta_1 + \theta_2)(e^t - 1)),$$

joten $X + Y \sim \text{Poi}(\theta_1 + \theta_2)$.

Poissonin prosessi. Yksi tilanne, josta syntyy Poissonin jakauma on Poissonin prosessi. Siinä mallinnetaan diskreettejä tapahtumia jatkuvalla aika- tai pituusvälillä tai tasossa tai avaruudessa. Saateetasiin mallintaa esimerkiksi sitä,

- kuinka monta asiakasta palvelupisteeseen saapuu tietyllä aikavälillä,
- kuinka monta fonia osuu tiettyyn filmin alueeseen,
- kuinka monta valkosolua löytyy tietyistä veripisarasta.

Mikäli välin pituus (tai tasoalueen pinta-ala, avaruuden osan tilavuus) on s yksikköä, niin Poissonin prosessissa siinä havaittavien tapahtumien lukumäärällä on Poissonin jakauma $\text{Poi}(s\lambda)$. Lisäksi erillisillä väleillä (tasoalueilla, avaruuden osilla) havaittavat lukumäärät ovat keskenään riippumattomia. Keskimääräinen tapahtumien lukumäärä yhtä (pituus-, pinta-ala- tai tilavuus-)yksikköä kohti on $\lambda s/s = \lambda$. Ts. $\lambda > 0$ eli Poissonin prosessin *intensiteetti* on tapahtumien odotusarvo yhtä (pituus-, pinta-ala-, tilavuus-)yksikköä kohti.

4.2 Gamma- ja beetafunktio

Eulerin *gammafunktio* Γ voidaan määritellä positiivisilla argumenteilla integraalilla

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx, \quad t > 0. \quad (4.3)$$

Gammafunktion määrittelevä integraali (4.3) on äärellinen jos ja vain jos $t > 0$.

Selvästi $\Gamma(1) = 1$, ja osittaisintegroinnilla nähdään, että

$$\Gamma(t+1) = \int_0^{\infty} x^t e^{-x} dx = - \left|_0^{\infty} x^t e^{-x} + t \int_0^{\infty} x^{t-1} e^{-x} dx,\right.$$

joten

$$\Gamma(t+1) = t\Gamma(t), \quad \text{kaikilla } t > 0. \quad (4.4)$$

Tämän takia kokonaislukuargumenteilla n pätee

$$\Gamma(n) = (n-1)!, \quad \text{kun } n = 1, 2, 3, \dots$$

Tässä mielessä gammafunktio on kertoman yleistyks.

Tarkastellaan seuraavaksi integraalia

$$I = \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx,$$

joka on standardinormaalijakauman normalisointivakio. Ottamalla huomioon, että integrandi on parillinen funktio ja tekemällä sitten muuttujanvaihto $t = \frac{1}{2}x^2$ integraali saadaan muotoon

$$I = 2 \int_0^{\infty} e^{-\frac{1}{2}x^2} dx = 2 \int_0^{\infty} e^{-t} \frac{1}{\sqrt{2t}} dt = \sqrt{2} \Gamma\left(\frac{1}{2}\right).$$

Pian näytetään vielä, että $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Tarkastellaan tuloa $\Gamma(a)\Gamma(b)$, jossa $a, b > 0$,

$$\Gamma(a)\Gamma(b) = \int_0^\infty u^{a-1} e^{-u} du \int_0^\infty v^{b-1} e^{-v} dv.$$

Tehdään muuttujanvaihdot $u = x^2$ ja $v = y^2$, jolloin saadaan

$$\Gamma(a)\Gamma(b) = 4 \int_0^\infty \int_0^\infty x^{2a-1} y^{2b-1} e^{-x^2-y^2} dx dy.$$

Tässä kahden integraalin tulo on tulkittu tasointegraaliksi. Seuraavaksi siirytään napakoordinaatteihin r ja θ ,

$$x = r \cos(\theta), \quad y = r \sin(\theta), \quad dx dy = r dr d\theta.$$

Edellä tasointegraali laskettiin yli ensimmäisen kvadrantin, jota napakoordinaateissa vastaa alue $0 < \theta < \pi/2$ ja $r > 0$, joten tasointegraali muuntuu muotoon

$$\begin{aligned} \Gamma(a)\Gamma(b) &= 4 \int_{r=0}^\infty \int_{\theta=0}^{\pi/2} (r \cos \theta)^{2a-1} (r \sin \theta)^{2b-1} e^{-r^2} r dr d\theta \\ &= \left(2 \int_0^\infty r^{2a+2b-1} e^{-r^2} dr \right) \left(2 \int_0^{\pi/2} (\cos \theta)^{2a-1} (\sin \theta)^{2b-1} d\theta \right). \end{aligned}$$

Muuttujanvaihdolla $t = r^2$ nähdään, että tässä

$$2 \int_0^\infty r^{2a+2b-1} e^{-r^2} dr = \int_0^\infty t^{a+b-1} e^{-t} dt = \Gamma(a+b),$$

ja muuttujanvaihdolla $u = \cos^2 \theta$ nähdään, että

$$\begin{aligned} 2 \int_0^{\pi/2} (\cos \theta)^{2a-1} (\sin \theta)^{2b-1} d\theta &= \int_0^{\pi/2} (\cos^2 \theta)^{a-1} (\sin^2 \theta)^{b-1} 2 \cos \theta \sin \theta d\theta \\ &= - \int_1^0 u^{a-1} (1-u)^{b-1} du = \int_0^1 u^{a-1} (1-u)^{b-1} du. \end{aligned}$$

Tässä funktiota

$$B(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du, \quad a, b > 0 \quad (4.5)$$

kutsutaan (Eulerin) *beetafunktioksi*. Beetafunktion määrittelevä integraali (4.5) on äärellinen jos ja vain jos $a > 0$ ja $b > 0$.

Edellä saatiin muuttujanvaihtojen avulla johdettua beetafunktiolle esitys gammafunktion avulla, nimittäin

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \quad (4.6)$$

Erityisesti

$$\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) = B\left(\frac{1}{2}, \frac{1}{2}\right)\Gamma(1) = B\left(\frac{1}{2}, \frac{1}{2}\right) = 2 \int_0^{\pi/2} (\cos \theta)^0 (\sin \theta)^0 d\theta = \pi,$$

joten

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi},$$

kuten edellä jo kerrottiin.

Näiden tarkastelujen sivutuotteena ollaan saatu todistettua, että

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi} \quad (4.7)$$

4.3 Jatkuvia jakaumia

4.3.1 Siirto ja skaalaus

Jos sm:lla Z on jatkuva jakauma tiheysfunktiolla f_0 , ja sm X määritellään kaavalla

$$X = \mu + \sigma Z, \quad \sigma > 0,$$

niin tiheysfunktion muuntokaavan (ks. kaava (2.6) tai muistisääntö (2.7)) nojalla sm:n X tf on

$$f_X(x | \mu, \sigma) = \frac{1}{\sigma} f_0\left(\frac{x - \mu}{\sigma}\right).$$

Jos erityisesti $\sigma = 1$ ja $\mu = 0$, niin X :n tf on f_0 . Tällä tavoin saadaan perhe jakaumia, jossa parametria μ kutsutaan sijaintiparametriksi (engl. *location parameter*) ja parametria $\sigma > 0$ skaalaparametriksi (engl. *scale parameter*).

Erityisesti, jos $\mu = 0$, saadaan jakaumaperhe, jonka tiheysfunktiot ovat muotoa

$$\frac{1}{\sigma} f_0(x/\sigma), \quad \sigma > 0.$$

Jotkin jakaumaperheet parametroidaan muodossa

$$\lambda f_0(\lambda x), \quad \lambda > 0,$$

jossa f_0 on tf. Tällöin $\sigma = 1/\lambda$ on skaalaparametri, ja parametria λ voidaan kutsua *rate*-parametriksi. (Englannin kielen sana *rate* voidaan kääntää, asiayhteydestä riippuen, esim. sanoilla intensiteetti, vauhti, osuus, aste, taajuus jne.)

4.3.2 Tasajakauma

Jos $a < b$, niin välin (a, b) tasajakaumalla, $X \sim U(a, b)$ on tiheysfunktio

$$f(x) = f(x | a, b) = \frac{1}{b - a} 1_{(a, b)}(x), \quad x \in \mathbb{R}.$$

Sen kaksi ensimmäistä momenttia lasketaan helposti integroimalla,

$$EX = \frac{1}{2}(a + b), \quad EX^2 = \frac{1}{3}(a^2 + ab + b^2),$$

ja tästä nähdään, että

$$\text{var } X = \frac{(b - a)^2}{12}.$$

4.3.3 Eksponenttijakauma

Sm X noudattaa eksponenttijakaumaa parametrilla $\lambda > 0$, eli $X \sim \text{Exp}(\lambda)$, jos sen tf on

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

Tässä kaavassa λ on *rate*-parametri. Jakauman odotusarvo ja varianssi ovat

$$EX = \frac{1}{\lambda}, \quad \text{var } X = \frac{1}{\lambda^2} = (EX)^2,$$

ja sen momenttiemäfunktio on

$$M(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

Nämä tulokset seuraavat gammajakauman vastaavista tuloksista, sillä eksponenttijakauma on gammajakauman erikoistapaus.

Eksponenttijakaumalle käytetään yleisesti myös sellaista parametrintia, jossa parametrina on $\theta = 1/\lambda$. Tällöin θ on eksponenttijakauman odotusarvo ja θ^2 sen varianssi.

Eksponenttijakaumaa käytetään usein komponentin eliniän mallina. Tällöin tehdään se oletus, että komponentti ei kulu käytössä. Eksponenttijakaumalla on nimittäin seuraava mielenkiintoinen ominaisuus, ns. muistinmenetysominaisuus. Oletetaan, että komponentin elinikä $X \sim \text{Exp}(\lambda)$, ja $x, h > 0$ ja lasketaan todennäköisyys, että elinikä on suurempi kuin $x + h$, kun tiedetään, että se on suurempi kuin x ,

$$\begin{aligned} P(X > x + h | X > x) &= \frac{P(X > x + h \text{ ja } X > x)}{P(X > x)} = \frac{P(X > x + h)}{P(X > x)} \\ &= \frac{1 - F(x + h)}{1 - F(x)} = \frac{e^{-\lambda(x+h)}}{e^{-\lambda x}} = e^{-\lambda h} = P(X > h). \end{aligned}$$

Tämä todennäköisyys ei siis riipu lainkaan siitä, kuinka kauan komponenttia on jo käytetty.

4.3.4 Gammajakauma

Jos $\alpha, \lambda > 0$, niin muuttujanvaihdolla $u = \lambda x$ nähdään, että

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}.$$

Tämän takia

$$f(x) = f(x | \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

on tiheysfunktio. Vastaavaa jakaumaa kutsutaan gammajakaumaksi parametrein $\alpha > 0$ (muotoparametri) ja $\lambda > 0$ (*rate*-parametri). Käytämme jakaumalle tunnusta $\text{Gam}(\alpha, \lambda)$.

Jakauman $\text{Gam}(\alpha, \lambda)$ momenttiemäfunktio saadaan laskettua, kun huomataan, että siinä tarvittava integraali on erään toisen gammajakauman normalisointivakio,

$$\begin{aligned} M(t) &= Ee^{Xt} = \int_0^\infty e^{xt} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda-t)^\alpha} \int_0^\infty \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx \\ &= \left(\frac{\lambda}{\lambda-t} \right)^\alpha, \quad t < \lambda. \end{aligned}$$

Tässä laskussa *integroimme kuten tilastotieteilijä*: tunnistamme että integrandi on vakiota vaille tutun jakauman tiheysfunktio, jota integroidaan kantajansa yli. Tällöin osaamme suoraan kirjoittaa lausekkeen integraalin arvolle katsomalla ko. tiheysfunktion normalisointivakiota.

Momenttiemäfunktion lausekkeesta nähdään helposti, että

$$EX = \frac{\alpha}{\lambda}, \quad \text{var } X = \frac{\alpha}{\lambda^2},$$

kun $X \sim \text{Gam}(\alpha, \lambda)$.

Yhteyksiä muihin jakaumiin.

- $\text{Exp}(\lambda)$ on $\text{Gam}(1, \lambda)$.
- χ_n^2 (khiin neliön jakauma n :llä vapausasteella) on $\text{Gam}(n/2, 1/2)$.

Yhteenlaskuominaisuus. Jos $X \sim \text{Gam}(\alpha_1, \lambda)$ ja $Y \sim \text{Gam}(\alpha_2, \lambda)$ (jälkimmäinen parametri sama) ja $X \perp Y$, niin momenttiemäfunktioita tarkastelemalla nähdään heti, että

$$X + Y \sim \text{Gam}(\alpha_1 + \alpha_2, \lambda).$$

Huomautus. Gammajakauma parametroidaan usein myös käyttämällä parametreja α ja skaalaparametria $s = 1/\lambda$. Kummassakin parametroinnissa jälkimmäistä parametria on tapana merkitä symbolilla β . Se, kummasta parametroinnista on kyse selviää helposti, mikäli jakauman tiheysfunktion tai odotusarvon kaava annetaan.

4.3.5 Beetajakauma

Kun $\alpha, \beta > 0$, niin beetafunktion määritelmän nojalla funktio

$$f(x) = f(x | \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1,$$

on tiheysfunktio. Käytämme sille tunnusta $\text{Be}(\alpha, \beta)$. Tämän jakauman momentit on helppo laskea, sillä ne saadaan suoraan erään toisen beetajakauman normalisointivakiosta (ts. integroimalla kuten tilastotieteilijä). Jos $r > 0$, on

$$EX^r = \frac{B(\alpha+r, \beta)}{B(\alpha, \beta)}.$$

Erityisesti

$$EX = \frac{\alpha}{\alpha + \beta}, \quad EX^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)},$$

josta

$$\text{var } X = EX^2 - (EX)^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Välin $(0, 1)$ tasajakauma $U(0, 1)$ on sama kuin $\text{Be}(1, 1)$.

4.3.6 Normaalijakauma

Standardinormaalijakaumaa noudattavan sm:n $Z \sim N(0, 1)$ tf on

$$f_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right), \quad z \in \mathbb{R}. \quad (4.8)$$

Standardinormaalijakauman tiheysfunktiolle käytetään yleisesti merkintää ϕ ja sen kertymäfunktiolle merkintää Φ . (Asiayhteydestä riippuen ϕ voi tarkoittaa joko karakteristista funktiota tai standardinormaalijakauman tiheysfunktiota tai jotakin muuta.) Se seikka, että kyseessä on tiheysfunktio seuraa siitä, että funktio on aidosti positiivinen ja että sen integraali koko reaaliakselin yli on kaavan (4.7) nojalla yksi.

Lasketaan seuraavaksi $N(0, 1)$:n momenttiemäfunktio.

$$M_Z(t) = Ee^{tZ} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2 + tz\right) dz.$$

Täydennetään eksponenttifunktion argumentti neliöksi,

$$-\frac{1}{2}z^2 + tz = -\frac{1}{2}(z - t)^2 + \frac{1}{2}t^2,$$

minkä jälkeen (sijoituksella $u = z - t$) nähdään, että

$$M_Z(t) = \exp\left(\frac{1}{2}t^2\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du = \exp\left(\frac{1}{2}t^2\right).$$

Laskemalla kumulanttiamäfunktion $K_Z(t) = \frac{1}{2}t^2$ derivaatat nähdään, että

$$EZ = 0, \quad \text{var } Z = 1.$$

Sm X noudattaa normaalijakaumaa parametrein μ, σ^2 , eli $X \sim N(\mu, \sigma^2)$, jos se voidaan esittää muodossa

$$X = \mu + \sigma Z, \quad Z \sim N(0, 1). \quad (4.9)$$

Tällöin

$$EX = \mu, \quad \text{var } X = \sigma^2 \text{var } Z = \sigma^2,$$

joten normaalijakauman parametrit ovat sen odotusarvo ja varianssi. Jos $\sigma > 0$, niin jakauman tf on

$$f_X(x) = f_X(x | \mu, \sigma^2) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right), \quad x \in \mathbb{R}. \quad (4.10)$$

Jakauman momenttiemäfunktio on

$$\begin{aligned} M_X(t) &= M_{\mu + \sigma Z}(t) = E \exp((\mu + \sigma Z)t) = \exp(\mu t) M_Z(\sigma t) \\ &= \exp(\mu t + \frac{1}{2}\sigma^2 t^2). \end{aligned} \quad (4.11)$$

Yhteenlaskuominaisuus. Jos $X \sim N(\mu_1, \sigma_1^2)$ ja $Y \sim N(\mu_2, \sigma_2^2)$, ja ne ovat riippumattomia, niin

$$M_{X+Y}(t) = M_X(t) M_Y(t) = \exp((\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2),$$

joten $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

4.3.7 Normaalijakaumasta johdettuja jakaumia

Khiin neliön jakauma. Jos $X_1, \dots, X_n \sim N(0, 1)$ ja ne ovat riippumattomia, niin niiden neliöiden summalla sanotaan olevan χ_n^2 -jakauma eli khiin neliön jakauma n :llä vapausasteella (tai vapausasteluvulla n).

Khiin neliön jakauma on erikoistapaus gammajakaumasta, mikä nähdään seuraavasti. Jos $X \sim N(0, 1)$, niin muuttujan $Y = X^2$ tiheysfunktio on (vrt. esim. 2.8)

$$f_Y(y) = (\phi(\sqrt{y}) + \phi(-\sqrt{y})) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{-1/2} \exp(-\frac{1}{2}y),$$

kun $y > 0$. Tästä nähdään, että $X^2 \sim \text{Gam}(\frac{1}{2}, \frac{1}{2})$. Gammajakauman yhteenlaskuominaisuuden perusteella $X_1^2 + X_2^2 \sim \text{Gam}(\frac{1}{2} + \frac{1}{2}, \frac{1}{2})$, ja samaa päättelyä jatkamalla

$$X_1^2 + \dots + X_n^2 \sim \text{Gam}(\frac{n}{2}, \frac{1}{2}).$$

Toisin sanoen χ_n^2 on sama jakauma kuin $\text{Gam}(\frac{n}{2}, \frac{1}{2})$. Jos $\nu > 0$ ei ole kokonaisluku, niin määrittelemme, että χ_ν^2 jakauma on sama kuin gammajakauma $\text{Gam}(\frac{\nu}{2}, \frac{1}{2})$.

t-jakauma. Jos $Z \sim N(0, 1)$ ja $V \sim \chi_\nu^2$ (jollekin $\nu > 0$) ja $Z \perp V$, niin satunnaismuuttujalla

$$T = \frac{Z}{\sqrt{V/\nu}}$$

on jakauma, jota kutsutaan (Studentin) t -jakaumaksi ν :llä vapausasteella (engl. *degrees of freedom, df*), eli $T \sim t_\nu$.

F-jakauma. Jos $U \sim \chi_k^2$ ja $V \sim \chi_m^2$ ja $U \perp V$, niin sm:illa

$$Y = \frac{U/k}{V/m}$$

on jakauma, jota kutsutaan F -jakaumaksi parametreilla k (osoittajan vapausasteluku) ja m (nimittäjän vapausasteluku), eli $Y \sim F_{k,m}$.

Tarkastelemme myöhemmin lähemmin joitakin t - ja F -jakauman ominaisuuksia.

Luku 5

Epäyhtälöitä

Epäyhtälöt ovat yksi matematiikan voimakkaimmista työvälineistä. Siinä missä yhtälö $a = b$ kertoo sen, että kaksi ehkä näennäisesti erilaista asiaa ovatkin samoja, epäyhtälö $a \leq b$ saattaa antaa keinon analysoida monimutkaista tilannetta paljon helpommin ymmärrettävän asian avulla.

Todennäköisyyslaskennan epäyhtälöitä käytetään tyypillisesti teorian apuna, kuten esim. raja-arvotulosten johtamiseen, mutta niille löytyy toisenlaisiakin sovelluksia. Esimerkiksi niitä voidaan käyttää työkaluna satunnaistettujen algoritmien vaativuuden analysoinnissa.

Kuten yhtälöiden kanssa, myös epäyhtälöiden kanssa työskennellessä täytyy kiinnittää huomiota paitsi epäyhtälön kummankin puolen muotoon myös siihen minkä ehtojen vallitessa se on voimassa. Näiden asioiden lisäksi epäyhtälöiden kohdalla pitää kiinnittää erityistä huomiota siihen, kumpi suuruusjärjestys puolien välillä vallitsee. Usein matemaattinen argumentointi perustuu siihen, että muotoa $a \leq b$ oleva tulos johdetaan pitkän epäyhtälöketjun

$$a \leq a_1 \leq \dots \leq a_k \leq b$$

avulla. Jos tässä ketjussa yksikin \leq -merkki on tullut kirjoitettua vahingossa väärin päin, niin koko argumentilta putoaa pohja pois.

5.1 Markovin ja Tšebyševin epäyhtälöt

Lause 5.1 (Markovin epäyhtälö). *Olkoon $X \geq 0$ sm, jonka odotusarvo on olemassa. Tällöin*

$$P(X \geq a) \leq \frac{EX}{a}, \quad \forall a > 0$$

Todistus. Määritellään sm Y kaavalla

$$Y = a1_{[a, \infty)}(X).$$

Sm Y on diskreetti, sen mahdolliset arvot ovat 0 ja a , ja lisäksi $Y \leq X$. Niinpä

$$EX \geq EY = 0 \cdot P(Y = 0) + a \cdot P(Y = a) = a P(X \geq a). \quad \square$$

Lause 5.2 (Tšebyševin epäyhtälö). *Olkoon X sm, jonka odotusarvo on μ ja varianssi σ^2 on olemassa Tällöin*

$$P[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2} \quad \forall t > 0.$$

Erityisesti, jos $\sigma^2 > 0$, niin

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

Todistus. Sovelletaan Markovin epäyhtälöä sm:aan $Y = (X - \mu)^2$. Jos $t > 0$, niin

$$P[|X - \mu| \geq t] = P[(X - \mu)^2 \geq t^2] \leq \frac{\sigma^2}{t^2}.$$

Toinen epäyhtälö seuraa valitsemalla $t = k\sigma$. □

Tšebyševin epäyhtälön sovelluksena seuraavaksi todistetaan heikko suurten lukujen laki (engl. *weak law of large numbers, WLLN*). Lauseessa esiintyvää satunnaismuuttujien suppenemisen lajia kutsutaan stokastiseksi konvergenssiksi (eli stokastiseksi suppenemiseksi).

Lause 5.3 (Heikko suurten lukujen laki). *Olkoon X_1, X_2, \dots jono riippumattomia sm:ia, joilla on sama odotusarvo $\mu = EX_i$ ja varianssi $\sigma^2 = \text{var } X_i < \infty$. Tällöin keskiarvo*

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

konvergoi stokastisesti kohti arvoa μ , eli

$$P(|\bar{X}_n - \mu| \geq \epsilon) \xrightarrow{n \rightarrow \infty} 0, \quad \text{kaikilla } \epsilon > 0.$$

Todistus. Nyt

$$E\bar{X}_n = \mu, \quad \text{var } \bar{X}_n = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Tšebyševin epäyhtälön nojalla

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}. \quad \square$$

Jos X_1, X_2, \dots on jono riippumattomia ja samoin jakautuneita sm:ia (engl. *independent, identically distributed; i.i.d.*), niin hieman monimutkaisemmilla tarkasteluilla voitaisiin todistaa, että ehto varianssin olemassaolosta on edellisessä lauseessa tarpeeton: mikäli $\mu = EX_i$ on olemassa, niin keskiarvo \bar{X}_n konvergoi stokastisesti kohti odotusarvoa μ . Tällöin pätee myös vahvempi tulos, ns. vahva suurten lukujen laki (engl. *strong law of large numbers, SLLN*), jonka mukaan

$$\bar{X}_n \rightarrow \mu$$

melkein varmasti eli todennäköisyydellä yksi. Tämä tarkoittaa sitä, että

$$\bar{X}_n(\omega) \rightarrow \mu$$

kaikilla ω , jotka eivät kuulu poikkeusjoukkoon $N \subset \Omega$, jonka tn on nolla.

5.2 Konveksit funktiot ja Jensenin epäyhtälö

Funktio on konvekksi, jos sen kuvaaja jää jokaisen jäniteensä alapuolelle. Funktio on konkaavi, jos sen kuvaaja on jokaisen jäniteensä yläpuolella. Tämä asia muotoillaan täsmällisemmin alla olevassa määritelmässä. Yleensä funktiot eivät ole konvekseja eikä konkaaveja.

Seuraavassa määritelmässä sana väli tarkoittaa mitä tahansa suljettua, avointa tai puoliavointa äärellistä tai ääretöntä väliä, ts. väli on muotoa (a, b) , $[a, b)$, $(a, b]$ tai $[a, b]$ oleva joukko, jossa $a \leq b$ ja tapaukset $a = -\infty$ tai $b = \infty$ sallitaan. Jos I on väli ja $x, y \in I$, niin myös pisteitä x ja y yhdistävän janan pisteet $\lambda x + (1 - \lambda)y$, $0 \leq \lambda \leq 1$ kuuluvat välille I . Konveksin joukon karakterisoi juuri tämä ominaisuus, ja reaaliakselin konveksit joukot ovat välejä.

Määritelmä 5.1 (Konvekssi funktio, konkaavi funktio). Olkoon $I \subset \mathbb{R}$ väli. Funktio $g : I \rightarrow \mathbb{R}$ on *konvekssi*, jos

$$g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y), \quad \text{kaikilla } x, y \in I \text{ ja } 0 \leq \lambda \leq 1.$$

Funktio $g : I \rightarrow \mathbb{R}$ on *konkaavi*, jos $-g$ on konvekssi.

Konveksisuuden määritelmässä voitaisiin yhtä hyvin rajoittaa tarkastelemaan väliä $0 < \lambda < 1$ koska tapaukset $\lambda = 0$ ja $\lambda = 1$ ovat triviaaleja. Funktiot x^2 , e^x ja $|x|$ ovat konvekseja koko reaaliakselilla; $1/x$ on konvekssi, kun $x > 0$, ja $\log x$ on konkaavi funktio, kun $x > 0$. Affiini funktio $a + bx$ on sekä konvekssi että konkaavi koko reaaliakselilla.

Seuraavassa lauseessa esiintyy erotusosamääriä, jotka kannattaa tulkita geometrisesti g :n jäniteiden kulmakertoimien avulla. Piirrä itse kuva. Lauseen tarkistus jätetään lukijalle.

Lause 5.4. *Olkoon I avoin väli. Funktio $g : I \rightarrow \mathbb{R}$ on konvekssi, jos ja vain jos kaikilla $a < b < c$, joille $a, b, c \in I$ pätee epäyhtälö*

$$\frac{g(b) - g(a)}{b - a} \leq \frac{g(c) - g(b)}{c - b}. \quad (5.1)$$

Lause 5.5 (Konveksisuustarkistimia). *Olkoon $I \subset \mathbb{R}$ avoin väli, ja $g : I \rightarrow \mathbb{R}$ funktio.*

- (a) *Jos g on jatkuvasti derivoituva ja g' on kasvava (ja molemmat ominaisuudet ovat voimassa koko I :llä), niin g on konvekssi.*
- (b) *Jos g on kahdesti derivoituva I :llä, ja $g''(x) \geq 0$ kaikilla $x \in I$, niin g on konvekssi.*

Todistus. (a): Jos $a, b, c \in I$ ja $a < b < c$, niin

$$\begin{aligned} \frac{g(b) - g(a)}{b - a} &= \frac{1}{b - a} \int_a^b g'(x) \, dx \leq g'(b) \\ &\leq \frac{1}{c - b} \int_b^c g'(x) \, dx = \frac{g(c) - g(b)}{c - b}. \end{aligned}$$

joten g on konvekssi.

(b): Koska $g'' \geq 0$, niin g' on kasvava I :llä. □

Jos $g : I \rightarrow \mathbb{R}$ on konvekksi, niin konveksin funktion määritelmästä saadaan helposti induktiolla johdettua epäyhtälö

$$g\left(\sum_{i=1}^n p_i x_i\right) \leq \sum_{i=1}^n p_i g(x_i), \quad (5.2)$$

joka pitää paikkansa, kun kukin $x_i \in I$ ja kukin $p_i \geq 0$ ja $\sum_i p_i = 1$. Tällaista lineaarikombinaatiota $x = \sum_i p_i x_i$, jossa painot p_i ovat ei-negatiivisia ja summautuvat ykköseksi, kutsutaan pisteiden x_1, \dots, x_n *konveksiksi kombinaatioksi*. Tietenkin myös konvekssi kombinaatio $x \in I$.

Jensenin epäyhtälö on edellisen epäyhtälön yleistys, kun epäyhtälössä esiintyvä diskreetti jakauma ptfn:llä $f(x_i) = p_i$ korvataan mielivaltaisella jakaumalla. Jensenin epäyhtälön todistaminen on helppoa, kunhan ensin todistamme seuraavan apulauseen. Siinä todetaan, että konveksin funktion kuvaaja on jokaisen tangenttinsa yläpuolella.

Lause 5.6. *Jos $I \subset \mathbb{R}$ on avoin väli ja $g : I \rightarrow \mathbb{R}$ on konvekssi funktio, ja $m \in I$, niin voidaan valita luku k siten, että*

$$g(x) \geq g(m) + k(x - m), \quad \forall x \in I.$$

Todistus. Rajoitutaan todistamaan väite siinä tapauksessa, jossa g on jatkuvasti derivoituva I :llä ja g' on kasvava funktio. Jos $x \geq m$ ja $x \in I$, niin

$$g(x) - g(m) = \int_m^x g'(u) du \geq g'(m)(x - m).$$

Jos taas $x < m$ ja $x \in I$, niin

$$g(m) - g(x) = \int_x^m g'(u) du \leq g'(m)(m - x).$$

Siis väite pitää paikkansa, kun valitaan kulmakertoimeksi $k = g'(m)$.

Tämä lause pitää paikkansa myös ilman oletusta g :n derivoituvuudesta. Todistus löytyy useista analyysin oppikirjoista, kuten esim. Billingsleyn teoksesta *Probability and Measure*. \square

Lause 5.7 (Jensenin epäyhtälö). *Olkoon $I \subset \mathbb{R}$ avoin väli, ja $g : I \rightarrow \mathbb{R}$ konvekssi funktio. Jos X on sm, jonka arvot ovat I :ssä (tn:llä yksi), ja EX sekä $Eg(X)$ ovat olemassa, niin*

$$g(EX) \leq Eg(X) \quad (5.3)$$

Todistus. Koska $\mu = EX \in I$, niin voidaan valita luku k siten, että

$$g(x) \geq g(\mu) + k(x - \mu), \quad \forall x \in \mu.$$

Siis todennäköisyydellä yksi

$$g(X) \geq g(\mu) + k(X - \mu),$$

ja Jensenin epäyhtälö seuraa ottamalla odotusarvo puolittain. \square

Esimerkki 5.1. Funktio x^2 on konvekssi, joten Jensenin epäyhtälön mukaan

$$(EX)^2 \leq E(X^2),$$

mikäli kyseiset momentit ovat olemassa. Tästä seuraa arvio

$$\text{var } X = E(X^2) - (EX)^2 \geq 0,$$

minkä tietenkin tiesimme ennestään. △

5.3 Hölderin epäyhtälö

Lause 5.8. Olkoot $p, q > 1$ sellaisia lukuja, että

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Tällöin

$$E|XY| \leq [E|X|^p]^{\frac{1}{p}} [E|Y|^q]^{\frac{1}{q}}$$

Todistus. Väite on triviaalisti tosi, jos $E|X|^p = \infty$ tai $E|Y|^q = \infty$ tai jos $E|X|^p = 0$ tai $E|Y|^q = 0$. Tämän takia oletamme, että nämä molemmat odotusarvot ovat äärellisiä ja nollaa suurempia.

Ensin todistetaan positiivisia reaalityyppisiä lukuja koskeva epäyhtälö. Olkoot $a, b > 0$ mielivaltaisia, ja määritellään s ja t siten, että

$$a = \exp\left(\frac{1}{p}s\right), \quad b = \exp\left(\frac{1}{q}t\right).$$

Koska $\exp(x) = e^x$ on konvekssi, on

$$ab = \exp\left(\frac{1}{p}s + \frac{1}{q}t\right) \leq \frac{1}{p}e^s + \frac{1}{q}e^t = \frac{a^p}{p} + \frac{b^q}{q}.$$

Sama epäyhtälö pätee myös, jos $a = 0$ tai $b = 0$, kun sovimme, että nolla korotettuna positiiviseen potenssiin on nolla.

Valitaan

$$u = [E|X|^p]^{\frac{1}{p}}, \quad v = [E|Y|^q]^{\frac{1}{q}},$$

ja otetaan puolittain odotusarvo epäyhtälöstä

$$\left| \frac{XY}{uv} \right| \leq \frac{1}{p} \left| \frac{X}{u} \right|^p + \frac{1}{q} \left| \frac{Y}{v} \right|^q,$$

jolloin saadaan

$$E \left| \frac{XY}{uv} \right| \leq \frac{1}{p} \frac{E|X|^p}{E|X|^p} + \frac{1}{q} \frac{E|Y|^q}{E|Y|^q} = 1. \quad \square$$

Hölderin epäyhtälön helppona sovelluksena (HT) voidaan todistaa, että

$$[E|X|^s]^{\frac{1}{s}}, \quad s > 0$$

on kasvava funktio $(0, \infty) \rightarrow [0, \infty]$. Tätä tulosta kutsutaan *Ljapunovin epäyhtälöksi*. Erityisesti

$$E|X|^s < \infty \quad \Rightarrow \quad E|X|^r < \infty \text{ kaikilla } 0 < r \leq s.$$

Minkowskin epäyhtälö toteaa, että

$$[E|X + Y|^p]^{\frac{1}{p}} \leq [E|X|^p]^{\frac{1}{p}} + [E|Y|^p]^{\frac{1}{p}}, \quad \text{kaikilla } p \geq 1. \quad (5.4)$$

(Jos jokin odotusarvoista ei ole olemassa, niin vastaavalle termille käytetään arvoa ∞). Minkowskin epäyhtälö seuraa kohtalaisen helposti kolmioepäyhtälöstä sekä Hölderin epäyhtälöstä (HT).

5.4 Cauchyn-Schwarzin epäyhtälö, kovarianssi ja korrelaatio

Cauchyn-Schwarzin epäyhtälöksi sanotaan Hölderin epäyhtälön sitä tapausta, jossa $p = q = 2$, jolloin saadaan

$$E|XY| \leq \sqrt{EX^2} \sqrt{EY^2}.$$

Tämä epäyhtälö voi joissakin tapauksissa olla voimassa muodossa $E|XY| \leq \infty$, mikä ei kerro mitään mielenkiintoista. Jos sekä $EX^2 < \infty$ että $EY^2 < \infty$, niin silloin yläraja on äärellinen, ja saadaan tulos

$$|E(XY)| \leq E|XY| \leq \sqrt{EX^2} \sqrt{EY^2}. \quad (5.5)$$

Tässä ensimmäinen epäyhtälö seurasi lauseesta 3.4.

Oletetaan nyt, että satunnaismuuttujilla X ja Y on äärelliset varianssit, $\sigma_X^2 = \text{var } X$ ja $\sigma_Y^2 = \text{var } Y$. Tällöin niiden **kovarianssi** $\text{cov}(X, Y)$ määritellään odotusarvona

$$\text{cov}(X, Y) = E[(X - EX)(Y - EY)]. \quad (5.6)$$

Tämä odotusarvo on äärellinen, sillä kun Cauchyn-Schwarzin epäyhtälöä sovelletaan satunnaismuuttujien $X - EX$ ja $Y - EY$ tuloon, niin saadaan epäyhtälö

$$|\text{cov}(X, Y)| \leq \sigma_X \sigma_Y.$$

Tässä ylärajana on muuttujien keskihajontojen tulo.

Mikäli molemmat keskihajonnat σ_X ja σ_Y ovat aidosti positiivisia, niin X :n ja Y :n **korrelaatiokerroin** (jota merkitään usein $\text{corr}(X, Y)$ tai ρ_{XY}) määritellään kaavalla

$$\text{corr}(X, Y) = \rho_{XY} = \rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}. \quad (5.7)$$

Cauchyn-Schwarzin epäyhtälön nojalla

$$|\text{corr}(X, Y)| \leq 1.$$

Jos $\text{corr}(X, Y) > 0$, jolloin myös $\text{cov}(X, Y) > 0$, niin sanotaan että X ja Y ovat *positiivisesti korreloituneita*. Jos $\text{corr}(X, Y) < 0$, jolloin myös $\text{cov}(X, Y) < 0$, niin sanotaan että X ja Y ovat *negatiivisesti korreloituneita*. Jos $\text{corr}(X, Y) = 0$, jolloin myös $\text{cov}(X, Y) = 0$, niin sanotaan, että X ja Y eivät korreloi (lineaarisesti) tai että ne ovat (lineaarisesti) korreloimattomia.

5.5 Momenttiemäfunktiota koskevia epäyhtälöitä

Palautetaan mieleen määritelmät. Satunnaismuuttujan X momenttiemäfunktio M_X ja sen kumulanttiemäfunktio ovat

$$M_X(t) = Ee^{tX}, \quad K_X(t) = \ln M_X(t)$$

niissä pisteissä, joissa odotusarvo on äärellinen.

Lause 5.9 (Emäfunktioiden ominaisuuksia). *a) Häntätodennäköisyyksille pätevät seuraavat ylärajat kaikilla a ,*

$$P(X \geq a) \leq \inf_{t>0} e^{-ta} M_X(t), \quad \text{ja} \quad P(X \leq a) \leq \inf_{t<0} e^{-ta} M_X(t).$$

b) Kumulanttiemäfunktio ja momenttiemäfunktio ovat konvekseja funktioita.

Todistus. Kohta (a): kun $t > 0$ niin funktio $x \mapsto e^{tx}$ on aidosti kasvava, joten Markovin epäyhtälön avulla

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{Ee^{tX}}{e^{ta}}.$$

Koska tämä arvio pätee kaikilla $t > 0$, niin se pätee vielä silloinkin, kun yläraja korvataan sen minimillä alueessa $t > 0$. Toinen epäyhtälöistä todistetaan soveltamalla ensimmäistä sm:an $Y = -X$.

Kumulanttiemäfunktion konveksisuus seuraa Hölderin epäyhtälöstä (HT). Tämän jälkeen momenttiemäfunktion konveksisuus seuraa siitä, että e^x on kasvava ja konvekssi funktio. \square

Luku 6

Kaksiulotteiset jakaumat

6.1 Kaksiulotteinen satunnaisvektori ja sen jakauma

Määritelmä 6.1 (Satunnaisvektori). Olkoot X ja Y samalla perusjoukolla Ω määriteltyjä satunnaismuuttujia, $X : \Omega \rightarrow \mathbb{R}$ ja $Y : \Omega \rightarrow \mathbb{R}$. Tällöin (X, Y) on kaksiulotteinen *satunnaisvektori* (lyh. sv) (engl. *two-dimensional random vector*; *bivariate random vector*).

Kaksiulotteinen sv (X, Y) on siis kuvaus (funktio)

$$(X, Y) : \Omega \rightarrow \mathbb{R}^2, \quad \text{sitte, että} \quad (X, Y)(\omega) = (X(\omega), Y(\omega)).$$

Satunnaisvektorista (X, Y) käytetään myös nimitystä kaksiulotteinen satunnaismuuttuja (engl. *two-dimensional random variable*) tai satunnaismuuttujapari.

Määritelmä 6.2 (Satunnaisvektorin jakauma, yhteisjakauma). Satunnaisvektorin (X, Y) jakauma eli satunnaismuuttujien X ja Y *yhteisjakauma* (engl. *joint distribution*) on \mathbb{R}^2 :n osajoukoilla A määritelty funktio

$$P((X, Y) \in A) = P(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\}).$$

Määritelmä 6.3 (Satunnaisvektorin kertymäfunktio, yhteiskertymäfunktio). Satunnaisvektorin (X, Y) kertymäfunktio eli satunnaismuuttujien X ja Y *yhteiskertymäfunktio* (lyh. ykf) (engl. *joint (cumulative) distribution function*, *joint cdf*) on

$$F_{X,Y}(x, y) = F(x, y) = P(X \leq x, Y \leq y), \quad x, y \in \mathbb{R}.$$

Kuten yksiulotteisessa tapauksessa, niin myös kaksi- ja useampiulotteisessa tapauksessa kyseessä olevat satunnaismuuttujat voidaan ilmoittaa kertymäfunktion (tai muun jakauman esityksen) alaindeksillä, ellei asiayhteyden perusteella muuten ole selvää, minkä muuttujien yhteisjakauman esityksestä on kyse.

Satunnaismuuttujien X ja Y yhteisjakauma tietenkin määrää niiden yhteiskertymäfunktion, sillä

$$P(X \leq x, Y \leq y) = P((X, Y) \in (-\infty, x] \times (-\infty, y]).$$

Kääntäen, mikäli tunnetaan ykf, niin on mahdollista osoittaa, että sen avulla voidaan laskea *kaikkien* joukkojen A todennäköisyydet kaavassa

$$P((X, Y) \in A), \quad A \subset \mathbb{R}^2,$$

eli yhteisjakauma määräytyy yhteiskertymäfunktioista.

Lause 6.1. *Yhteiskertymäfunktio määrää jakauman.*

Todistus. Todistus siivutetaan, sillä se vaatisi mittateoriaa. \square

Esimerkki 6.1. Olkoot $x_1 < x_2$ ja $y_1 < y_2$. Lasketaan ykf:n F avulla

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = P((X, Y) \in (x_1, x_2] \times (y_1, y_2]).$$

Nyt (piirrä kuva!)

$$\{X \leq x_2, Y \leq y_2\} = \{x_1 < X \leq x_2, y_1 < Y \leq y_2\} \cup C,$$

jossa joukot ovat erillisiä, ja $C = A \cup B$, jossa

$$A = \{X \leq x_1, Y \leq y_2\}, \quad B = \{X \leq x_2, Y \leq y_1\}.$$

Niinpä

$$\begin{aligned} P(x_1 < X \leq x_2, y_1 < Y \leq y_2) &= F(x_2, y_2) - P(C) \\ &= F(x_2, y_2) - (P(A) + P(B) - P(A \cap B)) \\ &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \end{aligned}$$

\triangle

Kun tarkastellaan kahta satunnaismuuttujaa X ja Y , niin yksittäisen sm:n X (tai Y) (yksiulotteista) jakaumaa kutsutaan sen *reunajakaumaksi* (engl. *marginal distribution*). Näiden jakaumien kertymäfunktioit eli X :n ja Y :n *reunakertymäfunktioit* (engl. *marginal cdf's*) saadaan yhteiskertymäfunktioista $F_{X,Y}$ raja-arvoina,

$$F_X(x) = P(X \leq x, Y \in \mathbb{R}) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y) = F_{X,Y}(x, \infty),$$

ja vastaavasti

$$F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y) = F_{X,Y}(\infty, y).$$

Ykf:n sijasta kaksiulotteisia jakaumia käsitellään yleensä niiden yhteispistetodennäköisyysfunktioiden tai yhteistilheysfunktioiden avulla. Yhteisjakauma kuvaillaan usein kertolaskusäännön eli ketjusäännön avulla, mitä varten pitää ensin kuvailla yhden muuttujan reunajakauma sekä toisen muuttujan ehdollinen jakauma.

6.2 Diskreetti kaksiulotteinen jakauma

Määritelmä 6.4 (Diskreetti (yhteis)jakauma, yhteispistetodennäköisyysfunktio). Satunnaisektori (X, Y) on diskreetti, eli satunnaismuuttujilla X ja Y on *diskreetti yhteisjakauma*, jos sv:lla (X, Y) on enintään numeroituva arvojoukko,

tai tarkemmin sanoen, on olemassa enintään numeroituva joukko $S \subset \mathbb{R}^2$ siten, että

$$P((X, Y) \in S) = 1.$$

Joukkoa S voidaan kutsua satunnaisvektorin (X, Y) arvojoukoksi tai sen jakauman kantajaksi. Tällöin sv:n (X, Y) pistetodennäköisyysfunktio eli satunnaismuuttujien X ja Y yhteispistetodennäköisyysfunktio (lyh. yptnf) (engl. *joint probability (mass) function, joint pmf*) on

$$f(x, y) = f_{X,Y}(x, y) = P(X = x, Y = y), \quad x, y \in \mathbb{R}.$$

Yptnf:llä on analogiset ominaisuudet yksiulotteisen jakauman ptnf:n kanssa.

Lause 6.2. *Olkoon f ja S kuten määritelmässä. Tällöin*

a) $0 \leq f(x, y) \leq 1$, ja $f(x, y) = 0$, kun $(x, y) \notin S$.

b) f määrää yhteisjakauman kaavalla

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

Todistus. Idea on sama kuin yksiulotteisessa tapauksessa. Kohdassa b) summataan yhteen korkeintaan numeroituva määrä nollassa poikkeavia termejä, jotka ovat kaikki positiivisia, minkä takia summausjärjestyksestä ei tarvitse välittää, ja merkintä on mielekäs. \square

Erityisesti X :n ja Y :n pistetodennäköisyydet saadaan summaamalla toinen muuttujista pois yptnf:stä,

$$f_X(x) = P(X = x) = P(X = x, Y \in \mathbb{R}) = \sum_y f(x, y),$$

$$f_Y(y) = P(Y = y) = P(X \in \mathbb{R}, Y = y) = \sum_x f(x, y).$$

Lisäksi yptnf summautuu ykköseksi, sillä

$$1 = P((X, Y) \in S) = \sum_{(x,y) \in S} f(x, y).$$

Seuraava lause on lauseen 2.5 kaksiulotteinen vastine. Sen todistus sivuutetaan.

Lause 6.3. *Olkoon $S \subset \mathbb{R}^2$ äärellinen tai numeroituvasti ääretön joukko, ja olkoon $f : S \rightarrow \mathbb{R}$ ei-negatiivinen funktio siten, että*

$$\sum_{(x,y) \in S} f(x, y) = 1.$$

Tällöin se on eräiden diskreettien satunnaismuuttujien X ja Y yptnf.

Seuraavassa lauseessa näytetään, että diskreetin yhteisjakauman tapauksessa sekä X ja Y ovat diskreettejä. Edellä on johdettu kaavat niiden (reuna)pistetodennäköisyysfunktioille.

Lause 6.4. *Satunnaismuuttujien X ja Y yhteisjakauma on diskreetti silloin ja vain silloin, kun reunajakaumat ovat diskreettejä.*

Todistus. Oletetaan ensin, että reunajakaumat ovat diskreettejä eli että sekä X että Y ovat diskreettejä satunnaismuuttujia. Olkoot niiden arvojoukot S_X ja S_Y , ts. $P(X \in S_X) = 1$ ja $P(Y \in S_Y) = 1$ ja $S_X \subset \mathbb{R}$ ja $S_Y \subset \mathbb{R}$ ovat korkeintaan numeroituvia joukkoja. Tällöin joukko $S = S_X \times S_Y \subset \mathbb{R}^2$ on myös korkeintaan numeroituva, ja helpolla laskulla nähdään, että

$$P((X, Y) \in S) = 1.$$

Siis yhteisjakauma on diskreetti.

Oletetaan seuraavaksi, että yhteisjakauma on diskreetti. Olkoon S korkeintaan numeroituva joukko siten, että $P((X, Y) \in S) = 1$. Tässä

$$S = \{(x'_1, y'_1), (x'_2, y'_2), \dots\}.$$

Olkoon nyt $S_X = \{x_1, x_2, \dots\}$, jossa x_1, x_2, \dots on jonossa x'_1, x'_2, \dots esiintyvät eri suuret arvot, ts. S_X on joukon S projektio x -akselille. Määritellään vastaavasti $S_Y = \{y_1, y_2, \dots\}$, jossa y_1, y_2, \dots ovat jonossa y'_1, y'_2, \dots esiintyvät eri suuret arvot. Nyt S_X ja S_Y ovat korkeintaan numeroituvia joukkoja, ja

$$P(X \notin S_X) \leq P(X \notin S_X \text{ tai } Y \notin S_Y) = P((X, Y) \notin S) = 0.$$

Siis $P(X \in S_X) = 1$, ja vastaavalla laskulla $P(Y \in S_Y) = 1$, joten X ja Y ovat diskreettejä. \square

Esimerkki 6.2. Reunajakaumat eivät määrää yhteisjakaumaa. Tarkastellaan diskreettiä yhteisjakaumaa, jonka yptnf:n kantaja S on

$$\{0, 1\} \times \{0, 1\}$$

ja yptnf on $f(x, y)$. Tällainen yptnf voidaan esittää taulukon muodossa seuraavasti,

$x \backslash y$	0	1	\sum_y
0	$f(0, 0)$	$f(0, 1)$	$f_X(0)$
1	$f(1, 0)$	$f(1, 1)$	$f_X(1)$
\sum_x	$f_Y(0)$	$f_Y(1)$	

Huomaa, että reunajakaumien ptnf:t ovat taulukon reunasummia; esim.

$$f_X(x) = \sum_y f(x, y) = f(x, 0) + f(x, 1), \quad x = 0, 1.$$

(Termi *reunajakauma* selittyy tällaisesta huomiosta.) Nyt tietenkin

$$X \sim \text{Bernoulli}(f_X(1)) \quad \text{ja} \quad Y \sim \text{Bernoulli}(f_Y(1)).$$

Moni erilainen yhteisjakauma tuottaa samat reunajakaumat. Esim. reunajakaumat $X \sim \text{Bernoulli}(1/2)$ ja $X \sim \text{Bernoulli}(1/2)$ saadaan valitsemalla $f(0, 0) = f(1, 1) = a$ ja $f(0, 1) = f(1, 0) = \frac{1}{2} - a$, jossa a on mikä tahansa

luku väliltä $0 \leq a \leq \frac{1}{2}$. Taulukon muodossa esitettynä yptnf näyttää seuraavalta,

$x \setminus y$	0	1	Σ
0	a	$\frac{1}{2} - a$	$\frac{1}{2}$
1	$\frac{1}{2} - a$	a	$\frac{1}{2}$
Σ	$\frac{1}{2}$	$\frac{1}{2}$	

△

6.3 Trinomijakauma

Käsitlemme nyt erästä esimerkkiä diskreetistä kaksiulotteisesta jakaumasta. Olkoot $p_1, p_2 > 0$ lukuja siten, että $p_1 + p_2 < 1$. Olkoon $n \geq 1$ positiivinen kokonaisluku. Tällöin

$$f(x, y) = \binom{n}{x, y, n-x-y} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}, \quad x, y \geq 0, \quad x+y \leq n.$$

on trinomijakauman $\text{Trin}(n, (p_1, p_2))$ yptnf. (Trinomijakaumaa on erikoistapaus myöhemmin esiin tulevasta multinomijakaumasta.) Se, että kyseessä on yptnf nähdään helposti multinomikaavan (1.8) avulla, sillä

$$\begin{aligned} 1 &= [p_1 + p_2 + (1-p_1-p_2)]^n \\ &= \sum_{x,y} \binom{n}{x, y, n-x-y} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}, \end{aligned}$$

jossa summa otetaan niiden kokonaislukuparien (x, y) yli, joille $x, y \geq 0$ ja $x+y \leq n$. Lisäksi jokainen summan termeistä on positiivinen.

Trinomijakaumalla on toistokokeen kautta syntyvä tulkinta. Tarkastellaan koetta, jonka yhdessä toistossa saadaan täsmälleen yksi tuloksista A, B tai C . Olkoon yhdessä toistossa

$$p_1 = P(A), \quad p_2 = P(B), \quad p_3 = P(C) = 1 - p_1 - p_2.$$

Toistetaan tätä koetta riippumattomasti n kertaa, ja annetaan satunnaismuuttujan Z_i kertoa lopputulos toistossa numero i siten, että

$$Z_i = \begin{cases} 1 & \text{jos tulos on } A, \\ 2 & \text{jos tulos on } B, \\ 3 & \text{jos tulos on } C. \end{cases}$$

Tällöin

$$P(Z_i = z) = p_z, \quad \text{kun } z = 1, 2, 3.$$

Määritellään, että sm X on niiden i lkm, joilla $Z_i = 1$ (eli lopputuloksen A lkm) ja sm Y on niiden i lkm, joilla $Z_i = 2$ (eli lopputuloksen B lkm). Tällöin niitä i , joilla $Z_i = 3$ (eli lopputuloksen C lkm) on $n - X - Y$.

Jos (z_1, z_2, \dots, z_n) on arvoista 1, 2 tai 3 koostuva jono, niin

$$\begin{aligned} P(Z_1 = z_1, Z_2 = z_2, \dots, Z_n = z_n) &= p_{z_1} p_{z_2} \dots p_{z_n} \\ &= p_1^x p_2^y p_3^{n-x-y}. \end{aligned}$$

Tässä x on lopputuloksen A lkm annetussa jonossa eli niiden i lkm, joilla $z_i = 1$, ja y on lopputuloksen B lkm eli niiden i lkm, joilla $z_i = 2$. Tällöin lopputuloksen C lkm eli niiden i lkm, joille $z_i = 3$ on $n - x - y$. Toisaalta, jos kiinnitetään kokonaisluvut x ja y siten, että $x, y \geq 0$ ja $x + y \leq n$, niin sellaisia jonoja (z_1, z_2, \dots, z_n) jotka koostuvat arvoista 1, 2 tai 3, joissa tulosta 1 on x kpl ja tulosta 2 on y kpl on multinomikertoimen

$$\binom{n}{x, y, n-x-y}$$

ilmaisema määrä, sillä kukin tällainen jono vastaa täsmälleen yhtä tapaa osittaa n -alkionen joukko kolmeen osaan siten, että ensimmäiseen osaan tulee x kpl alkioita ja toiseen osaan y kpl alkioita (jolloin kolmanteen osaan jää $n - x - y$ kpl alkioita). Jokaiselle tällaiselle jonolle

$$P(Z_1 = z_1, Z_2 = z_2, \dots, Z_n = z_n) = p_1^x p_2^y p_3^{n-x-y}.$$

Tällä tavalla nähdään, että

$$P(X = x, Y = y) = \binom{n}{x, y, n-x-y} p_1^x p_2^y p_3^{n-x-y},$$

jossa $p_3 = 1 - p_1 - p_2$. Toisin sanoen satunnaismuuttujien X ja Y yhteisjakauma on trinomijakauma otoskokoparametrilla n ja todennäköisyysparametreilla (p_1, p_2) .

Toistokoetulkinna valossa on ilmeistä, että sm:n X reunajakauma on $X \sim \text{Bin}(n, p_1)$. (Ajattele toistokoetta, jossa onnistutaan, kun saadaan lopputulos A ja muuten epäonnistutaan.) Vastaavasti, $Y \sim \text{Bin}(n, p_2)$.

Kuvassa 6.1 trinomijakauman yptnf on piirretty kahdella eri tavalla. Ensimmäisessä tavassa funktio $z = f(x, y)$ on esitetty perspektiivikuvana. Toisessa tavassa kolmas dimensio esitetään piirtosymbolin koon avulla: funktion $f(x, y)$ arvo on verrannollinen pisteeseen (x, y) piirretyn ympyrän pinta-alaan.

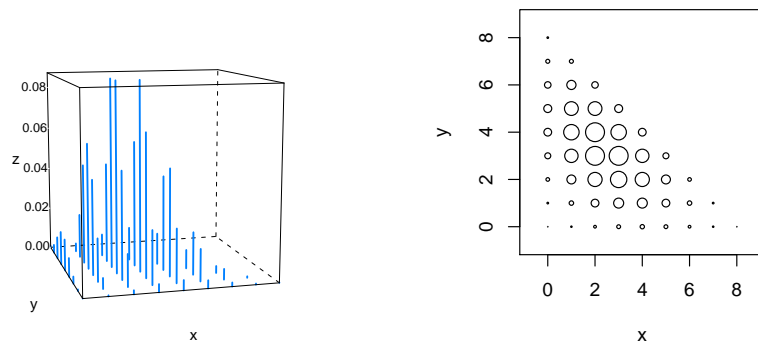
6.4 Jatkuva kaksiulotteinen jakauma

Määritelmä 6.5 (Jatkuva (yhteis)jakauma, yhteistiheysfunktio). Satunnaisvektorilla (X, Y) on jatkuva jakauma, eli satunnaismuuttujilla X ja Y on jatkuva yhteisjakauma, mikäli on olemassa \mathbb{R}^2 :lla määritelty funktio $f \geq 0$ siten, että

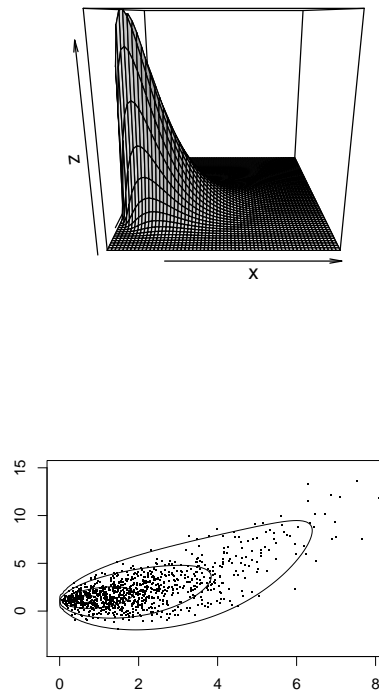
$$P((X, Y) \in B) = \iint_B f(x, y) dx dy, \quad \text{kaikille } B \subset \mathbb{R}^2. \quad (6.1)$$

Tällöin $f = f_{X,Y}$ on sv:n (X, Y) tiheysfunktio, eli satunnaismuuttujien X ja Y yhteistiheysfunktio (lyh. ytf) (engl. *joint (probability) density function, joint pdf*).

Kuva 6.1 Trinomijakauman $\text{Bin}(n, (p_1, p_2))$ yptfn, kun $n = 8$ ja $p_1 = 0.3$ ja $p_2 = 0.4$.



Kuva 6.2 Erilaisia tapoja havainnollistaa jatkuvaa jakaumaa: (a) perspektiivikuva, (b) tiheysfunktion tasa-arvokäyriä sekä jakaumasta simuloitu pisteparvi.



Kaava (6.1) ilmaistaan toisinaan heuristisella merkinnällä

$$P(X \in dx, Y \in dy) = f(x, y) dx dy.$$

Tässä merkinnän $P(X \in dx, Y \in dy)$ voidaan ajatella tarkoittavan samaa kuin $P(x \leq X \leq x + dx, y \leq Y \leq y + dy)$, jossa dx ja dy ovat hyvin pieniä positiivisia lukuja. Kuten yksiulotteisessa myös kaksiulotteisessa tapauksessa tiheysfunktion arvolle pisteessä (x, y) voidaan antaa frekvenssitulkinta.

Kaksiulotteista jatkuvaa jakaumaa voidaan havainnollistaa erilaisilla tavoilla. Yksi mahdollisuus on piirtää perspektiivikuva funktiosta $z = f(x, y)$. Toinen mahdollisuus on piirtää xy -tasoon funktion *tasa-arvokäyriä* (engl. *level curve*, *contour line*). Tasoon c liittyvä funktion f tasa-arvokäyrä koostuu niistä pisteistä (x, y) , joissa $f(x, y) = c$. Eri vakion c arvoilla saadaan erilaisia tasa-arvokäyriä. Kolmas mahdollisuus on simuloida pisteitä $(X_i, Y_i), i = 1, \dots, N$ kyseisestä jakaumasta ja piirtää pisteet tasoon. Kuvassa 6.2 esitellään näitä tapoja.

Kaksiulotteisen jatkuvan jakauman tapauksessa joudutaan laskemaan tasointegraaleja annettujen integrointijoukkojen yli. Joukon B yli laskettu tasointegraali funktiosta g tarkoittaa funktion $1_B g$ integraali koko tason \mathbb{R}^2 , ja sille käytetään mm. seuraavia merkintöjä,

$$\int_B g = \int_B g(x, y) d(x, y) = \iint_B g(x, y) dx dy = \iint_{\mathbb{R}^2} 1_B(x, y) g(x, y) dx dy.$$

Tavallisesti tasointegraalit lasketaan iteroituina integraaleina. Tässä yhteydessä sovelletaan Fubinin lausetta (jota ei todisteta tällä kurssilla). Seuraavan lauseen muotoilu on pätevä Lebesguen integraalille.

Lause 6.5 (Fubini). *Seuraavissa kahdessa tapauksissa tasointegraali voidaan laskea iteroituna integraalina, eli alla olevat yhtälöt ovat voimassa,*

$$\begin{aligned} \int_B g(x, y) d(x, y) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} 1_B(x, y) g(x, y) dx \right) dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} 1_B(x, y) g(x, y) dy \right) dx \end{aligned} \quad (6.2)$$

- (a) Jos $g \geq 0$, jolloin integraalien yhteinen arvo voi olla ∞ .
- (b) Jos $\int_B |g| < \infty$, jolloin integraalien yhteinen arvo on äärellinen. (Huomaa, että integraali $\int_B |g|$ voidaan aina laskea a -kohdan perusteella iteroituna integraalina.)

Esimerkiksi, jos B voidaan esittää muodossa

$$B = \{(x, y) : a < x < b, c(x) < y < d(x)\},$$

ja $\int_B |g| < \infty$ tai $g \geq 0$, niin

$$\int_B g = \iint_B g(x, y) dx dy = \int_a^b \left(\int_{c(x)}^{d(x)} g(x, y) dy \right) dx.$$

Sama tulos saadaan, vaikka joukon B määritelmässä yksi tai useampi relaatioista $<$ korvataan relaatiolla \leq . Tälle iteroidulle integraalille voidaan käyttää myös muita merkintöjä, kuten esim.

$$\int_B g = \int_{x=a}^b \int_{y=c(x)}^{d(x)} g(x, y) dx dy = \int_a^b dx \int_{c(x)}^{d(x)} g(x, y) dy$$

Jos taas B voidaan esittää muodossa

$$B = \{(x, y) : c < y < d, a(y) < x < b(y)\},$$

ja $\int_B |g| < \infty$ tai $g \geq 0$, niin tasointegraali voidaan laskea iteroituna integraalina

$$\begin{aligned} \iint_B g(x, y) \, dx \, dy &= \int_c^d \left(\int_{a(y)}^{b(y)} g(x, y) \, dx \right) dy \\ &= \int_{y=c}^d \int_{x=a(y)}^{b(y)} g(x, y) \, dx \, dy = \int_c^d dy \int_{a(y)}^{b(y)} g(x, y) \, dx \end{aligned}$$

Useimmat sovelluksissa esiintyvät integrointijoukot voidaan osittaa erillisiksi paloiksi, joille voidaan käyttää jompaa kumpaa esitystä. Koko alkuperäisen joukon yli laskettu tasointegraali voidaan laskea summana näiden palojen yli lasketuista tasointegraaleista.

Huomautus. Myös merkintä

$$\int_a^b \int_c^d g(x, y) \, dx \, dy.$$

on yleinen. Sen yhteydessä ei ole itsetään selvää, kumpaan muuttujaan kumpikin integrointijoukko liittyy: joissakin lähteissä merkinnällä tarkoitetaan joukon $(x, y) \in (a, b) \times (c, d)$ yli laskettua integraalia, ja toisissa taas joukon $(x, y) \in (c, d) \times (a, b)$ yli laskettua integraalia.

Kaksiulotteisen jakauman tiheysfunktio ei ole yksikäsitteinen, mutta melkein yksikäsitteinen. Jos f on tiheysfunktio, ja g on saman jakauman tiheysfunktio, niin $f = g$ melkein kaikkialla, eli $f(x, y) = g(x, y)$ kaikkialla muualla paitsi nollamittaisessa joukossa. Kaksiulotteisessa avaruudessa esim. kaikkien yksiulotteisten käyrien kuvaajat ovat nollamittaisia.

Lause 6.6. *Olkoon satunnaisvektorilla (X, Y) jatkuva jakauma. Tällöin reunajakaumat ovat myös jatkuvia, ja niiden tiheysfunktiot saadaan integroimalla toinen muuttuja pois yhteistiheysfunktiosta,*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx.$$

Todistus. Jos $B \subset \mathbb{R}$, niin

$$P(X \in B) = P((X, Y) \in B \times \mathbb{R}) = \int_B \left(\int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \right) dx.$$

Y :n tiheysfunktion lauseke johdetaan vastaavasti. □

Esimerkki 6.3. Siitä, että reunajakaumat ovat jatkuvia, ei seuraa, että yhteisjakauma olisi jatkuva. Otetaan sm X , jolla on jatkuva jakauma tf:lla f_X . Määritellään sm Y siten, että $Y = X$, ts.

$$Y(\omega) = X(\omega), \quad \text{kaikilla } \omega \in \Omega.$$

Tällöin Y :n jakauma on sama kuin X :n jakauma, joten molemmat reunajakaumat ovat jatkuvia. Sv:n (X, Y) jakauma on keskittynyt lävistäjälle

$$B = \{(x, y) : x = y\},$$

mutta mille tahansa funktiolle $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ on $\int_B f = 0$, sillä lävistäjä B on yksiulotteisena joukkona nollamittainen \mathbb{R}^2 :ssa. Tämän takia sv:lla (X, Y) ei voi olla jatkuva jakauma, eli yhteisjakauma ei ole jatkuva. \triangle

Jatkuvan yhteisjakauman tapauksessa ykf:llä on esitys,

$$F_{X,Y}(x, y) = P((X, Y) \in (-\infty, x] \times (-\infty, y]) = \int_{-\infty}^x ds \int_{-\infty}^y f_{X,Y}(s, t) dt.$$

Derivoidaan tämä yhtälö ensin muuttujan x suhteen,

$$\frac{\partial}{\partial x} F_{X,Y}(x, y) = \int_{-\infty}^y f_{X,Y}(x, t) dt,$$

ja sitten vielä muuttujan y suhteen,

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) = f_{X,Y}(x, y),$$

ja tämä tulos pätee ainakin niissä pisteissä (x, y) , joissa $f_{X,Y}$ on jatkuva. Seuraavan lauseen todistus vaatisi työkaluja, joita meillä ei ole käytettävissä.

Lause 6.7. *Jos satunnaisvektorilla (X, Y) on jatkuva jakauma, niin jakauman tiheysfunktioiksi voidaan valita*

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

Tämä toisen kertaluvun sekaderivaatta on olemassa melkein kaikilla (x, y) , ja niissä pisteissä, joissa sekaderivaatta ei ole määritelty, voidaan käyttää mieli-valtaista määrittelyä $f_{X,Y}$:lle.

Huomautus. Mielivaltaisesta annetusta kaksiulotteisen jakauman kertymäfunktioista voi olla hankala tunnistaa, onko se jatkuvan jakauman kertymäfunktio. Periaatteessa pitäisi tutkia, kelpaako ko. toisen kertaluvun sekaderivaattaa jakauman tiheysfunktioiksi.

Lause 6.8. *Jos $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ on ei-negatiivinen, ja sen integraali koko tason yli on yksi, niin se on jonkin kaksiulotteisen satunnaisvektorin tiheysfunktio.*

6.5 Tasajakauma alueessa

Määritelmä 6.6 (Alueen A tasajakauma). Olkoon $A \subset \mathbb{R}^2$ joukko, jonka pinta-ala

$$m(A) = \iint 1_A(x, y) dx dy$$

toteuttaa ehdot $0 < m(A) < \infty$. Sv (X, Y) noudattaa tasajakaumaa joukossa A , mikäli sillä on ytf

$$f(x, y) = 1_A(x, y)/m(A).$$

Tasajakauman tiheysfunktio häviää kyseisen alueen A ulkopuolella, ja sen pisteissä $(x, y) \in A$ sillä on nollasta poikkeava vakioarvo. Olkoon vektorilla (X, Y) tasajakauma alueessa A . Jos $B \subset \mathbb{R}^2$, niin

$$P((X, Y) \in B) = \iint_B f(x, y) dx dy = \frac{m(A \cap B)}{m(A)}.$$

Erityisesti $P((X, Y) \in \mathbb{R}^2) = 1$, kuten pitää ollakin.

Huomaa, että esimerkiksi suljetun neliön $[0, 1] \times [0, 1]$ ja avoimen neliön $(0, 1) \times (0, 1)$ tasajakaumat ovat samat, sillä neliön reunan pinta-ala on nolla eli se on nollamittainen.

Seuraavassa esimerkissä tarkastellaan tasajakaumaa ei-negatiivisen ja integroituvan funktion h kuvaajan ja x -akselin väliin jäävässä alueessa. Huomaa, että x -koordinaatin reunajakaumalla on tällöin tiheysfunktio, joka on verrannollinen funktion h . Toisin sanoen, sm:llä X on jatkuva jakauma, jonka määrää normalisoimaton tiheysfunktio h .

Esimerkki 6.4. Olkoon $h(x) = \exp(-\frac{1}{2}x^2)$, ja olkoon sv:lla (X, Y) tasajakauma h :n kuvaajan alla, eli joukossa

$$A = \{(x, y) : 0 < y < h(x)\}.$$

Tämän joukon pinta-ala on

$$m(A) = \iint_A 1_A(x, y) dx dy = \int_{-\infty}^{\infty} dx \int_0^{h(x)} 1 dy = \int_{-\infty}^{\infty} h(x) dx = \sqrt{2\pi},$$

joten ytf on

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} 1\{0 < y < h(x)\},$$

jossa käytetään merkintää

$$1\{0 < y < h(x)\} = 1_{\{(u,v):0 < v < h(u)\}}(x, y) = \begin{cases} 1, & \text{kun } 0 < y < h(x) \\ 0, & \text{muuten.} \end{cases}$$

X :n reunatiheysfunktio on

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_0^{h(x)} \frac{1}{\sqrt{2\pi}} dy = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$$

joten X :n reunajakauma on $N(0, 1)$.

Mikä on Y :n reunatiheysfunktio? Kaikilla x on $0 < h(x) \leq 1$, joten epäyhtälö

$$0 < y < h(x)$$

voi toteutua vain, kun $0 < y < 1$. Jos $0 < y < 1$, niin epäyhtälö

$$0 < y < \exp(-\frac{1}{2}x^2)$$

saadaan helposti ratkaistua x :n suhteen, ja tulos on

$$-\sqrt{-2 \ln y} < x < \sqrt{-2 \ln y}.$$

(Huomaa, että $-\ln y > 0$, kun $0 < y < 1$.) Siis ytf voidaan esittää myös kaavalla

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} 1\{0 < y < 1, -\sqrt{-2\ln y} < x < \sqrt{-2\ln y}\},$$

josta on helppo laskea Y :n reunatiheysfunktio,

$$f_Y(y) = \frac{2}{\sqrt{\pi}} \sqrt{-\ln y}, \quad 0 < y < 1.$$

△

6.6 Riippumattomuus

Palautetaan mieleen määritelmä. Sm:t X ja Y ovat riippumattomia (merkintä $X \perp Y$), jos

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \quad \text{kaikilla } A, B \subset \mathbb{R}.$$

Luvussa 2.9 todettiin, että

$$X \perp Y \iff F_{X,Y}(x,y) = F_X(x)F_Y(y) \text{ kaikilla } (x,y).$$

Myös yptnf tai ytf voidaan esittää reuna-jakaumien vastaavien funktioiden tulona, jos yhteisjakauma on joko diskreetti tai jatkuva.

Lause 6.9. *Olkoon satunnaisvektorilla (X, Y) joko diskreetti tai jatkuva yhteisjakauma, jonka yptnf tai ytf on $f_{X,Y}$.*

a) Jos $X \perp Y$, niin ytf:ksi kelpaa $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

b) Jos $f_{X,Y}$ voidaan esittää tulomuodossa

$$f_{X,Y}(x,y) = g(x)h(y), \quad \forall x,y,$$

jossa $g \geq 0$ ja $h \geq 0$ ovat yhden muuttujan funktioita, niin $X \perp Y$.

Todistus. (a): Diskreetti tapaus todistettiin luvussa 2.9, ja nyt todistetaan jatkuva tapaus. Oletetaan, että yhteisjakauma on jatkuva ja että $X \perp Y$. Tarkistamme, että ykf:n arvo mielivaltaisessa pisteessä (x,y) saadaan integroimalla väitteessä ilmoitettua ytf:a. Väite seuraa tästä sillä perusteella, että ykf määrää jakauman. Riippumattomuuden nojalla kaikilla $x,y \in \mathbb{R}$ on

$$\begin{aligned} F_{X,Y}(x,y) &= F_X(x)F_Y(y) = \int_{-\infty}^x f_X(s) ds \int_{-\infty}^y f_Y(t) dt \\ &= \int_{-\infty}^x f_X(s) \left(\int_{-\infty}^y f_Y(t) dt \right) ds \\ &= \int_{-\infty}^x ds \int_{-\infty}^y f_X(s) f_Y(t) dt. \end{aligned}$$

Siis $f_X(x)f_Y(y)$ on (X,Y) :n ytf.

(b): Esitetään todistus jatkuvassa tapauksessa; todistuksen idea pysyy samana diskreetissä tapauksessa. Nyt oletetaan, että ytf voidaan esittää tulomuodossa $f_{X,Y}(x, y) = g(x)h(y)$. Olkoon

$$c = \int_{-\infty}^{\infty} g(x) dx, \quad d = \int_{-\infty}^{\infty} h(y) dy.$$

Tällöin

$$1 = \iint_{\mathbb{R}^2} f_{X,Y}(x, y) dx dy = \int_{-\infty}^{\infty} g(x) dx \int_{-\infty}^{\infty} h(y) dy = cd.$$

X :n reunatiheysfunktio on

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = g(x)d = g(x)/c,$$

ja Y :n reunatiheysfunktio on

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = ch(y) = h(y)/d.$$

Kaikilla x, y ykf on reunajakaumien kertymäfunktioiden tulo, sillä

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{-\infty}^x ds \int_{-\infty}^y f_{X,Y}(s, t) dt = \int_{-\infty}^x ds \int_{-\infty}^y \frac{g(s)}{c} \frac{h(t)}{d} dt \\ &= \int_{-\infty}^x f_X(s) ds \int_{-\infty}^y f_Y(t) dt = F_X(x) F_Y(y), \end{aligned}$$

joten $X \perp Y$. □

6.7 Satunnaisvektorin muunnoksen odotusarvo

Lause 6.10. *Olkoon (X, Y) sv, jolla on diskreetti jakauma, ja olkoon $Z = g(X, Y)$ jokin sen reaaliarvoinen muunnos. Tällöin*

$$EZ = \sum_{x,y} g(x, y) f_{X,Y}(x, y),$$

mikäli summa suppenee itseisesti.

Todistus. Olkoon $A = \{(x_i, y_i) : i \geq 1\}$ sv:n (X, Y) korkeintaan numeroituva arvojoukko. Tällöin joukot

$$\{(X, Y) \notin A\}, \quad \{(X = x_i, Y = y_i)\}_{i \geq 1}$$

osittavat perusjoukon. Nyt

$$g(X, Y)(\omega) = g(X(\omega), Y(\omega)) = g(x_i, y_i), \quad \text{kun } \omega \in \{X = x_i, Y = y_i\},$$

joten

$$EZ = \sum_{i \geq 1} g(x_i, y_i) P(X = x_i, Y = y_i) = \sum_{i \geq 1} g(x_i, y_i) f_{X,Y}(x_i, y_i). \quad \square$$

Jatkuvan yhteisjakauman tapauksessa odotusarvo $Eg(X, Y)$ saadaan lasket-
tua integraalina. Sivuumme tämän tapauksen todistuksen.

Lause 6.11. *Olkoon (X, Y) sv, jolla on jatkuva yhteisjakauma, ja olkoon $Z = g(X, Y)$ jokin sen reaaliarvoinen muunnos. Tällöin*

$$EZ = \iint_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) dx dy$$

mikäli integraali suppenee itseisesti.

Esimerkki 6.5. Jos satunnaismuuttujilla X ja Y on jatkuva yhteisjakauma, ja ne ovat integroituvia, niin kaikilla vakioilla $a, b \in \mathbb{R}$

$$\begin{aligned} E(aX + bY) &= \iint (ax + by) f_{X,Y}(x, y) dx dy \\ &= a \int x \left(\int f_{X,Y}(x, y) dy \right) dx + b \int y \left(\int f_{X,Y}(x, y) dx \right) dy \\ &= a \int x f_X(x) dx + b \int y f_Y(y) dy = a EX + b EY. \end{aligned}$$

Tässä integraalimerkki tarkoittaa koko reaaliakselin yli laskettua integraali. Olemme johtaneet tässä tapauksessa sen ennestään tutun tosiseikan, että odotusarvo on lineaarinen operaattori. \triangle

Huomautus. Funktiolle $g(x, y) = ax + by$ pätee $Eg(X, Y) = g(EX, EY)$, mutta muille funktioille tällainen identiteetti ei yleensä päde.

Esimerkki 6.6. Jos $X \perp Y$, niin tällöin $g(X) \perp h(Y)$ kaikilla funktioilla $g, h : \mathbb{R} \rightarrow \mathbb{R}$, joten

$$E(g(X) h(Y)) = E(g(X)) E(h(Y)),$$

mikäli kyseiset odotusarvot ovat olemassa. Jos sv:lla (X, Y) on jatkuva jakauma, niin $f_{X,Y}(x, y) = f_X(x) f_Y(y)$, joten asian näkee myös laskulla

$$\begin{aligned} E(g(X) h(Y)) &= \iint g(x) h(y) f_X(x) f_Y(y) dx dy \\ &= \left(\int g(x) f_X(x) dx \right) \left(\int h(y) f_Y(y) dy \right) = E(g(X)) E(h(Y)). \end{aligned}$$

\triangle

6.8 Kovarianssi ja muita yhteismomentteja

Määritelmä 6.7. Olkoot $m, n \geq 0$ kokonaislukuja. Mikäli odotusarvo

$$E[X^m Y^n]$$

on olemassa, sitä kutsutaan kertalukujen (m, n) yhteismomentiksi (tai tulomomentiksi). Mikäli odotusarvo

$$E[(X - EX)^m (Y - EY)^n]$$

on olemassa, sitä sanotaan kertalukujen (m, n) keskusmomentiksi.

Näistä momenteista ehdottomasti tärkeimmät ovat raunajakaumien momentit EX^m sekä EY^n (eli kertalukujen $(m, 0)$ ja $(0, n)$ yhteismomentit) sekä kertaluvun $(1, 1)$ keskusmomentti, joka tunnetaan paremmin nimellä X :n ja Y :n kovarianssi,

$$\text{cov}(X, Y) = E[(X - EX)(Y - EY)].$$

Huomautus. Kovarianssia käsiteltiin jo jaksossa 3.6. Kovarianssi on symmetrinen, $\text{cov}(X, Y) = \text{cov}(Y, X)$, ja varianssi saadaan lausuttua kovarianssin avulla, sillä $\text{var } X = \text{cov}(X, X)$. Odotusarvon lineaarisuuden avulla nähtiin, että

$$\text{cov}(X, Y) = E(XY) - (EX)(EY).$$

Lause 6.12. *Kovarianssi on bilineaarinen, eli molempien argumenttiensa suhteen lineaarinen operaattori. Ts. jos $a_1, a_2 \in \mathbb{R}$ ovat vakioita ja X_1, X_2 ja Z ovat satunnaismuuttujia, niin*

$$\begin{aligned} \text{cov}(a_1 X_1 + a_2 X_2, Z) &= a_1 \text{cov}(X_1, Z) + a_2 \text{cov}(X_2, Z) \\ \text{cov}(Z, a_1 X_1 + a_2 X_2) &= a_1 \text{cov}(Z, X_1) + a_2 \text{cov}(Z, X_2) \end{aligned}$$

Jos lisäksi $b_1, b_2 \in \mathbb{R}$ ja Y_1, Y_2 ovat satunnaismuuttujia, niin

$$\text{cov}(a_1 X_1 + a_2 X_2, b_1 Y_1 + b_2 Y_2) = \sum_{i=1}^2 \sum_{j=1}^2 a_i b_j \text{cov}(X_i, Y_j).$$

Nämä kaavat pitävät paikkansa, mikäli kyseiset odotusarvot ovat olemassa.

Todistus. Odotusarvon lineaarisuuden nojalla

$$E(a_1 X_1 + a_2 X_2) = a_1 EX_1 + a_2 EX_2,$$

joten

$$a_1 X_1 + a_2 X_2 - E(a_1 X_1 + a_2 X_2) = a_1 (X_1 - EX_1) + a_2 (X_2 - EX_2).$$

Siis odotusarvon lineaarisuuden nojalla,

$$\begin{aligned} \text{cov}(a_1 X_1 + a_2 X_2, Z) &= E[(a_1 (X_1 - EX_1) + a_2 (X_2 - EX_2))(Z - EZ)] \\ &= a_1 E[(X_1 - EX_1)(Z - EZ)] + a_2 E[(X_2 - EX_2)(Z - EZ)] \\ &= a_1 \text{cov}(X_1, Z) + a_2 \text{cov}(X_2, Z). \end{aligned}$$

Kovarianssin symmetrisyyden nojalla

$$\text{cov}(Z, \sum_i a_i X_i) = \text{cov}(\sum_i a_i X_i, Z) = \sum_i a_i \text{cov}(X_i, Z) = \sum_i a_i \text{cov}(Z, X_i).$$

Kovarianssi on nyt todistettu molempien argumenttiensa suhteen lineaariseksi. Näin ollen

$$\begin{aligned} \text{cov}\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) &= \sum_i a_i \text{cov}\left(X_i, \sum_j b_j Y_j\right) = \sum_i a_i \sum_j b_j \text{cov}(X_i, Y_j) \\ &= \sum_i \sum_j a_i b_j \text{cov}(X_i, Y_j). \quad \square \end{aligned}$$

Esimerkki 6.7. Linearikombinaation varianssi.

$$\begin{aligned}\operatorname{var}(aX + bY) &= \operatorname{cov}(aX + bY, aX + bY) \\ &= a \operatorname{cov}(X, aX + bY) + b \operatorname{cov}(Y, aX + bY) \\ &= a^2 \operatorname{cov}(X, X) + ab \operatorname{cov}(X, Y) + ab \operatorname{cov}(Y, X) + b^2 \operatorname{cov}(Y, Y) \\ &= a^2 \operatorname{var}(X) + 2ab \operatorname{cov}(X, Y) + b^2 \operatorname{var}(Y).\end{aligned}$$

△

6.9 Paras lineaarinen ennuste

Joissakin tilanteissa $sm:n$ X arvo on mahdollista havaita, mutta $sm:n$ Y arvoa ei havaita. Tällöin saatetaan haluta ennustaa $Y:n$ arvo $X:n$ arvon avulla. Ennusteelle voidaan esimerkiksi valita lineaarinen muoto $a + bX$, mutta tällöin jää vielä ratkaistavaksi kysymys, miten kertoimet a ja b pitäisi valita jotta ennuste olisi mahdollisimman hyvä. Mikäli ennusteen hyvyyden mittana pidetään mahdollisimman pientä keskineliövirhettä

$$E[(Y - (a + bX))^2] = \min!$$

niin osoittautuu (HT), että keskineliövirheen mielessä paras lineaarinen ennuste on

$$EY + \frac{\operatorname{cov}(X, Y)}{\operatorname{var}(X)}(X - EX) \quad (6.3)$$

ts. $a:n$ ja $b:n$ optimaaliset arvot ovat

$$a^* = EY - \frac{\operatorname{cov}(X, Y)}{\operatorname{var} X} EX, \quad b^* = \frac{\operatorname{cov}(X, Y)}{\operatorname{var} X}.$$

Ratkaisemme myöhemmin vielä kunnianhimoisemman tehtävän, jossa kysytään, mikä on keskineliövirheen mielessä paras $Y:n$ ennuste lausekkeella $m(X)$, jossa funktio m saadaan valita vapaasti.

Ennustevirheen varianssille saadaan helpolla laskulla lauseke

$$\operatorname{var}(Y - a^* - b^*X) = (1 - \rho^2) \operatorname{var} Y, \quad (6.4)$$

jossa ρ on korrelaatiokerroin

$$\rho = \operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\operatorname{var} X} \sqrt{\operatorname{var} Y}},$$

jonka arvot ovat (Cauchyn–Schwarzin epäyhtälön mukaan) välillä $-1 \leq \rho \leq 1$. Ennustevirheen varianssin kaavasta nähdään, kuinka *korrelaatiokerroin mittaa muuttujien välisen lineaarisen riippuvuuden voimakkuutta*. Jos korrelaatiokerroin saavuttaa jommankumman ääriarvonsa $\rho = \pm 1$, niin tällöin $\operatorname{var}(Y - a^* - b^*X) = 0$, joten tällöin

$$Y = EY + \frac{\operatorname{cov}(X, Y)}{\operatorname{var} X}(X - EX) \quad (\text{m.v.}), \quad (6.5)$$

Edellä lyhenne m.v. on lyhenne sanoille melkein varmasti, joka taas tarkoittaa samaa kuin todennäköisyydellä yksi. Siis ominaisuudesta $|\rho| = 1$ seuraa, että

$$Y(\omega) = EY + \frac{\text{cov}(X, Y)}{\text{var } X}(X(\omega) - EX)$$

kaikilla alkeistapauksilla ω , paitsi sellaisilla, jotka kuuluvat poikkeusjoukkoon, jonka tn on nolla.

Huomaa myös, että korrelaatiokerroin on sama kuin kovarianssin $\text{cov}(X, Y)$ etumerkki ja se on edelleen sama kuin parhaan lineaarisen ennusteen kulmakertoimen etumerkki. Jos $\rho > 0$ eli $\text{cov}(X, Y) > 0$, niin Y :n arvolla on taipumus kasvaa X :n arvon kasvaessa, koska parhaan lineaarisen ennusteen kulmakertoimen on tällöin positiivinen. Jos taas $\rho < 0$ eli $\text{cov}(X, Y) < 0$, niin Y :n arvoilla on taipumus vähetä X :n arvon kasvaessa.

6.10 Odotusarvovektori ja kovarianssimatriisi

Silloin kuin satunnaisvektori $\mathbf{V} = (X, Y)$ esiintyy vektori- ja matriisilausekkeissa, sitä pidetään pystyvektorina. Varoitus: joissakin muissa lähteissä paria (X, Y) pidetään vaakavektorina. Näissä luentomuistiinpanoissa vektorit ja matriisit ovat paksunnettuja ja skalaarit paksuntamattomia. Edistyneemmissä esityksissä tällaista konventioita ei välttämättä käytetä, vaan lukijan pitää päätellä asiayhteydestä, minkätyyppistä oliota kukin symboli tarkoittaa.

Määritelmä 6.8 (Satunnaisvektorin odotusarvo). Satunnaisvektorin (X, Y) odotusarvo(vektori) määritellään ottamalla satunnaisvektorista odotusarvot komponentti komponentilta, eli

$$E\mathbf{V} = E(X, Y) = E \begin{bmatrix} X \\ Y \end{bmatrix} = (EX, EY) = \begin{bmatrix} EX \\ EY \end{bmatrix}$$

mikäli kaikkien komponenttien odotusarvot ovat olemassa.

Koska kaksikomponenttista vektoria (X, Y) voidaan pitää myös pitää 2×1 -matriisina, niin satunnaisvektorin odotusarvon määritelmä on erikoistapaus seuraavaksi esitettävästä satunnaismatriisin odotusarvon määritelmästä.

Määritelmä 6.9 (Satunnaismatriisi ja sen odotusarvo). Satunnaismatriisi on matriisi, jonka alkiot ovat satunnaismuuttujia. Sen odotusarvo on se vakiomatriisi, jonka alkio kohdassa (i, j) on ko. satunnaismatriisin alkion (i, j) odotusarvo (mikäli kaikkien alkoiden odotusarvot ovat olemassa).

Määritelmästä seuraa välittömästi, että

$$E(\mathbf{Z}^T) = (E\mathbf{Z})^T \quad (6.6)$$

mikäli \mathbf{Z} on satunnaismatriisi.

Odotusarvon lineaarisuuden avulla nähdään helposti, että

$$E(\mathbf{A}\mathbf{Z}\mathbf{B} + \mathbf{C}) = \mathbf{A}(E\mathbf{Z})\mathbf{B} + \mathbf{C}, \quad (6.7)$$

kun \mathbf{Z} on satunnaismatriisi ja \mathbf{A} , \mathbf{B} ja \mathbf{C} ovat vakiomatriiseja, joiden dimensiot ovat sellaiset, että lauseke on määritelty. Vakiomatriisit saadaan vetää ulos

odotusarvon alta, jos ne sijaitsevat matriisitulossa äärimmäisenä vasemmalla tai äärimmäisenä oikealla. Sen sijaan lausekkeen keskellä olevia vakiomatriiseja ei pystytä vetämään ulos tällä kaavalla. Kaava (6.7) pätee myös satunnaisvektorille \mathbf{Z} , sillä d -komponenttinen pystyvektori voidaan tulkita $d \times 1$ -matriisiksi.

Määritelmä 6.10 (Kovarianssimatriisi). Satunnaisvektorin $\mathbf{V} = (X, Y)$ kovarianssimatriisi määritellään kaavalla

$$\text{Cov}(\mathbf{V}) = E[(\mathbf{V} - E\mathbf{V})(\mathbf{V} - E\mathbf{V})^T]. \quad (6.8)$$

Huomautus: Sv:n \mathbf{V} kovarianssimatriisille käytetään kirjallisuudessa monia muitakin merkintöjä, kuten esim. $\text{var}(\mathbf{V})$. Englannin kielellä kovarianssimatriisista käytetään ainakin nimityksiä *covariance matrix*, *variance-covariance matrix*, *variance matrix*, *dispersion matrix*.

Määritelmän perusteella

$$\begin{aligned} \text{Cov}(\mathbf{V}) &= E \left\{ \begin{bmatrix} X - EX \\ Y - EY \end{bmatrix} \begin{bmatrix} X - EX & Y - EY \end{bmatrix} \right\} \\ &= E \begin{bmatrix} (X - EX)(X - EX) & (X - EX)(Y - EY) \\ (Y - EY)(X - EX) & (Y - EY)(Y - EY) \end{bmatrix} \\ &= \begin{bmatrix} \text{cov}(X, X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) \end{bmatrix} = \begin{bmatrix} \text{var}(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \text{var}(Y) \end{bmatrix} \end{aligned}$$

Kovarianssimatriisin pääälvistäjällä on muuttujien varianssit ja muualla muuttujien väliset kovarianssit. Tästä syystä kovarianssimatriisia kutsutaan myös kömpelöllä nimellä varianssi-kovarianssimatriisi.

Jos merkitään $\sigma_X^2 = \text{var}(X)$ ja $\sigma_Y^2 = \text{var}(Y)$ ja $\sigma_X, \sigma_Y > 0$, niin satunnaismuuttujien X ja Y välinen korrelaatio (kerroin) $\rho = \text{corr}(X, Y)$ määriteltiin kaavalla

$$\rho = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var } X} \sqrt{\text{var } Y}} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Tämän ansiosta sv:n $\mathbf{V} = (X, Y)$ kovarianssimatriisi voidaan kirjoittaa myös muodossa

$$\text{Cov}(\mathbf{V}) = \begin{bmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix}$$

Jos satunnaisvektori \mathbf{W} saadaan satunnaisvektorista \mathbf{V} affinilla muunnoksella,

$$\mathbf{W} = \mathbf{A}\mathbf{V} + \mathbf{b},$$

jossa \mathbf{A} on vakiomatriisi ja \mathbf{b} on vakiovektori, niin

$$E\mathbf{W} = \mathbf{A}(E\mathbf{V}) + \mathbf{b} \Rightarrow (\mathbf{W} - E\mathbf{W})(\mathbf{W} - E\mathbf{W})^T = \mathbf{A}(\mathbf{V} - E\mathbf{V})(\mathbf{V} - E\mathbf{V})^T \mathbf{A}^T,$$

joten kaavan (6.7) nojalla

$$\text{Cov}(\mathbf{A}\mathbf{V} + \mathbf{b}) = \mathbf{A} \text{Cov}(\mathbf{V}) \mathbf{A}^T. \quad (6.9)$$

Sovelletaan lopuksi kaavaa (6.9) tapaukseen $\mathbf{V} = (X, Y)$ ja $\mathbf{A} = \mathbf{u}^T$, jossa $\mathbf{u} = (a, b)$. Tällöin saadaan johdettua tuttu kaava, nimittäin

$$\begin{aligned} \text{var}(aX + bY) &= \text{var}(\mathbf{u}^T \mathbf{V}) = \mathbf{u}^T \text{Cov}(\mathbf{V}) \mathbf{u} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \text{var}(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \text{var}(Y) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \\ &= a^2 \text{var}(X) + 2ab \text{cov}(X, Y) + b^2 \text{var}(Y). \end{aligned}$$

6.11 Tiheysfunktion muuntokaava

Jos sv:illa (X, Y) on diskreetti jakauma, ja satunnaismuuttuja tai -vektori Z määritellään sen muunnoksena, $Z = g(X, Y)$, niin tällöin Z :lla on diskreetti jakauma, jonka ptnf voidaan laskea suoraan määritelmän perusteella, sillä

$$f_Z(z) = P(Z = z) = \sum_{(x,y):g(x,y)=z} f_{X,Y}(x,y).$$

Jatkuvan jakauman tapauksessa tarvitaan monimutkaisempia laskuja.

Tarkastellaan jatkuvasti jakautunutta sv:a (X, Y) , jolla on tf $f_{X,Y}$, sekä funktiota $\mathbf{g} : A \rightarrow B$, jossa $A, B \subset \mathbb{R}^2$, ja A on sellainen joukko, että

$$P((X, Y) \in A) = 1.$$

Määritellään kaksiulotteinen sv (U, V) kaavalla

$$\begin{bmatrix} U \\ V \end{bmatrix} = \mathbf{g}(X, Y) = \begin{bmatrix} g_1(X, Y) \\ g_2(X, Y) \end{bmatrix}$$

Oletamme, että $\mathbf{g} : A \rightarrow B$ on diffeomorfisimi, eli että

- joukot A ja B ovat avoimia,
- funktio $\mathbf{g} : A \rightarrow B$ on jatkuvasti derivoituva bijektio,
- sen käänteisfunktio $\mathbf{h} = \mathbf{g}^{-1}$ on jatkuvasti derivoituva bijektio $B \rightarrow A$.

Kuvausten \mathbf{g} ja \mathbf{h} komponenttikuvaukset ovat

$$\mathbf{g} = (g_1, g_2), \quad \mathbf{h} = (h_1, h_2).$$

Nyt ajattelemme bijektiivistä vastaavuutta

$$(u, v) = (g_1(x, y), g_2(x, y)) \iff (x, y) = (h_1(u, v), h_2(u, v)), \quad (6.10)$$

eli

$$(u, v) = (u(x, y), v(x, y)) \iff (x, y) = (x(u, v), y(u, v)). \quad (6.11)$$

Kuvauksen \mathbf{h} derivaatta(matriisi) eli Jacobin matriisi pisteessä (u, v) on

$$\begin{bmatrix} \frac{\partial h_1(u, v)}{\partial u} & \frac{\partial h_1(u, v)}{\partial v} \\ \frac{\partial h_2(u, v)}{\partial u} & \frac{\partial h_2(u, v)}{\partial v} \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} \quad (6.12)$$

Jacobin matriisin eli derivaattamatriisin determinanttia kutsutaan kuvauksen \mathbf{h} *funktionaalideterminantiksi* eli *Jacobin determinantiksi* eli *jacobiaaniksi* (engl. *Jacobian determinant; Jacobian*). (Myös nimiä Jakobin determinantti tai jakobiaani käytetään; käsitteen esitti 1800-luvulla saksalainen matemaatikko Carl Gustav Jacobi.) Varoitus: joissakin muissa lähteissä termi jacobiaani saattaa tarkoittaa derivaattamatriisia eikä sen determinanttia.

Jacobiaanille eli Jacobin matriisin determinantille käytetään kirjallisuudessa useita erilaisia merkintöjä. Me käytämme merkintöjä

$$J_{\mathbf{h}}(u, v), \quad \frac{\partial(x, y)}{\partial(u, v)},$$

jossa

$$J_{\mathbf{h}}(u, v) = \frac{\partial(x, y)}{\partial(u, v)} = \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u}.$$

Funktion \mathbf{g} ja sen käänteisfunktion \mathbf{h} derivaattamatriisit ovat tunnetusti toistensa käänteismatriiseja, josta seuraa jacobiaaneille yhteys

$$1 = J_{\mathbf{h}}(u, v) J_{\mathbf{g}}(x, y) = \frac{\partial(x, y)}{\partial(u, v)} \frac{\partial(u, v)}{\partial(x, y)}. \quad (6.13)$$

Tässä kaavassa (x, y) ja (u, v) vastaavat toisiaan bijektiivisesti yhtälön (6.11) mukaan. Kaava (6.13) seuraa siitä, että toisaalta identiteettimatriisin determinantti on yksi ja toisaalta kaavasta $\det(\mathbf{A} \mathbf{B}) = \det(\mathbf{A}) \det(\mathbf{B})$, joka pätee, kun \mathbf{A} ja \mathbf{B} ovat samankokoisia neliömatriiseja.

Lause 6.13. *Tämän jakson oletuksilla satunnaisvektorilla (U, V) on jatkuva jakauma tf:llä*

$$f_{U,V}(u, v) = \begin{cases} f_{X,Y}(h_1(u, v), h_2(u, v)) |J_{\mathbf{h}}(u, v)|, & \text{kun } (u, v) \in B \\ 0, & \text{muuten.} \end{cases} \quad (6.14)$$

Todistus. Oletusten nojalla $(U, V) \in B$ tn:llä yksi, sillä

$$P((U, V) \in B) = P(\mathbf{g}(X, Y) \in B) = P((X, Y) \in \mathbf{g}^{-1}(B)) = P((X, Y) \in A) = 1.$$

Olkoon $C \subset B$ mielivaltainen joukko. Nyt

$$\begin{aligned} P((U, V) \in C) &= P(\mathbf{g}(X, Y) \in C) = P((X, Y) \in \mathbf{h}(C)) \\ &= \iint_{\mathbf{h}(C)} f_{X,Y}(x, y) \, dx \, dy. \end{aligned}$$

Tehdään nyt muuttujanvaihto $(u, v) = \mathbf{g}(x, y)$, eli $(x, y) = \mathbf{h}(u, v)$, jolloin taso-integraali saadaan muotoon (ks. esim. Billingsleyn [1] Theorem 17.2)

$$\iint_{\mathbf{h}(C)} f_{X,Y}(x, y) \, dx \, dy = \iint_C f_{X,Y}(h_1(u, v), h_2(u, v)) |J_{\mathbf{h}}(u, v)| \, du \, dv$$

Nämä tarkastelut todistavat väitteen. □

Diffeomorfismin tapauksessa muuttujanvaihtokaava (6.14) kannattaa pitää mielessään muodossa

$$f_{X,Y}(x, y) |\partial(x, y)| = f_{U,V}(u, v) |\partial(u, v)|, \quad \text{kun} \quad (6.15)$$

$$(u, v) = \mathbf{g}(x, y) \Leftrightarrow (x, y) = \mathbf{h}(u, v), \quad (x, y) \in A, (u, v) \in B. \quad (6.16)$$

Tästä sitten tuntemattomat suureet ratkaistaan tunnettujen suureiden avulla. Jos $f_{X,Y}$ on annettu, niin

$$f_{U,V}(u, v) = f_{X,Y}(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| = f_{X,Y}(h_1(u, v), h_2(u, v)) |J_{\mathbf{h}}(u, v)|,$$

kun $(u, v) \in B$. Saman ytf:n voi ilmaista myös muodossa

$$f_{U,V}(u, v) = \frac{f_{X,Y}(x, y)}{\left| \frac{\partial(u, v)}{\partial(x, y)} \right|} = \frac{f_{X,Y}(h_1(u, v), h_2(u, v))}{|J_{\mathbf{g}}(h_1(u, v), h_2(u, v))|},$$

kun $(u, v) \in B$. Tämän tuloksen oikeellisuus perustuu kaavaan (6.13).

Tiheysfunktion muuntokaavaa (6.14) sovellettaessa on tärkeää pitää kirjaa siitä, missä joukossa B johdettu kaava on pätevä. Tämän kirjanpidon saa usein hoidettua automaattisesti joukkojen indikaattorien avulla, kuten seuraavassa esimerkissä tehdään. Valitettavasti jatkotarkasteluja varten (esim. reunajakau- man johtamista varten) on usein tarpeen esittää alue B esim. muodossa

$$B = \{(u, v) : a < u < b, c(u) < v < d(u)\}$$

tai muodossa

$$B = \{(u, v) : c < v < d, a(v) < u < b(v)\}.$$

Tästä voi aiheutua paljon lisätyötä.

Esimerkki 6.8. Olkoon sv:lla (X, Y) jatkuva yhteisjakauma tf:lla

$$f_{X,Y}(x, y) = \begin{cases} k(x, y) & \text{kun } x > 0 \text{ ja } y > 0 \\ 0 & \text{muuten.} \end{cases} \quad (6.17)$$

Esitetään tämä ytf tason ensimmäisen neljänneksen indikaattorin $1\{x > 0, y > 0\}$ ja lausekkeen $k(x, y)$ tulona,

$$f_{X,Y}(x, y) = 1\{x > 0, y > 0\} k(x, y).$$

Huomaa, että lauseke $k(x, y)$ ei ole välttämättä hyvin määritelty, jos $x \leq 0$ tai $y \leq 0$. Ylläoleva esitys pitää ymmärtää lyhennemerkintänä kaavalle (6.17). Käytämme nyt tällaista konventiota, jossa joukon indikaattori tarvittaessa nol- laa mahdollisesti huonosti määritellyn lausekkeen.

Tarkastellaan sv:a (U, V) , jossa $U = X + Y$ ja $V = X - Y$. Tätä muunnosta vastaava diffeomorfismi on

$$\begin{cases} u = x + y \\ v = x - y \end{cases} \iff \begin{cases} x = \frac{1}{2}(u + v) \\ y = \frac{1}{2}(u - v) \end{cases}$$

joka on diffeomorfismi koko tason ja koko tason välillä. Tämän takia satunnais- muuttujien U ja V ytf on

$$\begin{aligned} f_{U,V}(u, v) &= f_{X,Y}(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| \\ &= 1\{u + v > 0, u - v > 0\} k\left(\frac{1}{2}(u + v), \frac{1}{2}(u - v)\right) \frac{1}{2}. \end{aligned}$$

Lasketaan seuraavaksi sm:n U reunatiheys. Oikeiden integrointirajojen selvittämiseksi edellä esiintyvä epäyhtälöpari $u + v > 0$ ja $u - v > 0$ pitää onnistua ratkaisemaan muodossa

$$a < u < b, \quad c(u) < v < d(u).$$

Hetken pohtimisen jälkeen (piirrä kuva) nähdään, että ratkaisu on

$$u > 0, \quad -u < v < u.$$

Tämän takia U :n reunatiheys on

$$f_U(u) = \int_{-u}^u \frac{1}{2} k\left(\frac{1}{2}(u+v), \frac{1}{2}(u-v)\right) dv, \quad u > 0.$$

△

Huomaa, että diffeomorfismi on aina kuvaus kahden samandimensioisen euklidisen avaruuden välillä. Toisinaan tavoitteena on johtaa jatkuvasti jakautuneen sv:n (X, Y) jonkin skalaariarvoisen muunnoksen $U = g_1(X, Y)$ tiheysfunktio. Tämä voidaan tehdä kahdella erilaisella tekniikalla.

1. Lasketaan ensin U :n kertymäfunktio

$$F_U(u) = P(g_1(X, Y) \leq u) = \int_{\{(x,y):g_1(x,y)\leq u\}} f_{X,Y}(x,y) dx dy$$

integroimalla kyseisen joukon yli, ja sitten U :n tiheysfunktio lasketaan derivoimalla (mikäli U :n jakauma osoittautuu jatkuvaksi).

2. Voidaan yrittää täydentää muunnos (keinotekoisesti) bijektioksi valitsemalla $V = g_2(X, Y)$, sitten johdetaan sv:n (U, V) tiheysfunktio, ja lopuksi kiinnostuksen kohteena olevan muuttujan U tiheysfunktio lasketaan integroimalla

Näistä tavoista jälkimmäinen on usein suoraviivaisempi.

Esimerkki 6.9. Summan tiheysfunktio. Olkoon sv:lla (X, Y) tf $f_{X,Y}$. Olkoon

$$U = X + Y.$$

Johda sm:n U tf. Minkäläisen kaavan saat, jos $X \perp Y$?

Ratkaisu bijektioksi täydentämällä: Tässä tehtävässä voisimme täydentää kuvauksen bijektioksi monella tavalla, mutta nyt teemme valinnan $V = X$. Tällöin

$$\begin{cases} u = x + y \\ v = x \end{cases} \iff \begin{cases} x = v \\ y = u - v \end{cases}$$

Siis

$$f_{U,V}(u, v) = f_{X,Y}(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| = f_{X,Y}(v, u - v).$$

Sm:n U reunatiheysfunktio saadaan tästä integroimalla v pois,

$$f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(v, u - v) dv$$

Ratkaisu kertymäfunktio-tekniikalla: Kun lasketaan U :n kertymäfunktioita pisteessä u , niin integrointijoukko on

$$\{(x, y) : x + y \leq u\} = \{(x, y) : x \in \mathbb{R}, y \leq u - x\}.$$

Tämän takia

$$F_U(u) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{u-x} f_{X,Y}(x, y) dy.$$

Mikäli derivaatan vienti integraalin alle pystytään perustelemaan jotenkin, niin arvataan, että täytyy olla

$$\begin{aligned} f_U(u) &= F'_U(u) \\ &= \int_{-\infty}^{\infty} dx \frac{\partial}{\partial u} \int_{-\infty}^{u-x} f_{X,Y}(x, y) dy \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, u-x) dx. \end{aligned}$$

Jälkimmäisessä laskussa jää arveluttamaan se, mitä ehtoja ytf:lle pitäisi asettaa, jotta tulos voitaisiin perustella täsmällisesti. Ensimmäisessä tavassa nähtiin, että mitään ehtoja ei tarvita.

Jos $X \perp Y$, niin $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ identtisesti, ja

$$f_U(u) = \int_{-\infty}^{\infty} f_X(v) f_Y(u-v) dv,$$

eli f_U on tiheysfunktioiden f_X ja f_Y konvoluutio,

$$f_U = f_X * f_Y.$$

△

Kirjallisuutta

[1] P. Billingsley. *Probability and Measure*. John Wiley & Sons, Inc., 2nd edition, 1986.

6.12 t -jakauman ominaisuuksia

Studentin t -jakauma vapausasteluvulla $\nu > 0$ eli t_ν -jakauma määriteltiin niin, että se on sm:n

$$Y = \frac{Z}{\sqrt{X/\nu}} \tag{6.18}$$

jakauma, kun $X \sim \chi_\nu^2 = \text{Gam}(\frac{\nu}{2}, \frac{1}{2})$, $Z \sim N(0, 1)$, ja $X \perp Z$. Johdamme seuraavaksi t -jakauman tiheysfunktion esimerkkinä muuttujanvaihtotekniikan käytöstä. Lisäksi päättelemme jakauman odotusarvon ja varianssin käyttämällä tätä jakauman stokastista esitystä.

Riippumattomuuden nojalla

$$f_{X,Z}(x, z) = f_X(x) f_Z(z) = \frac{(\frac{1}{2})^{\nu/2}}{\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{1}{2}x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad x > 0.$$

Täydennetään kuvaus bijektioksi valitsemalla $U = X$. Tällöin tarkasteltavana on diffeomorfismi

$$\begin{cases} u = x \\ y = \frac{z}{\sqrt{x/\nu}} \end{cases} \iff \begin{cases} x = u \\ z = y\sqrt{u/\nu} \end{cases}$$

jossa $x > 0$ ja $u > 0$. Siis

$$f_{U,Y}(u, y) = f_{X,Z}(x, z) \left| \frac{\partial(x, z)}{\partial(u, y)} \right| = f_{X,Z}(u, y\sqrt{u/\nu}) \sqrt{\frac{u}{\nu}}, \quad u > 0.$$

Tästä saadaan Y :n reunatiheys integroimalla u pois,

$$f_Y(y) = \int_0^\infty f_{X,Z}(u, y\sqrt{u/\nu}) \sqrt{\frac{u}{\nu}} du = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{\nu\pi}} \frac{1}{(1+y^2/\nu)^{(\nu+1)/2}}.$$

Edellä oleva määrätty integraali pystytään laskemaan integroimalla kuten tilastotieteilijä: koska tiedämme, mikä on gammajakauman tiheysfunktion lauseke, osaamme suoraan kirjoittaa tämänmuotoisen integraalin arvon.

Arvolla $\nu = 1$ tf saa muodon

$$f_Y(y) = \frac{1}{\pi} \frac{1}{1+y^2},$$

joten t_1 on Cauchyn jakauma.

Studentin t -jakaumalla ei ole olemassa kaikkia absoluuttisia momenteja. Jos $a > 0$, niin

$$E|Y|^a < \infty \iff a < \nu.$$

Tämän seikan näkee esim. jakauman stokastisesta esityksestä (6.18), josta seuraa että

$$E|Y|^a = E|Z|^a \nu^a E X^{-a/2},$$

mikäli molemmat kaavan oikealla puolella olevat odotusarvot ovat äärellisiä. Normaalijakauman absoluuttiset momentit ovat kyllä äärellisiä, mutta tarkastelemalla jälkimmäistä odotusarvoa päädytään helposti ehtoon $a < \nu$. Esim. Cauchyn jakaumalla ei ole odotusarvoa; t -jakaumalla on olemassa varianssi vain silloin, kun $\nu > 2$. Siitä, että jakauman kaikki momentit eivät ole äärellisiä seuraa, että t -jakauman momenttiemäfunktio ei ole olemassa missään origon ympäristössä, joten momenttiemäfunktion on hyödytön työkalu tälle jakaumalle.

Studentin t -jakauman odotusarvo ja varianssi selviävät helposti jakauman stokastisesta esityksestä (6.18). Jos $\nu > 1$, niin t_ν -jakauman odotusarvo on nolla. Tämä nähdään siitä, että koska $Z \perp X$, niin

$$EY = EZ E \sqrt{\frac{\nu}{X}} = 0 \cdot E \sqrt{\frac{\nu}{X}} = 0.$$

Jos $\nu > 2$, niin t_ν -jakauman varianssi on

$$\text{var } Y = EY^2 = EZ^2 E \frac{\nu}{X} = 1 \cdot E \frac{\nu}{X}.$$

Tästä saadaan muutaman välivaiheen jälkeen (integrooi kuten tilastotieteilijä ja käytä gammafunktion funktionaaliyhtälöä):

$$\text{var } Y = \frac{\nu}{\nu-2}, \quad \text{kun } \nu > 2.$$

Luku 7

Ehdollinen jakauma

7.1 Ehdolliset jakaumat

Jos s.v:lla (X, Y) on diskreetti jakauma, niin sm:n X ehdollinen ptnf ehdolla $Y = y$ määritellään ehdollisen todennäköisyyden kaavan avulla,

$$f_{X|Y}(x | y) = P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad (7.1)$$

kun y on sellainen piste, jossa $f_Y(y) > 0$. Funktio $f_{X|Y}(\cdot | y)$ on satunnaismuuttujan X ptnf, kun tiedetään, että $Y = y$. Ehdollinen ptnf $f_{Y|X}(y | x)$ määritellään analogisesti.

Kertolaskusääntö eli *ketjusääntö* kertoo, että

$$f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y | x) = f_Y(y) f_{X|Y}(x | y),$$

ja *Bayesin kaava* kertoo, että

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x) f_{Y|X}(y | x)}{f_Y(y)}.$$

Näissä kaavoissa oletetaan, että x ja y ovat sellaisia pisteitä, joissa ei jouduta jakamaan nolllalla. Jos $X \perp Y$, niin

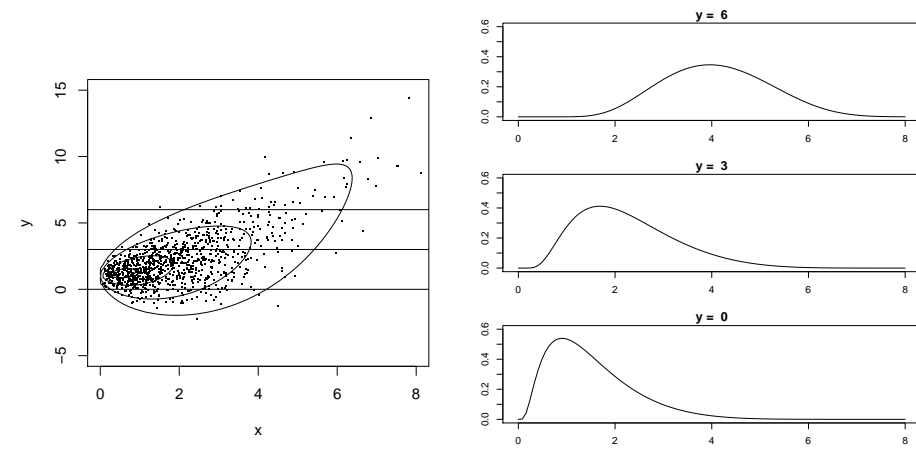
$$f_{X|Y}(x | y) = f_X(x), \quad f_{Y|X}(y | x) = f_Y(y)$$

kaikilla kyseeseen tulevilla argumenteilla. Kun $X \perp Y$, niin tieto sm:n Y arvosta ei anna tietoa X :n jakaumasta eikä tieto X :n arvosta anna tietoa Y :n jakaumasta.

Kaavaa (7.1) pidetään mallina, kun määritellään ehdollinen tiheysfunktio jatkuvan yhteisjakauman tapauksessa.

Määritelmä 7.1. Olkoon (X, Y) :llä jatkuva jakauma tiheysfunktiolla $f_{X,Y}$. Tällöin sm:n X ehdollinen tiheysfunktio ehdolla $Y = y$ on

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad x \in \mathbb{R},$$

Kuva 7.1 Yhteistiheysfunktio ja ehdollisia tiheysfunktioita $x \mapsto f_{X|Y}(x | y)$.

kun y on sellainen, että $f_Y(y) > 0$. Vastaavasti määritellään

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \quad y \in \mathbb{R},$$

kun $f_X(x) > 0$.

Huomaa, että $x \mapsto f_{X|Y}(x | y)$ on tiheysfunktio, sillä se on ei-negatiivinen ja

$$\int_{-\infty}^{\infty} f_{X|Y}(x | y) dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = 1.$$

Funktio $x \mapsto f_{X|Y}(x | y)$ saadaan yhteistiheysfunktioista $f_{X,Y}$ normalisoimalla sen vaakasuora "leike" $x \mapsto f_{X,Y}(x, y)$ tiheysfunktioiksi. Vastaavasti funktio $y \mapsto f_{Y|X}(y | x)$ saadaan normalisoimalla pystysuora leike $y \mapsto f_{X,Y}(x, y)$. Kuva 7.1 havainnollistaa sitä, miten ehdollinen tiheysfunktio saadaan yhteistiheysfunktioista.

Jatkuvan yhteisjakauman voi esittää useamman kuin yhden tiheysfunktion avulla kunhan ne vain yhtyvät melkein kaikkialla. Ehdollisen tf:n määrittelyyn saadaan käyttää mitä tahansa ytf:n määritelmää, joten myöskään ehdollinen tf ei ole yksikäsitteinen. Toisaalta osoittautuu, ettei tästä monikäsitteisyydestä synny sen suurempia ongelmia, minkä takia sitä ei oteta huomioon puheavassa: puhutaan ehdollisesta tiheysfunktioista (eikä esim. ehdollisen tiheysfunktion tietystä versiosta).

Jatkuvan yhteisjakauman tapauksessa ehdon $\{Y = y\}$ todennäköisyys on aina nolla. Tämän takia ehdolliselle tiheysfunktiolle ei voida antaa suoraan tulkintaa ehdollisen todennäköisyyden avulla, vaan kyseessä on määritelmä, jota yritämme seuraavaksi motivoida.

Ehdollisen tiheysfunktion kaavalle voidaan antaa infinitesimaalinen tulkinta

$$f_{X|Y}(x | y) dx = \frac{f_{X,Y}(x, y) dx dy}{f_Y(y) dy} = \frac{P(X \in dx, Y \in dy)}{P(Y \in dy)} = P(X \in dx | Y \in dy)$$

Tämä ei ole pelkästään formaalia leikkiä symboleilla, vaan sille voidaan antaa

tulkinta raja-arvojen kautta. Jos ytf on riittävän sileä, niin voidaan tulkita, että

$$P(X \in A \mid Y \in dy) = \lim_{h \rightarrow 0^+} P(X \in A \mid y \leq Y \leq y + h).$$

Jos $f_{X,Y}$ ja f_Y ovat riittävän sileitä, ja $h > 0$ on pieni, niin suurpiirteisesti laskien nähdään, että

$$\begin{aligned} P(X \in A \mid y \leq Y \leq y + h) &= \frac{\frac{1}{h} P(X \in A, y \leq Y \leq y + h)}{\frac{1}{h} P(y \leq Y \leq y + h)} \\ &= \frac{\int_A dx \frac{1}{h} \int_y^{y+h} f_{X,Y}(x, t) dt}{\frac{1}{h} \int_y^{y+h} f_Y(u) du} \approx \frac{\int_A f_{X,Y}(x, y) dx}{f_Y(y)} \end{aligned}$$

ja tässä tarkkuus kasvaa, kun $h \rightarrow 0^+$. Näin saadaan ehdolliselle tf:lle tulkinta raja-arvona, ts.

$$P(X \in A \mid y \leq Y \leq y + h) \xrightarrow{h \rightarrow 0^+} \int_A f_{X|Y}(x \mid y) dx, \quad \text{kaikilla } A \subset \mathbb{R}.$$

(Jätämme auki, mitä ehtoja ytf:lle ja reunatiheydelle f_Y pitäisi asettaa, jotta lasku saataisiin vietyä täsmällisesti läpi.) Jatkuvan sm:n arvo havaitaan vain tietyllä tarkkuudella, joten sovelluksissa tehdään tämän tapainen tulkinta, kun käsitellään ehdollista tf:a ehdolla $Y = y$, jossa y on havaittu arvo. Näiden tarkastelujen avulla myös ehdolliselle tiheysfunktiolle olisi mahdollista antaa frekvenssitulkinta.

7.2 Kertolaskusääntö eli ketjusääntö

Ehdollisen ptnf/tf:n määritelmästä saadaan *kertolaskusääntö* eli *ketjusääntö*,

$$f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y \mid x) = f_Y(y) f_{X|Y}(x \mid y), \quad \text{kaikilla } x, y. \quad (7.2)$$

Tämä kaava vaatii hieman tulkintoja jatkuvan jakauman tapauksessa, koska ytf ei ole yksikäsitteinen. Kaavan voidaan ajatella tarkoittavan sitä, että jos reunatiheysfunktio $f_X(x)$ ja ehdollinen tf $f_{Y|X}(y \mid x)$ johdetaan lähtemällä liikkeelle joistakin ytf:n (mahdollisesti toisistaan eriävistä) versioista, niin niiden tulo kelpaa yhteistiheysfunktiksi.

Lisäksi kaavassa on se ongelma, että esim. $f_{Y|X}(y \mid x)$ ei ole välttämättä määritelty kaikilla x . Sovitaan, että kertolaskusäännön yhteydessä ehdollisen tf:n tai ptnf:n määritelmää laajennetaan (tarvittaessa) jollakin tavalla niihin pisteisiin, joissa ehtomuuttujan tiheys on nolla, esim. sopimalla, että

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{f_{X,Y}(x, y)}{f_X(x)}, & \text{kun } f_X(x) > 0, \\ 0, & \text{muuten.} \end{cases} \quad (7.3)$$

(Yhtä hyvin voitaisiin käyttää jotakin muuta johdonmukaista sopimusta.) Toiselle ehdolliselle tf:lle $f_{X|Y}(x \mid y)$ käytetään tietenkin samanlaista laajennusta. Tämän jälkeen kertolaskusääntö pitää ongelmattomasti paikkansa kaikilla $x, y \in \mathbb{R}$. Nämä komplikaatiot eivät aiheuta todellisia ongelmia käytännön laskuissa.

Kertolaskusäännöstä voidaan ratkaista toinen ehdollisista tiheysfunktioista, jos reunatiheydet sekä toinen ehdollinen tf tunnetaan. Esim., kun $f_X(x) > 0$ (ja muut x -arvot eivät ole relevantteja), on

$$f_{Y|X}(x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_Y(y) f_{X|Y}(x|y)}{f_X(x)}. \quad (7.4)$$

Tämä on *Bayesin kaava* tiheysfunktioille.

Huomaa, että Bayesin kaavasta seuraa, että kiinteällä x

$$f_{Y|X}(y|x) = \frac{1}{f_X(x)} f_Y(y) f_{X|Y}(x|y) \propto f_Y(y) f_{X|Y}(x|y) = f_{X,Y}(x,y).$$

Verrannollisuus $g(y) \propto h(y)$ tarkoittaa sitä, että

$$g(y) = c h(y)$$

jollakin verrannollisuusvakiolla c (joka saa riippua parametreista, ja muista muuttujista paitsi muuttujasta y). Tämän ansiosta ehdollinen jakauma on toisinaan helppo tunnistaa tarkastelemalla ytf:n lauseketta.

Esimerkki 7.1. Esimerkissä 6.4 ytf:llä on lauseke

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} 1\{0 < y < \exp(-\frac{1}{2}x^2)\}.$$

Ehdolliset tf:t voidaan laskea joko jakamalla ytf esimerkissä johdetuilla reunatf:oilla, tai seuraavalla päättelyllä.

a) $f_{Y|X}$: Kiinteällä x on ytf:llä $f_{X,Y}(x,y)$ nollaa suurempi vakioarvo, kun $0 < y < \exp(-\frac{1}{2}x^2)$. Tämän takia ehdollinen jakauma on tasajakauma,

$$Y | (X = x) \sim U(0, \exp(-\frac{1}{2}x^2)).$$

Kun muistetaan, että reunajakaumassaan $X \sim N(0, 1)$, niin yhteisjakauma voidaan esittää muodossa

$$\begin{aligned} Y | X &\sim U(0, \exp(-\frac{1}{2}X^2)), \\ X &\sim N(0, 1). \end{aligned}$$

b) $f_{X|Y}$: Kiinteällä $0 < y < 1$ on ytf:llä nolla suurempi vakioarvo tietyllä x -akselin välillä, joka saadaan ratkaisemalla epäyhtälö

$$0 < y < \exp(-\frac{1}{2}x^2)$$

muuttujan x suhteen. Tämä lasku ratkaistiin esimerkissä 6.4. Tuloksesta tunnistetaan, että ehdollinen jakauma on tasajakauma

$$X | (Y = y) \sim U(-\sqrt{-2 \ln y}, \sqrt{-2 \ln y}), \quad 0 < y < 1.$$

△

7.3 Diskreetin ja jatkuvan muuttujan yhteisjakauma

Tarkastellaan nyt samalla perusjoukolla määriteltyjen sm:ien X ja Y yhteisjakaumaa siinä tapauksessa, kun X :n jakauma on diskreetti Y :n jatkuva. Olkoon X :n mahdolliset arvot x_1, x_2, \dots ja olkoon sen (reuna-)ptnf f_X . Voidaan osoittaa, että tällöin on olemassa ehdolliset tiheysfunktiot $y \mapsto f_{Y|X}(y | x_i), i \geq 1$ siten, että

$$P(X = x_i, Y \in B) = f_X(x_i) \int_B f_{Y|X}(y | x_i) dy, \quad \text{kaikilla } i \text{ ja kaikilla } B \subset \mathbb{R},$$

mutta todistuksessa tarvitaan mittateoriaa (tulos seuraa ns. Radonin-Nikodymin lauseesta). Tällöin yhteisjakauman esittää funktio

$$f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y | x)$$

siinä mielessä, että seuraaventyypiset todennäköisyydet saadaan laskettua summamalla diskreetin muuttujan ja integroimalla jatkuvan muuttujan suhteen,

$$P(X \in A, Y \in B) = \sum_{x \in A} \int_B f_{X,Y}(x, y) dy, \quad A, B \subset \mathbb{R}.$$

Voimme myös tässä tapauksessa kutsua yhteisjakauman esitystä $f_{X,Y}$ tiheydeksi tai tiheysfunktioksi, mutta on tärkeää pitää mielessä, että yhden muuttujan suhteen summataan ja toisen suhteen integroidaan.

Diskreetin muuttujan X ptnf ehdolla $Y = y$ määritellään Bayesin kaavalla, siis

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

kun $f_Y(y) > 0$. Tämän kaavan voisi myös motivoida rajankäynnin kautta samaan tapaan kuin jatkuvan yhteisjakauman tapauksessa tehtiin.

Kertolaskusääntö ja Bayesin kaava ovat voimassa myös tässä tapauksessa, ja tarvittaessa ehdollisen tiheyden $f_{Y|X}(y | x)$ ja ehdollisen ptnf:n $f_{X|Y}(x | y)$ määritelmää voidaan laajentaa myös sellaisille argumenteille, joilla ne eivät vielä tulleet määriteltyä.

Siinä tapauksessa jossa toisen jakauma on diskreetti ja toisen jatkuva, odotusarvoja laskettaessa diskreetin muuttujan suhteen summataan ja jatkuvan suhteen integroidaan.

Lause 7.1. *Olkoon (X, Y) sv, jossa X :llä on diskreetti ja Y :llä jatkuva jakauma, ja olkoon $Z = g(X, Y)$ jokin sen reaaliarvoinen muunnos. Tällöin*

$$EZ = \sum_x \int g(x, y) f_{X,Y}(x, y) dy = \int \sum_x g(x, y) f_{X,Y}(x, y) dy$$

mikäli

$$\sum_x \int |g(x, y)| f_{X,Y}(x, y) dy < \infty.$$

7.4 Ehdollinen odotusarvo

Määritelmä 7.2 (Ehdollinen odotusarvo ja ehdollinen varianssi ehdolla $sm:n$ arvo). Olkoon (X, Y) sv, $g(X, Y)$ jokin sen muunnos, ja oletetaan, että osaamme määrittellä $sm:n$ Y ehdollisen jakauman ehdolla $X = x$. $Sm:n$ $g(X, Y)$ ehdollinen odotusarvo ehdolla $X = x$,

$$E(g(X, Y) | X = x),$$

on satunnaismuuttujan $g(x, Y)$ odotusarvo, kun $Y:n$ jakaumana käytetään sen ehdollista jakaumaa ehdolla $X = x$. $Sm:n$ $g(X, Y)$ ehdollinen varianssi ehdolla $X = x$,

$$\text{var}(g(X, Y) | X = x),$$

on satunnaismuuttujan $g(x, Y)$ varianssi, kun $Y:n$ jakaumana käytetään sen ehdollista jakaumaa ehdolla $X = x$.

Jos $Y:n$ jakauma on jatkuva, on

$$E(g(X, Y) | X = x) = \int g(x, y) f_{Y|X}(y | x) dy,$$

ja

$$\text{var}(g(X, Y) | X = x) = \int [g(x, y) - E(g(X, Y) | X = x)]^2 f_{Y|X}(y | x) dy.$$

Jos jälkimmäisessä kaavassa binomin neliö kerrotaan auki ja järjestellään termit, nähdään että

$$\text{var}(g(X, Y) | X = x) = E[(g(X, Y))^2 | X = x] - \{E[g(X, Y) | X = x]\}^2. \quad (7.5)$$

Tämä on tutun kaavan $\text{var} Z = EZ^2 - (EZ)^2$ vastine ehdolliselle varianssille. Jos Y on diskreetti sm , ylläolevissa kaavoissa tarvitaan summia integraalien sijasta, ja identiteetti (7.5) pysyy voimassa.

Jos funktio g on muotoa

$$g(x, y) = g_1(x) g_2(x, y),$$

niin tietenkin

$$E(g_1(X) g_2(X, Y) | X = x) = g_1(x) E(g_2(X, Y) | X = x). \quad (7.6)$$

Tämä voidaan ilmaista sanomalla, että tunnetut tekijät saadaan vetää ulos ehdollisesta odotusarvosta.

Valintaa $g(x, y) = y$ vastaava ehdollinen odotusarvo on nimeltään $Y:n$ ehdollinen odotusarvo ehdolla $X = x$.

Määritelmä 7.3 ($Y:n$ ehdollinen odotusarvo ehdolla $X = x$; regressiofunktio). $Sm:n$ Y ehdollinen odotusarvo ehdolla $X = x$ on sen ehdollisen jakauman odotusarvo,

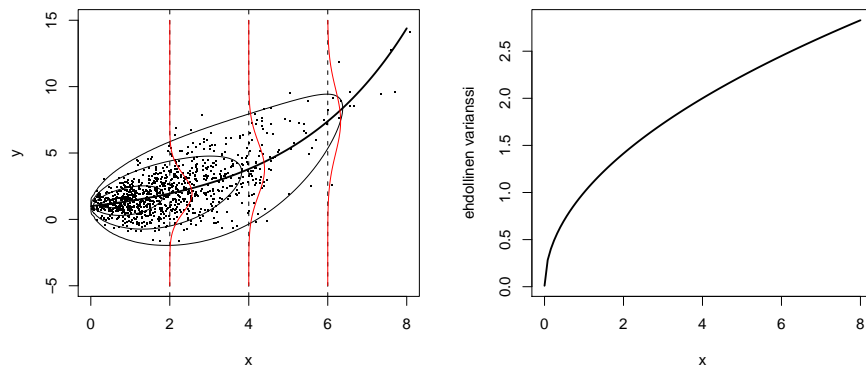
$$E(Y | X = x)$$

Funktiota

$$x \mapsto E(Y | X = x)$$

kutsutaan $Y:n$ regressiofunktioiksi $X:n$ suhteen (engl. *regression function of Y on X*).

Kuva 7.2 (a) Yhteisjakauma, ehdollisia tiheysfunktioita $y \mapsto f_{Y|X}(y | x)$ ja ehdollinen odotusarvo $E(Y | X = x)$ (b) ehdollinen varianssi $\text{var}(Y | X = x)$.



Jos Y :llä on jatkuva jakauma, on

$$E(Y | X = x) = \int y f_{Y|X}(y | x) dy,$$

ja jos Y :llä on diskreetti jakauma, on

$$E(Y | X = x) = \sum_y y f_{Y|X}(y | x).$$

Kuvassa 7.2 on piirretty ehdollinen odotusarvo $E(Y | X = x)$ eli regressiofunktio sekä ehdollinen varianssi $\text{var}(Y | X = x)$ kuvan 7.1 yhteisjakaumalle.

Määritelmä 7.4 (Ehdollinen odotusarvo ehdolla satunnaismuuttuja). Merkitään väliaikaisesti

$$m(x) = E(g(X, Y) | X = x).$$

Sovitaan, että $m(x) = 0$ niillä x , joilla $m(x)$ se ei muuten tule määritellyksi, eli joilla ehdollinen jakauma $Y | (X = x)$ ei ole määritelty. Tämän jälkeen voidaan vapaasti puhua satunnaismuuttujasta $m(X)$. Sitä kutsutaan sm:n $g(X, Y)$:n ehdolliseksi odotusarvoksi ehdolla sm X . Sille käytetään merkintää

$$E(g(X, Y) | X) = m(X).$$

Huomautus. Mieleen saattaa juolahtaa merkintä $E(g(X, Y) | X = X)$, mutta se on järjetön. Ehdollinen odotusarvo $E(g(X, Y) | X)$ on sellainen sm, joka saa arvon $E(g(X, Y) | X = x)$ (todennäköisyydellä yksi) silloin, kun X saa arvon x .

Lause 7.2. Jos $E[g(X, Y)^2] < \infty$, niin $E(g(X, Y) | X)$ on keskineliövirheen mielessä paras sm:n $g(X, Y)$ ennuste sm:n X funktion avulla, ts.

$$E[(g(X, Y) - E(g(X, Y) | X))^2] \leq E[(g(X, Y) - h(X))^2]$$

valitaan funktio $h : \mathbb{R} \rightarrow \mathbb{R}$ miten tahansa.

Todistus. Harjoitustehtävä. □

Erityisesti, jos tehtävänä on ennustaa $sm:n$ Y arvo jollakin $sm:n$ X arvon funktiolla, niin keskineliövirheen mielessä paras mahdollinen ennuste $m(x)$, jossa m on regressiofunktio $m(x) = E[Y | X = x]$.

Lause 7.3. *Odotusarvo voidaan laskea iteroituna odotusarvona, eli*

$$Eg(X, Y) = EE(g(X, Y) | X),$$

mikäli odotusarvo $Eg(X, Y)$ on olemassa laajennettuna reaali-lukuna.

Todistus. Esitetään perustelu diskreetissä tapauksessa.

$$\begin{aligned} Eg(X, Y) &= \sum_{x,y} g(x, y) f_{X,Y}(x, y) = \sum_{x,y} g(x, y) f_X(x) f_{Y|X}(y | x) \\ &= \sum_x f_X(x) \sum_y g(x, y) f_{Y|X}(y | x) \\ &= EE(g(X, Y) | X). \end{aligned} \quad \square$$

Esimerkki 7.2. Tarkastellaan yhteisjakaumaa

$$\begin{aligned} X | Y &\sim \text{Bin}(Y, \theta), \\ Y &\sim \text{Poi}(\lambda) \end{aligned}$$

jossa $\lambda > 0$ ja $0 < \theta < 1$. Nyt

$$E(X | Y) = Y\theta,$$

joten

$$EX = EE(X | Y) = E(Y\theta) = \theta EY = \theta\lambda.$$

△

Kuten ehdollinen odotusarvo, myös ehdollinen varianssi voidaan laskea ehdon satunnaismuuttuja X (eikä ehdolla sen arvo $X = x$). Ensinnäkin määritellään funktio

$$v(x) = \text{var}(g(X, Y) | X = x),$$

ja määritelmää jatketaan koko reaaliakselille sopimalla, että $v(x) = 0$ niillä argumenteilla, joilla ehdollinen jakauma $Y | (X = x)$ ei ole luonnostaan määritelty. Tämän jälkeen määritellään, että

$$\text{var}(g(X, Y) | X) = v(X).$$

Lause 7.4. *$Sm:n$ varianssi on yhtä kuin sen ehdollisen varianssin odotusarvon sekä ehdollisen odotusarvon varianssin summa, eli*

$$\text{var}(g(X, Y)) = E \text{var}(g(X, Y) | X) + \text{var} E(g(X, Y) | X). \quad (7.7)$$

Todistus. Harjoitustehtävä. □

Huomautus. Mittateoriaan pohjautuvassa todennäköisyysteoriassa ehdollinen odotusarvo $E(g(X, Y) | X)$ määritellään täysin erilaisella tekniikalla, kuin millä me sen teemme. Lisäksi oppikirjoissa se tavallisesti määritellään vain siinä tapauksessa, jossa $g(X, Y)$ on integroitava eli kun $Eg(X, Y) \in \mathbb{R}$, mikä on yhtäpitävää sen kanssa, että $E|g(X, Y)| < \infty$. Integroitavuus on tässä kuitenkin tarpeeton rajoitus; kvasi-integroitavuus, eli odotusarvon $Eg(X, Y)$ olemassaolo laajennettuna reaalityyppinä on riittävää mielekkään ehdollisen odotusarvon käsitteen olemassaololle. Tätä laajennettua teoriaa on hieman hankala löytää oppikirjoista, mutta se esitellään esim. Ashin [1], tai Chown ja Teicherin [2] tai Jacodin ja Protterin teoksessa [3]. Tämän laajennetun teorian perusteella satunnaismuuttujan $g(X, Y)$ integroitavuuden saa tarkistaa laskemalla odotusarvon $E|g(X, Y)|$ kaavalla $EE(|g(X, Y)| | X)$. Mikäli tulos on äärellinen, niin sitten $g(X, Y)$ on integroitava, eli $Eg(X, Y)$ on reaalityyppi, jonka puolestaan saa laskea kaavalla $EE(g(X, Y) | X)$. Integraalin äärellisyyden tarkistamisessa voidaan siis soveltaa samaa tekniikkaa kuin mitä käytetään Fubinin lauseen kohdalla.

Kirjallisuutta

- [1] Robert B. Ash. *Probability and Measure Theory*. Academic Press, 2nd edition, 2000.
- [2] Y. S. Chow and H. Teicher. *Probability Theory: Independence, Interchangeability, Martingales*. Springer-Verlag, 2nd edition, 1988.
- [3] Jean Jacod and Philip Protter. *Probability Essentials*. Springer, 2nd edition, 2002.

7.5 Yhteisjakauman määrittely kertolaskukaavan avulla

Tilastollisia malleja spesifoidaan usein kertomalla, mikä on yhden sm:n reuna-jakauma ja toisen ehdollinen jakauma. Tällöin voidaan puhua *hierarkkisesta mallista*. Tarkastellaan tästä esimerkkejä.

Esimerkki 7.3. (Diskreetti ja diskreetti.) Hyönteisen munimien munien lkm $Y \sim \text{Poi}(\lambda)$, jossa $\lambda > 0$ on jakauman odotusarvo. Kukin munista kehittyy toukaksi toisistaan riippumatta tn:llä $0 < \theta < 1$. Olkoon X toukaksi kehittyvien munien lkm. Tällöin

$$X | (Y = y) \sim \text{Bin}(y, \theta).$$

Yhteisjakauma on diskreetti, ja sen yptnf on

$$f_{X,Y}(x, y) = f_Y(y) f_{X|Y}(x | y) = e^{-\lambda} \frac{\lambda^y}{y!} \binom{y}{x} \theta^x (1 - \theta)^{y-x}, \quad 0 \leq x \leq y,$$

jossa x ja y ovat kokonaislukuja $0, 1, 2, \dots$

Tämä malli voitaisiin määrittellä myös sanomalla, että

$$\begin{aligned} X | Y &\sim \text{Bin}(Y, \theta), \\ Y &\sim \text{Poi}(\lambda) \end{aligned}$$

jossa $\lambda > 0$ ja $0 < \theta < 1$ ovat vakioita.

Tässä mallissa voidaan esim. kysyä, mikä on X :n reunajakauma. Sen ptnf voidaan esittää summana

$$f_X(x) = \sum_y f_{X,Y}(x, y).$$

Joidenkin laskujen jälkeen havaitaan, että reunajakauma on $X \sim \text{Poi}(\lambda\theta)$. \triangle

Esimerkki 7.4. (Jatkuva ja diskreetti.) Olkoon sm:lla Θ betajakauma $\text{Be}(\alpha, \beta)$, jossa $\alpha, \beta > 0$ ovat vakioita. Kolikkoa heitetään n kertaa, ja lasketaan kuinka monta kertaa saadaan klaava, kun klaavan todennäköisyys on Θ . Ehdolla $\Theta = \theta$ klaavojen lukumäärällä X on jakauma $\text{Bin}(n, \theta)$. Tällöin yhteisjakauma voidaan esittää funktiolla

$$f_{\Theta, X}(\theta, x) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \binom{n}{x} \theta^x (1-\theta)^{n-x},$$

jossa $0 < \theta < 1$ ja $x = 0, 1, \dots, n$.

Tämä malli voitaisiin spesifioida myös sanomalla, että

$$\begin{aligned} X \mid \Theta &\sim \text{Bin}(n, \Theta), \\ \Theta &\sim \text{Be}(\alpha, \beta) \end{aligned}$$

jossa $\alpha, \beta > 0$ ja $n \geq 0$ ovat vakioita.

Mikä on sm:n Θ jakauma, kun havaitaan, että $X = x$? Thomas Bayes esitti 1700-luvulla ratkaisun tähän kysymykseen. Me ratkaisemme kysymyksen tarkastelemalla hetken yhteisjakauman esitystä muuttujan θ funktiona, minkä jälkeen on selvää, että

$$\Theta \mid (X = x) \sim \text{Be}(\alpha + x, \beta + n - x).$$

Tämä lasku on esimerkki *Bayes-päätelystä* (eng. *Bayesian inference*). Tässä binomijakauman parametria pidettiin satunnaismuuttujana, jolla oli tietty priorijakauma (lat. *a priori*, ennen [havaintoa]). Havainnon jälkeen parametrilla on sen ehdollinen jakauma ehdolla havainto, ja tätä jakaumaa kutsutaan parametrin posteriorijakaumaksi (lat. *a posteriori*, [havainnon] jälkeen). Tässä esimerkissä kävi niin onnellisesti, että sekä priorijakauma että posteriorijakauma ovat samassa jakaumaperheessä: ne ovat molemmat beetajakaumia. Tällöin sanotaan, että sekä priorin että posteriorin ovat *konjugaatti-* eli *liittoperheessä* (engl. *conjugate family*) tarkasteltavan uskottavuusfunktion suhteen. Asia voidaan ilmaista myös sanomalla, että beetajakauma on binomiuskottavuuden *liitto-* eli *konjugaattipriori*. \triangle

Esimerkki 7.5. (Jatkuva ja jatkuva.) Useasta ohjelmistoista löytyy satunnaislukugeneraattori normaalijakaumalle. Tarkastellaan seuraavaa algoritmia, jossa $\sigma_X > 0$, μ_X ja $\sigma_Z > 0$ ovat annettuja lukuja ja m on jokin funktio, joka palauttaa reaalityyppisen reaalilukuargumentilla.

1. Simuloi $X \sim N(\mu_X, \sigma_X^2)$.
2. Simuloi $Z \sim N(0, \sigma_Z^2)$.
3. Aseta $Y = m(X) + Z$.

Kuvaile sm:ien X ja Y yhteisjakauma.

Huomautus. Kun satunnaislukugeneraattoria kutsutaan monta kertaa (kuten edellä askelissa 1 ja 2), niin eri kerroilla palautettavia lukuja voidaan pitää keskenään riippumattomien satunnaismuuttujien arvoina, sillä todelliset satunnaislukugeneraattorit toimivat tällä tavoin. Tätä ominaisuutta pidetään kirjallisuudessa niin itsestään selvänä asiana, että sitä harvoin vaivaudutaan selittämään.

Ratkaisu: Yhteisjakauma voidaan esittää kaavoilla

$$\begin{aligned} Y | X &\sim N(m(X), \sigma_Z^2) \\ X &\sim N(\mu_X, \sigma_X^2). \end{aligned}$$

Ehdollinen jakauma $[Y | X = x]$ nähdään siitä, että simuloinnissa $X \perp Z$, joten X :n arvolla x ehdollistaminen ei muuta Z :n jakaumaa. Kyseisessä ehdollisessa jakaumassa sm:n X arvo on vakio x , ja lopuksi vakion x :n tilalle kirjoitetaan satunnaismuuttuja X .

Koska X :n jakauma on jatkuva, ja Y :n ehdollinen jakauma ehdolla $X = x$ on jatkuva kaikilla x , on myös yhteisjakauma jatkuva, ja sen ytf on

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x) f_{Y|X}(y|x) \\ &= \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x - \mu_X)^2}{\sigma_X^2}\right) \frac{1}{\sigma_Z \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(y - m(x))^2}{\sigma_Z^2}\right) \end{aligned}$$

Huomioita: X :n reunajakauma on $N(\mu_X, \sigma_X^2)$ ja regressiofunktio on

$$E[Y | X = x] = m(x), \quad \text{var}(Y | X = x) = \sigma_Z^2,$$

mutta esim. Y :n reunajakauma ei ole mikään tuttu jakauma, ellei regressiofunktion m muotoa rajoiteta. \triangle

Esimerkki 7.6. (Jatkoa edelliselle esimerkille.) Valitaan regressiofunktiolle m lineaarinen muoto. Tällöin $m(X)$ antaa keskineliövirheen mielessä parhaan ennusteen satunnaismuuttujan Y arvolle. Koska $m(X)$ on lineaarinen, sen täytyy myös olla keskineliövirheen mielessä paras lineaarinen ennuste, joten jakson 6.9 kaavojen mukaan täytyy olla

$$m(x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X),$$

jossa μ_Y ja $\sigma_Y > 0$ (oletus) ovat sm:n Y odotusarvo ja keskihajonta ja $-1 < \rho < 1$ (oletus) on X :n ja Y :n korrelaatiokerroin. Lisäksi pätee

$$\sigma_Z^2 = \sigma_Y^2 (1 - \rho^2).$$

Pitkähköjen mutta suoraviivaisten laskujen jälkeen ytf voidaan kirjoittaa muotoon

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \frac{1}{\sqrt{\det \mathbf{C}}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{z} - \boldsymbol{\mu})\right), \quad \mathbf{z} = \begin{bmatrix} x \\ y \end{bmatrix}$$

jossa

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Tässä $\boldsymbol{\mu}$ on satunnaisvektorin (X, Y) odotusarvovektori, ja \mathbf{C} sen kovarianssimatriisi.

Johdettu jakauma on kaksiulotteinen normaalijakauma parametreilla $\boldsymbol{\mu}$ ja \mathbf{C} , mikä merkitään

$$(X, Y) \sim N(\boldsymbol{\mu}, \mathbf{C}).$$

Käsitlemme myöhemmin tarkemmin moniulotteista normaalijakaumaa eli multinormaalijakaumaa, joka on normaalijakauman yleistys mielivaltaiseen dimensioon. \triangle

Luku 8

Moniulotteiset jakaumat

Moniulotteinen satunnaisvektori on muuten samanlainen kuin kaksiulotteinen satunnaisvektori, paitsi että sillä voi olla useampi kuin kaksi komponenttia. Sen jakaumaa kuvaillaan samaan tapaan kuin kaksiulotteisen satunnaisvektorin jakaumaa. Tämän takia suurin osa tämän luvun ideoista on ennestään tuttuja, eikä tuttuja tuloksia enää perustella uudestaan. Tilastotieteilijä tarvitsee moniulotteisia jakaumia sen takia, että hän voisi määritellä ja analysoida tilastollisia malleja.

Tässä luvussa uutta on se, että reunajakauma voidaan määritellä mille tahansa komponenttien osajoukolle, ja ehdollisia jakaumia voidaan laskea ehdollistamalla millä tahansa komponenttien osajoukolla. Ehdollinen riippumattomuus on käsite, joka ei ole mielekäs kuin vasta dimensioissa kolmesta eteenpäin.

Moniulotteisten jakaumien havainnollistaminen on hankalaa. Kuvien piirtäminen ei onnistu korkeissa dimensioissa, vaan niiden sijasta pitää luottaa kaavojen manipulointiin.

8.1 Satunnaisvektori

Avaruuden \mathbb{R}^n vektoria (x_1, \dots, x_n) merkitään $\mathbf{x} = (x_1, \dots, x_n)$. Silloin, kun se esiintyy matriisien kanssa samassa lausekkeessa, se ymmäretään pystyvektoriksi eli $n \times 1$ -matriisiksi,

$$\mathbf{x} = (x_1, \dots, x_n) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Määritelmä 8.1. Jos X_1, \dots, X_n ovat samalla perusjoukolla Ω määriteltyjä satunnaismuuttujia, niin $\mathbf{X} = (X_1, \dots, X_n)$ on n -ulotteinen satunnaisvektori (lyh. sv). Se on kuvaus $\Omega \rightarrow \mathbb{R}^n$ siten, että

$$\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)) = \begin{bmatrix} X_1(\omega) \\ \vdots \\ X_n(\omega) \end{bmatrix}$$

Määritelmä 8.2 (Yhteiskertymäfunktio). Satunnaismuuttujien X_1, \dots, X_n yhteiskertymäfunktio (ykf) eli satunnaisvektorin $\mathbf{X} = (X_1, \dots, X_n)$ kertymäfunktio

on

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leq x_1, \dots, X_n \leq x_n),$$

jossa $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Satunnaisvektorin jakauma määritellään samoin kuin aikaisemmin, eli se on funktio

$$B \mapsto P(\mathbf{X} \in B), \quad B \subset \mathbb{R}^n.$$

Kuten kaksiolotteisessa tapauksessa, ykf määrää jakauman myös dimensiossa n . Korkeissa dimensioissa ykf:a käytetään harvoin konkreettisissa laskuissa, vaan se on pikemminkin teoreettinen apuväline.

Jos kaikki satunnaisvektorin \mathbf{X} komponentit X_i ovat diskreettejä sm:ia, niin sen ptnf eli sm:ien X_i yhteispistetodennäköisyysfunktio (yptnf) on

$$f_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}) = P(X_1 = x_1, \dots, X_n = x_n). \quad (8.1)$$

Jos $g(\mathbf{X})$ on sv:n \mathbf{X} reaaliarvoinen muunnos, niin sen odotusarvo saadaan kaavalla

$$Eg(\mathbf{X}) = \sum_{\mathbf{x}} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}), \quad (8.2)$$

mikäli kyseinen summa suppenee itseisesti.

Satunnaisvektorilla \mathbf{X} on jatkuva jakauma, mikäli sillä on tiheysfunktio $f_{\mathbf{X}}$, mikä tarkoittaa sitä, että

$$P(\mathbf{X} \in B) = \int_B f_{\mathbf{X}} = \int_B f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \quad \text{kaikilla } B \subset \mathbb{R}^n. \quad (8.3)$$

Tällöin sen komponenteilla X_1, \dots, X_n on jatkuva yhteisjakauma, ja tiheysfunktioita

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{\mathbf{X}}(\mathbf{x}), \quad \mathbf{x} = (x_1, \dots, x_n)$$

kutsutaan sm:ien X_1, \dots, X_n yhteistiheysfunktioiksi. Edellä kyseessä on n -kertainen integraali,

$$\int_B f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} = \int \dots \int_B f_{\mathbf{X}}(\mathbf{x}) \, dx_1 \dots dx_n,$$

joka voidaan laskea iteroituna integraalina käyttämällä mielivaltaista integrointijärjestystä.

Jos $g(\mathbf{X})$ on sv sv:n \mathbf{X} reaaliarvoinen muunnos, niin sen odotusarvo saadaan kaavalla

$$Eg(\mathbf{X}) = \int g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} \quad (8.4)$$

mikäli kyseinen integraali suppenee itseisesti. Tällöin integraali voidaan laskea iteroituna integraalina käyttäen mielivaltaista integrointijärjestystä. Kun integrointijoukkoa ei merkitä näkyviin, tarkoitetaan koko avaruuden (tässä \mathbb{R}^n) yli laskettua integraalia. (Jos dimensio $n \geq 2$, niin tämä merkintä ei voi sekaantua määräämättömän integraalin (engl. *indefinite integral*) merkinnän kanssa, sillä määräämättömän integraalin käsitettä käytetään vain dimensiossa yksi.)

Jatkuvan (yhteis)jakauman tapauksessa ykf on

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{s_1=-\infty}^{x_1} \dots \int_{s_n=-\infty}^{x_n} f_{\mathbf{X}}(s_1, \dots, s_n) \, ds_1 \dots ds_n.$$

Kun tätä funktiota derivoidaan osittain jokaisen argumenttinsa suhteen, nähdään että

$$\frac{\partial^n F_{\mathbf{X}}(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n} = f_{\mathbf{X}}(x_1, \dots, x_n),$$

ainakin niissä pisteissä (x_1, \dots, x_n) , joissa $f_{\mathbf{X}}$ on jatkuva. Voidaan osoittaa, että tämä funktio kelpaa jatkuvasti jakautuneen sv:n tiheysfunktioiksi (kun se jatketaan mielivaltaisesti niissä pisteissä, joissa ko. derivaatta ei ole olemassa).

Jos \mathbf{X} on diskreetti sv, ja \mathbf{Y} on jatkuvasti jakautunut sv, ja ne on määritelty samalla perusjoukolla Ω , niin silloin niiden yhteisjakauma voidaan esittää funktiolla

$$f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}),$$

jossa $f_{\mathbf{X}}$ on sv:n \mathbf{X} reunajakauman (yhteis-)ptnf, ja $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})$ on sv:n \mathbf{Y} ehdollinen (yhteis-)tf ehdolla $\mathbf{X} = \mathbf{x}$. Jos $g(\mathbf{X}, \mathbf{Y})$ on reaaliarvoinen sm, niin sen odotusarvo saadaan kaavalla

$$Eg(\mathbf{X}, \mathbf{Y}) = \sum_{\mathbf{x}} \int g(\mathbf{x}, \mathbf{y}) f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) d\mathbf{y}, \quad (8.5)$$

mikäli kyseinen odotusarvo on olemassa.

Kahden satunnaisvektorin \mathbf{X} ja \mathbf{Y} riippumattomuus määritellään täysin vastaavasti kuin satunnaismuuttujien kohdalla, ks. jakso 2.9. Samalla päättelyllä kuin aikaisemmin,

$$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \quad \Rightarrow \quad g(\mathbf{X}) \perp\!\!\!\perp h(\mathbf{Y}),$$

kun g ja h ovat mielivaltaisia vektoriargumentin reaaliarvoisia funktioita.

8.2 Odotusarvovektori ja kovarianssimatriisi

Määritelmä 8.3. Asetamme seuraavat määritelmät, mikäli kyseessä olevat odotusarvot ovat olemassa.

- Satunnaisvektorin $\mathbf{X} = (X_1, \dots, X_n)$ odotusarvo(vektori) on n -komponenttinen vakiovektori

$$E\mathbf{X} = E(\mathbf{X}) = (EX_1, \dots, EX_n) = \begin{bmatrix} EX_1 \\ \vdots \\ EX_n \end{bmatrix}.$$

- Dimensioltaan $m \times n$ olevan satunnaismatriisin $\mathbf{Z} = [Z_{ij}]$ odotusarvo(matriisi) on se $m \times n$ -vakiomatriisi, jonka alkio (i, j) on alkion Z_{ij} odotusarvo, ts.

$$E\mathbf{Z} = E \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mn} \end{bmatrix} = \begin{bmatrix} EZ_{11} & EZ_{12} & \cdots & EZ_{1n} \\ EZ_{21} & EZ_{22} & \cdots & EZ_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ EZ_{m1} & EZ_{m2} & \cdots & EZ_{mn} \end{bmatrix}.$$

- Satunnaisvektorien $\mathbf{X} = (X_1, \dots, X_n)$ ja $\mathbf{Y} = (Y_1, \dots, Y_k)$ (välinen) *kovarianssi* on $n \times k$ -matriisi

$$\begin{aligned} \text{cov}(\mathbf{X}, \mathbf{Y}) &= E[(\mathbf{X} - E\mathbf{X})(\mathbf{Y} - E\mathbf{Y})^T] \\ &= \begin{bmatrix} \text{cov}(X_1, Y_1) & \text{cov}(X_1, Y_2) & \dots & \text{cov}(X_1, Y_k) \\ \text{cov}(X_2, Y_1) & \text{cov}(X_2, Y_2) & \dots & \text{cov}(X_2, Y_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, Y_1) & \text{cov}(X_n, Y_2) & \dots & \text{cov}(X_n, Y_k) \end{bmatrix}. \end{aligned}$$

Mikäli $\text{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$, niin sanotaan, että satunnaisvektorit \mathbf{X} ja \mathbf{Y} *eivät korreloi* eli että ne ovat *korreloimattomat* (engl. *uncorrelated*).

- Satunnaisvektorin $\mathbf{X} = (X_1, \dots, X_n)$ *kovarianssimatriisi* on $n \times n$ -matriisi

$$\begin{aligned} \text{Cov}(\mathbf{X}) &= \text{cov}(\mathbf{X}, \mathbf{X}) = E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T] \\ &= \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \dots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & \text{cov}(X_n, X_n) \end{bmatrix}. \end{aligned}$$

Määritelmästä seuraa suoraan, että

$$E(\mathbf{Z}^T) = (E\mathbf{Z})^T. \quad (8.6)$$

Laskusääntö

$$E[\mathbf{A}\mathbf{Z}\mathbf{B} + \mathbf{C}] = \mathbf{A}(E\mathbf{Z})\mathbf{B} + \mathbf{C}, \quad (8.7)$$

on voimassa kun \mathbf{Z} on satunnaismatriisi ja \mathbf{A} , \mathbf{B} ja \mathbf{C} ovat vakiomatriiseja, joiden dimensiot ovat sellaiset, että lauseke on määritelty. Ts. vakiomatriisit saadaan vetää ulos odotusarvosta, jos ne sijaitsevat matriisitulossa äärimmäisenä vasemmalla tai äärimmäisenä oikealla.

Mikäli \mathbf{X} koostuu osavektoreista $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$, ja $\mathbf{G}(\mathbf{Y})$ ja $\mathbf{H}(\mathbf{Z})$ ovat matriisiarvoisia lausekkeita, niin

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \quad \Rightarrow \quad \mathbf{G}(\mathbf{Y}) \perp\!\!\!\perp \mathbf{H}(\mathbf{Z}). \quad (8.8)$$

(Koska tämä kaava pätee matriisiarvoisille funktioille, se pätee tietenkin myös vektori- tai skalaariarvoisille funktioille.) Toisin sanoen *riippumattomien satunnaisvektorien funktiot ovat keskenään riippumattomia*.

Tässä matriisilausekkeiden riippumattomuus tarkoittaa tietenkin sitä, että kaikki matriisin $\mathbf{G}(\mathbf{Y})$ komponentit ovat riippumattomia kaikista matriisin $\mathbf{H}(\mathbf{Z})$ komponenteista. Riippumattomuudesta ja matriisikertolaskun määritelmästä seuraa, että

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} \quad \Rightarrow \quad E[\mathbf{G}(\mathbf{Y})\mathbf{H}(\mathbf{Z})] = E[\mathbf{G}(\mathbf{Y})]E[\mathbf{H}(\mathbf{Z})] \quad (8.9)$$

mikäli dimensiot ovat matriisitulossa $\mathbf{G}(\mathbf{Y})\mathbf{H}(\mathbf{Z})$ yhteensopivat ja mikäli odotusarvot ovat olemassa.

Lause 8.1 (Kovarianssin ominaisuuksia).

(a) Jos $\mathbf{X} \perp \mathbf{Y}$, ja \mathbf{X} on m -komponenttinen ja \mathbf{Y} on n -komponenttinen, niin

$$\text{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}_{m \times n}.$$

(b) Suvien \mathbf{X} ja \mathbf{Y} kovarianssi voidaan laskea myös kaavalla

$$\text{cov}(\mathbf{X}, \mathbf{Y}) = E[\mathbf{X}\mathbf{Y}^T] - (E\mathbf{X})(E\mathbf{Y})^T.$$

(c) Jos $\mathbf{v} \in \mathbb{R}^m$ on vakiovektori, ja \mathbf{X} on n -komponenttinen sv, niin

$$\text{cov}(\mathbf{v}, \mathbf{X}) = \mathbf{0}_{m \times n}, \quad \text{cov}(\mathbf{X}, \mathbf{v}) = \mathbf{0}_{n \times m}.$$

(d) $\text{cov}(\mathbf{X}, \mathbf{Y}) = \text{cov}(\mathbf{Y}, \mathbf{X})^T$.

(e) Jos $\mathbf{X}_1, \mathbf{X}_2$ ja \mathbf{Y} ovat satunnaisvektoreita, niin

$$\begin{aligned} \text{cov}(\mathbf{X}_1 + \mathbf{X}_2, \mathbf{Y}) &= \text{cov}(\mathbf{X}_1, \mathbf{Y}) + \text{cov}(\mathbf{X}_2, \mathbf{Y}), \\ \text{cov}(\mathbf{Y}, \mathbf{X}_1 + \mathbf{X}_2) &= \text{cov}(\mathbf{Y}, \mathbf{X}_1) + \text{cov}(\mathbf{Y}, \mathbf{X}_2). \end{aligned}$$

(f) Jos \mathbf{A} ja \mathbf{B} ovat vakiomatriiseja ja \mathbf{v} ja \mathbf{w} ovat vakiovektoreja, niin

$$\text{cov}(\mathbf{A}\mathbf{X} + \mathbf{v}, \mathbf{B}\mathbf{Y} + \mathbf{w}) = \mathbf{A} \text{cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}^T. \quad (8.10)$$

Erityisesti

$$\text{Cov}(\mathbf{A}\mathbf{X} + \mathbf{v}) = \mathbf{A} \text{Cov}(\mathbf{X}) \mathbf{A}^T. \quad (8.11)$$

Todistus. Kaikki kohdat seuraavat helpoilla laskuilla kovarianssin määritelmästä sekä laskusäännöistä (8.7) ja (8.9). \square

Huomaa, että satunnaisvektorin kovarianssimatriisi on symmetrinen ja että sen päälävistäjällä on komponenttisatunnaisuuttujen varianssit.

Palautetaan mieleen, että neliömatriisi \mathbf{B} on

- säännöllinen eli kääntyvä eli *ei-singulaarinen*, jos sillä on olemassa käänteismatriisi \mathbf{B}^{-1} ;
- symmetrinen, jos $\mathbf{B}^T = \mathbf{B}$;
- positiivisesti semidefiniitti, jos $\mathbf{v}^T \mathbf{B} \mathbf{v} \geq 0$ kaikilla \mathbf{v} ;
- positiivisesti definiitti, jos $\mathbf{v}^T \mathbf{B} \mathbf{v} > 0$ kaikilla $\mathbf{v} \neq \mathbf{0}$.

Lause 8.2. *Sv:n \mathbf{X} kovarianssimatriisi on symmetrinen ja positiivisesti semidefiniitti matriisi. Jos se ei ole positiivisesti definiitti, niin sv:n \mathbf{X} komponenttien välillä on lineaarisia sidosehtoja, eli \mathbf{X} saa todennäköisyydellä yksi arvoja vain tietyltä kiinteältä hypertasolta.*

Todistus. Symmetrisyys seuraa kovarianssioperaattorin symmetrisyydestä. Positiivinen semidefiniittisyys seuraa siitä, että

$$\begin{aligned} \mathbf{v}^T \text{Cov}(\mathbf{X}) \mathbf{v} &= \mathbf{v}^T E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T] \mathbf{v} \\ &= E[\mathbf{v}^T (\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T \mathbf{v}] = E\left[\left(\mathbf{v}^T (\mathbf{X} - E\mathbf{X})\right)^2\right] \geq 0, \end{aligned}$$

sillä ei-negatiivisen sm:n odotusarvo on ei-negatiivinen. Jos kovarianssimatriisi ei ole positiivisesti definiitti, niin on olemassa $\mathbf{v} \neq \mathbf{0}$ siten, että

$$0 = \mathbf{v}^T \text{Cov}(\mathbf{X}) \mathbf{v} = E \left[(\mathbf{v}^T (\mathbf{X} - E\mathbf{X}))^2 \right],$$

joten todennäköisyydellä yksi on

$$\mathbf{v}^T (\mathbf{X} - E\mathbf{X}) = 0. \quad \square$$

Huomautus. Jos $\text{Cov} \mathbf{X}$ ei ole positiivisesti definiitti, niin \mathbf{X} :llä ei voi olla jatkuva jakauma, sillä hypertaso on alempiulotteisena pintana nollamittainen.

8.3 Ehdolliset jakaumat, kertolaskusääntö ja ehdollinen odotusarvo

Tarkastellaan sv:ia \mathbf{Z} , jonka r ensimmäistä koordinaattia muodostavat sv:n \mathbf{X} ja loput s koordinaattia sv:n \mathbf{Y} ,

$$\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) = (X_1, \dots, X_r, Y_1, \dots, Y_s).$$

Oletetaan, että sv:n \mathbf{Z} yhteisjakauman esittää tiheys

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}), \quad \mathbf{z} = (\mathbf{x}, \mathbf{y}).$$

Nyt sallimme sen tilanteen, että jotkut \mathbf{Z} :n koordinaatit ovat diskreettejä sm:ia ja lopuilla on jatkuva yhteisjakauma. Tällöin \mathbf{X} :n reunajakauman esittää tiheys $f_{\mathbf{X}}$, joka saadaan summaamalla pois tiheydestä $f_{\mathbf{X}, \mathbf{Y}}$ sv:n \mathbf{Y} diskreetit komponentit ja integroimalla pois sen jatkuvat komponentit. Merkintöjä väärinkäyttämällä tämän idean voi esittää kaavalla

$$f_{\mathbf{X}}(\mathbf{x}) = \int f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) \, d\mathbf{y}, \quad (8.12)$$

missä integraalimerkki tarkoittaa summaamista diskreettien komponenttien kohdalla. Tässä yhteydessä voidaan sanoa, että \mathbf{X} :n jakauma marginalisoidaan yhteisjakaumasta. Vastaavasti $f_{\mathbf{Y}}$ saadaan kaavalla

$$f_{\mathbf{Y}}(\mathbf{y}) = \int f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) \, d\mathbf{x}.$$

Se tilanne, missä pitää esittää mielivaltaiselle osalle komponenteista reunajakauma saadaan aina palautettua edelliseen tilanteeseen permutoimalla koordinaatteja.

Esimerkki 8.1. Olkoot U ja V diskreettejä sm:ia ja (X, Y) jatkuvasti jakautunut sv. Tällöin esim. sm:ien V ja Y yhteisjakauman esittää tiheys

$$f_{V, Y}(v, y) = \sum_u \int f_{U, V, X, Y}(u, v, x, y) \, dx,$$

sillä (permutoidaan V ja Y ensimmäisiksi)

$$f_{V, Y, U, X}(v, y, u, x) = f_{U, V, X, Y}(u, v, x, y),$$

ja (summataan tai integroidaan muut pois)

$$f_{V,Y}(v, y) = \sum_u \int f_{V,Y,U,X}(v, y, u, x) dx.$$

△

Sv:n \mathbf{Y} ehdollisen jakauman ehdolla $\mathbf{X} = \mathbf{x}$ esittää tiheys

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) = \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})},$$

kun $f_{\mathbf{X}}(\mathbf{x}) > 0$. Tarvittaessa tämä lauseke voidaan määrittellä esim. nollassa niillä \mathbf{x} , joilla $f_{\mathbf{X}}(\mathbf{x}) = 0$. Vastaavasti määritellään

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y}) = \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}.$$

Kertolaskukaava (eli ketjusääntö) on voimassa, eli

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) = f_{\mathbf{Y}}(\mathbf{y}) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y}).$$

Kertolaskukaavaa voidaan iteroida. Oletetaan, että sv $\mathbf{X} = (\mathbf{U}, \mathbf{V})$. Tällöin

$$f_{\mathbf{U},\mathbf{V}}(\mathbf{u}, \mathbf{v}) = f_{\mathbf{U}}(\mathbf{u}) f_{\mathbf{V}|\mathbf{U}}(\mathbf{v} | \mathbf{u}),$$

minkä takia

$$f_{\mathbf{U},\mathbf{V},\mathbf{Y}}(\mathbf{u}, \mathbf{v}) = f_{\mathbf{U}}(\mathbf{u}) f_{\mathbf{V}|\mathbf{U}}(\mathbf{v} | \mathbf{u}) f_{\mathbf{Y}|\mathbf{U},\mathbf{V}}(\mathbf{y} | \mathbf{u}, \mathbf{v}).$$

Tätä prosessia voidaan jatkaa, kunnes ollaan päästy skalaarikomponentteihin asti. Esimerkiksi neljän sm:n U, V, X, Y tiheys voidaan esittää muodossa

$$f_{U,V,X,Y}(u, v, x, y) = f_U(u) f_{V|U}(v | u) f_{X|U,V}(x | u, v) f_{Y|X,U,V}(y | x, u, v).$$

Kertolaskusääntöä voidaan soveltaa yhtä hyvin käyttämällä jotakin muuta sm:ien permutaatiota.

Kertolaskusääntö pätee myös ehdollisille jakaumille. Esim.

$$\begin{aligned} f_{\mathbf{U},\mathbf{V}|\mathbf{Y}}(\mathbf{u}, \mathbf{v} | \mathbf{y}) &= \frac{f_{\mathbf{U},\mathbf{V},\mathbf{Y}}(\mathbf{u}, \mathbf{v}, \mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{f_{\mathbf{U},\mathbf{Y}}(\mathbf{u}, \mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} \frac{f_{\mathbf{U},\mathbf{V},\mathbf{Y}}(\mathbf{u}, \mathbf{v}, \mathbf{y})}{f_{\mathbf{U},\mathbf{Y}}(\mathbf{u}, \mathbf{y})} \\ &= f_{\mathbf{U}|\mathbf{Y}}(\mathbf{u} | \mathbf{y}) f_{\mathbf{V}|\mathbf{U},\mathbf{Y}}(\mathbf{v} | \mathbf{u}, \mathbf{y}). \end{aligned}$$

Ehdollisen jakauman reunajakauman voi laskea marginalisoimalla ehdollista yhteisjakaumaa. Esim.

$$\begin{aligned} f_{\mathbf{U}|\mathbf{Y}}(\mathbf{u} | \mathbf{y}) &= \frac{f_{\mathbf{U},\mathbf{Y}}(\mathbf{u}, \mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{1}{f_{\mathbf{Y}}(\mathbf{y})} \int f_{\mathbf{U},\mathbf{V},\mathbf{Y}}(\mathbf{u}, \mathbf{v}, \mathbf{y}) d\mathbf{v} \\ &= \int f_{\mathbf{U},\mathbf{V}|\mathbf{Y}}(\mathbf{u}, \mathbf{v} | \mathbf{y}) d\mathbf{v}. \end{aligned}$$

Huomautus. Kun jakaumien välisiä yhteyksiä johdetaan tähän tapaan kertolaskukaavan ja marginalisoinnin kautta, niin tiheysfunktioiden alaindeksit jätetään usein kirjoittamatta, koska ne selviävät tiheysfunktion argumenteista. Tämä merkintöjen väärinkäyttö ei johda sekaannukseen, kunhan pidetään mielessä, että esim. $f(x)$, $f(y)$, $f(x | y)$ ja $f(y | x)$ ovat tyypillisesti kaikki eri funktioita.

Satunnaisvektorin $\mathbf{g}(\mathbf{X}, \mathbf{Y})$ ehdollinen odotusarvo(vektori) ehdolla $\mathbf{X} = \mathbf{x}$ määritellään siten, että se on satunnaisvektorin $\mathbf{g}(\mathbf{x}, \mathbf{Y})$ odotusarvovektori, kun \mathbf{Y} :n jakaumana käytetään ehdollista jakaumaa $f_{\mathbf{Y}|\mathbf{X}}(\cdot | \mathbf{x})$. Merkintöjä väärinkäyttämällä

$$E(\mathbf{g}(\mathbf{X}, \mathbf{Y}) | \mathbf{X} = \mathbf{x}) = \int \mathbf{g}(\mathbf{x}, \mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) d\mathbf{y}.$$

Tässä integraalimerkki voi tarkoittaa summausta joidenkin vektorin \mathbf{y} komponenttien suhteen. Jos ehdollistetaan satunnaisvektorilla \mathbf{X} , niin satunnaisvektori $E(\mathbf{g}(\mathbf{X}, \mathbf{Y}) | \mathbf{X})$ määritellään niin, että se on $\mathbf{m}(\mathbf{X})$, kun

$$\mathbf{m}(\mathbf{x}) = E(\mathbf{g}(\mathbf{X}, \mathbf{Y}) | \mathbf{X} = \mathbf{x}),$$

ja funktion \mathbf{m} määritelmä jatketaan tarvittaessa nollavektoriksi niillä argumentin arvoilla, jolla se ei luonnostaan ole määritelty.

Satunnaisvektorin $\mathbf{g}(\mathbf{X}, \mathbf{Y})$ ehdollinen kovarianssimatriisi ehdolla $\mathbf{X} = \mathbf{x}$ määritellään vastaavasti satunnaisvektorin $\mathbf{g}(\mathbf{x}, \mathbf{Y})$ kovarianssimatriisina, kun \mathbf{Y} :n jakaumana käytetään ehdollista jakaumaa $f_{\mathbf{Y}|\mathbf{X}}(\cdot | \mathbf{x})$, ja sille voidaan käyttää merkintää

$$\text{Cov}(\mathbf{g}(\mathbf{X}, \mathbf{Y}) | \mathbf{X} = \mathbf{x}).$$

Jos ehdollistetaan satunnaisvektorilla \mathbf{X} , niin ehdollinen kovarianssimatriisi on satunnaismatriisi, ja sitä merkitään

$$\text{Cov}(\mathbf{g}(\mathbf{X}, \mathbf{Y}) | \mathbf{X}).$$

8.4 Ehdollinen riippumattomuus

Satunnaisvektorit \mathbf{X} ja \mathbf{Y} ovat riippumattomia, jos

$$f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y}).$$

kaikilla argumenteilla \mathbf{x}, \mathbf{y} . Vastaavasti satunnaisvektorit $\mathbf{X}_1, \dots, \mathbf{X}_n$ ovat riippumattomia, jos

$$f_{\mathbf{X}_1, \dots, \mathbf{X}_n}(\mathbf{x}_1, \dots, \mathbf{x}_n) = f_{\mathbf{X}_1}(\mathbf{x}_1) \cdots f_{\mathbf{X}_n}(\mathbf{x}_n)$$

kaikilla argumenteilla.

Määritelmä 8.4. Satunnaisvektorit \mathbf{X} ja \mathbf{Y} ovat riippumattomia ehdolla \mathbf{Z} , jos ne ovat riippumattomia niiden ehdollisessa yhteisjakaumassa ehdolla $\mathbf{Z} = \mathbf{z}$ kaikilla \mathbf{z} , ts. mikäli

$$f_{\mathbf{X}, \mathbf{Y}|\mathbf{Z}}(\mathbf{x}, \mathbf{y} | \mathbf{z}) = f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x} | \mathbf{z}) f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y} | \mathbf{z}), \quad \text{kaikilla } \mathbf{x}, \mathbf{y} \text{ ja } \mathbf{z} \quad (8.13)$$

Vastaavasti, satunnaisvektorit $\mathbf{X}_1, \dots, \mathbf{X}_n$ ovat riippumattomia ehdolla \mathbf{Z} , mikäli

$$f_{\mathbf{X}_1, \dots, \mathbf{X}_n|\mathbf{Z}}(\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{z}) = \prod_{i=1}^n f_{\mathbf{X}_i|\mathbf{Z}}(\mathbf{x}_i | \mathbf{z})$$

kaikilla $\mathbf{x}_1, \dots, \mathbf{x}_n$ ja kaikilla \mathbf{z} .

Ehdollisesti riippumattomat satunnaisvektorit eivät tyypillisesti ole marginaalisesti riippumattomia, eli riippumattomia niiden (reuna-)yhteisreunajakaumassa. Esim. jos $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}) \mid \mathbf{Z}$, niin

$$f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \int f_{\mathbf{X}, \mathbf{Y} \mid \mathbf{Z}}(\mathbf{x}, \mathbf{y} \mid \mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} = \int f_{\mathbf{X} \mid \mathbf{Z}}(\mathbf{x} \mid \mathbf{z}) f_{\mathbf{Y} \mid \mathbf{Z}}(\mathbf{y} \mid \mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z},$$

eikä tämä tyypillisesti enää faktoroidu.

8.5 Tilastollisia malleja

Tilastolliseen päättelyyn on kaksi pääasiallista lähestymistapaa: ns. frekventistinen päättely ja Bayes-päättely. Ne molemmat perustuvat uskottavuusfunktion käyttöön. Suurpiirteisesti selittäen kysymyksenasettelu on seuraava.

Meillä on käsillä numeerinen aineisto vektorin $\mathbf{y} = (y_1, \dots, y_n)$ muodossa. Ennen havaintojen tekoa aineiston arvot ovat epävarmoja (mittausvirheiden, populaation luonnollisen vaihtelun tms. syyn takia). Tämän takia mallinamme tilanteen niin, että \mathbf{y} on jollakin perusjoukolla määritellyn satunnaisvektorin \mathbf{Y} havaittu arvo, ts.

$$\mathbf{y} = \mathbf{Y}(\omega^{\text{act}}),$$

jossa ω^{act} on todennäköisyysmallissa aktualisoitunut alkeistapaus.

Tyypillisesti vektorin \mathbf{Y} jakauma mallinnetaan parametrisella mallilla, jossa on yksi tai useampia parametreja $\theta_1, \dots, \theta_p$, jotka kootaan parametrivektoriksi $\boldsymbol{\theta}$. Kun parametrivektorin arvo on kiinnitetty, niin sv:n \mathbf{Y} jakauman esittää tiheys

$$\mathbf{y} \mapsto f(\mathbf{y} \mid \boldsymbol{\theta}).$$

Kiinnostuksen kohteena on sv:n \mathbf{Y} jakauma. Sitä voidaan arvoida, jos ensin arvioidaan eli estimoidaan tuntematonta parametrivektoria $\boldsymbol{\theta}$.

Kun aineisto \mathbf{y} on havaittu, ja havaittua arvoa käytetään funktion $f(\mathbf{y} \mid \boldsymbol{\theta})$ ensimmäisenä argumenttina, niin funktiota

$$\boldsymbol{\theta} \mapsto f(\mathbf{y} \mid \boldsymbol{\theta})$$

kutsutaan uskottavuusfunktiksi (engl. *likelihood function*). Tässä yhteydessä tiheysfunktioista saatetaan jättää pois kertoimia, jotka eivät riipu parametrivektorista $\boldsymbol{\theta}$, ja silti kyseistä funktiota edelleen kutsutaan uskottavuusfunktiksi.

Ns. klassisessa eli frekventistisessä tilastotieteessä parametrivektoria $\boldsymbol{\theta}$ pidetään tuntemattomana vakiona, josta tiedetään vain, missä joukossa (eli parametriavaruudessa) sen arvot voivat olla. Tällöin merkintään $f(\mathbf{y} \mid \boldsymbol{\theta})$ ei liitetä tulkintaa ehdollisena jakaumana, koska parametrivektorille ei ole olemassa mitään todennäköisyysjakaumaa. Tyypillisempi merkintä tässä tilanteessa olisikin $f(\mathbf{y}; \boldsymbol{\theta})$. Tunnetuin estimointiperiaate on ns. *suurimman uskottavuuden*, eli SU-periaate (engl. *maximum likelihood*, *ML*), jonka mukaan parametrivektorin parhaana estimaattina pidetään sitä arvoa $\hat{\boldsymbol{\theta}}^{\text{ML}}$ parametriavaruudessa, joka maksimoi uskottavuusfunktion. Sitä kutsutaan suurimman uskottavuuden estimaatiksi (eli SU-estimaatiksi) (engl. *maximum likelihood estimate*, *MLE*).

Bayes-päättelyssä parametrivektoria pidetään satunnaisvektorin $\boldsymbol{\Theta}$ arvona. Funktio $f(\mathbf{y} \mid \boldsymbol{\theta})$ ja ehdollinen jakauma $f_{\mathbf{Y} \mid \boldsymbol{\Theta}}(\mathbf{y} \mid \boldsymbol{\theta})$ samastetaan. Bayes-päättelyssä

tilastollisessa mallissa tarvitaan uskottavuusfunktion lisäksi reunajakauma vektorille Θ . Tätä kutsutaan priorijakaumaksi. Parametrivektorin ja havaintovektorin yhteisjakauman esittää tiheys

$$(\mathbf{y}, \boldsymbol{\theta}) \mapsto f_{\Theta, \mathbf{Y}}(\boldsymbol{\theta}, \mathbf{y}) = f_{\Theta}(\boldsymbol{\theta}) f_{\mathbf{Y}|\Theta}(\mathbf{y} | \boldsymbol{\theta}).$$

Kun on saatu havainto $\mathbf{Y} = \mathbf{y}$, niin tilastolliset päätelmät tehdään karakterisoidulla parametrivektorin posteriorijakaumaa, joka on ehdollinen jakauma

$$f_{\Theta|\mathbf{Y}}(\boldsymbol{\theta} | \mathbf{y}) = \frac{f_{\Theta, \mathbf{Y}}(\boldsymbol{\theta}, \mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{f_{\Theta}(\boldsymbol{\theta}) f_{\mathbf{Y}|\Theta}(\mathbf{y} | \boldsymbol{\theta})}{\int f_{\Theta}(\mathbf{t}) f_{\mathbf{Y}|\Theta}(\mathbf{y} | \mathbf{t}) d\mathbf{t}}$$

Kummassakin tilastollisen päättelyn lähestymistavassa tarvitaan konkreettinen kaava uskottavuusfunktiolle. Tämän takia tilastotieteilijän pitää osata johtaa moniulotteisten satunnaisvektoreiden tiheyksiä.

Esimerkki 8.2. Tehdas valmistaa komponentteja, jotka ovat joko toimivia tai viallisia. Määritellään

$$Y_i = \begin{cases} 1, & \text{jos } i\text{:s komponentti on viallinen,} \\ 0, & \text{jos } i\text{:s komponentti toimii.} \end{cases}$$

Tässä tilanteessa tilastollinen malli voisi yksinkertaisimmillaan olla seuraava. Kun parametrin arvo on $0 \leq \theta \leq 1$, niin satunnaismuuttujat Y_i ovat riippumattomia ja samoin jakautuneita, ja niillä on jakauma Bernoulli(θ). (“Onnistuminen” tarkoittaa nyt sitä, että komponentti ei toimi.) Tällöin uskottavuusfunktio on

$$f(y_1, \dots, y_n | \theta) = \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1-y_i}, \quad \mathbf{y} = (y_1, \dots, y_n) \in \{0, 1\}^n.$$

Tässä tilanteessa SU-estimaatti löytyy helposti: se on k/n , jossa $k = \sum y_i$ on viallisten komponenttien lukumäärä.

Bayes-päätelyä harrastava tilastotieteilijä antaisi parametrille Θ priorijakauman, esim. välin $(0, 1)$ tasaajakauman $\text{Be}(1, 1)$, ja johtaisi sitten posteriorijakauman käyttämällä samaa uskottavuusfunktiota. Tällöin siis sm:t Y_i ovat riippumattomia ehdolla Θ , ja niillä on jakauma Bernoulli(Θ). Yhteisjakauma on

$$f_{\Theta, \mathbf{Y}}(\theta, \mathbf{y}) = f(y_1, \dots, y_n | \theta), \quad 0 \leq \theta \leq 1, \quad \mathbf{y} \in \{0, 1\}^n$$

ja posteriorijakauma on $\text{Be}(1+k, 1+n-k)$. Tämän laskun olemme jo ratkaisseet esimerkissä 7.4.

Ajatellaan seuraavaksi, että komponentin i valmistuksen yhteydessä mitataan jokin tieto x_i , joka on jossakin yhteydessä sen seikan kanssa, tuleeko komponentista i toimiva vai viallinen. Tällöin x_i :tä kutsutaan selittäväksi muuttujaksi (engl. *explanatory variable*) tai kovariaatiksi (engl. *covariate*). Eräs tapa ottaa kovariaatit mukaan malliin on seuraava.

Otetaan parametreiksi α ja β , ja yritetään selittää onnistumistodennäköisyyttä i :nnessä Bernoullin kokeessa lineaarisen lausekkeen $\alpha + \beta x_i$ kautta. Tämä lauseke voi saada mielivaltaisia arvoja, joten se ei suoraan sovi Bernoullin jakauman todennäköisyysparametriksi. Sen sijaan kaavalla

$$p(\alpha, \beta, x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)},$$

saadaan luku joka on nollan ja yhden välillä oli arvot α , β ja x mitä tahansa. Kun parametrit ovat α, β , niin mallin mukaan

$$Y_i \sim \text{Bernoulli}(p(\alpha, \beta, x_i)), \quad i = 1, \dots, n$$

riippumattomasti. Tällöin uskottavuusfunktio on

$$f(y_1, \dots, y_n | \alpha, \beta) = \prod_{i=1}^n \left(\frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right)^{y_i} \left(\frac{1}{1 + \exp(\alpha + \beta x_i)} \right)^{1-y_i}.$$

Tämä on esimerkki logistisesta regressiomallista, joka puolestaan on erikoistapaus yleistetystä lineaarisesta mallista (engl. *generalized linear model, GLM*). Monesta tilastollisesta ohjelmistosta löytyy rutiini, joka ratkaisee SU-estimaatit iteratiivisesti logistisessa regressiossa.

Bayes-päätelyssä parametreille tarvitaan priorijakauma $f_{A,B}(\alpha, \beta)$, joka voisi olla esim. jokin kaksiulotteinen normaalijakauma. Mallin mukaan sm:t Y_i ovat riippumattomia ehdolla $(A, B) = (\alpha, \beta)$, ja sm:n Y_i ehdollinen jakauma on Bernoulli($p(\alpha, \beta, x_i)$). Parametrien ja aineiston yhteisjakauma on

$$f_{A,B}(\alpha, \beta) f(y_1, \dots, y_n | \alpha, \beta),$$

josta posteriorijakauman ominaisuudet joudutaan käytännössä selvittämään numeerisesti. \triangle

Esimerkki 8.3. Jossain tapauksissa havaittu aikasarja y_1, \dots, y_n voidaan mallintaa satunnaismuuttujilla Y_0, Y_1, \dots, Y_n , jossa $Y_0 = y_0$ on jokin tunnettu vakio, ja muuten pätee autoregressiivinen malli

$$Y_t = h(Y_{t-1}, \beta) + \epsilon_t, \quad t \geq 1.$$

Tässä $h(y, \beta)$ on jokin tunnettu funktio, ja sm:t $\epsilon_t \sim N(0, \sigma^2)$ riippumattomasti. Parametreja ovat β sekä σ .

Tässä mallissa on voimassa Markov-ominaisuus

$$f(y_t | y_1, \dots, y_{t-1}) = f(y_t | y_{t-1}), \quad \forall t \geq 1.$$

Kertolaskusääntö antaa yhteisjakaumalle esityksen

$$\begin{aligned} f(y_1, \dots, y_n) &= f(y_1) f(y_2 | y_1) f(y_3 | y_1, y_2) \cdots f(y_n | y_1, \dots, y_{n-1}) \\ &= f(y_1) f(y_2 | y_1) f(y_3 | y_2) \cdots f(y_n | y_{n-1}). \end{aligned}$$

Tästä saadaan uskottavuusfunktioiksi

$$f(y_1, \dots, y_n | \beta, \sigma) = \prod_{t=1}^n \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(y_t - h(y_{t-1}, \beta))^2}{\sigma^2}\right)$$

Tilanteessa voidaan soveltaa joko frekventistä tai Bayes-päätelyä. Jos erityisesti $h(y, \beta) = \beta y$, niin tilanteeseen löytyy paljon teoriaa ja valmisohjelmia, mutta yleisessä tapauksessa joudutaan itse kirjoittamaan ohjelmat mallin estimointia varten. \triangle

Esimerkki 8.4. (Puuttuva aineisto) Käytännön tilastotieteilijä joutuu usein tekemisiin sellaisten aineistojen kanssa, joista puuttuu jokin alunperin suunniteltu havainto. Toisinaan taas todennäköisyysmalli on kätevämpi muotoilla siten, että se sisältää satunnaismuuttujia, joiden arvoja ei voida havaita. Kummassakin tapauksessa ei-havaittuja satunnaismuuttujia voidaan kutsua puuttuvaksi aineistoksi (engl. *missing data*) tai niitä voidaan kutsua latenteiksi muuttujiksi (engl. *latent variables*) tai voidaan puhua apumuuttujista (engl. *auxiliary variables*). Tällaisessa mallissa havaitun aineiston uskottavuusfunktio saadaan marginalisoimalla havaitun aineiston \mathbf{Y} ja puuttuvan aineiston \mathbf{U} yhteistiheyttä

$$f_{\mathbf{Y}|\Theta}(\mathbf{y} | \boldsymbol{\theta}) = \int f_{\mathbf{U},\mathbf{Y}|\Theta}(\mathbf{u}, \mathbf{y} | \boldsymbol{\theta}) d\mathbf{u}.$$

Usein tätä integraalia ei osata käsitellä analyttisesti, minkä takia tähän puuttuvan aineiston tilanteen analysointiin on kehitetty erityismenetelmiä (esim. ns. EM-algoritmi). \triangle

8.6 Multinomijakauma

Olkoon (p_1, \dots, p_n) todennäköisyysvektori, eli kukin $p_i \geq 0$, ja

$$p_1 + \dots + p_n = 1.$$

Multinomijakauman $\text{Mult}(k, (p_1, \dots, p_n))$ ptnf on

$$f(x_1, \dots, x_n) = \binom{k}{x_1, \dots, x_n} p_1^{x_1} \dots p_n^{x_n}, \quad (8.14)$$

kun $x_1 + \dots + x_n = k$ ja nolla muuten.

Multinomijakauma syntyy, kun tarkastellaan k -kertaista toistokoetta, jossa kussakin toistossa toteutuu yksi n :sta vaihtoehdosta, jossa toistot ovat toisistaan riippumattomia. Muuttujat x_1, \dots, x_n ovat eri lopputulosten frekvenssit k :ssa toistossa. Kaava (8.14) perustellaan samalla tavalla kuin trinomijakauman tapauksessa (vrt. jakso 6.3); viimekädessä vedotaan jaksossa 1.5 esitettyyn multinomikertoimien kombinatoriseen luonnehdintaan.

Esimerkki 8.5. (Ryhmitellyt normaaliset havainnot.) Olkoot $Z_1, \dots, Z_k \sim N(\mu, \sigma^2)$ riippumattomasti. Tarkastellaan katkaisupisteitä $t_1 < t_2 < \dots < t_{n-1}$ sekä välejä

$$B_1 = (-\infty, t_1], \quad B_n = (t_{n-1}, \infty), \quad B_j = (t_{j-1}, t_j], \quad j = 2, \dots, n-1.$$

Määritellään sm X_j siten, että se kertoo, kuinka moni sm:sta Z_i saa arvon joukosta B_j . Sv:n (X_1, \dots, X_n) jakauma on multinomijakauma otoskoolla k ja todennäköisyysvektorilla (p_1, \dots, p_n) , jossa

$$p_1 = \Phi(t_1 | \mu, \sigma), \quad p_n = 1 - \Phi(t_{n-1} | \mu, \sigma), \\ p_j = \Phi(t_j | \mu, \sigma) - \Phi(t_{j-1} | \mu, \sigma), \quad j = 2, \dots, n-1.$$

Tässä $\Phi(x | \mu, \sigma)$ on jakauman $N(\mu, \sigma^2)$ kf.

Satunnaismuuttuja Z_i voisi esim. edustaa miespuolisen yliopisto-opiskelijan pituutta. Frekvenssit X_j saattaisivat olla peräisin kyselylomaketutkimuksesta, jossa pyydetään laittamaan rasti ruutuun sen mukaan, onko oma pituus pienempi kuin 160, välillä 160–170, välillä 170–180, välillä 180–190 vai suurempi kuin 190 senttimetriä. \triangle

Useimmiten multinomijakaumaa pidetään n -ulotteisena jakauma, mutta joissakin yhteyksissä on mielekkäämpää pitää sitä $(n-1)$ -ulotteisena jakauma (jolloin $X_n = k - X_1 - \dots - X_{n-1}$). Tällä konventiolla binomijakauma ja trinomijakauma ovat multinomijakauman erikoistapauksia.

Se, että (8.14) on ptnf seuraa suoraan *multinomikaavasta*, jonka mukaan kaikilla reaaliluvuilla a_1, \dots, a_n pätee

$$(a_1 + \dots + a_n)^k = \sum \binom{k}{x_1, \dots, x_n} a_1^{x_1} \dots a_n^{x_n}, \quad (8.15)$$

jossa summataan yli kaikkien (x_1, \dots, x_n) , joille $x_i \geq 0$ ovat kokonaislukuja, joiden summa on k . Kaikki multinomijakauman reunajakaumat ovat multinomijakaumia: esimerkiksi yksiulotteiset reunajakaumat ovat binomijakaumia. Jos joidenkin komponenttien arvoilla ehdollistetaan, niin multinomijakauman ehdolliset jakaumat ovat multinomijakaumia.

Multinomijakauman odotusarvovektori on

$$k(p_1, \dots, p_n),$$

ja sen kovarianssimatriisin alkiot ovat

$$\text{var } X_i = kp_i(1 - p_i), \quad \text{cov}(X_i, X_j) = -kp_i p_j, \quad i \neq j.$$

8.7 Tiheysfunktion muuntokaava

Tarkastellaan ensin diffeomorfisimia $\mathbf{g} : A \rightarrow B$, jossa $A, B \subset \mathbb{R}^n$ ovat avoimia joukkoja. Tällöin \mathbf{g} :llä on käänteisfunktio $\mathbf{h} : B \rightarrow A$, ja sekä \mathbf{g} että \mathbf{h} ovat jatkuvasti derivoituvia. Näiden kuvausten komponenttifunktiot ovat

$$\mathbf{g} = (g_1, \dots, g_n), \quad \mathbf{h} = (h_1, \dots, h_n).$$

Otetaan käyttöön merkintä $D_i h_j(\mathbf{y})$ tarkoittamaan reaaliarvoisen funktion h_j osittaisderivaattaa sen i :nnen argumentin suhteen pisteessä \mathbf{y} , ts.

$$D_i h_j(\mathbf{y}) = \frac{\partial}{\partial y_i} h_j(\mathbf{y}).$$

Tarkastelemme bijektiivistä vastaavuutta

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \iff \mathbf{x} = \mathbf{h}(\mathbf{y}),$$

eli

$$\mathbf{y} = \mathbf{y}(\mathbf{x}) \iff \mathbf{x} = \mathbf{x}(\mathbf{y}).$$

Kuvauksen $\mathbf{h} = (h_1, \dots, h_n)$ derivaatta(matriisi) (eli Jacobin matriisi) pisteessä $\mathbf{y} \in B$ on

$$\begin{bmatrix} D_1 h_1(\mathbf{y}) & D_2 h_1(\mathbf{y}) & \dots & D_n h_1(\mathbf{y}) \\ D_1 h_2(\mathbf{y}) & D_2 h_2(\mathbf{y}) & \dots & D_n h_2(\mathbf{y}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 h_n(\mathbf{y}) & D_2 h_n(\mathbf{y}) & \dots & D_n h_n(\mathbf{y}) \end{bmatrix}$$

Kuvauksen \mathbf{h} jacobiaani, eli Jacobin determinantti eli funktionaalideterminantti

$$J_{\mathbf{h}}(\mathbf{y}) = \frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)}$$

on Jacobin matriisin determinantti.

Tiheysfunktion muuntokaava n -ulotteisessa tilanteessa todistetaan täsmälleen samalla tekniikalla kuin kaksiulotteisessa tilanteessa (vrt. lause 6.13).

Lause 8.3. Jos $\mathbf{g} : A \rightarrow B$ on diffeomorfismi, $\mathbf{h} : B \rightarrow A$ on sen käänteiskuvaus, eli $\mathbf{h} = \mathbf{g}^{-1}$, ja sv:lla \mathbf{X} on jatkuva jakauma, ja $P(\mathbf{X} \in A) = 1$, niin sv:lla $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ on jatkuva jakauma tf:lla

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{h}(\mathbf{y})) |J_{\mathbf{h}}(\mathbf{y})|, \quad \text{kun } \mathbf{y} \in B,$$

ja nolla muualla.

Tämä tulos kannattaa pitää mielessään muodossa

$$f_{\mathbf{X}}(\mathbf{x}) |\partial \mathbf{x}| = f_{\mathbf{Y}}(\mathbf{y}) |\partial \mathbf{y}|, \quad \text{kun} \quad (8.16)$$

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \iff \mathbf{x} = \mathbf{h}(\mathbf{y}), \quad \mathbf{x} \in A, \mathbf{y} \in B \quad (8.17)$$

Esimerkki 8.6. (Affiini muunnos.) Jos \mathbf{A} on vakiomatriisi ja \mathbf{b} on vakiovektori siten, että seuraavassa kaavassa dimensiot ovat yhteensopivia, niin funktiota

$$\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}, \quad \mathbf{x} \in \mathbb{R}^n.$$

kutsutaan affiiniksi muunnokseksi. Jos edellä \mathbf{A} on säännöllinen matriisi (eli kääntyvä matriisi), jolloin sen täytyy olla neliömatriisi, niin tämän kuvauksen käänteiskuvaus on myös affiini muunnos, sillä

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} \iff \mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) = \mathbf{h}(\mathbf{y})$$

Kuvauksen \mathbf{h} jacobiaaniksi saadaan helpolla laskulla

$$J_{\mathbf{h}}(\mathbf{y}) = \frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}.$$

Jos \mathbf{X} on n -ulotteinen sv, jolla on jatkuva jakauma, ja sv \mathbf{Y} määritellään kaavalla $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, niin muistisäännöstä

$$f_{\mathbf{X}}(\mathbf{x}) |\partial \mathbf{x}| = f_{\mathbf{Y}}(\mathbf{y}) |\partial \mathbf{y}|$$

saadaan tulos

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| = f_{\mathbf{X}}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})) \frac{1}{|\det(\mathbf{A})|}.$$

△

Muistathan, että diffeomorfismi on aina samandimensioisten avaruuksien välinen kuvaus. Jos tavoitteena on johtaa alempidimensioisen sv:n tf, niin sitten kuvaus ensin täydennetään bijektioksi (mikäli tämä on mahdollista), johdetaan muunnoksen tf, ja lopuksi kiinnostuksen kohteena olevan osavektorin tf johdetaan integroimalla.

Toisinaan tiheysfunktion muuntokaavaa tarvitaan tilanteessa, jossa $\mathbf{g} : A \rightarrow B$ ei ole diffeomorfismi, mutta A voidaan osittaa paloihin A_0 ja $A_i, i \geq 1$ siten, että seuraavat ehdot ovat voimassa

- $P(\mathbf{X} \in A_0) = 0$.
- Kun $i \geq 1$, niin kuvaus \mathbf{g} rajoitettuna joukolle A_i , eli $\mathbf{g}|_{A_i}$, on diffeomorfini $A_i \rightarrow B_i$, jossa B_i on kyseisen funktion kuvajoukko. Olkoon $\mathbf{h}_i : B_i \rightarrow A_i$ tämän funktion käänteisfunktio.

Tällöin sv:lla $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ on jatkuva jakauma tf:lla

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i \geq 1} 1_{B_i}(\mathbf{y}) f_{\mathbf{X}}(\mathbf{h}_i(\mathbf{y})) |J_{\mathbf{h}_i}(\mathbf{y})|. \quad (8.18)$$

Tämä tulos on yksiulotteisen tuloksen (lause 2.15) moniulotteinen yleistys.

8.8 Satunnaisvektorin momenttiemäfunktio

Määritelmä 8.5. Satunnaisvektorin $\mathbf{X} = (X_1, \dots, X_n)$ momenttiemäfunktio, eli satunnaismuuttujien X_1, \dots, X_n yhteismomenttiemäfunktio on

$$M(\mathbf{t}) = M_{\mathbf{X}}(\mathbf{t}) = E \exp(\mathbf{t}^T \mathbf{X}) = E \exp \left(\sum_{j=1}^n t_j X_j \right),$$

niillä $\mathbf{t} = (t_1, \dots, t_n) \in \mathbb{R}^n$, joilla odotusarvo on määritelty. Sv:n \mathbf{X} kumulantiemäfunktio (eli sm:ien X_1, \dots, X_n yhteiskumulantiemäfunktio) on

$$K(\mathbf{t}) = K_{\mathbf{X}}(\mathbf{t}) = \ln M_{\mathbf{X}}(\mathbf{t}),$$

niillä $\mathbf{t} \in \mathbb{R}^n$, joilla $M(\mathbf{t})$ on määritelty.

Määritelmä 8.6. Satunnaisvektorin $\mathbf{X} = (X_1, \dots, X_n)$ karakteristinen funktio, eli satunnaismuuttujien X_1, \dots, X_n yhteisjakauman karakteristinen funktio on

$$\phi(\mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t}) = E \exp(i\mathbf{t}^T \mathbf{X}) = E \exp \left(i \sum_{j=1}^n t_j X_j \right),$$

jossa $i = \sqrt{-1}$.

Merkintä $D_i g$ tarkoittaa funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ osittaisderivaattaa sen i :nnen argumentin suhteen,

$$D_i g(\mathbf{x}) = \frac{\partial}{\partial x_i} g(x_1, \dots, x_n).$$

Merkintä D_i^k tarkoittaa k :nnetta osittaisderivaattaa i :nnen argumentin suhteen, ja jos k_1, \dots, k_n ovat ei-negatiivisia kokonaislukuja, niin

$$D_1^{k_1} D_2^{k_2} \dots D_n^{k_n} g(\mathbf{x}) = \frac{\partial^{k_1+k_2+\dots+k_n}}{\partial x_1^{k_1} \partial x_2^{k_2} \dots \partial x_n^{k_n}} g(\mathbf{x}).$$

Jos g on äärettömän monta kertaa jatkuvasti derivoituva, niin edellä osittaisderivaattojen laskujärjestyksellä ei ole väliä.

Momenttiemäfunktiolla ja kumulantiemäfunktiolla on samanlaiset ominaisuudet kuin skalaaritapauksessa.

Lause 8.4. Jos $M_{\mathbf{X}}(\mathbf{t})$ on olemassa jossakin epätyhjässä origon ympäristössä, niin $M_{\mathbf{X}}$ on origon ympäristössä äärettömän monta kertaa jatkuvasti derivoituva ja $M_{\mathbf{X}}$ voidaan esittää origon ympäristössä suppenevana potenssisarjana. Lisäksi

a) Kertaluvun (k_1, \dots, k_n) momentti saadaan derivaattana

$$E\left(X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}\right) = D_1^{k_1} D_2^{k_2} \dots D_n^{k_n} M_{\mathbf{X}}(\mathbf{0}),$$

jossa (k_1, \dots, k_n) on vektori, jonka komponentit ovat ei-negatiivisia kokonaislukuja.

b) $M_{\mathbf{X}}$ määrää $sv:n$ \mathbf{X} jakauman.

c) Karakteristinen funktio saadaan sijoituksella

$$\phi_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}}(it).$$

Määritelmä 8.7 (Gradientti, Hessen matriisi). (Pysty)vektori

$$\nabla g(\mathbf{x}) = (D_1 g(\mathbf{x}), D_2 g(\mathbf{x}), \dots, D_n g(\mathbf{x}))$$

on funktion $g: \mathbb{R}^n \rightarrow \mathbb{R}$ pisteessä \mathbf{x} laskettu *gradientti* (engl. *gradient*). Funktion g pisteessä \mathbf{x} laskettu Hessen matriisi $\mathbf{H}_g(\mathbf{x})$ (engl. *Hessian (matrix)*) on sen toisen kertaluvun osittaisderivaatoista koottu $n \times n$ -matriisi

$$\mathbf{H}_g(\mathbf{x}) = \begin{bmatrix} D_1 D_1 g(\mathbf{x}) & D_1 D_2 g(\mathbf{x}) & \dots & D_1 D_n g(\mathbf{x}) \\ D_2 D_1 g(\mathbf{x}) & D_2 D_2 g(\mathbf{x}) & \dots & D_2 D_n g(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ D_n D_1 g(\mathbf{x}) & D_n D_2 g(\mathbf{x}) & \dots & D_n D_n g(\mathbf{x}) \end{bmatrix}$$

Huomautuksia. Hessen matriisin käsitteen esitti 1800-luvulla preussilainen matemaatikko Ludwig Otto Hesse. Sitä varten on virheellistä kutsua sitä Hessen matriisiksi (jota nimitystä myös käytetään). Monissa lähteissä gradientti ymmärretään vaakavektoriksi, mutta tässä monisteessa se on pystyvektori. Hessen matriisille käytetään usein myös merkintöjä $\nabla^2 g(\mathbf{x})$ tai $g''(\mathbf{x})$.

Lause 8.5. Olkoot $sv:n$ \mathbf{X} momenttiemäfunktio M ja kumulanttiemäfunktio K määritellyjä jossakin origon ympäristössä. Tällöin

$$E\mathbf{X} = \nabla M(\mathbf{0}) = \nabla K(\mathbf{0}), \quad \text{Cov}(\mathbf{X}) = \mathbf{H}_K(\mathbf{0}).$$

Todistus. Tulokset johdetaan yhdistetyn funktion derivointikaavalla, aivan kuten yksiulotteisessa tapauksessa. \square

Esimerkki 8.7. Multinomijakaumalle $\mathbf{X} \sim \text{Mult}(k, \mathbf{p})$ momenttiemäfunktion saa laskettua multinomikaavan avulla,

$$\begin{aligned} M(\mathbf{t}) &= E \exp(\mathbf{t}^T \mathbf{X}) = \sum \binom{k}{x_1, \dots, x_n} (p_1 e^{t_1})^{x_1} \dots (p_n e^{t_n})^{x_n} \\ &= (p_1 e^{t_1} + \dots + p_n e^{t_n})^k \end{aligned}$$

Tätä tai sen logaritmia derivoimalla nähdään helposti, että

$$E\mathbf{X} = k\mathbf{p}, \quad \text{Cov}(\mathbf{X}) = k(\text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^T).$$

Tässä $\text{diag}(\mathbf{p})$ on lävistäjämatriisi, jonka lävistäjällä on luvut (p_1, \dots, p_n) . \triangle

Lause 8.6. Olkoon $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$ sellainen sv, että sen momenttiemäfunktio $M_{\mathbf{X}}(\mathbf{t})$ on olemassa jossakin origon ympäristössä. Olkoon $\mathbf{t} = (\mathbf{u}, \mathbf{v})$ ositettu samandimensioisiin osiin kuin \mathbf{Y} and \mathbf{Z} . Tällöin

(a) Sv:ien \mathbf{Y} ja \mathbf{Z} momenttiemäfunktiot ovat

$$M_{\mathbf{Y}}(\mathbf{u}) = M_{\mathbf{X}}(\mathbf{u}, \mathbf{0}), \quad M_{\mathbf{Z}}(\mathbf{v}) = M_{\mathbf{X}}(\mathbf{0}, \mathbf{v}).$$

(b) $\mathbf{Y} \perp \mathbf{Z}$ jos ja vain jos

$$M_{\mathbf{X}}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{Y}}(\mathbf{u}) M_{\mathbf{Z}}(\mathbf{v}).$$

kaikilla riittävän pienillä \mathbf{u} ja \mathbf{v} .

Todistus. (a):

$$M_{\mathbf{Y}}(\mathbf{u}) = E \exp(\mathbf{u}^T \mathbf{Y}) = E \exp(\mathbf{u}^T \mathbf{Y} + \mathbf{0}^T \mathbf{Z}) = M_{\mathbf{X}}(\mathbf{u}, \mathbf{0}),$$

ja $M_{\mathbf{Z}}(\mathbf{v})$:n kaava nähdään oikeaksi samalla tavalla.

(b): Olkoon ensin $\mathbf{Y} \perp \mathbf{Z}$. Nyt

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{u}, \mathbf{v}) &= E \exp(\mathbf{u}^T \mathbf{Y} + \mathbf{v}^T \mathbf{Z}) = E (\exp(\mathbf{u}^T \mathbf{Y}) \exp(\mathbf{v}^T \mathbf{Z})) \\ &= M_{\mathbf{Y}}(\mathbf{u}) M_{\mathbf{Z}}(\mathbf{v}), \end{aligned}$$

missä käytettiin tietoa $\exp(\mathbf{u}^T \mathbf{Y}) \perp \exp(\mathbf{v}^T \mathbf{Z})$. Käänteistä implikaatiota varten oletetaan, että $M_{\mathbf{X}}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{Y}}(\mathbf{u}) M_{\mathbf{Z}}(\mathbf{v})$ kaikilla riittävän pienillä argumenteilla. Olkoot \mathbf{Y}' ja \mathbf{Z}' sellaisia sv:eita, että

$$\mathbf{Y}' \stackrel{d}{=} \mathbf{Y}, \quad \mathbf{Z}' \stackrel{d}{=} \mathbf{Z}, \quad \mathbf{Y}' \perp \mathbf{Z}'.$$

Tällöin

$$M_{\mathbf{Y}', \mathbf{Z}'}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{Y}'}(\mathbf{u}) M_{\mathbf{Z}'}(\mathbf{v}) = M_{\mathbf{Y}}(\mathbf{u}) M_{\mathbf{Z}}(\mathbf{v}) = M_{\mathbf{Y}, \mathbf{Z}}(\mathbf{u}, \mathbf{v}),$$

joten sv:n $(\mathbf{Y}', \mathbf{Z}')$ ja sv:n (\mathbf{Y}, \mathbf{Z}) momenttiemäfunktiot yhtyvät jossakin origon ympäristössä, joten niillä on sama jakauma. \square

Luku 9

Moniulotteinen normaalijakauma

Tässä luvussa tarkastellaan normaalijakauman moniulotteista yleistystä eli moniulotteista (eli monimuuttujaista) normaalijakaumaa (engl. *multivariate normal distribution*). Sitä kutsutaan myös multinormaalijakaumaksi. Paitsi että tässä luvussa tutustutaan tähän sovelluksissa usein esiintyvään moniulotteiseen jakaumaan, tarkoituksena on lisäksi demonstroida, miten moniulotteisia jakauksia käsitellään vektori- ja matriisimerkinnöillä.

9.1 Standardinormaalijakauma $N_n(\mathbf{0}, \mathbf{I})$

Määritelmä 9.1. Sv:lla $\mathbf{U} = (U_1, \dots, U_n)$ on n -ulotteinen standardinormaalijakauma eli normaalijakauma $N_n(\mathbf{0}, \mathbf{I})$ täsmälleen silloin, kun sen komponentit ovat riippumattomia $N(0, 1)$ -jakaumaa noudattavia satunnaismuuttujia.

Satunnaisvektorin $\mathbf{U} \sim N_n(\mathbf{0}, \mathbf{I})$ tf on

$$\begin{aligned} f_{\mathbf{U}}(\mathbf{u}) &= \prod_{i=1}^n f_{U_i}(u_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_i^2} \\ &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}(u_1^2 + \dots + u_n^2)\right) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\mathbf{u}^T \mathbf{u}\right). \end{aligned} \quad (9.1)$$

Sv:n \mathbf{U} odotusarvovektori on n -dimensionen nollavektori, ja sen kovarianssimatriisi on dimensiota $n \times n$ oleva yksikkömatriisi

$$E\mathbf{U} = \mathbf{0}_n, \quad \text{Cov } \mathbf{U} = \mathbf{I}_n. \quad (9.2)$$

Sv:n \mathbf{U} momenttiemäfunktio on

$$M_{\mathbf{U}}(\mathbf{t}) = E \exp(t^T \mathbf{U}) = E \prod_{i=1}^n \exp(t_i U_i) = \prod_{i=1}^n e^{\frac{1}{2}t_i^2} = \exp\left(\frac{1}{2}\mathbf{t}^T \mathbf{t}\right). \quad (9.3)$$

Jos sv \mathbf{X} määritellään kaavalla

$$\mathbf{X} = \mathbf{A}\mathbf{U} + \boldsymbol{\mu}, \quad (9.4)$$

jossa \mathbf{A} on $m \times n$ -vakiomatriisi, $\boldsymbol{\mu} \in \mathbb{R}^m$ on vakiovektori, ja $\mathbf{U} \sim N_n(\mathbf{0}, \mathbf{I}_n)$, niin

$$E\mathbf{X} = \boldsymbol{\mu}, \quad \text{Cov } \mathbf{X} = \mathbf{A} \mathbf{I}_n \mathbf{A}^T = \mathbf{A} \mathbf{A}^T. \quad (9.5)$$

Merkitään \mathbf{X} :n kovarianssimatriisia $\boldsymbol{\Sigma} = \text{Cov } \mathbf{X} = \mathbf{A} \mathbf{A}^T$. Sv:n \mathbf{X} momenttiemäfunktio saadaan helpolla laskulla, sillä

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) &= E \exp(\mathbf{t}^T \mathbf{X}) = E \exp(\mathbf{t}^T (\mathbf{A} \mathbf{U} + \boldsymbol{\mu})) = \exp(\mathbf{t}^T \boldsymbol{\mu}) M_{\mathbf{U}}(\mathbf{A}^T \mathbf{t}) \\ &= \exp(\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^T \mathbf{A} \mathbf{A}^T \mathbf{t}) = \exp(\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}) \end{aligned} \quad (9.6)$$

Määrittelemme kohta multinormaalijakauman esityksen (9.4) avulla. Sitä ennen tarkastelimme kysymystä, kuinka mielivaltainen kovarianssimatriisi $\boldsymbol{\Sigma}$ voidaan esittää tulona $\mathbf{A} \mathbf{A}^T$.

Eräs (mutta ei suinkaan ainoa) mahdollisuus hajotelman $\boldsymbol{\Sigma} = \mathbf{A} \mathbf{A}^T$ löytämiseksi on käyttää *Choleskyn hajotelmaa* (engl. *Cholesky decomposition*). Choleskyn hajotelmassa symmetrinen ja positiivisesti semidefiniitti matriisi $\boldsymbol{\Sigma}$ esitetään tulona

$$\boldsymbol{\Sigma} = \mathbf{L} \mathbf{L}^T, \quad (9.7)$$

jossa \mathbf{L} on alakolmiomatriisi. (Alakolmiomatriisi on neliömatriisi, jonka yläkolmion alkiot (i, j) , $j > i$ ovat kaikki nolliä.) Choleskyn hajotelma on saatavilla matriisilaskennan ohjelmakirjastoissa (mutta tyypillisesti ohjelmat palauttavat jälkimmäisen tekijän eli yläkolmiomatriisin \mathbf{L}^T).

Huomautus. Jos $\boldsymbol{\Sigma}$ on positiivisesti definiitti matriisi, niin se on kääntyvä matriisi. Jos se esitetään muodossa

$$\boldsymbol{\Sigma} = \mathbf{A} \mathbf{A}^T,$$

jossa \mathbf{A} on neliömatriisi, niin matriisi \mathbf{A} on välttämättä kääntyvä matriisi.

9.2 Yleinen multinormaalijakauma

Määritelmä 9.2. Sv:lla \mathbf{X} on multinormaalijakauma, jos se voidaan esittää muodossa

$$\mathbf{X} = \mathbf{A} \mathbf{U} + \boldsymbol{\mu} \quad (9.8)$$

jossa \mathbf{A} on $m \times n$ -vakiomatriisi, $\boldsymbol{\mu} \in \mathbb{R}^m$ on vakiovektori, ja $\mathbf{U} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ jollakin n .

Yksiulotteinen normaalijakauma $N(\mu, \sigma^2)$ on multinormaalijakauman erikoistapaus, sillä

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad (X = \mu + \sigma U, \quad U \sim N(0, 1)).$$

Määritelmästä sekä kaavoista (9.5) ja (9.6) seuraa, että

$$\begin{aligned} E\mathbf{X} &= \boldsymbol{\mu}, \quad \boldsymbol{\Sigma} = \text{Cov } \mathbf{X} = \mathbf{A} \mathbf{A}^T, \\ M_{\mathbf{X}}(\mathbf{t}) &= \exp(\mathbf{t}^T \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}) \end{aligned}$$

Lause 9.1. Määritelmästä 9.2 seuraa, että $E\mathbf{X} = \boldsymbol{\mu}$ ja $\text{Cov } \mathbf{X} = \mathbf{A}\mathbf{A}^T$. Sv:n \mathbf{X} jakauma riippuu vain sen odotusarvovektorista ja kovarianssimatriisista, ei siis esitysdimensiosta n eikä esitysmatriisin \mathbf{A} muista ominaisuuksista.

Kääntäen, jos $\boldsymbol{\mu} \in \mathbb{R}^m$ on vakiovektori ja $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$ on positiivisesti semidefiniitti matriisi, niin on olemassa sv \mathbf{X} , jolla on multinormaalijakauma odotusarvovektorilla $\boldsymbol{\mu}$ ja kovarianssimatriisilla $\boldsymbol{\Sigma}$.

Todistus. Odotusarvovektori ja kovarianssimatriisi on jo johdettu aikaisemmin. Koska \mathbf{X} :n momenttiemäfunktio riippuu vain sen odotusarvovektorista ja kovarianssimatriisista, niin myös sen jakauma riippuu vain näistä parametreista.

Käänteisen tuloksen todistamiseksi positiivisesti semidefiniitti matriisi $\boldsymbol{\Sigma}$ jaetaan tekijöihin $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$, ja sen jälkeen jakaumaa noudattava sv konstruoidaan kaavalla (9.8). \square

Tästä lähtien multinormaalijakaumaa merkitään tunnuksella $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa $\boldsymbol{\mu}$ on jakauman odotusarvo ja $\boldsymbol{\Sigma}$ sen kovarianssimatriisi. Alaindeksi m voidaan jättää pois, jos satunnaisvektorin \mathbf{X} dimensiosta ei tarvitse pitää kirjaa.

Määritelmä 9.2 voidaan tulkita simulointireseptiksi. Jos tahdotaan simuloida jakaumaa $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa $\boldsymbol{\Sigma}$ on annettu symmetrinen ja positiivisesti semidefiniitti matriisi, niin ensin etsitään kovarianssimatriisille $\boldsymbol{\Sigma}$ hajotelma muodossa

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T.$$

Tekijöihinjako voidaan tehdä esim. Choleskyn hajotelmalla. Tämän jälkeen simuloidaan (riippumattomasti) m arvoa (u_1, \dots, u_m) standardinormaalijakaumasta $N(0, 1)$, ja lopuksi lasketaan

$$\mathbf{x} = \mathbf{A}\mathbf{u} + \boldsymbol{\mu}, \quad \text{jossa } \mathbf{u} = (u_1, \dots, u_m).$$

9.3 Affinin muunnoksen jakauma

Lause 9.2 (Multinormaalisuus säilyy affiinissa muunnoksessa). Olkoon $\mathbf{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, ja olkoon $\mathbf{B} \in \mathbb{R}^{p \times m}$ vakiomatriisi ja $\mathbf{b} \in \mathbb{R}^p$ vakiovektori. Tällöin

$$\mathbf{B}\mathbf{X} + \mathbf{b} \sim N_p(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T).$$

Todistus. Käytetään esitystä (9.8):

$$\mathbf{B}\mathbf{X} + \mathbf{b} = \mathbf{B}(\mathbf{A}\mathbf{U} + \boldsymbol{\mu}) + \mathbf{b} = (\mathbf{B}\mathbf{A})\mathbf{U} + (\mathbf{B}\boldsymbol{\mu} + \mathbf{b}).$$

Määritelmän 9.2 mukaan sv:lla $\mathbf{B}\mathbf{X} + \mathbf{b}$ on multinormaalijakauma. Lisäksi

$$\begin{aligned} E(\mathbf{B}\mathbf{X} + \mathbf{b}) &= \mathbf{B}E(\mathbf{X}) + \mathbf{b} = \mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \\ \text{Cov}(\mathbf{B}\mathbf{X} + \mathbf{b}) &= \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T. \quad \square \end{aligned}$$

Tämä lause selvittää myös kaikki reunajakaumat. Jos määritellään i :s yksikkövektori \mathbf{e}_i siten, että sen i :s koordinaatti on yksi ja muut koordinaatit ovat nollia, niin

$$X_i = \mathbf{e}_i^T \mathbf{X}.$$

Jos sv \mathbf{Y} koostuu sv:n \mathbf{X} komponenteista $(X_{i_1}, \dots, X_{i_k})$, niin \mathbf{Y} saadaan kertomalla sv:a \mathbf{X} vasemmalta vakiomatriisilla, jonka j :s vaakarivi on \mathbf{e}_{i_j} ,

$$\mathbf{Y} = \begin{bmatrix} X_{i_1} \\ \vdots \\ X_{i_k} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{i_1}^T \\ \vdots \\ \mathbf{e}_{i_k}^T \end{bmatrix} \mathbf{X}.$$

Tästä seuraa, että \mathbf{Y} :llä on multinormaalijakauma, joten multinormaalijakauman kaikki reunajakaumat ovat multinormaalisia.

Sv:n \mathbf{Y} jakauman parametrit saadaan joko lauseen 9.2 avulla, tai vaihtoehtoisesti poimimalla asiaankuuluva osa \mathbf{X} :n jakauman parametreista, sillä

$$E\mathbf{Y} = \begin{bmatrix} EX_{i_1} \\ \vdots \\ EX_{i_k} \end{bmatrix} = \begin{bmatrix} \mu_{i_1} \\ \vdots \\ \mu_{i_k} \end{bmatrix}, \quad (\text{Cov } \mathbf{Y})(p, q) = \text{cov}(X_{i_p}, X_{i_q}) = \Sigma(i_p, i_q).$$

9.4 Tiheysfunktio

Lause 9.3 (Multinormaalijakauman tiheysfunktio). *Jos $\mathbf{X} \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa $\boldsymbol{\Sigma}$ on positiivisesti definiitti matriisi, niin*

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-m/2} (\det(\boldsymbol{\Sigma}))^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right). \quad (9.9)$$

Todistus. Olkoon \mathbf{A} säännöllinen $m \times m$ -matriisi siten, että $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$. Tällainen \mathbf{A} on mahdollista löytää, koska $\boldsymbol{\Sigma}$ on positiivisesti definiitti. Käytetään esitystä

$$\mathbf{X} = \mathbf{A}\mathbf{U} + \boldsymbol{\mu}, \quad \text{jossa } \mathbf{U} \sim N(\mathbf{0}, \mathbf{I}_m),$$

sekä muuttujanvaihtoa

$$\mathbf{x} = \mathbf{A}\mathbf{u} + \boldsymbol{\mu} \iff \mathbf{u} = \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}).$$

(Jotta voitaisiin soveltaa tiheysfunktion muuntokaavaa, tarvitaan säännöllinen kerroinmatriisi \mathbf{A} , ja multinormaalijakauman esityksessä vektoreilla \mathbf{X} ja \mathbf{U} täytyy olla sama dimensio.) Tiheysfunktion muuntokaavalla saadaan

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{U}}(\mathbf{u}) \left| \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right| = f_{\mathbf{U}}(\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})) |\det(\mathbf{A}^{-1})| \\ &= (2\pi)^{-m/2} |\det(\mathbf{A}^{-1})| \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{A}^{-1})^T \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \end{aligned}$$

Väite seuraa tästä kaavasta sekä seuraavista laskuista,

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &= (\mathbf{A}\mathbf{A}^T)^{-1} = (\mathbf{A}^T)^{-1} \mathbf{A}^{-1} = (\mathbf{A}^{-1})^T \mathbf{A}^{-1}, \\ \det(\boldsymbol{\Sigma}) &= \det(\mathbf{A}\mathbf{A}^T) = \det(\mathbf{A}) \det(\mathbf{A}^T) = (\det(\mathbf{A}))^2 > 0, \\ |\det(\mathbf{A}^{-1})| &= \left| \frac{1}{\det(\mathbf{A})} \right| = \frac{1}{\sqrt{\det(\boldsymbol{\Sigma})}}. \quad \square \end{aligned}$$

Huomautus. Multinormaalijakaumalla $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ on tiheysfunktio täsmälleen silloin, kun $\boldsymbol{\Sigma}$ on positiivisesti definiitti. Jos $\boldsymbol{\Sigma}$ on pelkästään positiivisesti semidefiniitti, mutta ei positiivisesti definiitti, niin lauseen 8.2 mukaan \mathbf{X} saa arvoja tietyltä hypertasolta todennäköisyydellä yksi, joten tällöin \mathbf{X} :llä ei voi olla tiheysfunktiota. Myös tällainen *singulaarinen multinormaalijakauma* on tärkeä jakauma sovelluksissa: esim. lineaaristen mallien yhteydessä toisaalta sovitevektorin jakauma ja toisaalta residuaalivektorin eli jäännösvektorin jakauma ovat kumpikin singulaarisia multinormaalijakaumia.

9.5 Tiheysfunktion tasa-arvopinnat

Moniulotteisen normaalijakauman tiheysfunktion (9.9) tasa-arvopinnat ovat m -ulotteisia ellipsoideja. Tämä nähdään soveltamalla kovarianssimatriisiin ominaisarvohajotelmaa, jonka ensin kertaamme.

Olkoon $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ symmetrinen ja positiivisesti semidefiniitti matriisi. Symmetrisenä matriisina sillä on reaaliset ominaisarvot λ_i ja ominaisvektorit \mathbf{v}_i . Ominaisvektorit voidaan valita ortonormaaleiksi, jolloin

$$\boldsymbol{\Sigma} \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, \dots, n, \quad (9.10)$$

$$\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 1, & \text{kun } i = j, \\ 0, & \text{kun } i \neq j. \end{cases} \quad (9.11)$$

Ominaisarvojen ja ominaisvektoreiden haku sisältyy kaikkiin matriisilaskennan ohjelmakirjastoihin.

Koska $\boldsymbol{\Sigma}$ on positiivisesti semidefiniitti, on

$$0 \leq \mathbf{v}_i^T \boldsymbol{\Sigma} \mathbf{v}_i = \mathbf{v}_i^T (\lambda_i \mathbf{v}_i) = \lambda_i.$$

Siis kaikki ominaisarvot ovat ei-negatiivisia. Jos $\boldsymbol{\Sigma}$ on peräti positiivisesti definiitti, niin kaikki ominaisarvot ovat aidosti positiivisia.

Kootaan nyt ominaisvektoreista matriisi \mathbf{V} laittamalla ne matriisiin pystyrieksi, sekä muodostetaan ominaisarvoista lävistäjämatriisi $\boldsymbol{\Lambda}$,

$$\mathbf{V} = [\mathbf{v}_1 \ \dots \ \mathbf{v}_n], \quad \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n).$$

Tällöin ominaisarvon ja ominaisvektorin määritelmästä seuraa kaava

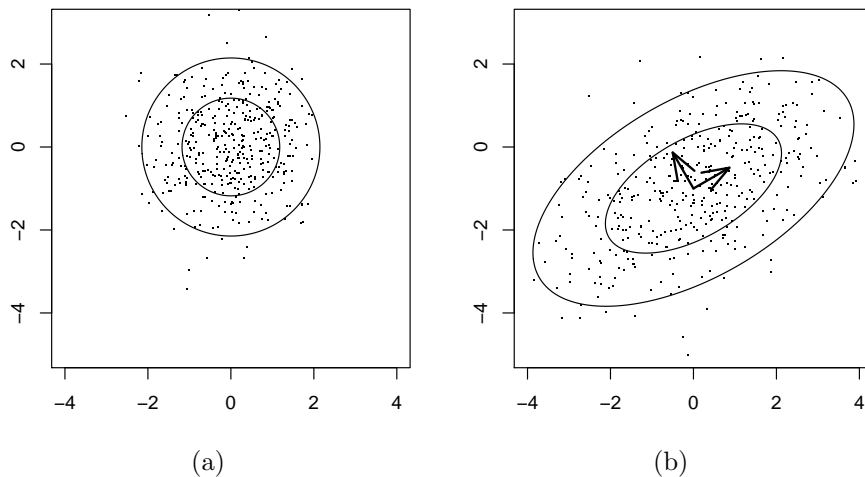
$$\boldsymbol{\Sigma} \mathbf{V} = \mathbf{V} \boldsymbol{\Lambda}$$

Siitä, että ominaisvektorit ovat ortonormaaleja seuraa, että $\mathbf{V}^T \mathbf{V} = \mathbf{I}_n$. Koska \mathbf{V} on lisäksi neliömatriisi, niin $\mathbf{V}^{-1} = \mathbf{V}^T$. Tällaista matriisia, jonka käänteismatriisi on sen transpoosi, kutsutaan *ortogonaaliseksi matriisiksi*. Nyt ollaan saatu johdettua matriisin $\boldsymbol{\Sigma}$ ominaisarvohajotelma,

$$\boldsymbol{\Sigma} = \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^T, \quad \text{jossa } \mathbf{V}^{-1} = \mathbf{V}. \quad (9.12)$$

Olkoon nyt $\boldsymbol{\Sigma}$ multinormaalijakauman kovarianssimatriisi, ja oletetaan että se on positiivisesti definiitti. Tällöin sen ominaisarvohajotelmasta saadaan hajotelma sen käänteismatriisille, nimittäin

$$\boldsymbol{\Sigma} = \mathbf{V} \boldsymbol{\Lambda} \mathbf{V}^T \Rightarrow \boldsymbol{\Sigma}^{-1} = \mathbf{V} \boldsymbol{\Lambda}^{-1} \mathbf{V}^T.$$

Kuva 9.1 Kaksiulotteinen normaalijakauma: (a) $N(\mathbf{0}, \mathbf{I})$, (b) $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Normaalijakauman $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ tiheysfunktio $f_{\mathbf{X}}(\mathbf{x})$ (9.9) saa vakioarvon niillä argumenteilla \mathbf{x} , joilla eksponenttifunktion argumentissa oleva lauseke saa vakioarvon, mutta

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V} \boldsymbol{\Lambda}^{-1} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) = \frac{y_1^2}{\lambda_1} + \dots + \frac{y_m^2}{\lambda_m},$$

jossa vektori \mathbf{y} määritellään kaavalla

$$\mathbf{y} = \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \iff \mathbf{x} = \boldsymbol{\mu} + \mathbf{V} \mathbf{y}.$$

Vektori \mathbf{y} saadaan vektorista \mathbf{x} koordinaatistonmuutoksella, jossa uudeksi origoksi valitaan $\boldsymbol{\mu}$ ja uusiksi koordinaattiakseleiksi vektorit $\mathbf{V} \mathbf{e}_i$ (jossa \mathbf{e}_i on avaruuden \mathbb{R}^n standardikannan i :s yksikkövektori). Nämä vektorit ovat ortonormaaleja, sillä

$$(\mathbf{V} \mathbf{e}_i)^T (\mathbf{V} \mathbf{e}_j) = \mathbf{e}_i^T \mathbf{V}^T \mathbf{V} \mathbf{e}_j = \mathbf{e}_i^T \mathbf{e}_j.$$

Koska kaikki $\lambda_i > 0$, tasa-arvopinnat toteuttavat yhtälön

$$\frac{y_1^2}{(\sqrt{\lambda_1})^2} + \dots + \frac{y_m^2}{(\sqrt{\lambda_m})^2} = c, \quad c > 0.$$

Tämä on m -ulotteisen ellipsoidin yhtälö uusille koordinaateille \mathbf{y} . Ellipsoidin puoliakselien pituudet ovat $\sqrt{c\lambda_1}, \dots, \sqrt{c\lambda_m}$.

Esimerkki 9.1. (Kaksiulotteisen normaalijakauman havainnollistus.) Olkoon

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

Kuvassa 9.1 a on piirretty otosvektoreita \mathbf{u}_i kaksiulotteisesta normaalijakaumasta $N(\mathbf{0}, \mathbf{I})$ sekä piirretty sen tiheysfunktion tasa-arvokäyriä. Kuvassa 9.1 b tämä otos on muunnettu kaavalla

$$\mathbf{x}_i = \mathbf{A} \mathbf{u}_i + \boldsymbol{\mu}, \quad \mathbf{A} = \mathbf{V} \boldsymbol{\Lambda}^{1/2}$$

niin, että on saatu otos kaksiulotteisesta normaalijakaumasta $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa

$$\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T,$$

sekä piirretty jakauman tiheysfunktion tasa-arvokäyriä, kun $\theta = 30^\circ$. Kuvaan on piirretty pisteeseen $\boldsymbol{\mu}$ myös vektorit $\mathbf{V}\mathbf{e}_1$ ja $\mathbf{V}\mathbf{e}_2$. \triangle

9.6 Korreloimattomuus ja riippumattomuus

Riippumattomat satunnaisvektorit $\mathbf{X} \perp \mathbf{Y}$ eivät korreloi, sillä

$$\begin{aligned} \text{cov}(\mathbf{X}, \mathbf{Y}) &= E[(\mathbf{X} - E\mathbf{X})(\mathbf{Y} - E\mathbf{Y})^T] \\ &= E[\mathbf{X} - E\mathbf{X}] E[(\mathbf{Y} - E\mathbf{Y})^T] = \mathbf{0}. \end{aligned}$$

Yleisesti ottaen korreloimattomuudesta ei seuraa riippumattomuus. Jos yhdistetyn satunnaisvektorin (\mathbf{X}, \mathbf{Y}) yhteisjakauma on multinormaalinen, niin tässä tapauksessa korreloimattomuudesta seuraa riippumattomuus, mikä asia todistetaan seuraavassa lauseessa.

Kirjataan myöhempää käyttöä varten, minkälaiset ovat osavektorien $\mathbf{X} = (X_1, \dots, X_k)$ ja $\mathbf{Y} = (Y_1, \dots, Y_m)$ multinormaalijakaumien parametrit, jos yhdistetty vektori (\mathbf{X}, \mathbf{Y}) noudattaa multinormaalijakaumaa $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Odotusarvovektori $\boldsymbol{\mu}$ koostuu osista

$$\boldsymbol{\mu} = E\mathbf{Z} = \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix} = \begin{bmatrix} E\mathbf{X} \\ E\mathbf{Y} \end{bmatrix}, \quad (9.13)$$

ja kovarianssimatriisi $\boldsymbol{\Sigma} = \text{Cov } \mathbf{Z}$ voidaan myös osittaa, sillä

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{bmatrix} \text{cov}(X_1, X_1) & \dots & \text{cov}(X_1, X_k) & \text{cov}(X_1, Y_1) & \dots & \text{cov}(X_1, Y_m) \\ \vdots & & \vdots & \vdots & & \vdots \\ \text{cov}(X_k, X_1) & \dots & \text{cov}(X_k, X_k) & \text{cov}(X_k, Y_1) & \dots & \text{cov}(X_k, Y_m) \\ \text{cov}(Y_1, X_1) & \dots & \text{cov}(Y_1, X_k) & \text{cov}(Y_1, Y_1) & \dots & \text{cov}(Y_1, Y_m) \\ \vdots & & \vdots & \vdots & & \vdots \\ \text{cov}(Y_m, X_1) & \dots & \text{cov}(Y_m, X_k) & \text{cov}(Y_m, Y_1) & \dots & \text{cov}(Y_m, Y_m) \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{bmatrix} = \begin{bmatrix} \text{cov}(\mathbf{X}, \mathbf{X}) & \text{cov}(\mathbf{X}, \mathbf{Y}) \\ \text{cov}(\mathbf{Y}, \mathbf{X}) & \text{cov}(\mathbf{Y}, \mathbf{Y}) \end{bmatrix} \end{aligned} \quad (9.14)$$

Tällöin reunajakaumat ovat jakson 9.3 kaavojen mukaan

$$\mathbf{X} \sim N(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_{XX}), \quad \mathbf{Y} \sim N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_{YY}). \quad (9.15)$$

Lause 9.4. Jos vektorilla (\mathbf{X}, \mathbf{Y}) on multinormaalijakauma, niin

$$\mathbf{X} \perp \mathbf{Y} \iff \text{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}.$$

Todistus. Koska multinormaalisen yhteisjakauman momenttiemäfunktio on määritelty kaikilla argumenteilla, niin satunnaisvektorit \mathbf{X} ja \mathbf{Y} ovat riippumattomat silloin ja vain silloin, kun niiden yhteismomenttiemäfunktio faktoroituu komponenttien \mathbf{X} ja \mathbf{Y} momenttiemäfunktioiden tuloksi, eli silloin kun

$$M_{\mathbf{X}, \mathbf{Y}}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{X}}(\mathbf{u}) M_{\mathbf{Y}}(\mathbf{v}) \quad \text{kaikilla } \mathbf{u}, \mathbf{v}.$$

Kun käytetään osituksia (9.13) ja (9.14), niin

$$M_{\mathbf{X}, \mathbf{Y}}(\mathbf{u}, \mathbf{v}) = \exp \left(\mathbf{u}^T \boldsymbol{\mu}_X + \mathbf{v}^T \boldsymbol{\mu}_Y + \frac{1}{2} (\mathbf{u}^T \boldsymbol{\Sigma}_{XX} \mathbf{u} + \mathbf{u}^T \boldsymbol{\Sigma}_{XY} \mathbf{v} + \mathbf{v}^T \boldsymbol{\Sigma}_{YX} \mathbf{u} + \mathbf{v}^T \boldsymbol{\Sigma}_{YY} \mathbf{v}) \right).$$

Toisaalta reunajakaumien momenttiemäfunktioiden tulo on

$$M_{\mathbf{X}}(\mathbf{u}) M_{\mathbf{Y}}(\mathbf{v}) = \exp \left(\mathbf{u}^T \boldsymbol{\mu}_X + \frac{1}{2} \mathbf{u}^T \boldsymbol{\Sigma}_{XX} \mathbf{u} + \mathbf{v}^T \boldsymbol{\mu}_Y + \frac{1}{2} \mathbf{v}^T \boldsymbol{\Sigma}_{YY} \mathbf{v} \right).$$

Edellä

$$\mathbf{v}^T \boldsymbol{\Sigma}_{YX} \mathbf{u} = (\mathbf{v}^T \boldsymbol{\Sigma}_{YX} \mathbf{u})^T = \mathbf{u}^T \boldsymbol{\Sigma}_{XY} \mathbf{v},$$

joten funktiot $M_{\mathbf{X}, \mathbf{Y}}$ ja $M_{\mathbf{X}} M_{\mathbf{Y}}$ ovat samat silloin ja vain silloin, kun

$$\mathbf{u}^T \boldsymbol{\Sigma}_{XY} \mathbf{v} = 0 \quad \text{kaikilla } \mathbf{u}, \mathbf{v},$$

eli silloin, kun

$$\boldsymbol{\Sigma}_{XY} = \text{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}. \quad \square$$

Huomautus. Jos oletetaan multinormaaliset reunajakaumat

$$\mathbf{X} \sim N(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_{XX}), \quad \mathbf{Y} \sim N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_{YY}),$$

niin yhdistetyn vektorin (\mathbf{X}, \mathbf{Y}) jakauma ei välttämättä ole multinormaalinen. Ei edes vaikka oletettaisiin esim. lisäehto $\text{cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$. Jos sen sijaan oletetaan, että $\mathbf{X} \perp \mathbf{Y}$, niin sen jälkeen momenttiemäfunktiosta

$$M_{\mathbf{X}, \mathbf{Y}}(\mathbf{u}, \mathbf{v}) = M_{\mathbf{X}}(\mathbf{u}) M_{\mathbf{Y}}(\mathbf{v})$$

nähdään (ks. todistuksen kaavoja), että $(\mathbf{X}, \mathbf{Y}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{YY} \end{bmatrix}.$$

9.7 Ehdolliset jakaumat

Lause 9.5. *Olkoon $(\mathbf{X}, \mathbf{Y}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, ja käytetään odotusarvovektorille $\boldsymbol{\mu}$ ositusta (9.13), ja kovarianssimatriisille $\boldsymbol{\Sigma}$ ositusta (9.14). Jos osamatriisi $\boldsymbol{\Sigma}_{XX}$ on säännöllinen, niin sv:n \mathbf{Y} jakauma ehdolla $\mathbf{X} = \mathbf{x}$ on multinormaalinen, ja ehdollisen jakauman parametrit ovat*

$$\mathbf{Y} \mid (\mathbf{X} = \mathbf{x}) \sim N(\boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX} \boldsymbol{\Sigma}_{XX}^{-1} (\mathbf{x} - \boldsymbol{\mu}_X), \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX} \boldsymbol{\Sigma}_{XX}^{-1} \boldsymbol{\Sigma}_{XY}).$$

Todistus. Vaikka kaavat ovat monimutkaisia, todistuksen idea on yksinkertainen. Ensin muodostetaan sv \mathbf{V} kaavalla

$$\mathbf{V} = \mathbf{Y} - \mathbf{B}\mathbf{X}, \quad (9.16)$$

jossa \mathbf{B} on vakiomatriisi, joka valitaan kohta. Yhdistetty vektori (\mathbf{V}, \mathbf{X}) saadaan multinormaalisesesta vektorista (\mathbf{X}, \mathbf{Y}) lineaarisella muunnoksella, joten sillä on multinormaalinen yhteisjakauma. Kun valitaan

$$\mathbf{B} = \boldsymbol{\Sigma}_{YX} \boldsymbol{\Sigma}_{XX}^{-1},$$

niin \mathbf{V} ja \mathbf{X} eivät korreloi, sillä

$$\text{cov}(\mathbf{V}, \mathbf{X}) = \text{cov}(\mathbf{Y} - \mathbf{B}\mathbf{X}, \mathbf{X}) = \boldsymbol{\Sigma}_{YX} - \mathbf{B}\boldsymbol{\Sigma}_{XX} = \mathbf{0}.$$

Koska vektorilla (\mathbf{V}, \mathbf{X}) on multinormaalijakauma, niin korreloimattomuudesta seuraa riippumattomuus, joten $\mathbf{V} \perp \mathbf{X}$. Esityksen (9.16) mukaan

$$\mathbf{Y} = \mathbf{V} + \mathbf{B}\mathbf{X},$$

jossa \mathbf{V} ja \mathbf{X} ovat riippumattomia. Tällöin \mathbf{Y} :n jakauma ehdolla $\mathbf{X} = \mathbf{x}$ sama kuin sv:n

$$\mathbf{V} + \mathbf{B}\mathbf{x}$$

jakauma, sillä ehto $\mathbf{X} = \mathbf{x}$ ei muuta sv:n \mathbf{V} jakaumaa, koska $\mathbf{V} \perp \mathbf{X}$. Tästä nähdään, että ehdollinen jakauma on multinormaalinen. Tämän jälkeen ehdollisen jakauman parametrit saadaan laskemalla sv:n $\mathbf{V} + \mathbf{B}\mathbf{x}$ jakauman parametrit. Ehdollisen jakauman odotusarvovektori on

$$E(\mathbf{V} + \mathbf{B}\mathbf{x}) = \boldsymbol{\mu}_Y + \mathbf{B}(\mathbf{x} - \boldsymbol{\mu}_X) = \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}(\mathbf{x} - \boldsymbol{\mu}_X),$$

ja kovarianssimatriisi on

$$\begin{aligned} \text{Cov}(\mathbf{V} + \mathbf{B}\mathbf{x}) &= \text{Cov}(\mathbf{V}) = \text{cov}(\mathbf{Y} - \mathbf{B}\mathbf{X}, \mathbf{Y} - \mathbf{B}\mathbf{X}) \\ &= \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\mathbf{B}^T - \mathbf{B}\boldsymbol{\Sigma}_{XY} + \mathbf{B}\boldsymbol{\Sigma}_{XX}\mathbf{B}^T \\ &= \boldsymbol{\Sigma}_{YY} - \mathbf{B}\boldsymbol{\Sigma}_{XX}^{-1}\mathbf{B}^T. \end{aligned} \quad \square$$

9.8 Kaksiulotteinen normaalijakauma

Tarkastellaan kaksiulotteisen normaalijakauman $(X, Y) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ominaisuuksia. Johdettavat kaavat ovat pitkiä, eikä niitä kannata yrittää painaa muistiinsa. Ne voi tarvittaessa johtaa itse tai tarkistaa kirjallisuudesta.

Olkoon $-1 < \rho = \text{corr}(X, Y) < 1$, ja merkitään jakauman parametrien komponentteja seuraavasti,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Kovarianssimatriisin käänteismatriisi on

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\det(\boldsymbol{\Sigma})} \begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}, \quad \text{jossa} \quad \det(\boldsymbol{\Sigma}) = (1 - \rho^2)\sigma_X^2\sigma_Y^2.$$

Jakauman ytf on

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \right]\right).$$

Tässä eksponenttifunktion argumentin kaava saadaan kertomalla auki tulo

$$-\frac{1}{2} \begin{bmatrix} (x - \mu_X) & (y - \mu_Y) \end{bmatrix} \boldsymbol{\Sigma}^{-1} \begin{bmatrix} (x - \mu_X) \\ (y - \mu_Y) \end{bmatrix}.$$

Reunajakaumat ovat

$$X \sim N(\mu_X, \sigma_X^2), \quad Y \sim N(\mu_Y, \sigma_Y^2).$$

Ehdolliset jakaumat saadaan lauseen 9.5 perusteella:

$$Y | (X = x) \sim N\left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X), (1 - \rho^2)\sigma_Y^2\right),$$

$$X | (Y = y) \sim N\left(\mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y), (1 - \rho^2)\sigma_X^2\right).$$

Tässä esim. jakauman $Y | (X = x)$ parametrit saadaan laskuilla

$$\begin{aligned} \mu_Y + \Sigma_{YX} \Sigma_{XX}^{-1}(\mathbf{x} - \mu_X) &= \mu_Y + \frac{\rho \sigma_X \sigma_Y}{\sigma_X^2}(x - \mu_X) \\ \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} &= \sigma_Y^2 - \frac{(\rho \sigma_X \sigma_Y)^2}{\sigma_X^2} = (1 - \rho^2)\sigma_Y^2. \end{aligned}$$

On suoraviivaista mutta työlästä tarkistaa, että nämä ehdolliset jakaumat saadaan johdettua myös kaavoilla

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \quad f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

9.9 Normaalijakauman otoskeskiarvon ja otosvarianssin yhteisjakauma

Tarkastellaan sm:ia X_1, \dots, X_n . Niiden otoskeskiarvo \bar{X} ja otosvarianssi S^2 määritellään kaavoilla

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (9.17)$$

Jos sm:t X_i ovat riippumattomia, ja niillä on yhteinen odotusarvo μ ja yhteinen varianssi σ^2 , niin helpohkoilla laskuilla voidaan näyttää, että

$$E\bar{X} = \mu, \quad ES^2 = \sigma^2.$$

Ts. otoskeskiarvo ja otosvarianssi ovat vastaavien populaatioparametrien μ ja σ^2 harhattomia estimaattoreita.

Oletetaan tästä lähtien, että sm:t X_1, \dots, X_n ovat riippumattomia, ja niillä kaikilla on normaalijakauma $N(\mu, \sigma^2)$. Tällöin vektorilla (X_1, \dots, X_n) on multinormaalijakauma $N_n(\mathbf{m}, \Sigma)$ parametreilla

$$\mathbf{m} = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \quad \Sigma = \sigma^2 \mathbf{I}_n.$$

Seuraavaksi johdetaan vektorin (\bar{X}, S^2) yhteisjakauma.

Otoskeskiarvon reunajakauma on helppo johtaa, nimittäin $\bar{X} \sim N(\mu, \frac{1}{n}\sigma^2)$, sillä se on multinormaalijakautuneen vektorin lineaarimuunnos, ja osaamme laskea sm:n \bar{X} odotusarvon ja varianssin. Seuraavan lauseen todistuksessa osoitetaan, että otosvarianssilla S^2 on sopivan skaalauksen jälkeen khiin neliön jakauma vapausasteluvulla $n - 1$.

Tehdään ennen lauseen muotoilua se huomio, että mikäli satunnaisvektori \mathbf{Z} on k -ulotteinen standardinormaalijakauma, ts. $\mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I})$, niin sen pituuden neliöllä $\|\mathbf{Z}\|^2 = \mathbf{Z}^T \mathbf{Z}$ on khiin neliön jakauma vapausasteluvulla k , sillä

$$\|\mathbf{Z}\|^2 = \mathbf{Z}^T \mathbf{Z} = \sum_{i=1}^k Z_i^2,$$

jossa $Z_i \sim N(0, 1)$ riippumattomasti. Tällöin $E\|\mathbf{Z}\|^2 = k$.

Lause 9.6. Olkoot $X_1, \dots, X_n \sim N(\mu, \sigma^2) \perp$, ja määritellään \bar{X} ja S^2 kaavoilla (9.17). Tällöin

- a) $\bar{X} \sim N(\mu, \sigma^2/n)$,
- b) $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$, joten $ES^2 = \sigma^2$,
- c) $\bar{X} \perp S^2$.

Todistus. Kohta a todistettiin jo edellä, joten todistamme vain kohdat b ja c.

Jakaumaoletus sv:lle $\mathbf{X} = (X_1, \dots, X_n)$ voidaan lausua muodossa

$$\mathbf{X} \sim N_n(\mu \mathbf{1}, \sigma^2 \mathbf{I}_n),$$

jossa $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$ on n -komponenttinen vakiovektori, jonka kaikki komponentit ovat ykkösiä.

Määritellään n -komponenttinen vektori \mathbf{u} kaavalla

$$\mathbf{u} = \frac{1}{\sqrt{n}} \mathbf{1},$$

jolloin \mathbf{u} on ykkösvektorin $\mathbf{1}$ suuntainen yksikkövektori, ts. $\mathbf{u}^T \mathbf{u} = 1$. Huomaa, että otoskeskiarvo voidaan esittää kaavalla

$$\bar{X} = \frac{1}{n} \mathbf{1}^T \mathbf{X} = \frac{1}{\sqrt{n}} \mathbf{u}^T \mathbf{X}.$$

Määrittelemme residuaalivektorin (eli jäännösvektorin) \mathbf{R} kaavalla

$$\mathbf{R} = \begin{bmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{bmatrix} = \mathbf{X} - \bar{X} \mathbf{1} = \mathbf{X} - \mathbf{u} \mathbf{u}^T \mathbf{X}.$$

(Tässä vektori $\bar{X} \mathbf{1}$ on nimeltään sovitevektori.) Otosvarianssi S^2 voidaan esittää kaavalla

$$(n-1)S^2 = \mathbf{R}^T \mathbf{R} = \|\mathbf{R}\|^2, \quad (9.18)$$

jossa residuaalivektori voidaan esittää kaavalla

$$\mathbf{R} = \mathbf{X} - \mathbf{u} \mathbf{u}^T \mathbf{X} = (\mathbf{I}_n - \mathbf{u} \mathbf{u}^T) \mathbf{X}, \quad (9.19)$$

Yhdistetty vektori (\bar{X}, \mathbf{R}) saadaan lineaarisella muunnoksella vektorista \mathbf{X} , sillä

$$\begin{bmatrix} \bar{X} \\ \mathbf{R} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{n} \mathbf{u}^T \\ \mathbf{I} - \mathbf{u} \mathbf{u}^T \end{bmatrix} \mathbf{X},$$

minkä takia sillä on multinormaalinen yhteisjakauma. Lisäksi komponentit \bar{X} ja \mathbf{R} ovat korreloimattomia, sillä

$$\text{cov}(\bar{X}, \mathbf{R}) = \frac{1}{\sqrt{n}} \text{cov}(\mathbf{u}^T \mathbf{X}, (\mathbf{I} - \mathbf{u}\mathbf{u}^T)\mathbf{X}) = \frac{1}{\sqrt{n}} \mathbf{u}^T \sigma^2 \mathbf{I} (\mathbf{I} - \mathbf{u}\mathbf{u}^T) = \mathbf{0}.$$

Lauseen 9.4 perusteella \bar{X} ja \mathbf{R} ovat riippumattomia. Koska S^2 saadaan lasketua kaavan (9.18) mukaan residuaalivektorin \mathbf{R} funktiona, ovat myös \bar{X} ja S^2 riippumattomia, mikä todistaa kohdan c.

Otosvarianssin jakauman selvittämiseksi määritellään $n \times n$ neliömatriisi \mathbf{Q} siten, että sen ensimmäinen pystyriivi on \mathbf{u} . Muut matriisin \mathbf{Q} pystyriivit $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$ valitaan siten, että matriisin \mathbf{Q} pystyriivit muodostavat yhdessä \mathbb{R}^n :n ortonormeeratun kannan. Tällöin $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_n$, ja koska \mathbf{Q} on neliömatriisi, niin se on ortogonaalinen matriisi, eli $\mathbf{Q}^{-1} = \mathbf{Q}^T$. Tästä seuraa, että $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}_n$. Kun käytetään \mathbf{Q} :n ositusta $\mathbf{Q} = [\mathbf{u} \ \mathbf{V}]$ jossa $n \times (n-1)$ -matriisin \mathbf{V} pystyriivit ovat $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$, niin saadaan kaava

$$\mathbf{I}_n = \mathbf{Q}\mathbf{Q}^T = \mathbf{u}\mathbf{u}^T + \mathbf{V}\mathbf{V}^T.$$

Tästä nähdään kaavan (9.19) avulla, että

$$\mathbf{R} = (\mathbf{I}_n - \mathbf{u}\mathbf{u}^T)\mathbf{X} = \mathbf{V}\mathbf{V}^T \mathbf{X}.$$

Koska matriisin \mathbf{V} pystyriivit ovat ortonormaalaisia, on $\mathbf{V}^T \mathbf{V} = \mathbf{I}_{n-1}$, joten

$$\mathbf{R}^T \mathbf{R} = (\mathbf{V}\mathbf{V}^T \mathbf{X})^T (\mathbf{V}\mathbf{V}^T \mathbf{X}) = \mathbf{X}^T \mathbf{V}\mathbf{V}^T \mathbf{X} = \|\mathbf{V}^T \mathbf{X}\|^2. \quad (9.20)$$

Vektorilla $\mathbf{V}^T \mathbf{X}$ on $(n-1)$ -ulotteinen normaalijakauma, jonka odotusarvo on

$$E(\mathbf{V}^T \mathbf{X}) = \mathbf{V}^T (\mu \mathbf{1}) = \mathbf{0}$$

sillä perusteella, että \mathbf{u} on kohtisuorassa jokaista matriisin \mathbf{V} pystyriiviä vastaan, ja $\mathbf{1} = \sqrt{n}\mathbf{u}$. Kovarianssimatriisi on

$$\text{Cov}(\mathbf{V}^T \mathbf{X}) = \mathbf{V}^T (\sigma^2 \mathbf{I}_n) \mathbf{V} = \sigma^2 \mathbf{I}_{n-1}.$$

Tällöin

$$\frac{1}{\sigma} \mathbf{V}^T \mathbf{X} \sim N_{n-1}(\mathbf{0}, \mathbf{I}_{n-1}),$$

joten sen pituuden neliöllä on khiin neliön jakauma vapausasteluvulla $n-1$. Kaavojen (9.18) ja (9.20) avulla nähdään lopulta, että

$$\frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \mathbf{R}^T \mathbf{R} = \left(\frac{1}{\sigma} \mathbf{V}^T \mathbf{X}\right)^T \left(\frac{1}{\sigma} \mathbf{V}^T \mathbf{X}\right) = \left\| \frac{1}{\sigma} \mathbf{V}^T \mathbf{X} \right\|^2 \sim \chi_{n-1}^2.$$

Tästä seuraa, että $E[(n-1)S^2/\sigma^2] = n-1$, josta $ES^2 = \sigma^2$. \square

Palautetaan mieleen, että t -jakauma vapausasteluvulla ν määritellään siten, että se on sm:n

$$\frac{Z}{\sqrt{Y/\nu}}$$

jakauma, kun $Z \sim N(0, 1)$, $Y \sim \chi_\nu^2$ ja $Z \perp Y$. Sen neliöllä

$$\frac{Z^2/1}{Y/\nu}$$

on F -jakauma $F_{1,\nu}$ osoittajan vapausasteilla 1 ja nimittäjän vapausasteilla ν .

Satunnaismuuttujalla $\bar{X} - \mu$ on normaalijakauma odotusarvolla nolla ja varianssilla σ^2/n , joten satunnaismuuttujalla

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

on standardinormaalijakauma. Jos tässä tuntemattoman keskihajontaparametrin σ tilalle sijoitetaan sen otosestimää, saadaan ns. t -testisuure

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}.$$

Tämä t -testisuure voidaan esittää yhtäpitävästi kaavalla

$$T = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}}.$$

Edellisen lauseen mukaan tässä osoittaja ja nimittäjä ovat riippumattomia, osoittajan jakauma on $N(0,1)$ ja nimittäjässä on neliöjuuri satunnaismuuttujasta, jossa χ^2 -jakaumaa noudattava satunnaismuuttuja jaetaan vapausasteluvullaan $n-1$. Tästä seuraa, että T :llä on t -jakauma $n-1$ vapausasteella.

Edeltävän lauseen tarkastelu on vähällä vaivalla laajennettavissa myös tilanteeseen, jossa tarkastellaan lineaarista mallia

$$\mathbf{X} \sim N_n(\mathbf{m}, \sigma^2 \mathbf{I}),$$

jossa odotusarvovektorin \mathbf{m} tiedetään kuuluvan tunnettuun \mathbb{R}^n :n lineaariseen aliavaruuteen L , jonka dimensio $p \leq n$. Muodostetaan tälle aliavaruudelle ortonormaalit kantavektorit $\mathbf{u}_1, \dots, \mathbf{u}_p$, ja asetetaan nämä vektorit matriisiin \mathbf{U} pystyriveiksi, $\mathbf{U} = [\mathbf{u}_1 \ \dots \ \mathbf{u}_p]$. Tällöin matriisi

$$\mathbf{H} = \mathbf{U}\mathbf{U}^T$$

on projektiomatriisi aliavaruudelle L , eli \mathbf{H} on symmetrinen ja idempotentti matriisi (eli $\mathbf{H}\mathbf{H} = \mathbf{H}$) ja $\mathbf{H}\mathbf{v} = \mathbf{v}$ aina kun $\mathbf{v} \in L$. Tämän jakson lauseen todistusta matkimalla on helppo tarkistaa, että

$$\mathbf{H}\mathbf{X} \perp (\mathbf{I} - \mathbf{H})\mathbf{X},$$

ja että

$$\frac{1}{\sigma^2} \|\mathbf{X} - \mathbf{H}\mathbf{X}\|^2 \sim \chi_{n-p}^2.$$

Tämän takia satunnaismuuttuja

$$\frac{1}{n-p} \|\mathbf{X} - \mathbf{H}\mathbf{X}\|^2$$

on varianssin σ^2 harhaton estimattori, ja sen jakauma on skaalattu khiin neliö. Tässä tilanteessa sovitevektori $\mathbf{H}\mathbf{X}$ ja residuaalivektori $\mathbf{X} - \mathbf{H}\mathbf{X} = (\mathbf{I} - \mathbf{H})\mathbf{X}$ ovat riippumattomia satunnaisvektoreita.

Luku 10

Raja-arvolauseita ja approksimaatioita

Tässä luvussa esitellään sellaisia kuuluisia todennäköisysteorian raja-arvolauseita, joita sovelletaan usein tilastollisessa päättelyssä. Näiden raja-arvolauseiden tunteminen kuuluu jokaisen tilastotieteilijän yleissivistykseen. Valitettavasti päätulosten todistuksia ei voida käydä läpi tämän kurssin puitteissa.

10.1 Suurten lukujen laki

Palautetaan mieleen stokastisen suppenemisen ja melkein varman suppenemisen määritelmät.

Määritelmä 10.1. Jono satunnaismuuttujia X_1, X_2, \dots *suppenee stokastisesti* eli *konvergoi stokastisesti* (engl. *converges in probability*) kohti satunnaismuuttujaa Y , jos

$$P(|X_n - Y| \geq \epsilon) \xrightarrow{n \rightarrow \infty} 0,$$

kaikilla $\epsilon > 0$.

Stokastista suppenemistä merkitään usein seuraavaan tapaan,

$$X_n \xrightarrow{P} Y, \quad \text{tai} \quad \text{plim}_{n \rightarrow \infty} X_n = Y.$$

(Määritelmässä voitaisiin yhtäpitävästi vaatia, että $P(|X_n - Y| > \epsilon) \rightarrow 0$ kaikilla $\epsilon > 0$.)

Määritelmä 10.2. Jono satunnaismuuttujia X_1, X_2, \dots *suppenee* (eli *konvergoi*) *melkein varmasti* (engl. *converges almost surely*) kohti satunnaismuuttujaa Y , jos $X_n(\omega) \rightarrow Y(\omega)$ perusjoukon osajoukossa, jonka tn on yksi, eli jos

$$P\left(\lim_{n \rightarrow \infty} X_n = Y\right) = 1.$$

Melkein varmaa suppenemistä merkitään esim. seuraavilla tavoilla,

$$X_n \xrightarrow{\text{m.v.}} Y, \quad \text{tai} \quad X_n \xrightarrow{\text{a.s.}} Y.$$

On mahdollista osoittaa, että melkein varmasta suppenemisestä seuraa stokastinen suppeneminen, eli että

$$X_n \xrightarrow{\text{a.s.}} Y \quad \Rightarrow \quad X_n \xrightarrow{P} Y \quad (10.1)$$

Todistimme luvussa 3.7 *i.i.d.*-jonoja (eli jonoa riippumattomia ja samoin jakautuneita satunnaismuuttujia) koskevan ns. heikon suurten lukujen lain (engl. *weak law of large numbers, WLLN*). Sen mukaan, tiettyjen momenttietojen ollessa voimassa, n ensimmäisen satunnaismuuttujan aritmeettinen keskiarvo suppenee stokastisesti kohti yhden jonon jäsenen keskiarvoa. Heikon suurten lukujen lain todistus oli suoraviivainen Tšebyševin epäyhtälön sovellus.

Kuuluisan vahvan suurten lukujen lain mukaan keskiarvon suppeneminen on peräti melkein varmaa.

Lause 10.1 (Vahva suurten lukujen laki). (*Engl. strong law of large numbers, SLLN.*) *Olkoon X_1, X_2, \dots jono riippumattomia ja samoin jakautuneita satunnaismuuttujia, joiden odotusarvo $\mu = EX_1$ on olemassa. Tällöin keskiarvo*

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

suppenee melkein varmasti kohti arvoa μ , eli $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$.

Vahvan suurten lukujen lain seurauksena jono (\bar{X}_n) toteuttaa tietenkin myös heikon suurten lukujen lain.

Tällaisia (vahvoja tai heikkoja) suurten lukujen lakeja löytyy kirjallisuudesta myös muunkin tyyppisille (ei välttämättä *i.i.d.*-oletukset täyttävillä) satunnaismuuttujien, satunnaisvektoreitten ja muitten satunnaisalkioitten muodostamille jonoille.

10.2 Jakaumasuppeneminen

Määritelmä 10.3. Olkoon X_1, X_2, \dots jono satunnaismuuttujia, joiden kertymäfunktiot ovat F_1, F_2, \dots . Olkoon Y satunnaismuuttuja, jonka kertymäfunktio on G , ts. kaikilla x

$$F_n(x) = P(X_n \leq x), \quad G(x) = P(Y \leq x).$$

Jono (X_n) suppenee jakaumaltaan (engl. *converges in distribution* tai *converges in law*) kohti Y :tä, mikäli

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

kaikissa rajajakauman kertymäfunktion G jatkuvuuspisteissä x .

Usein jakaumasuppenemisessä rajajakauma on normaalijakauma $N(0, 1)$ tai jokin muu jatkuva jakauma. Jatkuvan jakauman kertymäfunktio on jatkuva koko reaaliakselilla, joten tälläisen rajajakauman kohdalla kertymäfunktioit suppenevät jokaisessa reaaliakselin pisteessä.

Jakaumasuppenemistä merkitään usein esim. seuraavasti:

$$X_n \xrightarrow{d} Y, \quad \text{tai} \quad X_n \xrightarrow{\mathcal{L}} Y, \quad \text{tai} \quad X_n \xrightarrow{w} Y.$$

Edellä yläindeksi d tulee sanasta *distribution*, \mathcal{L} sanasta *law* ja yläindeksi w sanasta *weak*. Viimeinen termi johtuu siitä, että jakaumasuppenemista kutsutaan toisinaan jakaumien (tai niiden kertymäfunktioiden) heikoksi suppenemiseksi. Jos rajajakauma on jokin tuttu jakauma, kuten $N(0, 1)$, niin jakaumasuppenemista voidaan merkitä siten, että satunnaismuuttujan sijasta käytetään rajajakauman tunnusta, seuraavaan tapaan,

$$X_n \xrightarrow{d} N(0, 1).$$

Voidaan osoittaa, että seuraavat implikaatiot pitävät paikkansa:

$$X_n \xrightarrow{\text{a.s.}} Y \quad \Rightarrow \quad X_n \xrightarrow{P} Y \quad \Rightarrow \quad X_n \xrightarrow{d} Y. \quad (10.2)$$

Yksikään implikaatioista ei päde yleisesti käänteiseen suuntaan. Jos rajamuuttuja Y on deterministinen eli vakio c , jolloin Y :n jakauman sanotaan olevan degeneroitunut, niin tällöin osoittautuu, että jakaumasuppenemisestä seuraa stokastinen suppeneminen, eli

$$X_n \xrightarrow{d} c \quad (c \text{ vakio}) \quad \Rightarrow \quad X_n \xrightarrow{P} c. \quad (10.3)$$

Tilastotieteessä jakaumasuppenemista käytetään usein jakaumien approksimointiin. Jos tiedetään, että

$$X_n \xrightarrow{d} Y, \quad \text{kun } n \rightarrow \infty$$

niin jollakin äärellisellä indeksin n arvolla voidaan satunnaismuuttujan X_n jakaumaa approksimoida rajamuuttujan Y jakaumalla, mitä voidaan merkitä symbolisesti

$$X_n \overset{d}{\approx} Y.$$

Edellisessä merkinnässä Y :n tilalla voidaan käyttää sen jakauman tunnusta.

Jos rajamuuttujalla Y on jatkuva jakauma, niin suppenemisestä $X_n \xrightarrow{d} Y$ seuraa esimerkiksi, että

$$P(X_n \in I) \xrightarrow{n \rightarrow \infty} P(Y \in I),$$

kun $I \subset \mathbb{R}$ on mikä tahansa väli. Tämän ansiosta suurilla n

$$P(X_n \in I) \approx P(Y \in I).$$

Usein tilastollisessa päättelyssä joudutaan tarkkojen luottamusvälien sijasta soveltamaan tähän ajatukseen perustuvia asymptoottisia luottamusvälejä. Tämän takia jakaumasuppeneminen on tärkeä käsite tilastotieteessä.

10.3 Suppenemistuloksia

Kukin edellä määritellystä suppenemislajista voidaan helposti yleistää myös koskemaan jonoa satunnaisvektoreita (\mathbf{X}_n) , joiden raja on satunnaisvektori \mathbf{Y} . Stokastinen suppeneminen $\mathbf{X}_n \xrightarrow{P} \mathbf{Y}$ määritellään vaatimalla, että

$$P(\|\mathbf{X}_n - \mathbf{Y}\| \geq \epsilon) \rightarrow 0, \quad \text{kaikilla } \epsilon > 0,$$

eli korvaamalla skalaaritapauksen määritelmässä erotuksen itseisarvo erotusvektorin normilla. Melkein varma suppeneminen määritellään satunnaisvektorien muodostamalle jonolle täysin samoin kuin skalaaritapauksessa. Jakaumasuppenemisen kohdalla vaaditaan moniulotteisten kertymäfunktioiden pisteittäistä suppenemista kaikissa rajamuuttujan kertymäfunktion jatkuvuusasteissa.

Jatkuvan kuvauksen soveltaminen suppenevaan jonoon säilyttää suppenemisen kaikille kolmelle suppenemisen lajille.

Lause 10.2 (Jatkuvan kuvauksen lause). *Jos k -ulotteisten satunnaisvektorien muodostama jono $\mathbf{X}_n \rightarrow \mathbf{Y}$ joko melkein varmasti, stokastisesti tai jakaumaltaan, niin $\mathbf{g}(\mathbf{X}_n)$ suppenee samalla tavalla kohti satunnaisvektoria $\mathbf{g}(\mathbf{Y})$, mikäli funktio $\mathbf{g} : \mathbb{R}^k \rightarrow \mathbb{R}^d$ on jatkuva funktio.*

Tämä lause pitää paikkansa myös silloin, jos \mathbf{g} on jatkuva pelkästään sellaisessa joukossa $A \subset \mathbb{R}^k$, jossa rajasatunnaisvektorin \mathbf{Y} arvot ovat melkein varmasti, eli jos $P(\mathbf{Y} \in A) = 1$.

Otetaan käyttöön merkintä satunnaisvektoreiden komponenteille,

$$\mathbf{X}_n = (X_{n,1}, \dots, X_{n,k}) \quad \text{ja} \quad \mathbf{Y} = (Y_1, \dots, Y_k).$$

Osoittautuu, että sekä melkein varmalle että stokastiselle suppenemiselle satunnaisvektoreille komponenttien suppeneminen (eli marginaalinen suppeneminen) ja koko satunnaisvektorin suppeneminen (eli yhteissuppeneminen) ovat yhtäpitäviä, ts.

$$\begin{aligned} \mathbf{X}_n \xrightarrow{\text{a.s.}} \mathbf{Y} &\Leftrightarrow (X_{n,j} \xrightarrow{\text{a.s.}} Y_j, \quad \forall j = 1, \dots, k), \\ \mathbf{X}_n \xrightarrow{P} \mathbf{Y} &\Leftrightarrow (X_{n,j} \xrightarrow{P} Y_j, \quad \forall j = 1, \dots, k). \end{aligned}$$

Jakaumasuppenemisessä yhteissuppenemisen ja marginaalisen suppenemisen välinen yhteys on mutkikkaampaa, sillä marginaalisesta jakaumasuppenemisestä ei seuraa yhteissuppeneminen. Esimerkiksi, jos $X_n \xrightarrow{d} X$ ja $Y_n \xrightarrow{d} Y$, niin tavallisesti vektori (X_n, Y_n) ei suppene jakaumaltaan eikä täten missään muussakaan mielessä. Tämä ongelma liittyy siihen seikkaan, että reunajakaumat eivät määrää yhteisjakaumaa. Toisaalta, jos $(X_n, Y_n) \xrightarrow{d} (X, Y)$, niin tällöin jatkuvan kuvauksen lauseen perusteella kyllä $X_n \xrightarrow{d} X$ ja $Y_n \xrightarrow{d} Y$.

Seuraavassa tärkeässä erikoistapauksessa marginaalisesta jakaumasuppenemisestä seuraa yhteissuppeneminen. Tätä tulosta (tai sen sovellusta peruslaskutoimituksiin) kutsutaan Slutskyn lauseeksi tai Slutskyn lemmaksi.

Lause 10.3 (Slutsky). *Jos $X_n \xrightarrow{d} X$ ja $Y_n \xrightarrow{d} c$, jossa c on vakio, niin $(X_n, Y_n) \xrightarrow{d} (X, c)$. Tällöin myös*

a) $X_n + Y_n \xrightarrow{d} X + c$,

b) $X_n Y_n \xrightarrow{d} cX$,

c) $X_n / Y_n \xrightarrow{d} X/c$, mikäli $c \neq 0$.

Todistus. Suppenemisen $(X_n, Y_n) \xrightarrow{d} (X, c)$ todistus sivuutetaan. Loppuosa väitteistä seuraa jatkuvan kuvauksen lauseesta. \square

10.4 Keskeinen raja-arvolause

Jos X_1, X_2, \dots on *i.i.d.*-jono, ja \bar{X}_n on n ensimmäisen jonon muuttujan keskiarvo, niin tiedämme suurten lukujen lain nojalla, että \bar{X}_n suppenee kohti muuttujien yhteistä odotusarvoa $\mu = EX_1$. Lisäksi tiedämme, että

$$E\bar{X}_n = \mu,$$

ja

$$\text{var } \bar{X}_n = \frac{\sigma^2}{n}, \quad \text{jossa } \sigma^2 = \text{var } X_1.$$

Keskiarvon \bar{X}_n jakauma keskittyy yhä tiiviimmin arvon μ ympärille, kun n kasvaa. Keskittymisvauhtia kuvaa tietyllä tavalla keskiarvon \bar{X}_n keskihajonta σ/\sqrt{n} , joka suppenee nollaa kohti vauhdilla $1/\sqrt{n}$.

Jos keskiarvo \bar{X}_n standardoidaan vähentämällä siitä keskiarvon odotusarvo ja jakamalla keskiarvon keskihajonnalla, saadaan lauseke

$$\frac{\bar{X}_n - E\bar{X}_n}{\sqrt{\text{var } \bar{X}_n}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}.$$

Tämän standardoidun keskiarvon odotusarvo on tietenkin nolla, ja sen varianssi on yksi kaikilla n . Keskeisen raja-arvolauseen mukaan standardoitu keskiarvo suppenee jakaumaltaan kohti standardinormaalijakaumaa.

Lause 10.4 (Keskeinen raja-arvolause). (*Engl.* central limit theorem, CLT.)
Olkkoon X_1, X_2, \dots jono riippumattomia ja samoin jakautuneita satunnaismuuttujia siten, että $0 < \sigma^2 < \infty$, jossa $\sigma^2 = \text{var } X_1$. Merkitään

$$\mu = EX_1, \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Tällöin

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

Jatkuvan kuvauksen lauseen mukaan edellisen lauseen oletuksilla pätee myös

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

sillä kuvaus $x \mapsto \sigma x$ on jatkuva.

Tämä tulos antaa karakterisoinnin stokastisen suppenemisen $\bar{X}_n \xrightarrow{P} \mu$ vauhdille. Jos kohti nollaa suppeneva erotus $\bar{X}_n - \mu$ kerrotaan kohti ääretöntä hajaantuvalla jonolla \sqrt{n} , niin yhä saadaan jotakin stabiilia, eli satunnaismuuttujia, jolla on (degeneroitumaton) rajajakauma. Jos kertoimena käytetään mitä tahansa hitaammin kasvavaa lukujonoa $a_n \sqrt{n}$, jossa $a_n \rightarrow 0$, niin Slutskyn lauseen perusteella

$$a_n \sqrt{n} (\bar{X}_n - \mu) \xrightarrow{P} 0.$$

Keskeiselle raja-arvolauseelle löytyy kirjallisuudesta lukuisia yleistyksiä. Tilastotieteessä tarvitaan usein keskeisen raja-arvolauseen muotoilua satunnaisvektorien muodostamalle *i.i.d.*-jonolle. Jos $\mathbf{X}_1, \mathbf{X}_2, \dots$ on *i.i.d.*-jono satunnaisvektoreita, joiden yhteinen odotusarvovektori on $\boldsymbol{\mu}$ ja kovarianssimatriisi on $\boldsymbol{\Sigma}$, niin keskeinen raja-arvolause on voimassa muodossa

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}), \quad (10.4)$$

jossa $\bar{\mathbf{X}}_n$ on n ensimmäisen satunnaisvektorin keskiarvo $\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$.

10.5 Normaaliapproksimaatio

Kun (X_i) on *i.i.d.*-jono, niin keskeiseen raja-arvolauseeseen tukeutuen toisinaan approksimoidaan äärellisellä otoskoolla n

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

eli standardoidun otoskeskiarvon jakaumaa approksimoidaan sen rajajakaumalla. Tämä on sama asia kuin se, että käytetään approksimaatiota

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

jossa keskiarvon \bar{X}_n jakaumaa approksimoidaan normaalijakaumalla, jonka odotusarvo ja varianssi ovat samat kuin keskiarvon \bar{X}_n odotusarvo ja varianssi. Edelleen, tämä on sama asia kuin se, että käytetään approksimaatiota

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

jossa summan jakaumaa approksimoidaan normaalijakaumalla, jonka odotusarvo ja varianssi ovat samat kuin ko. summan odotusarvo ja varianssi.

Keskeinen raja-arvolause takaa, että nämä normaaliapproksimaatiot (eli normaaliset approksimaatiot tai normaalijakauma-approksimaatiot) saadaan mieltävaltaisen tarkoiksi, kun otoskoko n valitaan riittävän suureksi. Milloin otoskoko n sitten on riittävän suuri? Tämä asia riippuu toisaalta halutusta tarkkuudesta ja approksimaation käyttötarkoituksesta ja toisaalta satunnaismuuttujien X_i yhteisen jakauman luonteesta. Symmetrisille ja yksihuippuisille jakaumille saavutetaan pienellä (muutamana kymmenenä) otoskoolla useisiin tarpeisiin riittävä tarkkuus, mutta vinojen jakaumien kohdalla vastaavaan tarkkuuteen tarvittava otoskoko voi olla monta kertaluokkaa suurempi.

Esimerkki 10.1. (Binomijakauman normaaliapproksimaatio.) Olkoot X_1, X_2, \dots, X_n riippumattomia Bernoulli(p)-jakaumaa noudattavia sm:ia, jossa $0 < p < 1$. Tällöin

$$E\bar{X}_n = p, \quad \text{var } \bar{X}_n = \frac{1}{n} p(1-p).$$

Normaaliapproksimaatiolla saadaan

$$\bar{X}_n \stackrel{d}{\approx} N\left(p, \frac{1}{n} p(1-p)\right), \quad \text{ja} \quad \sum_{i=1}^n X_i \stackrel{d}{\approx} N(np, np(1-p)).$$

Tässä tapauksessa pätee tietenkin tarkka tulos

$$\sum_{i=1}^n X_i \sim \text{Bin}(n, p).$$

Esimerkiksi, kun $n = 25$ ja $p = 0.6$, niin $np = 15$ ja $np(1-p) = 6$. Tällöin saadaan approksimaatio

$$P\left(\sum_1^n X_i \leq 13\right) \approx P(15 + \sqrt{6} Z \leq 13) = P\left(Z \leq \frac{13-15}{\sqrt{6}}\right) \approx 0.207,$$

jossa $Z \sim N(0, 1)$. Kun käytetään binomijakaumaa, eikä sen approksimaatiota, saadaan tulos

$$P\left(\sum_1^n X_i \leq 13\right) \approx 0.268.$$

Näin pienellä otoskoolla normaalin approksimaatio ei ole kovin tarkka, mutta sitä voidaan parantaa huomattavasti käyttämällä ns. *jatkuvuuskorjausta*. Koska summa on kokonaislukuarvoinen, niin seuraavat tapahtumat ovat samoja,

$$\left\{\sum_1^n X_i \leq 13\right\} = \left\{\sum_1^n X_i \leq 13 + 0.5\right\}.$$

Kun normaaliapproksimaatiossa yläraja 13 korvataan luvulla 13.5, niin saadaan jo alkuperäistä approksimaatiota paljon tarkempi approksimaatio

$$P\left(\sum_1^n X_i \leq 13\right) \approx P\left(Z \leq \frac{13.5 - 15}{\sqrt{6}}\right) \approx 0.270.$$

Nykyaikana tällaiset tunnettujen jakaumien approksimaatiot eivät ole niin tärkeitä kuin ennen. Nykyään standardijakaumien kertymäfunktioit pystytään helposti laskemaan tietokoneella. \triangle

10.6 Deltamenetelmä

Seuraava lause on yksinkertainen esimerkki ns. deltamenetelmästä (engl. *delta method*). Sen todistus käydään läpi lähinnä siksi, että lukija saisi kuvan siitä, miten suppenemiseen liittyvää koneistoa käytetään hyväksi.

Lause 10.5. *Jos*

$$\sqrt{n}(X_n - a) \xrightarrow{d} N(0, \sigma^2),$$

jossa a on vakio, ja funktio g on derivoituva pisteessä a, niin

$$\sqrt{n}(g(X_n) - g(a)) \xrightarrow{d} N(0, \sigma^2(g'(a))^2). \quad (10.5)$$

Todistus. Sovelletaan ensimmäisen kertaluvun Taylorin kehitelmää

$$g(X_n) - g(a) = g'(a)(X_n - a) + r(X_n)(X_n - a),$$

jossa jäännöstermin funktio r on

$$r(x) = \begin{cases} \frac{g(x) - g(a)}{x - a} - g'(a), & x \neq a \\ 0 & x = a \end{cases}$$

Huomaa, että funktio r on jatkuva pisteessä $x = a$ (derivoitavuuden nojalla).

Nyt $X_n \xrightarrow{p} a$, sillä

$$X_n - a = \frac{1}{\sqrt{n}} \sqrt{n}(X_n - a),$$

jossa termi $1/\sqrt{n} \rightarrow 0$ ja toinen termi suppenee jakaumaltaan, joten (Slutsdyn lause) $X_n - a \xrightarrow{d} 0$, ja koska rajamuuttuja on vakio, niin suppeneminen on peräti stokastista (ks. kaava (10.3)). Siis $X_n - a \xrightarrow{P} 0$, eli $X_n \xrightarrow{P} a$. Lisäksi

$$\sqrt{n}(g(X_n) - g(a)) = g'(a) \sqrt{n}(X_n - a) + r(X_n) \sqrt{n}(X_n - a)$$

Tässä $r(X_n) \xrightarrow{P} 0$ (jatkuvan kuvauksen lause), ja koska jono $\sqrt{n}(X_n - a)$ suppenee jakaumaltaan, niin $r(X_n) \sqrt{n}(X_n - a) \xrightarrow{d} 0$ (Slutsdyn lause). Tämä jäännöstermin suppeneminen on peräti stokastista, koska rajamuuttuja on vakio. Kun sovelletaan toistamiseen Slutsdyn lausetta, niin nähdään, että

$$g'(a) \sqrt{n}(X_n - a) + r(X_n) \sqrt{n}(X_n - a) \xrightarrow{d} g'(a)Y,$$

jossa $Y \sim N(0, \sigma^2)$. Toisin sanoen jakaumasuppenemisen kannalta jäännöstermin saa unohtaa, ja rajajakauma määräytyy ensimmäisen kertaluvun Taylorin approksimaatiosta. Väite seuraa lopulta siitä, että $g'(a)Y \sim N(0, (g'(a))^2 \sigma^2)$. \square

Tätä raja-arvotulosta sovelletaan usein siten, että äärellisellä otoskoolla n approksimoidaan

$$\sqrt{n}(g(X_n) - g(a)) \stackrel{d}{\approx} N(0, \sigma^2 (g'(a))^2).$$

Esimerkki 10.2. Tarkastellaan riippumattomia muuttujia $X_i \sim \text{Bernoulli}(p)$, $i = 1, \dots, n$, jossa $0 < p < 1$. Tällöin parametrin p SU-estimaatti on $\bar{X}_n = s/n$, jossa s on onnistumisten lukumäärä. Oletetaan, että tahdotaan estimoida ns. *log-odds* -suuretta (vedonlyöntisuhteen logaritmia),

$$\theta = \ln \frac{p}{1-p}.$$

Sen luonteva estimaatti on

$$\hat{\theta}_n = \ln \frac{\bar{X}_n}{1 - \bar{X}_n} = \ln \frac{s}{n - s},$$

jonka jakauma ei ole mikään tunnettu jakauma.

Kun delta-menetelmässä valitaan

$$g(u) = \ln \frac{u}{1-u}, \quad 0 < u < 1,$$

saadaan helposti approksimaatio

$$\sqrt{n}(\hat{\theta}_n - \theta) \stackrel{d}{\approx} N\left(0, \frac{1}{p(1-p)}\right).$$

Tällä perusteella on esim. mahdollista johtaa asymptoottinen luottamusväli parametrille θ . \triangle