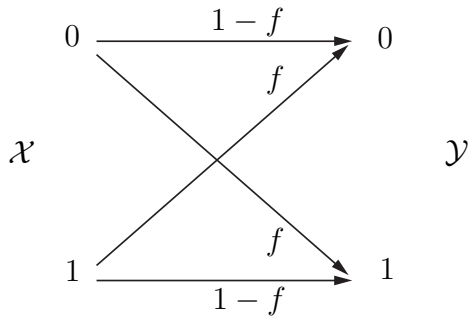
Informaatioteoria

Lasse Holmström
Sovelletun matematiikan ja tilastotieteen yksikkö
Oulun yliopisto

Kevät 2016



Sisältö

Luku 1

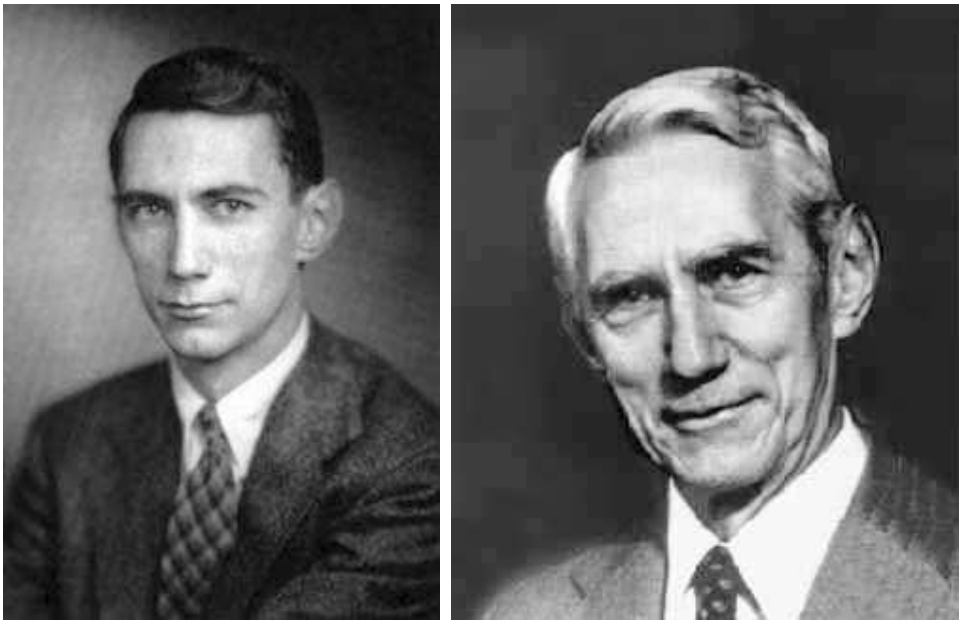
Johdanto

1.1 Historiaa

1.1.1 Informaatioteorian synty

Nykymuotoisen informaatioteorian perustaja on Claude Shannon (1916 - 2001). Shannon oli Yhdysvaltalainen matemaatikko, sähköinsinööri ja keksijä. Shannonin informaatioteorian perusteita käsittelevä raportti ”A Mathematical Theory of Communication” ilmestyi vuonna 1948 Bell Systems Technical Journalissa. Tämän raportin tuloksiin perustuva Shannonin ja Warren Weaverin kirja ilmestyi vuonna 1949 ja siitä on saatavissa vuonna 1998 julkaistu uusintapainos [6].

Shannonia pidetään ns. digitaalisen vallankumuoksen aloittajana. Shannon ymmärsi, että kaikkea informaatiota voidaan kommunikoida bitteinä ja hän johti tiedonsiirron tehokkuuden rajat. Shannonin läpimurtotyö käynnisti myös koodausteorian kehittelyn. Tehokkaat koodausmenetelmät ovat nykyisin keskeisen tärkeitä mm. mobiililaitteissa, CD- ja DVD-soittmimissa, erilaisissa



Kuva 1.1: Claude Elwood Shannon (1916 - 2001).

muistilaitteissa, internetin toiminnassa jne.

Informaation olemusta oli ennen Shannonia tutkittu myös tilastollisen fysiikan piirissä (mm. Ludwig Boltzmann ja John von Neumann). Leo Szilard lanseerasi ”bitin” käsitteen informaation mittauksessa. Termi ”bit” tosin on peräisin matemaatikko John Tukeyltä (mm. Tukeyn lemma, Explorative Data Analysis (EDA),...).

Shannonin tutkimussaralla oli edeltäjiä myös itse Bellin tutkimuslaboratoriossa, mm. Harry Nyquist ja Ralph (Vynton Leon) Hartley. Bellin tutkimuslaboratoriot ovat informaatioteorian lisäksi monen keskeisen keksinnön koti: laboratorioilla lasketaan olevan yli 26 000 patenttia ja 40 000 keksintöä mm. stereofoninen ääni, äänielokuva, telefax, UNIX käyttöjärjestelmä, sellaiset ohjelmointikielet kuin C ja C++ jne. Puhelimen keksijän Abraham Bellin mukaan nimetty tutkimuslaboratorio perustettiin vuonna 1925 ja nykyään Bellin laboratorioissa työskentelee yli 9000 henkilöä useissa maissa.

Laboratorion työntekijöiden joukossa on ollut mm. 11 nobelistia. Shannon itse toimi 15 vuotta Bellillä. Vuonna 1956 hänestä tuli MIT:n (Massachusetts Institute of Technology) professori. MIT oli ensimmäisiä yliopistoja, jossa informaatioteoriaa alettiin säännöllisesti opettaa.

Toisen maailmansodan aikaisilla sotaponnisteluilla oli tärkeä merkitys informaatioteorian ja sen sovellusten siivittäjänä. Sotilastutkimusta tukemaan koottiin poikkitieteellinen ryhmä eri alojen huippututkijoita ratkomaan informaatioon ja sen käsittelyyn liittyviä peruskysymyksiä (koneet, biologia). Tähän ryhmään kuuluivat mm. Claude Shannon, Norbert Wiener, Warren McCulloch, Walter Pitts, Alan Turing ja John von Neumann. Myös elektroniikan ja viestintätekniikan voimakas kehitys sodan aikana ja luotettavan ja turvallisen kommunikaation tarve suuntasi kiinnostusta informaatioteoreettisiin kysymyksiin.

1.1.2 Informaatioteorian vaiheita vuodesta 1948

Shannonin informaatioteorian läpimurtojulkaisua (1948) seurasi Norbert Wienerin esittämä teoria vuonna 1949. Seuraavaa vuosikymmentä luonnehti voimakas kiinnostuksen kasvu informaatioteoriaa kohtaan:

- Järjestettiin lukuisia yliopistoseminaareja, kursseja ja konferensseja.
- IRE (Institute of Radio Engineers) ryhtyi julkaisemaan IRE Transactions on Information Theory lehteä vuonna 1955. Vuonna 1963 IRE:stä tuli tunnettu ja monella tutkimusalueella nykyisin toimiva IEEE (Institute of Electrical and Electronics Engineers).
- Informaatioteorian keskeiseksi yhteistyöverkostoksi perustettiin PGIT (Professional Group on Information Theory), joka toimi alan tärkeänä koordinoivana sisäpiirinä.

- Keskeisiä nimiä olivat mm. Peter Elias, Norbert Wiener, Robert Fano, David Huffman, Richard Hamming ja Edgar Gilbert (kummatkin virheitä korjaavien koodien uranuurtajia).
- Matemaattista informaatioteoriaa edustivat Aleksandr Khintšin, Amiel Feinstein ja Jacob Wolfowitz (mm. IMF:n pääjohtajana jonkin aikaa toimineen Paul Wolfowitzin isä).

Kuten usein käy voimakkaasti kehittyvien alojen kohdalla, niin myös informaatioteorian suosion räjähdysmäinen kasvu johti ”ylikuumenemiseen” ja ”hypeen”. Vuoden 1952 informaatioteorian konferenssissa melkein puolet papereista olivat psykologiaa ja neurofysiologiaa ja vuoden 1956 konferenssissa edustettuina olevien alojen kirjo oli sitten jo todella suuri: anatomia, antropologia,..., lingvistiikka, matematiikka,..., politiikan teoria, tilastotiede. Syynä informaatioteorian ideoiden ylikäyttöön mitä erilaisimmilla aloilla oli usein se, että ”informaatioteorian” esiintyminen määrärahahakemuksissa arveltiin (ilmeisesti osittain perustellusti) lisäävän hankkeen uskottavuutta ja siten rahoitusmahdollisuuksia. Tässä tilanteessa PGIT katsoi välttämättömäksi informaatioteorian ”puhdistamisen” erilaisista lieveilmiöistä. Tämän operaation käynnisti Shannonin itsensä laatima kirjoitus vuonna 1956 ja järjestyksen katsotaan palanneen vuoteen 1958 mennessä.

1950-luvulla elettiin intensiivistä kylmän sodan aikaa ja rahoitusta informaatioteorian tutkimukseen saatiin erityisesti Yhdysvaltojen asevoimilta. Kolme päätutkimussuuntaa tällöin olivat

- Hajaspektriteknologia. 1980-luvun puoleen väliin asti tämä tutkimus oli sotilaallista ja siten salaista. Nykyinen CDMA-tekniikka on saanut alkunsa tästä tutkimuksesta.
- Kompressio, informaation pakkaaminen. Tämä oli itseasiassa tutkimuksen alkuvuosein pääasiallinen kiinnostuksen kohde kun tiedon *siirtoa* ei vielä pidetty niin keskeisenä ongelmana.

- Koodaus tiedon siirtoa varten häiriöisessä kanavassa. Shannonin lause kertoi tällöin mihin tehokkuuteen tiedon siirrossa teoriassa voidaan päästä.

Koodausta tiedon siirrossa ei aluksi pidetty kiinnostavana, koska voitiin ajatella aina lisättävän lähetystehoa häiriöiden voittamiseksi. Tilanteen muutti Neuvostoliiton Sputnik vuonna 1957: Yhdysvaltojen ja Neuvostoliiton kilpajuoksu avaruuteen alkoi. Lähetystehoa oli kallista tai jopa mahdotonta lisätä avaruudessa, koska jokainen avaruuten lähettävä gramma maksoi todella paljon. Tehokkaasta koodauksesta oli saatavissa ratkaisevaa kustannushyötyä ja Shannonin Gaussin kanavan malli sopi hienosti kuvaamaan satelliitin ja maa-aseman välistä viestintää. 1960-luvulla kiinnostus koodaukseen kasvoi nopeasti. Tältä ajalta voidaan mainita esimerkiksi Irving Reed ja Gustave Solomon. Koodausta käytettiin ensimmäistä kertaa virallisesti 1969 Yhdysvaltain Mariner VI Mars-luotaimessa. Se lähetti mm. värikuvia Marsin kiertoradalta. Viestinnässä käytettiin jo 1954 kehitettyä, virheitä korjaavaa Reed-Müller koodia. Tosin koodausta epävirallisesti käytti avaruudessa itseasiassa ensimmäisenä vuonna 1968 Pioneer IX, Yhdysvaltain aurinkoa kiertävä fysikaalisia perusmittauksia tekevä satelliitti.

1960-luvun lopussa kasvoi kuitenkin epävarmuus koodausmenetelmien kehittelyn käytännön merkityksestä. Algoritmit olivat kalliita implementoida ja vain avaruustutkimuksella oli siihen varaa. Informaatioteorian tutkimusryhmät alkoivatkin hajota tutkijoiden siirtyessä muihin lupaavimpiin projekteihin. Floridan St. Petersburgissa vuonna 1971 pidetty ”Future Directions” konferenssi päätyi hyvin pessimistisiin tunnelmiin koodausteorian tulevaisuuden suhteen. Vallalle oli noussut tunne siitä, että oli parempi itse asiassa lyödä niin sanotusti hanskat naulaan. Siinä missä Sputnik oli aikaisemmin muuttanut kaiken, saman teki kuitenkin kertaheitolla Intelin ensimmäinen mikroprosessori vuonna 1971. Nyt uusi halvempi ja tehokkaampi teknologia mahdollisti uusien ja parhaimpien koodausalgoritmien käytön. Kuvaan tuli-

vat mukaan myös kaupalliset, ei-sotilaalliset ja avaruustekniikkaan suoraan liittymättömät sovellukset, modeemi ja telefax ensimmäisten joukossa.

Tänä päivänä kehittyvä teknologia on informaatioteorian kehitystä ja hyödyntämistä ylläpitävä voima. Esimerkiksi Gallagerin 1960 väitöskirjassaan esittämä koodi (low-density parity-check codes) on tullut vasta nyt käyttöön! Jatkuvia uusia haasteita ja sovellusmahdollisuuksia tarjoavat mobiili tiedonsiirto, erilaiset muistitekniikat (RAM, kiintolevyt), CD-, DVD-, ja MP3-soittimet, tietokoneverkot, internet jne.

Mikä sitten on ollut Shannonin teorian merkitys koodausteknologian kehitykselle? Voidaan sanoa, että se määritti tiedonsiirron tehokkuudelle rajat, joita ei voinut ylittää. Kun rajat olivat tiedossa, syntyi motivaatio pyrkiä niitä kohti ja joka vaiheessa tiedettiin kuinka paljon parantamisen varaa vielä oli. Parhailla nykyisillä koodeilla päästään jo Shannonin rajalle tietyissä kanavissa (Gaussin kanava).

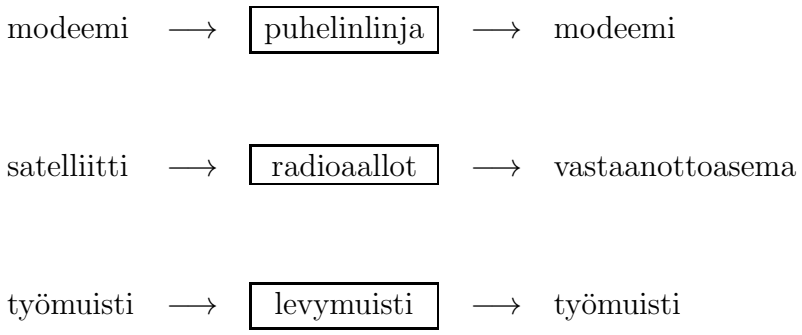
1.2 Peruskysymyksiä

Tarkastellaan tiedonsiirtoa seuraavan yksinkertaisen mallin mukaisesti:

$$\text{informaatiolähde} \longrightarrow \boxed{\text{kanava}} \longrightarrow \text{vastaanottaja}$$

Konkreettisia esimerkkejä tiedonsiirrosta kanavien läpi on esitetty kuvassa 1.2.

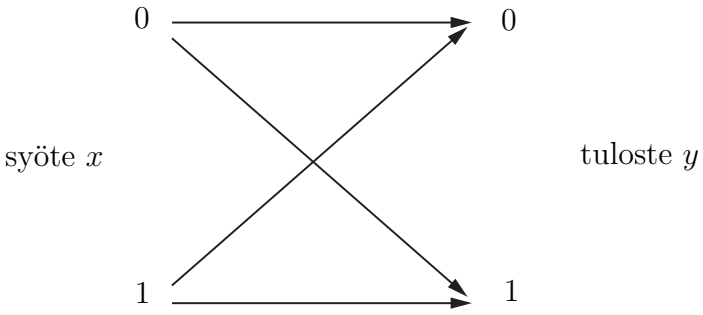
Kanavassa, jonka läpi informaatiota siirretään on useinmiten *häiriötä* (”kohinaa”). Tällöin keskeinen kysymys on se miten vähentää häiriöiden aiheuttamia virheitä.



Kuva 1.2: Eräitä esimerkkejä tiedonsiirrosta kanavien läpi.

1.2.1 Binäärinen symmetrinen kanava (BSK)

Kuvassa 1.3 on esitetty ns. binäärinen symmetrinen kanava. Syötteinä ja tulosteina ovat bitit 0 ja 1.



Kuva 1.3: Binäärinen symmetrinen kanava

Olkoon tiedon siirrossa tapahtuvan virheen todennäköisyys (ns. kohinataso) $0 < f < 1$: virheettömän bitin siirtymisen todennäköisyys on

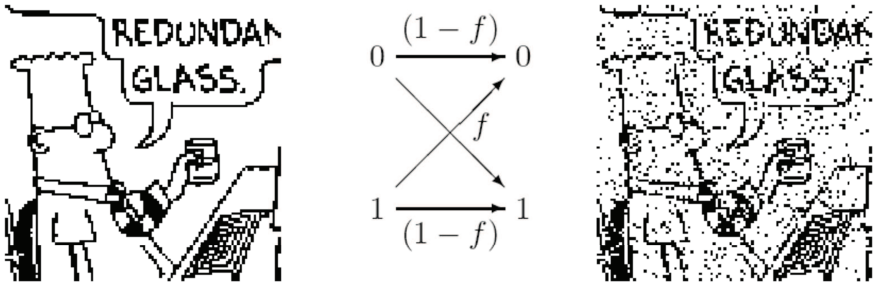
$$\mathbb{P}\{y = 0 \mid x = 0\} = \mathbb{P}\{y = 1 \mid x = 1\} = 1 - f,$$

ja virheen todennäköisyys on

$$\mathbb{P}\{y = 0 \mid x = 1\} = \mathbb{P}\{y = 1 \mid x = 0\} = f.$$

Tässä

$$\mathbb{P}\{y = 0 \mid x = 0\} = \frac{\mathbb{P}\{y = 0 \text{ ja } x = 0\}}{\mathbb{P}\{x = 0\}} \text{ jne.}$$



Kuva 1.4: Binäärinen symmetrinen kanava kohinatasolla $f = 0.1$ (esimerkki lähteestä [5]).

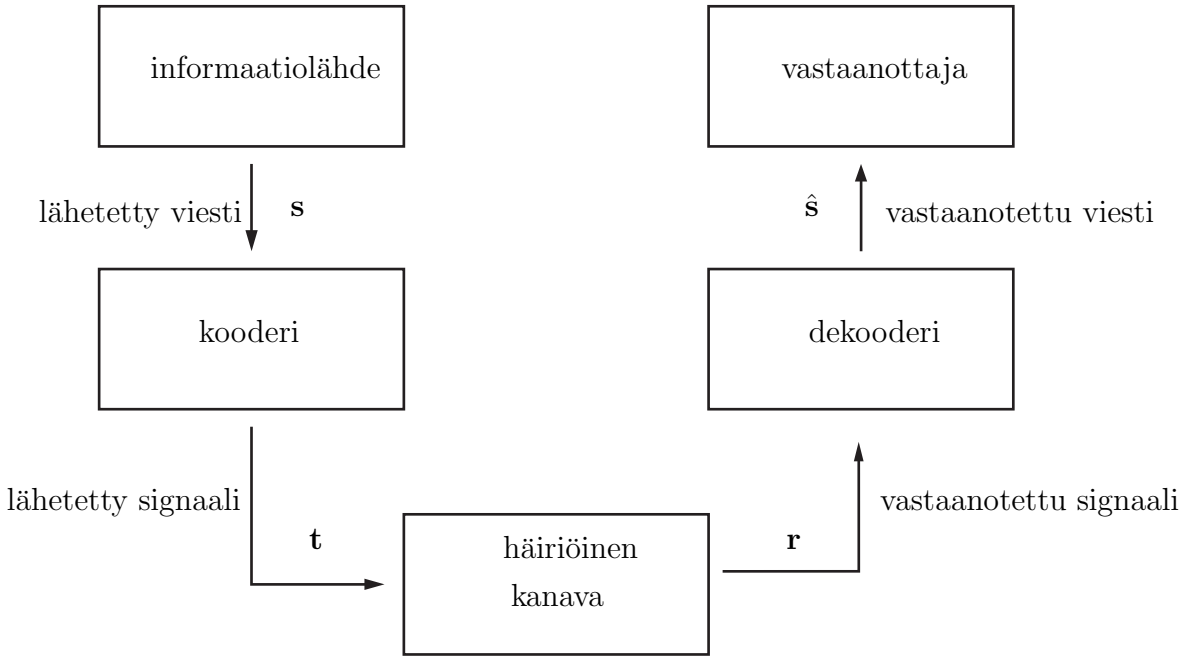
Kuvassa 1.4 on esimerkki binäärisestä symmetrisestä kanavasta, jossa syötteenä on 100×100 digitaalinen kuva. Vasemman puoleisen kuvan pikselit on syötetty yksi kerrallaan toisistaan riippumatta binääriseen symmetriseen kanavaan, jonka kohinataso on $f = 0.1$.

Ajatellaan toisena esimerkkinä tietokoneen kiintolevyä, jolle luetaan ja kirjoitetaan 1 GB päivässä 10 vuoden ajan ja ajatellaan BSK-mallin kuvaavan bittien siirtymistä lukemisessa ja kirjoittamisessa. Mikä tällöin on kohtuullinen f? Kohtuullista on selvästikin odottaa kiintolevyltä lähes virheetöntä toimintaa.

Luku/kirjoitusoperaatioita on yhteensä $\approx 10^9 \cdot 8 \cdot 10 \cdot 365 \approx 3 \cdot 10^{13} = n$ kappaletta. Olkoon $f = 10^{-15}$, jolloin

$$\mathbb{P}\{\text{"virheetön toiminta"}\} \approx (1-f)^n \approx 1 - nf \approx 0.97.$$

Kysymys kuuluu: miten näin pieneen virhetodennäköisyyteen f päästään? Voidaan ensinnäkin ajatella tehtävän parannuksia itse fyysiseen laitteeseen. Tämä voi kuitenkin johtaa kustannusten jyrkkään nousuun. Vaihtoehtona



Kuva 1.5: Tiedonsiirto koodamalla ja dekoodaamalla viesti.

on koodata/dekoodata bittejä sopivasti jolloin vain tarvittava *laskentatyö* lisääntyy (ks. kuva 1.5).

Informaatioteoria kertoo tämän koodaukseen/dekoodaukseen perustuvan tiedonsiirtotavan mahdollisuudet ja rajat. *Koodausteoriassa* kehitetään käytäntöön sopivia koodereita ja dekoodereita.

1.2.2 Toistokoodit

Eräs yksinkertainen koodausmenetelmä on ns. *toistokoodi*. Toistokoodissa R_m kukin bitti toistetaan m kertaa.

Esimerkki 1.1. Toistotkoodi R_3.

Koodaus tapahtuu siis seuraavan kaavion mukaisesti:

$$\begin{array}{ccc} 0 & \overset{\text{kooderi}}{\longrightarrow} & 000 \\ 1 & \longrightarrow & 111 \end{array}$$

Olkoon nyt lähetety viesti

$$\mathbf{s} = 0010110,$$

jolloin kooderi tekee siitä lähetettävän signaalin

$$\mathbf{t} = 000\,000\,111\,000\,111\,111\,000.$$

Olkoon edelleen häiriöinen kanava muotoa

$$\mathbf{r} = \mathbf{t} + \mathbf{n} \pmod 2,$$

missä $\mathbf{n}$ on häiriö. Esimerkiksi

$\mathbf{t}$	000	000	111	000	111	111	000
$\mathbf{n}$	000	001	000	000	101	000	000
$\mathbf{r}$	000	001	111	000	010	111	000

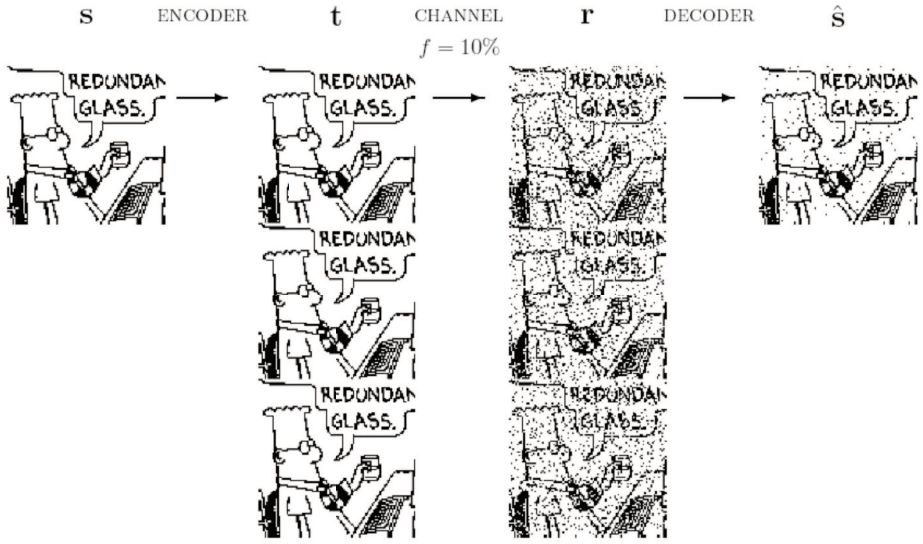
Dekooderi tekee enemmistöpäätöksen kolmen ryhmissä, jolloin vastaanotettu viesti on

$$\begin{array}{rcccccccc} \hat{\mathbf{s}} & = & 0 & 0 & 1 & 0 & 0 & 1 & 0. \\ & & & \uparrow & & & \uparrow & & \\ & & & \text{virhe} & & & \text{virhe} & & \\ & & & \text{korjattu} & & & \text{ei korjattu} & & \end{array}$$

Voidaan osoittaa (harjoitustehtävä), että tämä dekooderi on tietyin edellytyksin *optimaalinen*. ‖

Harjoitustehtävänä osoitetaan myös, että kohinatasolla $0 < f < 1/2$ toimivassa BSK:ssa edellisen esimerkin dekooderin virheen todennäköisyys on pienempi kuin f. Kuitenkin tiedonsiirtonopeus on vain 1/3 alkuperäisestä,

$$R = \frac{1}{3} \quad \text{(Rate)} \quad \text{(bittiä/kanavan käyttö)}.$$



Kuva 1.6: Binäärinen symmetrinen kanava kohinatasolla $f = 0.1$, kun käytetään toistokoodia R_3. Bittivirheen todennäköisyys on nyt noin 0.03 (esimerkki lähteestä [5]).

Jos esimerkiksi kiintolevyn nopeus on 1 Gbit/s, on se toistokoodin R_3 jälkeen $\frac{1}{3}$ Gbit/s.

Tarkastellaan sitten yleistä toistokoodia R_m, missä $m = 2n + 1$ on pariton. Olkoon kanava binäärinen symmetrinen kanava, $0 < f < 1/2$, ja oletetaan, että bitit siirtyvät kanavan läpi toisistaan riippumatta. Kooderi on nyt siis

$$\begin{array}{ccc} \mathbf{s} & & \mathbf{t} \\ 0 & \overset{\text{kooderi}}{\longrightarrow} & 00\cdots0 \\ 1 & \longrightarrow & \underbrace{11\cdots1}_{2n+1} \end{array}$$

Olkoon

$$\begin{aligned} p_b &= \mathbb{P}\{\text{"virhe bitissä"}\} \\ &= \mathbb{P}\{\text{"vähintään } n+1 \text{ koodibittiä vaihtuu kanavassa"}\}. \end{aligned}$$

Vaihtuvien bittien lukumäärän jakauma on $\mathrm{Bin}(2n+1, f)$, jolloin siis

$$\mathbb{P}\{\text{"}k \text{ bittiä vaihtuu"}\} = \binom{2n+1}{k} f^k (1-f)^{2n+1-k}$$

ja siten

$$p_b = \sum_{k=n+1}^{2n+1} \binom{2n+1}{k} f^k (1-f)^{2n+1-k}.$$

Olkoon S_{2n+1} vaihtuvien bittien lukumäärä. Silloin heikon suurten lukujen lain (ns. Bernoullin lause) mukaan

$$\frac{S_{2n+1}}{2n+1} \to f \quad \text{stokastisesti,}$$

eli kaikilla $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{ \left| \frac{S_{2n+1}}{2n+1} - f \right| \geq \varepsilon \right\} = 0.$$

Bernoullin lauseen sisältöhän on se, että toistokokeessa esiintyvän tapahtuman suhteellinen esiintymisfrekvenssi lähenee tapahtuman todennäköisyyttä toistokokeiden määrän kasvaessa. Nyt

$$\begin{aligned} p_b &= \mathbb{P}\left\{ S_{2n+1} \geq n+1 \right\} = \mathbb{P}\left\{ \frac{S_{2n+1}}{2n+1} \geq \frac{n+1}{2n+1} \right\} \\ &= \mathbb{P}\left\{ \frac{S_{2n+1}}{2n+1} \geq f + \frac{n+1}{2n+1} - f \right\}. \end{aligned}$$

Tässä

$$\frac{n+1}{2n+1} - f \xrightarrow[n\to\infty]{} \frac{1}{2} - f > 0.$$

Siis: jos $0 < \varepsilon < \frac{1}{2} - f$ ja n on niin suuri, että $\frac{n+1}{2n+1} - f > \varepsilon$, pätee heikon suurten lukujen lain mukaan

$$p_b \leq \mathbb{P}\left\{ \frac{S_{2n+1}}{2n+1} \geq f + \varepsilon \right\} \leq \mathbb{P}\left\{ \left| \frac{S_{2n+1}}{2n+1} - f \right| \geq \varepsilon \right\} \longrightarrow 0,$$

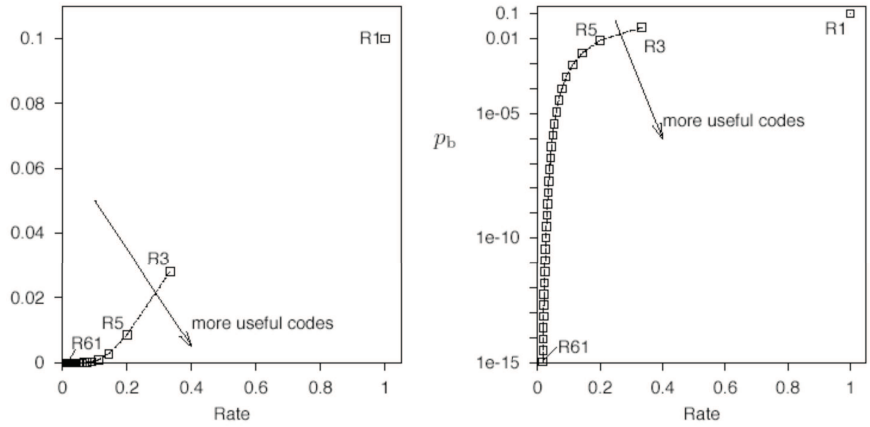
kun $n \to \infty$.

Siten bittivirhe p_b saadaan mielivaltaisen pieneksi, kun $n \to \infty$ eli $m \to \infty$ toistokoodissa R_m. Mutta samalla tiedonsiirtonopeudelle saadaan

$$R = \frac{1}{2n+1} \longrightarrow 0,$$

kun $n \to \infty$. Siksi $p_b \to 0$ vain, jos samalla $R \to 0$.

Kuvassa 1.7 on esitetty bittivirheen p_b riippuvuus tiedonsiirtonopeudesta R eräille toistokoodeille.



Kuva 1.7: Bittivirheen p_b riippuvuus tiedonsiirtonopeudesta R eräille toistokoodeille binäärinessä symmetrisessä kanavassa kohinatasolla $f = 0.1$. Oikean puoleisessa paneelissa on logaritminen skaala (kuva lähteestä [5]).

1.2.3 Virheenpaljastavat ja -korjaavat koodit

Parempiin koodeihin päästään koodaamalla yksittäisten bittien sijaan kokonaisia *bittilohkoja*. Yksinkertainen virheenpaljastava koodi saadaan lisäämällä lohkoon pariteetin tarkastusbitti.

Lohkon $s_1 \dots s_n$ *pariteetti* on

$$\sum_{i=1}^{n} s_i \mod 2$$

eli

$$\text{pariteetti} = \begin{cases} 0, & \text{jos ykkösien lukumäärä on parillinen} \\ 1, & \text{jos ykkösien lukumäärä on pariton.} \end{cases}$$

Esimerkki 1.2. Tarkastellaan seuraavia tapauksia:

001011	pariteetti 1	(pariton)
101000	pariteetti 0	(parillinen)

Koodaus tapahtuu seuraavasti:

s		**t**
001011	$\longrightarrow$	0010111
101000	$\longrightarrow$	1010000

Lopputuloksen pariteetti on aina 0. ‖

Nyt pystytään havaitsemaan, jos kanavassa on tapahtunut *pariton* määrä virheitä.

Esimerkki 1.3. Jos $\mathbf{r} = 01000$, tiedetään, että virhe tai virheitä on tapahtunut, mutta ei tiedetä *missä*. ‖

Hammingin koodi pystyy korjaamaan *yhden* virheellisen bitin. Hammingin $(7, 4)$-koodi on:

$$\begin{array}{ccccc} \mathbf{s} & & \mathbf{t}(\mathbf{s}) & & \mathbf{r} \\ s_1s_2s_3s_4 & \xrightarrow{\text{kooderi}} & t_1t_2t_3t_4t_5t_6t_7 & \xrightarrow{\text{kanava}} & r_1r_2r_3r_4r_5r_6r_7 \end{array}$$

Tässä

- $t_i = s_i$, kun $i = 1, 2, 3, 4$,
- t_5, t_6, t_7 asetetaan siten, että lohkoilla $s_1s_2s_3t_5$, $s_2s_3s_4t_6$ ja $s_1s_3s_4t_7$ on parillinen pariteetti.

Saadaan $2^4 = 16$ ”koodisanaa”, joiden pituus on seitsemän. Esimerkiksi $0010 \to 0010111$. Koodi on esitetty kuvassa 1.8.

s	t	s	t	s	t	s	t
0000	0000000	0100	0100110	1000	1000101	1100	1100011
0001	0001011	0101	0101101	1001	1001110	1101	1101000
0010	0010111	0110	0110001	1010	1010010	1110	1110100
0011	0011100	0111	0111010	1011	1011001	1111	1111111

Kuva 1.8: Hammingin (7,4)-koodi.

Tässä koodissa koodisanat eroavat vähintään kolmessa bitissä. Mikä mahtaa olla optimaalinen dekooderi? Tämän selvittämiseksi lasketaan $\mathbf{t}$:n ja $\mathbf{r}$:n *Hammingin etäisyys*,

$$d_H(\mathbf{t}, \mathbf{r}) = \sum_{i=1}^{7} |t_i - r_i| = |\{i \mid t_i \neq r_i\}|,$$

Missä $|\{\cdots\}|$ tarkoittaa joukon alkioiden lukumäärää. Kun kyseessä on binäärinen symmetrinen kanava, jolle $0 < f < 1/2$ ja kaikki viestit $\mathbf{s} \in \{0,1\}^4$ ovat yhtä todennäköisiä, optimaalinen dekooderi on valita sellainen $\hat{\mathbf{s}}$, että

$$d_H(\mathbf{t}(\hat{\mathbf{s}}), \mathbf{r}) = \min_{\mathbf{s} \in \{0,1\}^4} d_H(\mathbf{t}(\mathbf{s}), \mathbf{r}).$$

(Optimaalisuuden todistus on harjoitustehtävänä). Koodisanojen $\mathbf{t}(\mathbf{s})$ etäisyydet ≥ 3, joten yhden bitin virhe korjaantuu!

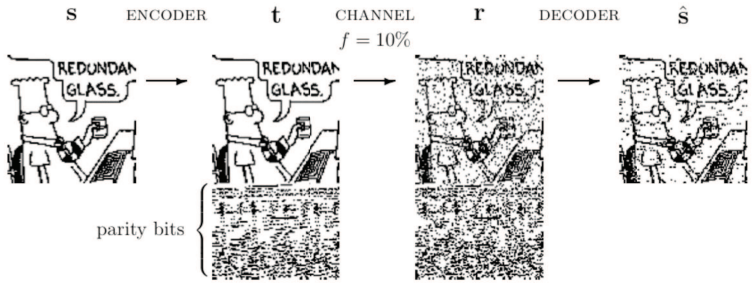
Käytännössä dekoodausta *ei* tarvitse tehdä minimoimalla Hammingin etäisyyttä, vaan laskennallisesti tehokkaampikin tapa löytyy (lineaarialgebra kunnassa $\mathbb{Z}_2$, koodausteoria,...).

Nyt

$$\mathbb{P}\{\text{"virhe"}\} = \mathbb{P}\{\hat{\mathbf{s}} \neq \mathbf{s}\}$$

ja bittivirheen todennäköisyys määritellään kaavalla

$$p_b = \frac{1}{4} \sum_{i=1}^{4} \mathbb{P}\{\hat{s}_i \neq s_i\},$$



Kuva 1.9: Hammingin (7,4)-koodin käyttö binäärisessä symmetrisessä kanavassa, jonka kohinataso on $f = 0.1$. Bittivirhe p_b on nyt noin 0.07 (esimerkki lähteestä [5]).

missä $\mathbf{s} = s_1 s_2 s_3 s_4$ ja $\hat{\mathbf{s}} = \hat{s}_1 \hat{s}_2 \hat{s}_3 \hat{s}_4$.

Kuvassa 1.9 on esimerkki Hammingin (7,4)-koodin käytöstä binäärisessä symmetrisessä kanavassa, jonka kohinataso on $f = 0.1$.

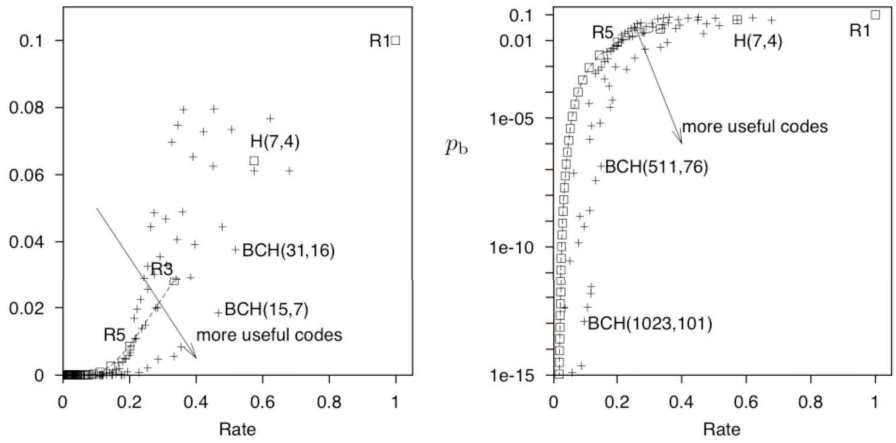
On helppo nähdä, että binäärisessä symmetrisessä kanavassa $\mathbb{P}\{\hat{\mathbf{s}} \neq \mathbf{s}\} = \mathcal{O}(f^2)$ eli samaa suuruusluokkaa kuin toistokoodissa R_3 (vrt. harjoitustehtävät). Mutta *nopeus* on nyt parempi:

$$R = \frac{4}{7} > \frac{1}{3}.$$

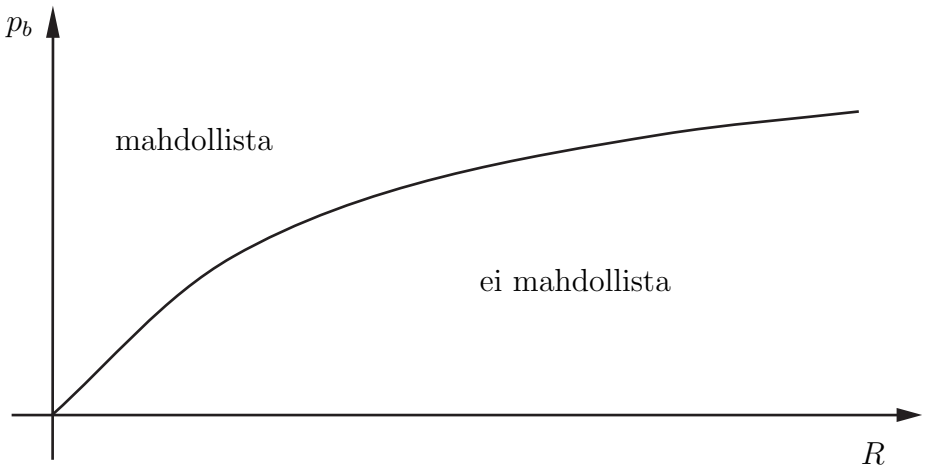
Kuvasssa 1.10 on vielä lisää esimerkkejä eri koodien suorituskyvystä.

Kuitenkin tiedonsiirron nopeus edelleen näyttää melko huonolta! Voidaankin kysyä, että mitkä (R, p_b)-yhdistelmät ovat ylipäänsä (edes periaatteessa) mahdollisia? Ennen vuotta 1948 uskottiin tilanteen olevan kuvan 1.11 kaltainen, eli virheetön tiedon siirto ei ole mahdollista. Shannon osoitti kuitenkin vuonna 1948 tilanteen olevankin itse asiassa kuvan 1.12 kaltainen. Tässä kuvassa C on kanavan *kapasiteetti*. Kun $R < C$, on siis mahdollista saavuttaa mielivaltaisen pieni bittivirhe p_b. Tilannetta on vielä havainnollisettu eräiden konkreettisten koodien osalta kuvassa 1.13.

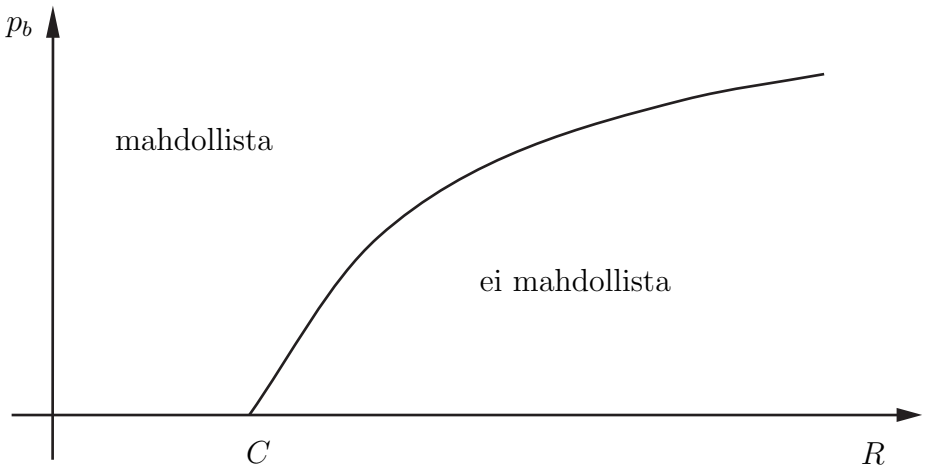
Shannonin keskeinen tulos vuodelta 1948 on *Informaatioteorian peruslause*. Tämä lause kertoo tiedonsiirron mahdollisuudet ($R < C$) ja rajat ($R > C$)



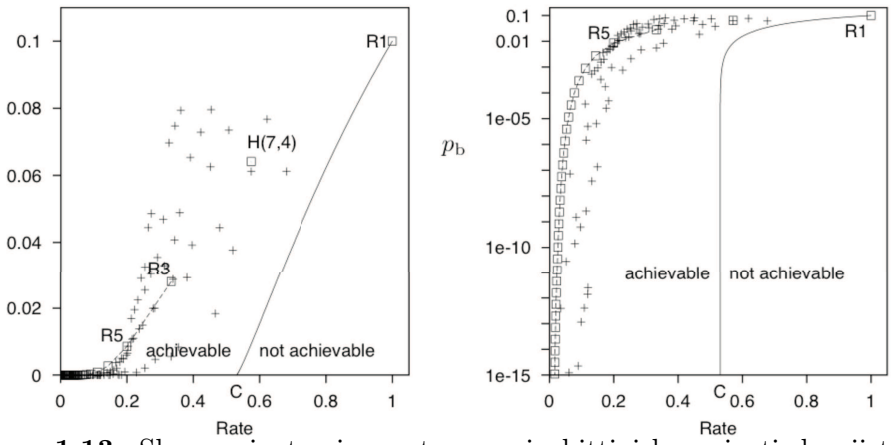
Kuva 1.10: Bittivirheen p_b riippuvuus tiedonsiirtonopeudesta R eräille toistokoodeille, Hammingin (7,4)-koodille ja BCH-koodeille (Bose-Chaudhuri-Hocquenhem). Kyseessä on binäärinen symmetrinen kanava kohinatasolla $f = 0.1$. Oikean puoleisessa paneelissa on logaritminen skaala (kuva lähteestä [5]).



Kuva 1.11: Käsitys bittivirheen p_b ja tiedosiirtonopeuden R riippuvuudesta ennen Shannonin teoriaa.



Kuva 1.12: Bittivirheen p_b ja tiedonsiirtonopeuden$R$ riippuvuus Shannonin teorian mukaan. Tässä C on kanavan kapasiteetti.



Kuva 1.13: Shannonin teorian antama raja bittivirheen ja tiedonsiirtonopeuden mahdollisille yhdistelmille (yhteinäinen käyrä) ja eräiden koodien suorituskyky binääriselle symmetriselle kanavalle kohinatasolla $f = 0.1$. Oikean puoleisessa paneelissa on logaritminen skaala (kuva lähteestä [5]).

ja se motivoi seuraavien vuosikymmenien koodausteorian kehitystä. Voidaan väittää, että informaatioteoria itse asiassa rakentuu tämän lauseen ja sen seurausten ympärille.

Luku 2

Informaatio ja sen mittaaminen

2.1 Tapahtuman sisältämä informaatio

Perusidea tapahtuman sisältämän informaatioon määrittelemisessä on, että epävarma tai odottamaton tapahtuma on informatiivinen. Tapahtuman epävarmuutta mitataan *poistuneella epävarmuudella*, kun tapahtuman tiedetään sattuneen. Kun epävarma tapahtuma sattuu, siihen liittynyt suuri epävarmuus poistuu ja näin on saatu paljon informaatiota. Jos taas melko varma tapahtuma sattuu, vain vähän epävarmuutta poistuu ja näin on saatu vain vähän informaatiota.

Esimerkki 2.1. Tarkastellaan 100 palloa, jotka on numeroitu 1,2,...,100. Pallot 1,...,10 ovat valkoisia ja pallot 11,...,100 ovat mustia. Nostetaan yksi pallo umpimähkään. Olkoon A =”valkoinen” ja B =”musta”, jolloin

$$\mathbb{P}(A) = \frac{1}{10} \quad \text{ja} \quad \mathbb{P}(B) = \frac{9}{10}.$$

Jos tapahtuma A sattuu, tiedetään, että kyseessä on pallo 1,...,10. Jos taas

tapahtuma B sattuu, tiedetään, että kyseessä on pallo 11,...,100.

Selvästi tapahtuma A vähentää epävarmuutta enemmän kuin tapahtuma B, eli tapahtuma A on informatiivisempi. Tapahtuman A jälkeen tiedetään siis enemmän kuin tapahtuman B jälkeen. Tapahtuma A on epävarmempi, sillä $\mathbb{P}(A) < \mathbb{P}(B)$. ‖

Miten mitata jonkun tapahtuman epävarmuutta tai informatiivisuutta täsmällisesti? Epävarmuus selvästi liittyy tapahtuman todennäköisyyteen. Olkoon siis A tapahtuma ja $\mathbb{P}(A) = p > 0$. Pyritään määrittelemään sellainen funktio h, että

$$h(p) = \text{"tapahtuman } A \text{ epävarmuus, informaatiosisältö"}.$$

Olkoon $A \perp\!\!\!\perp B$ (riippumattomat), $\mathbb{P}(A) = p_1$ ja $\mathbb{P}(B) = p_2$. Silloin

$$\mathbb{P}\{\text{"}A \text{ ja } B\text{"}\} = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = p_1 p_2,$$

joten leikkauksen "A ja B" epävarmuus on $h(p_1p_2)$. Luonteva vaatimus tällöin on, että

$$h(p_1p_2) - h(p_1) = h(p_2).$$

Toinen luonteva vaatimus on, että $p \mapsto h(p)$ on aidosti vähenevä ja jatkuva.

Lause 2.1. *Olkoon* $h :]0,1] \to \mathbb{R}$ *ja*

(*i*) $h(p_1p_2) = h(p_1) + h(p_2)$, $p_1, p_2 \in]0,1]$,

(*ii*) h *on aidosti vähenevä ja jatkuva.*

Silloin $h(p) = -C \log_b p$, *missä* $b > 1$ *ja* $C > 0$ *riippuu vakiosta* b.

Huomautus. $p \mapsto -C \log_b p$ selvästi toteuttaa ehdot (*i*) ja (*ii*).

Todistus. Olkoon

$$g(n) = h\Big(\frac{1}{n}\Big), \quad n \in \mathbb{N}_+.$$

Ehdon (i) nojalla

$$h\left(\frac{1}{nm}\right) = h\left(\frac{1}{n} \cdot \frac{1}{m}\right) = h\left(\frac{1}{n}\right) + h\left(\frac{1}{m}\right)$$

eli

$$g(nm) = g(n) + g(m), \quad n, m \in \mathbb{N}_+. \tag{2.1}$$

Oletetaan, että $n < m$. Ehdon (ii) nojalla saadaan

$$g(n) < g(m), \quad n, m \in \mathbb{N}_+.$$

Osoitetaan, että

$$g(n) = C \log_b n, \tag{2.2}$$

jollain $C > 0$ ja $b > 1$.

Osoitetaan ensin induktiolla, että

$$g(n^k) = k\, g(n), \quad n, k \in \mathbb{N}_+. \tag{2.3}$$

Väite on selvä, kun $k = 1$. Oletetaan, että väite pätee arvolla k. Silloin

$$\begin{aligned} g(n^{k+1}) = g(n \cdot n^k) &\overset{(2.1)}{=} g(n) + g(n^k) \\ &\overset{\text{ind.ol}}{=} g(n) + kg(n) = (k+1)g(n). \end{aligned}$$

Edelleen,

$$g(1) = g(1 \cdot 1) = g(1) + g(1),$$

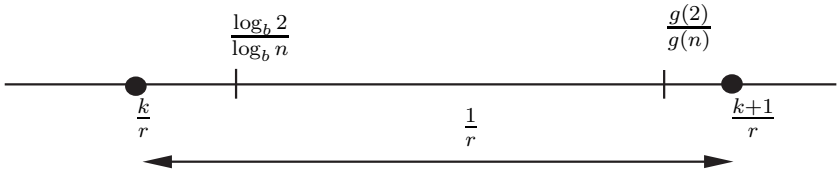
joten

$$g(1) = 0. \tag{2.4}$$

Olkoon $n \in \mathbb{N}$, $n > 1$, kiinteä ja $r \in \mathbb{N}_+$. Valitaan (ks. kuva 2.1) sellainen $k = k(r) \in \mathbb{N}$, että

$$n^k \leq 2^r < n^{k+1}. \tag{2.5}$$

Kuva 2.1: Indeksin k valinta lauseen 2.1 todistuksessa.



Kuva 2.2: Lauseen 2.1 todistuksen havainnollistus.

Nyt g on aidosti kasvava, joten

$$g(n^k) \leq g(2^r) < g(n^{k+1}).$$

Tuloksen (2.3) nojalla saadaan

$$kg(n) \leq rg(2) < (k+1)g(n),$$

eli

$$\frac{k}{r} \leq \frac{g(2)}{g(n)} < \frac{k+1}{r}. \tag{2.6}$$

Huomaa, että g on aidosti kasvava, joten $g(n) > g(1) = 0$.

Edelleen $b > 1$, joten $\log_b$ on aidosti kasvava. Kaavasta (2.5) saadaan siten

$$k \log_b n \leq r \log_b 2 < (k+1) \log_b n,$$

josta edelleen

$$\frac{k}{r} \leq \frac{\log_b 2}{\log_b n} < \frac{k+1}{r}.$$

Huomioidaan tulos (2.6), jolloin (ks. kuva 2.2)

$$\left| \frac{\log_b 2}{\log_b n} - \frac{g(2)}{g(n)} \right| < \frac{1}{r}.$$

Luku $r \in \mathbb{N}_+$ on mielivaltainen, joten

$$\frac{\log_b 2}{\log_b n} = \frac{g(2)}{g(n)},$$

eli

$$g(n) = \frac{g(2)}{\log_b 2} \cdot \log_b n,$$

mikä pätee myös, kun $n = 1$. Siten ehdossa (2.2) voidaan ottaa $C = \frac{g(2)}{\log_b 2}$.

Olkoon sitten $p = \frac{r}{s} \in \mathbb{Q} \cap]0, 1]$, $r, s > 0$. Nyt

$$h\left(\frac{1}{s}\right) = h\left(\frac{r}{s} \cdot \frac{1}{r}\right) \overset{(i)}{=} h\left(\frac{r}{s}\right) + h\left(\frac{1}{r}\right),$$

josta edelleen saadaan

$$\begin{aligned} h\left(\frac{r}{s}\right) &= h\left(\frac{1}{s}\right) - h\left(\frac{1}{r}\right) = g(s) - g(r) \\ &= C \log_b s - C \log_b r = -C \log_b \frac{r}{s}. \end{aligned} \tag{2.7}$$

Lauseen väite pätee siis rationaalisilla p. Lauseen väite mielevaltaiselle $p \in]0, 1]$ seuraa nyt funktioiden h ja $\log_b$ jatkuvuudesta: kun $p_k \to p$, $p_k \in]0, 1] \cap \mathbb{Q}$, saadaan

$$h(p) = \lim_{k \to \infty} h(p_k) \overset{(2.7)}{=} \lim_{k \to \infty} [-C \log_b p_k] = -C \log_b p.$$

□

Jatkossa otetaan $b = 2$ ja merkitään $\log_2 = \log$. Tämä valinta vaikuttaa vain vakioon C, koska jos $a, b > 1$, niin

$$\log_a p = \log_a b \log_b p.$$

Otamme myös jatkossa $C = 1$, mikä vaikuttaa vain mitta-asteikkoon. Kun $p = \frac{1}{2}$, niin $h(p) = -C \log \frac{1}{2} = C \log 2 = C$. Näin valinta $C = 1$ tarkoittaa, että symmetrisen lantin heiton antama informaatio on ”1” yksikköä. Näin tapahtuman A, $\mathbb{P}(A) = p > 0$, *epävarmuus* tai *informaatiosisältö* määritellään kaavalla

$$h(p) = -\log p.$$

Epävarmuuden tai informaatiosisällön yksikkö on *bitti*.

2.2 Satunnaismuuttujat ja informaatio

Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus. Siis,

- Ω on alkeistapausten joukko eli perusjoukko
- $\mathcal{F}$ on tapahtumien joukko (Ω:n osajoukkojen σ-algebra)
- $\mathbb{P}$ on todennäköisyys eli $\mathbb{P}$ on kuvaus $\mathcal{F} \to [0, 1]$

Esimerkki 2.2. Tarkastellaan nopan heittoa. Alkeistapausten joukko on nyt $\Omega = \{1, 2, 3, 4, 5, 6\}$ ja tapausten joukkona $\mathcal{F}$ on Ω:n kaikki osajoukot. Kun $A \subset \Omega$, määritellään

$$\mathbb{P}(A) = \frac{|A|}{6},$$

missä $|A|$ =joukon A alkioiden lukumäärä. ‖

Jatkossa käsitellään satunnaismuuttujia (sm), joiden arvojoukko on äärellinen, eli satunnaismuuttujat voivat saada vain äärellisen monta eri arvoa. Tällainen satunnaismuuttuja on kuvaus

$$X : \Omega \to \mathcal{X},$$

missä $\mathcal{X}$ on äärellinen joukko ja X:lle pätee

$$\{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\} \in \mathcal{F}$$

kaikilla $x \in \mathcal{X}$. Merkitään

$$p(x) = \mathbb{P}\{X = x\}, \quad x \in \mathcal{X},$$

missä $p(x)$:t ovat satunnaismuuttujan X *pistetodennäköisyyksiä*. Merkitsemme tavallisesti myös $p(x)$:llä itse *pistetodennäköisyysfunktiota* (ptnf) $p : \mathcal{X} \to [0, 1]$. Myös merkintää $X \sim p(x)$ käytetään toisinaan.

Edelleen, jos $Y : \Omega \to \mathcal{Y}$ on toinen satunnaismuuttja, merkitään tavallisesti $p(y)$:llä satunnaismuuttujan Y pistetodennäköisyysfunktiota. Tässä hieman huolimattomassa merkintätavassa siis vain argumentin nimi (x tai y) kertoo sen, että kyseessä on yleensä eri funktiot $p(x)$ ja $p(y)$.

$\mathcal{X}$ voi periaatteessa olla mikä äärellinen joukko hyvänsä:

$$\{0, 1\},\ \{a, b, c, d\},\ \{\bigcirc, \triangle, \square\}.$$

Toisaalta, nimeämällä alkiot uudestaan, voitaisiin yhtä hyvin olettaa, että $\mathcal{X} = \{1, \ldots, m\}$, jos $|\mathcal{X}| = m$.

Kuvassa 2.3 on erään Linux-oppaan perusteella laadittu taulukko englannin kielen kirjainten esiintymistodennäköisyyksistä. Nämä ovat siis sellaisen satunnaismuuttujan arvojen todennäköisyydet, joka kuvaa umpimähkään valittua kirjainta kyseisestä oppaasta.

Tapahtuman $\{X = x\}$ epävarmuus tai informaatiosisältö on edellisen luvun mukaan

$$-\log\big(\mathbb{P}\{X = x\}\big) = -\log p(x).$$

Määritelmä 2.2. Satunnaismuuttujan X *entropia* on

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

Huomautus. Sovitaan, että $0 \log 0 = 0$ (koska $\lim\limits_{t \to 0^+} t \log t = 0$).

Huomautus. $H(X)$ on itseasiassa *odotusarvo*

$$H(X) = -\mathbb{E} \log p(X) = \mathbb{E}\big(-\log p(X)\big).$$

Tässä $\log p(X)$ on satunnaismuuttuja

$$\omega \mapsto \log p\big(X(\omega)\big) = \log\big(\mathbb{P}\{X = X(\omega)\}\big).$$

Siten $H(X)$ on satunnaismuuttujan X *arvojen keskimääräinen epävarmuus tai informaatiosisältö.*

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928

Kuva 2.3: Eräs arvio englannin kielen kirjainten esiintymistodennäköisyyksistä. Oikean puoleinen sarake havainnollistaa todennäköisyyksiä vielä graafisesti (esimerkki lähteestä [5])

Huomautus. Vain todennäköisyydet $p(x)$ ovat tässä tärkeitä ja satunnaismuuttujan X varsinaiset *arvot* ovat täysin epäoleellisia.

Huomautus. Vaikka funktion $h(p)$ ja sitä kautta entropian $H(X)$ määritelmä pyrittiin perustelemaan intuitiivisesti, on asetettujen määritelmien *todellinen* motivaatio se, että ne johtavat hyvään ja hyödylliseen tiedonsiirron teoriaan, jota voi menestyksellä soveltaa mm. koodien konstruktioon.

Lause 2.3. *$H(X) \geq 0$ ja $H(X) = 0$ jos ja vain jos X on vakio (todennäköisyydellä 1).*

Todistus. Kaikilla $x \in \mathcal{X}$ on $0 \leq p(x) \leq 1$, joten $p(x) \log p(x) \leq 0$. Siten

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \geq 0.$$

Edelleen, jos $H(X) = 0$ on $p(x) \log p(x) = 0$ kaikilla $x \in \mathcal{X}$, eli $p(x) = 0$ tai 1 kaikilla $x \in \mathcal{X}$. Mutta $\sum_{x \in \mathcal{X}} p(x) = 1$, joten tällöin $p(x) = 1$ täsmälleen yhdellä $x \in \mathcal{X}$, jolle siis pätee $p(x) = \mathbb{P}\{X = x\} = 1$. Kääntäen, jos X on vakio (todennäköisyydellä 1), on yksi luvuista $p(x)$ arvoltaan 1 ja muut 0, jolloin $H(X) = 0$. □

Siis: Satunnaismuuttujassa X ei ole epävarmuutta $\Leftrightarrow H(X) = 0 \Leftrightarrow X$ on vakio.

Esimerkki 2.3. Olkoon $\mathcal{X} = \{0, 1\}$, $0 \leq p \leq 1$,

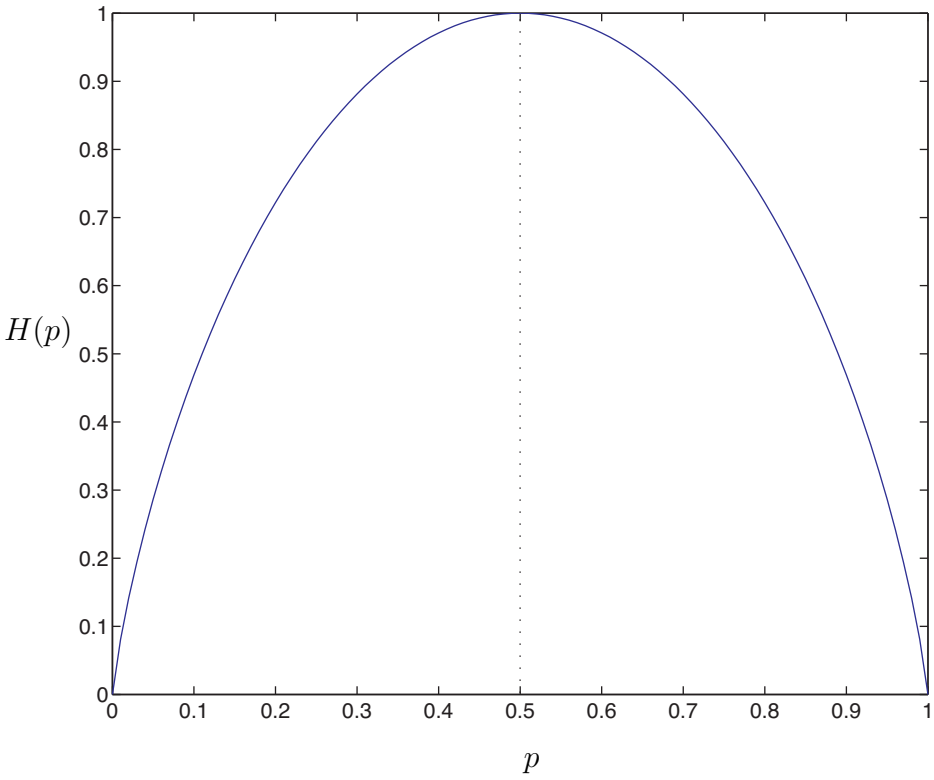
$$X = \begin{cases} 1, & \text{todennäköisyydellä } p \\ 0, & \text{todennäköisyydellä } 1 - p. \end{cases}$$

Silloin

$$H(X) = -p \log p - (1 - p) \log(1 - p) \equiv H(p).$$

Kuvassa 2.4 on esitetty tämän funktion kuvaaja.

Havaitaan, että kun $p = 0$ tai $p = 1$, ei satunnaismuuttujassa X ole lainkaan epävarmuutta: $H(0) = H(1) = 0$. Tällöin X on vakio (todennäköisyydellä



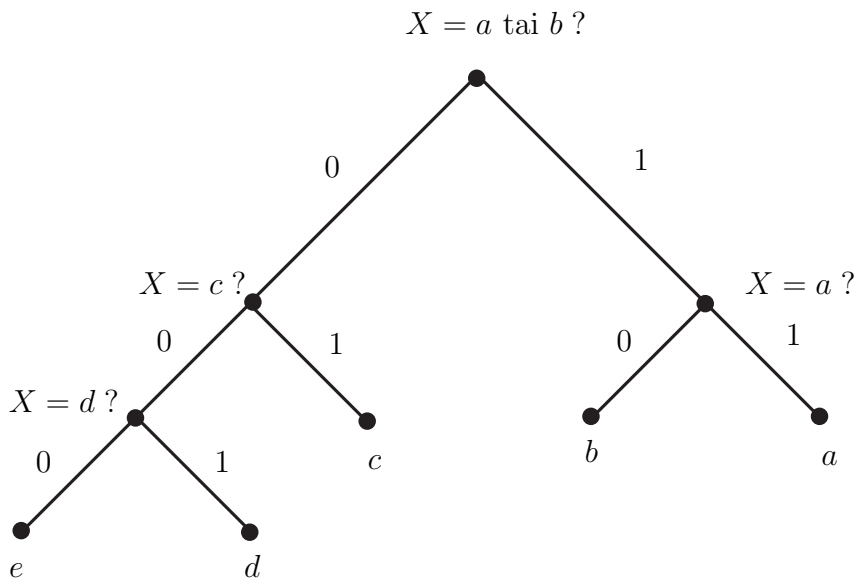
Kuva 2.4: Funktio $H(p)$.

1). Suurin epävarmuus saadaan arvolla $p = 1/2$, jolloin $H(1/2) = 1$. Tämä vastaa symmetrisen lantin heittoa. Saatu informaatio heiton tuloksesta on 1 bitti. ‖

Jos $|\mathcal{X}| = m$ ja $p_1, \ldots, p_m$ ovat arvojen $x \in \mathcal{X}$ todennäköisyydet, merkitään jatkossa joskus myös

$$H(X) = -\sum_{i=1}^{m} p_i \log p_i \equiv H(p_1, \ldots, p_m).$$

Entropian voi ajatella liittyvän myös satunnaismuuttujan X arvon määräämiseen ”binäärisillä” ei/kyllä vastauksilla.



Kuva 2.5: Satunnaismuuttujaa X vastaava binääripuu.

Esimerkki 2.4. X saa arvot a, b, c, d ja e todennäköisyyksillä 0.3, 0.2, 0.2, 0.15 ja 0.15. Kuvassa 2.5 on X:ää vastaava binääripuu, missä ei = 0 ja kyllä = 1. Keskimääräinen kysymysten lukumäärä X:n arvo selvittämiseksi on

$$0.3 \cdot 2 + 0.2 \cdot 2 + 0.2 \cdot 2 + 0.15 \cdot 3 + 0.15 \cdot 3 = 2.3.$$

Binääripuusta saadaan koodaus

$$\begin{aligned} a &\longrightarrow 11 \\ b &\longrightarrow 10 \\ c &\longrightarrow 01 \\ d &\longrightarrow 001 \\ e &\longrightarrow 000 \end{aligned}$$

Keskimääräinen koodin pituus $L = 2.3$ bittiä, sama kuin keskimääräinen kysymysten lukumäärä. Toisaalta,

$$\begin{aligned} H(X) &= -0.3 \log 0.3 - 0.2 \log 0.2 - 0.2 \log 0.2 - 0.15 \log 0.15 - 0.15 \log 0.15 \\ &\approx 2.27. \end{aligned}$$

$\|$

Ei ole itse asiassa sattumaa, että $L = H(X) + \varepsilon$, missä $\varepsilon > 0$. Myöhemmin tullaan osoittamaan, että tietyn tyyppisten binääristen koodien joukossa keskimäärin lyhimmälle koodille pätee

$$H(X) \leq L < H(X) + 1.$$

Edelleen, koodaamalla *jonoja* $(x_1, \ldots, x_n)$, $x_i \in \{a, b, c, d, e\}$ yksittäisten alkioiden sijaan saadaan tietyissä tilanteissa keskimäärin lyhimmälle koodille L, että

$$H(X) \leq \frac{L}{n} < H(X) + \frac{1}{n},$$

eli optimikoodin keskimääräinen pituus *per symboli* $\approx H(X)$.

Näin olemme saaneet entropialle toisen tulkinnan:

$$H(X) = \text{keskimäärin pienin binääristen kysymysten lukumäärä}$$
$$\text{satunnaismuuttujan } X \text{ arvon selvittämiseksi.}$$

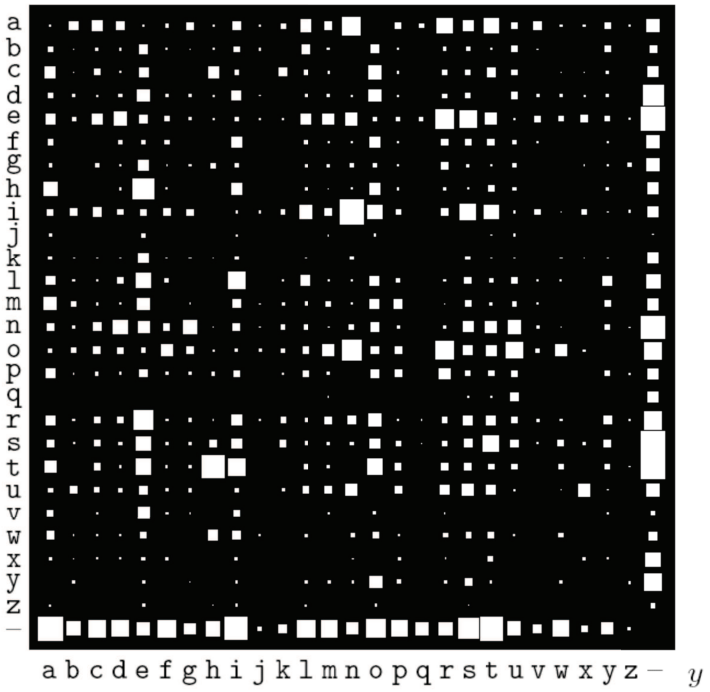
Tarkastellaan sitten satunnaismuuttajaparia (X, Y). Satunnaismuuttujan X arvojoukko on $\mathcal{X}$ ja satunnaismuuttujan Y arvojoukko on $\mathcal{Y}$. Parin (X, Y) arvojoukko on siten $\mathcal{X} \times \mathcal{Y}$ (myös äärellinen). Pistetodennäköisyydet ovat $p(x, y) = \mathbb{P}\{X = x \text{ ja } Y = y\}$ ja merkitsemme $(X, Y) \sim p(x, y)$.

Kuvassa 2.6 on samasta tekstiaineistosta kuin kuvassa 2.3 lasketut kirjainparien pistetodennäköisyydet graafisesti havainnollistettuna.

Määritelmä 2.4. Parin (X, Y) *yhteisentropia* on satunnaismuuttujan (X, Y) entropia,

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y).$$

Huomautus. Siis $H(X, Y) = -\mathbb{E} \log p(X, Y)$.



Kuva 2.6: Arvio englannin kielen kirjainten kirjainparien esiintymistodennäköisyyksistä kuvan 2.3 esimerkissä.

Määritelmä 2.5. Satunnaismuuttujan Y *entropia ehdolla* $X = x$ on

$$H(Y \mid X = x) = -\sum_{y \in \mathcal{Y}} p(y \mid x) \log p(y \mid x).$$

Tässä käytetään siis satunnaismuuttujan Y jakaumaa ehdolla $X = x$,

$$p(y \mid x) = \frac{p(x, y)}{p(x)},$$

kun $p(x) > 0$. Sovitaan, että $p(y \mid x) = 0$, kun $p(x) = 0$. $H(Y \mid X = x)$ kuvaa satunnaismuutujan Y epävarmuutta, kun $X = x$.

Määritelmä 2.6. Satunnaismuuttujan Y *entropia ehdolla* X on

$$H(Y \mid X) = \sum_{x \in \mathcal{X}} p(x)\, H(Y \mid X = x).$$

Huomautus. Siis $H(Y \mid X)$ on $H(Y \mid X = x)$:n *keskimääräinen* arvo, kun $x \in \mathcal{X}$. Tämä voidaan tietysti taas tulkita odotusarvona.

Myöhemmin osoitetaan, että $H(Y \mid X) \leq H(Y)$. Kuitenkin joillain x *voi* olla $H(Y \mid X = x) > H(Y)$.

Esimerkki 2.5. Olkoon $p(x, y)$ muotoa

		Y	
		1	2
X	1	0	3/4
	2	1/8	1/8

Esimerkiksi siis $p(1, 2) = 3/4$. Nyt

$$p(y) = \sum_{x \in \mathcal{X}} p(x, y) = \begin{cases} 1/8, & y = 1, \\ 7/8, & y = 2. \end{cases}$$

Siten

$$H(Y) = H\left(\frac{1}{8}, \frac{7}{8}\right) = -\frac{1}{8}\log\frac{1}{8} - \frac{7}{8}\log\frac{7}{8} \approx 0.544.$$

Edelleen,

$$p(x) = \sum_{y\in\mathcal{Y}} p(x,y) = \begin{cases} 3/4, & x = 1, \\ 1/4, & x = 2. \end{cases}$$

Nyt helposti saadaan, että

$$H(Y \mid X = 1) = -p(1 \mid 1)\log p(1 \mid 1) - p(2 \mid 1)\log p(2 \mid 1) = 0.$$

ja

$$H(Y \mid X = 2) = -p(1 \mid 2)\log p(1 \mid 2) - p(2 \mid 2)\log p(2 \mid 2) = 1.$$

Siten

$$H(Y \mid X) = p(1)H(Y \mid X = 1) + p(2)H(Y \mid X = 2) = \frac{3}{4}\cdot 0 + \frac{1}{4}\cdot 1 = 0.25.$$

Näin $H(Y \mid X) < H(Y)$, mutta $H(Y \mid X = 2) > H(Y)$. ‖

Lause 2.7 (Ketjusääntö)**.** *Pätee*

$$H(X,Y) = H(X) + H(Y \mid X).$$

Todistus.

$$\begin{aligned} H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log[p(x)p(y \mid x)] \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y \mid x) \\ &= -\sum_{x\in\mathcal{X}}\Big[\sum_{y\in\mathcal{Y}} p(x,y)\Big]\log p(x) - \sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(y \mid x)\log p(y \mid x) \\ &= H(X) + H(Y \mid X). \end{aligned}$$

□

Huomautus. Tässä tuli osoitettua myös kaava

$$H(Y \mid X) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y \mid x). \tag{2.8}$$

Edelleen,

$$p(x,y) = \mathbb{P}\{X = x \text{ ja } Y = y\} = \mathbb{P}\{Y = y \text{ ja } X = x\} = p(y,x),$$

joten

$$H(X,Y) = H(Y,X).$$

Siten ketjusääntö voidaan myös kirjoittaa muodossa

$$H(X,Y) = H(Y) + H(X \mid Y).$$

Lause 2.8. *Olkoon* $0 < p_i, q_i \leq 1$*,* $i = 1, \ldots, m$*,* $\sum_{i=1}^{m} p_i = \sum_{i=1}^{m} q_i = 1$*. Silloin*

$$-\sum_{i=1}^{m} p_i \log p_i \leq -\sum_{i=1}^{m} p_i \log q_i \tag{2.9}$$

ja yhtäsuuruus pätee jos ja vain jos $p_i = q_i$ *kaikilla* i*.*

Todistus. Käytetään todistuksessa poikkeuksellisesti luonnollista logaritmia ln; koska $\log_2 x = \log_2 e \ln x$, $\log_2 e > 0$, ei tämä vaikuta väitteeseen (2.9).

Tunnetusti $\ln x \leq x - 1$ kaikilla $x > 0$ ja yhtäsuuruus on voimassa jos ja vain jos $x = 1$ (kuva 2.7). Siten

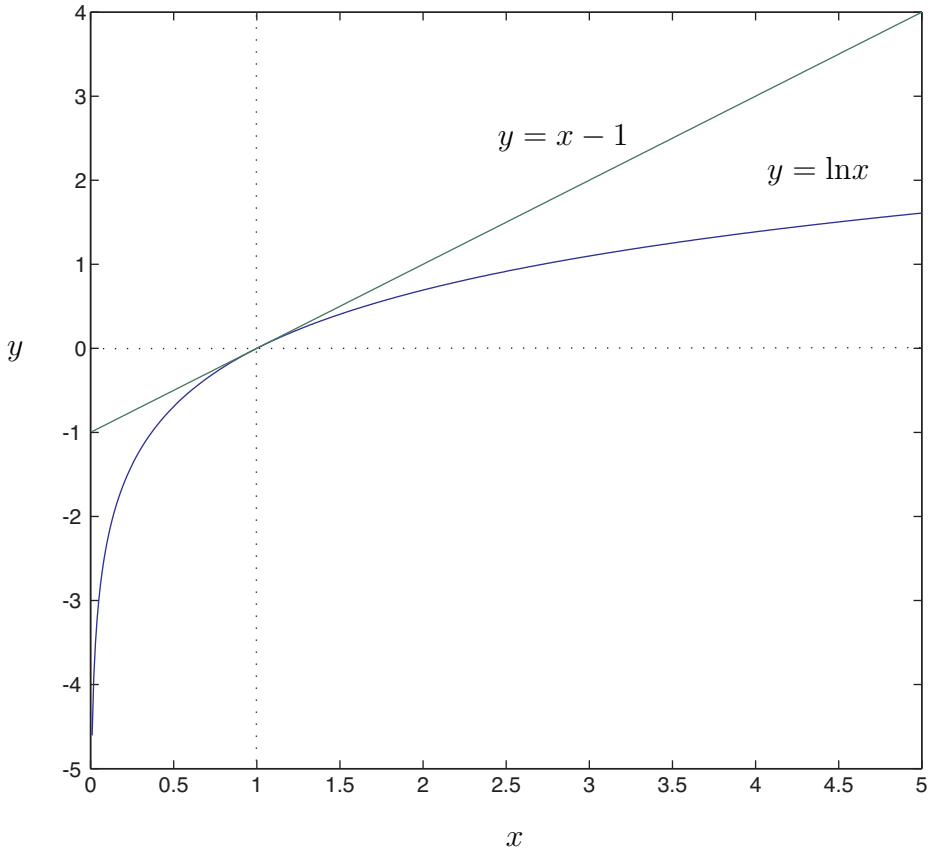
$$\ln \frac{q_i}{p_i} \leq \frac{q_i}{p_i} - 1$$

ja yhtäsuuruus pätee jos ja vain jos $p_i = q_i$ $(i = 1, \ldots, m)$. Edelleen,

$$\sum_{i=1}^{m} p_i \ln \frac{q_i}{p_i} \leq \sum_{i=1}^{m} p_i \left(\frac{q_i}{p_i} - 1 \right) = \sum_{i=1}^{m} q_i - \sum_{i=1}^{m} p_i = 0$$

ja yhtäsuuruus pätee jos ja vain jos $p_i = q_i$ kaikilla i. Väite seuraa nyt helposti soveltamalla vasemmalla puolella logaritmin laskusääntöjä. □

Huomautus. Sovitaan "$0 \log 0 = 0$":n lisäksi, että "$a \log 0 = -\infty$", kun $a > 0$. Silloin lause pätee myös, kun $p_i = 0$ tai $q_i = 0$ joillain i.



Kuva 2.7: Epäyhtälön $\ln x \leq x - 1$ $(x > 0)$ havainnollistaminen. Suora $y = x - 1$ on tangentti pisteessä $x = 1$. Käyrä $y = \ln x$ on alaspäin kupera, siis aina tangentin alapuolella.

Huomautus. Luvut p_i, q_i voivat erityisesti olla pistetodennäköisyyksiä $p(x)$, $q(x)$, $x \in \mathcal{X}$. Silloin kaavan (2.9) mukaan

$$\sum_{x\in\mathcal{X}} p(x)\log p(x) - \sum_{x\in\mathcal{X}} p(x)\log q(x) \geq 0$$

eli

$$\sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \geq 0. \tag{2.10}$$

Kun $p(x) = 0$ tai $q(x) = 0$, otetaan tässä

$$p(x)\log\frac{p(x)}{q(x)} = p(x)\log p(x) - p(x)\log q(x)$$

ja sovelletaan aikaisemmin sovittuja sääntöjä.

Määritelmä 2.9. Olkoot $p(x)$ ja $q(x)$ pistetodennäköisyysfunktioita joukossa $\mathcal{X}$. Silloin $p(x)$:n ja $q(x)$:n *Kullbackin-Leiblerin etäisyys* on

$$D(p \parallel q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}.$$

Myös nimitystä *suhteellinen entropia* käytetään.

Kaavan (2.10) ja lauseen 2.8 nojalla saadaan

Lause 2.10 (Informaatioepäyhtälö, Gibbsin epäyhtälö)**.** *Olkoot $p(x)$ ja $q(x)$ pistetodennäköisyysfunktioita joukossa $\mathcal{X}$. Silloin*

$$D(p \parallel q) \geq 0$$

ja yhtäsuuruus pätee jos ja vain jos $p(x) = q(x)$ kaikilla $x \in \mathcal{X}$.

Huomautus. $D(p \parallel q)$ *ei* ole ”kunnon etäisyys”, esimerkiksi metriikka, mutta se kuitenkin mittaa pistetiheysfunktioiden $p(x)$ ja $q(x)$ välistä erilaisuutta tietyllä (jatkossa hyödyllisellä) tavalla.

Lause 2.11. *Ehdollistaminen vähentää entropiaa, eli*

$$H(Y \mid X) \leq H(Y)$$

ja yhtäsuuruus on voimassa jos ja vain jos $X \perp\!\!\!\perp Y$.

Todistus. Sovelletaan lausetta 2.10 pistetodennäköisyysfunktioihin $p(x, y)$ ja $p(x)p(y)$ joukossa $\mathcal{X} \times \mathcal{Y}$, jolloin

$$\begin{aligned}
0 \leq D\big(p(x,y) \parallel p(x)p(y)\big) &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \left[\log \frac{p(x,y)}{p(x)} - \log p(y)\right] \\
&= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log p(y) + \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log p(y \mid x) \\
&= H(Y) - H(Y \mid X).
\end{aligned}$$

Yhtälö on voimassa jos ja vain jos $p(x, y) = p(x)p(y)$ eli $X \perp\!\!\!\perp Y$. □

Millaisella satunnaismuuttujalla on suurin mahdollinen entropia?

Lause 2.12. *$H(X) \leq \log |\mathcal{X}|$ ja yhtäsuuruus pätee jos ja vain jos $p(x) = 1/|\mathcal{X}|$ kaikilla $x \in \mathcal{X}$.*

Todistus. Otetaan lauseessa 2.8 $p_i = p(x)$, $q_i = 1/|\mathcal{X}|$, $m = |\mathcal{X}|$. Silloin tuloksen (2.9) nojalla saadaan

$$\begin{aligned}
H(X) &= -\sum_{x\in\mathcal{X}} p(x) \log p(x) \\
&\leq -\sum_{x\in\mathcal{X}} p(x) \log \frac{1}{|\mathcal{X}|} \\
&= \sum_{x\in\mathcal{X}} p(x) \log |\mathcal{X}| = \log |\mathcal{X}|.
\end{aligned}$$

Edelleen, lauseen 2.8 nojalla yhtäsuuruus on voimassa jos ja vain jos $p(x) = 1/|\mathcal{X}|$ kaikilla $x \in \mathcal{X}$. □

Huomautus. $p(x) = 1/|\mathcal{X}|$ on *tasainen jakauma* joukossa $\mathcal{X}$; kaikki arvot x ovat yhtä todennäköisiä.

2.3 Keskinäisinformaatio

Miten paljon informaatiota satunnaismuuttuja Y antaa satunnaismuuttujasta X?

Määritelmä 2.13. Satunnaismuuttujien X ja Y *keskinäisinformaatio* on

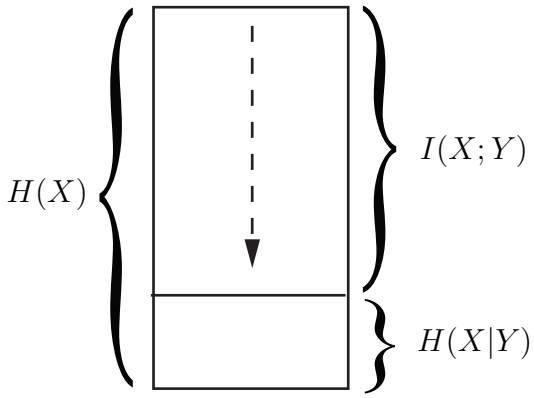
$$I(X;Y) = H(X) - H(X \mid Y).$$

Huomautus. Lauseen 2.11 mukaan $I(X;Y) \geq 0$ ja $I(X;Y) = 0$ jos ja vain jos $X \perp\!\!\!\perp Y$. Suure $I(X;Y)$ siis mittaa jonkinlaista riippuvuutta.

Entropialle saadaan nyt esitys

$$H(X) = I(X;Y) + H(X \mid Y).$$

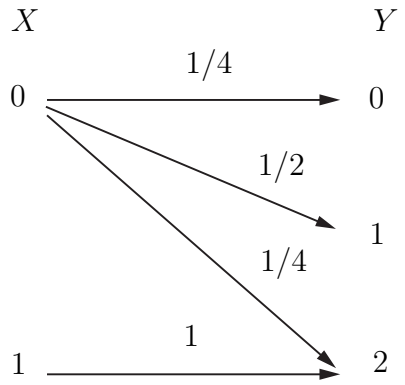
Keskinäisinformaatio kuvaa epävarmuuden vähentymistä X:stä, kun Y tunnetaan (kuva 2.8).



Kuva 2.8: Y:n tunteminen vähentää X:n epävarmuutta keskinäisinformaation $I(X;Y)$ verran.

Esimerkki 2.6. Tarkastellaan kahta lanttia, toinen symmetrinen (0) ja toinen, jonka kummallakin puolella on kruuna (1):

$$\begin{cases} 0: & \text{symmetrinen, } \mathbb{P}\{\text{kruuna}\} = \mathbb{P}\{\text{klaava}\} = 1/2 \\ 1: & \text{kaksi kruunua} \end{cases}$$



Kuva 2.9: Kahden erilaisen lantin heitto esimerkissä 2.6.

Toimitaan nyt seuraavasti.

1. Valitaan lantti X umpimähkään, $X \in \{0, 1\}$.
2. Heitetään sitä kaksi kertaa.
3. Lasketaan kruunujen lukumäärä $Y \in \{0, 1, 2\}$.

Miten paljon nyt Y paljastaa X:stä eli vähentää X:n epävarmuutta? Tilanne on esitetty kuvassa 2.9.

Nyt $H(X) = H(1/2) = 1$ (bitti). Edelleen,

$$\mathbb{P}\{X = 0\} = \mathbb{P}\{X = 1\} = 1/2,$$

$$\begin{aligned} \mathbb{P}\{Y = 0\} &= 1/8, \quad \mathbb{P}\{X = 0 \mid Y = 0\} = 1, \\ \mathbb{P}\{Y = 1\} &= 1/4, \quad \mathbb{P}\{X = 0 \mid Y = 1\} = 1, \\ \mathbb{P}\{Y = 2\} &= 5/8, \quad \mathbb{P}\{X = 0 \mid Y = 2\} = 1/5. \end{aligned}$$

Tässä viimeinen todennäköisyys laskettiin Bayesin kaavasta,

$$\mathbb{P}\{X = 0 \mid Y = 2\} = \frac{\mathbb{P}\{X = 0\}\,\mathbb{P}\{Y = 2 \mid X = 0\}}{\mathbb{P}\{Y = 2\}} = \frac{\frac{1}{2} \cdot \frac{1}{4}}{\frac{5}{8}} = \frac{1}{5}.$$

Näin ollen

$$\begin{aligned} H(X \mid Y) &= \mathbb{P}\{Y=0\}\, H(X \mid Y=0) + \mathbb{P}\{Y=1\}\, H(X \mid Y=1\} \\ &\quad + \mathbb{P}\{Y=2\}\, H(X \mid Y=2) \\ &= \frac{1}{8}\cdot 0 + \frac{1}{4}\cdot 0 + \frac{5}{8}\cdot\left(-\frac{1}{5}\log\frac{1}{5} - \frac{4}{5}\log\frac{4}{5}\right) \approx 0.45. \end{aligned}$$

Siten $I(X;Y) \approx 1 - 0.45 = 0.55$ (bittiä). ‖

Yleisesti $H(X,Y) = H(Y,X)$. Siten lauseen 2.7 nojalla

$$H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$$

ja

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) = I(Y;X).$$

Siten $I(X;Y) = I(Y;X)$ eli X ja Y kertovat toisistaan *yhtä paljon.*

Huomautus. Joskus laskuissa on helpompi laskea $I(Y;X)$ kuin $I(X;Y)$, koska usein $p(y \mid x)$ on tiedossa ja $p(x \mid y)$ pitäisi erikseen johtaa ($H(Y \mid X)$ on siis helpompi laskea kuin $H(X \mid Y)$).

Lauseen 2.11 todistuksessa havaittiin, että

$$D\big(p(x,y) \parallel p(x)p(y)\big) = H(Y) - H(Y \mid X).$$

Siten

$$I(X;Y) = D\big(p(x,y) \parallel p(x)p(y)\big).$$

Ketjusäännön mukaan saadaan

$$\begin{aligned} I(X;Y) &= H(X) - H(X \mid Y) \\ &= H(X) + H(Y) - H(X,Y). \end{aligned}$$

Vielä

$$I(X;X) = H(X) - H(X \mid X) = H(X),$$

koska $H(X|X) = 0$ (harjoitustehtävä). Kootaan tulokset yhteen seuraavaan lauseeseen (ks. kuva 2.10).

Lause 2.14.

(i) $I(X;Y) = H(X) - H(X|Y)$

(ii) $I(X;Y) = H(Y) - H(Y|X)$

(iii) $I(X;Y) = H(X) + H(Y) - H(X,Y)$

(iv) $I(X;Y) = I(Y;X)$

(v) $I(X;X) = H(X)$

Kaikki edellä olevat käsitteet ja tulokset yleistyvät suoraan satunnaisvektoreille (sv) $(X_1, \ldots, X_n)$ ja $(Y_1, \ldots, Y_m)$, jotka koostuvat (äärellisen monta arvoa saavista) satunnaismuuttujista X_i, Y_j. Näin siksi, että satunnaisvektorit ovat itsekin itseasiassa vain äärellisen monta arvoa saavia satunnaismuuttujia.

Esimerkiksi,

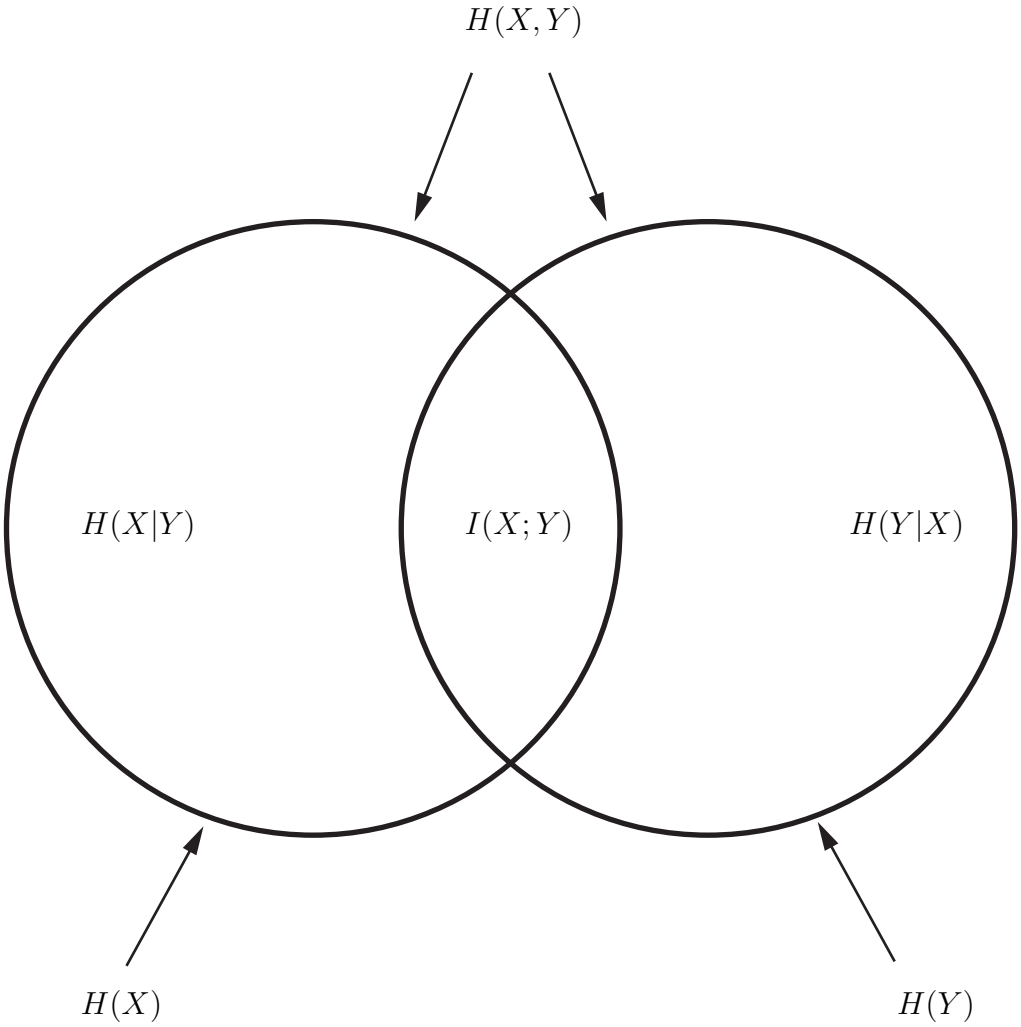
$$H(X_1, \ldots, X_n) = -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n) \log p(x_1, \ldots, x_n),$$

$$\begin{aligned} &H(Y_1, \ldots, Y_m \mid X_1, \ldots, X_n) \\ &\quad = -\sum_{x_1,\ldots,x_n} p(x_1, \ldots, x_n)\, H(Y_1, \ldots, Y_m \mid X_1 = x_1, \ldots, X_n = x_n) \\ &\quad = -\sum_{x_1,\ldots,x_n,y_1,\ldots,y_m} p(x_1, \ldots, x_n, y_1, \ldots, y_m) \log p(y_1, \ldots, y_m \mid x_1, \ldots, x_n), \end{aligned}$$

$$I(X_1, \ldots, X_n; Y_1, \ldots, Y_m) = H(X_1, \ldots, X_n) - H(X_1, \ldots, X_n \mid Y_1, \ldots, Y_m).$$

Lauseen 2.7 mukaisesti (ketjusääntö)

$$\begin{aligned} H(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = &H(X_1, \ldots, X_n) + \\ &H(Y_1, \ldots, Y_m \mid X_1, \ldots, X_n) \end{aligned} \tag{2.11}$$



Kuva 2.10: Entropiaan liittyvien peruskäsitteden väliset yhteydet lauseen 2.14 mukaan.

ja niin edelleen. Esimerkiksi lauseen 2.11 vastine nyt on:

$$H(Y_1, \ldots, Y_m \mid X_1, \ldots, X_n) \leq H(Y_1, \ldots, Y_m) \tag{2.12}$$

ja yhtäsuuruus pätee jos ja vain jos $(X_1, \ldots, X_n) \perp\!\!\!\perp (Y_1, \ldots, Y_m)$.

Lause 2.15 (Ketjusääntö)**.** *Pätee*

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1).$$

Tässä sovitaan, että $H(X_1 \mid X_0, \ldots, X_1) \equiv H(X_1)$.

Todistus. Induktio. Väite on selvästi tosi arvolla $n = 1$ (määritelmän mukaan). Oletetaan, että väite on tosi arvolla n. Siten

$$\begin{aligned} &H(X_1, \ldots, X_n, X_{n+1}) \\ &\quad \overset{(2.11)}{=} H(X_1, \ldots, X_n) + H(X_{n+1} \mid X_1, \ldots, X_n) \\ &\quad \overset{\text{ind.ol}}{=} \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1) + H(X_{n+1} \mid X_n, \ldots, X_1) \\ &\quad = \sum_{i=1}^{n+1} H(X_i \mid X_{i-1}, \ldots, X_1). \end{aligned}$$

□

Lause 2.16. *Pätee*

$$H(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} H(X_i)$$

ja yhtäsuuruus on voimassa jos ja vain jos $X_1, \ldots, X_n$ *ovat riippumattomia.*

Todistus. Epäyhtälö seuraa suoraan ketjusäännöstä (lause 2.15) ja kaavasta (2.12). Yhtäsuuruus on voimassa jos ja vain jos

$$H(X_i \mid X_{i-1}, \ldots, X_1) = H(X_i) \quad \text{kaikilla } i$$

eli $X_i \perp\!\!\!\perp \{X_{i-1}, \ldots, X_1\}$ kaikilla i eli $X_1, \ldots, X_n$ ovat riippumattomia (kaavan (2.12) jälkeinen huomio). □

2.4 Fanon epäyhtälö

Olkoot X ja Y satunnaismuuttujia arvojoukkoina $\mathcal{X}$ ja $\mathcal{Y}$, $|\mathcal{X}| \geq 2$, ja olkoon $(X, Y) \sim p(x, y)$. Oletetaan, että havaitaan Y ja yritetään arvata X sen perusteella. Arvaus X:ksi on

$$\widehat{X} = g(Y),$$

missä $g : \mathcal{Y} \to \mathcal{X}$. Virheellisen arvauksen todennäköisyys on

$$P_e = \mathbb{P}\{\widehat{X} \neq X\},$$

missä e ="error".

Kuten tavallisesti, merkitään $H(P_e) = -P_e \log P_e - (1 - P_e) \log(1 - P_e)$.

Lause 2.17 (Fanon epäyhtälö)**.**

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X \mid Y).$$

Todistus. Olkoon E tapahtuman $\{\widehat{X} \neq X\}$ indikaattori,

$$E = \begin{cases} 1, & \text{kun } \widehat{X} \neq X \\ 0, & \text{kun } \widehat{X} = X. \end{cases}$$

"Ehdollisen ketjusäännön" (harjoitustehtävä) nojalla saadaan

$$\begin{aligned} H(E, X \mid Y) &= H(X \mid Y) + H(E \mid X, Y) \\ &= H(E \mid Y) + H(X \mid E, Y). \end{aligned} \tag{2.13}$$

Edelleen pätee

$$\begin{aligned} H(X \mid E, Y) &= \sum_{e,y} p(e,y) H(X \mid E = e, Y = y) \\ &= -\sum_{e,y} p(e,y) \sum_{x} p(x \mid e, y) \log p(x \mid e, y) \\ &= -\sum_{y} p(0,y) \sum_{x} p(x \mid 0, y) \log p(x \mid 0, y) \qquad (a) \\ &\quad -\sum_{y} p(1,y) \sum_{x} p(x \mid 1, y) \log p(x \mid 1, y) \qquad (b) \end{aligned}$$

Termissä (a) on $E = 0$, joten $X = \widehat{X} = g(Y)$. Siten kiinteällä y satunnaismuuttuja X voi saada vain arvon $g(y)$ ja siis

$$p(x \mid 0, y) = \begin{cases} 0, & \text{kun } x \neq g(y), \\ 1, & \text{kun } x = g(y). \end{cases}$$

Näin ollen $\sum_x p(x \mid 0, y) \log p(x \mid 0, y) = 0$, joten $(a) = 0$.

Termissä (b) on $E = 1$, joten $g(Y) \neq X$. Siten kiinteällä y satunnaismuuttuja X voi saada vain arvot $\mathcal{X} \setminus \{g(y)\}$, jolloin

$$\sum_{x \neq g(y)} p(x \mid 1, y) = 1.$$

Nyt siis

$$\begin{aligned} (b) &= -\sum_{y} p(1,y) \sum_{x} p(x \mid 1, y) \log p(x \mid 1, y) \\ &= \sum_{y} p(1,y) \Big[-\sum_{x \neq g(y)} p(x \mid 1, y) \log p(x \mid 1, y) \Big] \\ &\leq P_e \log(|\mathcal{X}| - 1), \end{aligned}$$

missä lopussa käytettiin yhtälöä $\sum_y p(1,y) = \mathbb{P}\{E = 1\} = P_e$ ja lausetta 2.12. Kaavassa (2.13) on selväsi $H(E|X,Y) = 0$ ja lauseen 2.11 nojalla

$H(E|Y) \leq H(E) = H(P_e)$. Siten

$$\begin{aligned} H(X \mid Y) &\leq H(P_e) + H(X \mid E, Y) \\ &\leq H(P_e) + P_e \log(|\mathcal{X}| - 1). \end{aligned}$$

□

Huomautus. Lauseen nojalla, kun $P_e = 0$, myös $H(X \mid Y) = 0$. Tämä on luontevaa, koska tällöin X on Y:n funktio (todennäköisyydellä 1).

Huomautus. Koska $H(P_e) \leq 1$, on $H(X \mid Y) \leq 1 + P_e \log |\mathcal{X}|$ eli

$$P_e \geq \frac{H(X \mid Y) - 1}{\log |\mathcal{X}|}.$$

Luku 3

Tyypillisyys

3.1 AEP

Olkoon satunnaismuuttujan $X \sim p(x)$ arvojoukko $\mathcal{X}$. Tarkastellaan riippumattomia satunnaismuuttujia $X_1, \ldots, X_n$, joilla on sama jakauma kuin satunnaismuuttujalla X,

$$\mathbb{P}\{X_i = x\} = \mathbb{P}\{X = x\} = p(x),$$

missä $x \in \mathcal{X}$, $i = 1, \ldots, n$. Ilmaisemme tämän merkinnällä

$$X_1, \ldots, X_n \overset{iid}{\sim} p(x),$$

missä iid tarkoittaa ”independently and identically distributed”.

Merkitään jatkossa myös

$$x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n \quad (x_i \in \mathcal{X},\, i = 1, \ldots, n),$$

$$X^n = (X_1, \ldots, X_n).$$

Riippumattomuudesta seuraa, että

$$p(x^n) = \mathbb{P}\{X^n = x^n\} = \mathbb{P}\{X_i = x_i,\, i = 1, \ldots, n\}$$
$$\overset{\perp\!\!\!\perp}{=} \prod_{i=1}^{n} \mathbb{P}\{X_i = x_i\} = \prod_{i=1}^{n} p(x_i).$$

Olkoon $x \in \mathcal{X}$ ja $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$. Merkitään

$$m(x, x^n) = \big|\{i \mid x_i = x,\, i = 1, \ldots, n\}\big|,$$

jolloin $m(x, x^n)$ on siis niiden x_i:den lukumäärä, joille $x_i = x$. Vastaava satunnaismuuttuja on

$$m(x, X^n) = \big|\{i \mid X_i = x,\, i = 1 \ldots, n\}\big|.$$

Esimerkki 3.1. Olkoon $\mathcal{X} = \{0, 1\}$, $n = 7$, $x^n = (1, 1, 0, 0, 0, 0, 1) \in \{0, 1\}^7$. Siten

$$m(0, x^n) = 4, \quad m(1, x^n) = 3.$$

$\|$

Symboli $x \in \mathcal{X}$ esiintyy kohdassa i todennäköisyydellä $\mathbb{P}\{X_i = x\} = p(x)$. Kyseessä voidaan ajatella olevan riippumaton toistokoe, jossa tapahtuman "$X_i = x$" todennäköisyys on $p(x)$. Siten

$$m(x, X^n) \sim \mathrm{Bin}\big(n, p(x)\big).$$

Näin ollen

$$\begin{cases} \mathbb{E}[m(x, X^n)] = n\, p(x), \\ D^2[m(x, X^n)] = n\, p(x)\big(1 - p(x)\big). \end{cases}$$

Oletetaan, että $0 < p(x) < 1$ kaikilla $x \in \mathcal{X}$. Olkoon $a > 0$ ja tarkastellaan jonoja $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$, joille pätee

$$\left| \frac{m(x, x^n) - n\, p(x)}{\sqrt{n\, p(x)(1 - p(x))}} \right| < a$$

kaikilla $x \in \mathcal{X}$, eli

$$\begin{aligned} n\,p(x) - a\sqrt{n}\sqrt{p(x)(1-p(x))} &< m(x, x^n) \\ &< n\,p(x) + a\sqrt{n}\sqrt{p(x)(1-p(x))}. \end{aligned} \tag{3.1}$$

Koska $\sqrt{n} \ll n$ suurilla n, on tällaisissa jonoissa x^n noin $n\,p(x)$ symbolia x, eli odotusarvon verran.

Osoitetaan nyt, että tällaisen jonon todennäköisyydelle pätee

$$2^{-nH(X)-c\sqrt{n}} < p(x^n) < 2^{-nH(X)+c\sqrt{n}}, \tag{3.2}$$

missä $c > 0$ on vakio.

Nyt

$$p(x^n) = \prod_{i=1}^{n} p(x_i) = \prod_{x \in \mathcal{X}} p(x)^{m(x,x^n)}.$$

Seuraava konkreettinen esimerkki valaisee tätä kaavaa.

Esimerkki 3.2. Olkoon $\mathcal{X} = \{0, 1\}$. Silloin

$$\begin{aligned} p(1,1,0,0,0,0,1) &= p(1)p(1)p(0)\cdots p(1) = p(0)^4 p(1)^3 \\ &= p(0)^{m(0,x^n)} p(1)^{m(1,x^n)}. \end{aligned}$$

‖

Siten saadaan

$$\log p(x^n) = \sum_{x \in \mathcal{X}} m(x, x^n) \log p(x)$$

ja siis kaavan (3.1) vasemman puolen nojalla

$$\begin{aligned} \log p(x^n) &< \sum_{x \in \mathcal{X}} \left[np(x) - a\sqrt{n}\sqrt{p(x)(1-p(x))} \right] \log p(x) \\ &= n \sum_{x \in \mathcal{X}} p(x) \log p(x) + \sqrt{n} \left[-a \sum_{x \in \mathcal{X}} \sqrt{p(x)(1-p(x))} \right] \log p(x). \end{aligned}$$

Siten

$$p(x^n) < 2^{-nH(X)+c\sqrt{n}}, \tag{3.3}$$

missä

$$c = -a \sum_{x \in \mathcal{X}} \sqrt{p(x)(1-p(x))}\, \log p(x) > 0.$$

Samoin osoitetaan, että

$$p(x^n) > 2^{-nH(X)-c\sqrt{n}}. \tag{3.4}$$

Siten (3.2) pätee.

Nyt epäyhtälöissä (3.3) ja (3.4)

$$-nH(X) \pm c\sqrt{n} = -n\left(H(X) \mp \frac{c}{\sqrt{n}}\right)$$

ja $c/\sqrt{n}$ on pieni, kun n on suuri. Tämä motivoi seuraavan määritelmän. Olkoon $p(x^n) = \prod_{i=1}^{n} p(x_i)$.

Määritelmä 3.1. Olkoon $n \in \mathbb{N}_+$ ja $\varepsilon > 0$. *Tyypillinen joukko* $A_\varepsilon^{(n)}$ on

$$A_\varepsilon^{(n)} = \left\{ x^n = (x_1, \ldots, x_n) \mid 2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)} \right\}.$$

Siten ehdon (3.1) täyttävät jonot x^n ovat tyypillisiä annetulla $\varepsilon > 0$, kun n on riittävän suuri, $c/\sqrt{n} < \varepsilon$. Kaikki tyypilliset jonot *eivät* täytä ehtoa (3.1). Voi olla

$$-\frac{1}{n} \sum \log p(x_i) \approx H(X)$$

vaikka $x \in \mathcal{X}$ symboleja *ei* esiinny ”oikeassa” suhteessa.

Huomautus. Tässä määritelmässä ei enää oleteta, että $0 < p(x) < 1$ (eikä jatkossakaan tätä oleteta).

Esimerkki 3.3. Kuvassa 3.1 on vaakariveillä satunnaisvektorin

$$X^{100} = (X_1, \ldots, X_{100})$$

saamia ”tyypillisiä” arvoja, kun $X_i \sim \mathrm{Bin}(100, 0.1)$. Alimmaisena on todennäköisin ja epätodennäköisin jono. Satunnaisvektorille X^{100} pätee $H(X^{100}) = 46.9$. ‖

$\mathbf{x}$	$\log_2(P(\mathbf{x}))$
...1.....1.....1...1.1.....1.....1.....1.....1.....11...	−50.1
.....1.....1.....1.....1...1.....1.....1....	−37.3
.....1....1..1...1....11..1.1.....11.....1...1.1..1...1.....1.	−65.9
1.1...1.....1.....11.1..1.....1.....1.....1..1.11.....	−56.4
...11.....1...1.....1.1.....1.....1....1...1.....1.....	−53.2
.....1.....1.....1.1.....1.....1.....1...1.....1.....	−43.7
.....1.....1.....1...1.....1.....1.....1.....1..11.....	−46.8
.....1..1..1.....111.....1.....1.....1.....1.1...1...1.....1	−56.4
.....1.....1.....1.....1.....1....1.....1...	−37.3
.....1.....1.....1.....1....1..1.1.1..1.....1.	−43.7
1.....1.....1...1.....1....1....1.....1..11..1.1...1.....	−56.4
.....11.1.....1.....1.....1.....1.....	−37.3
.1.....1...1.1.....1.....11.....1.1...1.....1.....11.....	−56.4
.....1...1..1.....1..11.1.1.1...1.....1.....1.....1..1.....	−59.5
.....11.1.....1....1..1.....1.....1.....1.....1.....	−46.8
.....	−15.2
11111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111	−332.1

Kuva 3.1: Tyypillisiä jonoja satunnaisvektorin $X^{100} = (X_1, \ldots, X_{100})$ arvoiksi, kun $X_i \sim \text{Bin}(100, 0.1)$. Kuvassa piste vastaa arvoa 0. Oikealla kunkin jonon todennäköisyyden logaritmi (esimerkki lähteestä [5]).

Kuvassa 3.2 on vielä lisää havainnollistettu tyypillisen joukon käsitettä, kun $X_i \sim \text{Bin}(n, 0.1)$ ja $n = 100$ tai $n = 1000$. Kuvassa on n:ää itseasiassa merkitty N:llä.

Palautetaan mieliin heikko suurten lukujen laki:

Olkoot $X_1, X_2, \ldots$ riippumattomia satunnaismuuttujia, $E(X_i) = \mu$, $D^2(X_i) = \sigma^2 < \infty$ kaikilla i. Silloin

$$\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mu \text{ stokastisesti, kun } n \to \infty,$$

eli kaikilla $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right\} = 0.$$

Meidän tilanteessamme oletetaan itseasiassa, että $X_1, X_2, \ldots \overset{iid}{\sim} p(x)$ ja odotusarvon sekä varianssin olemassaolo on selvä, koska arvojoukko $\mathcal{X}$ on äärellinen.

Lause 3.2. *Olkoon $\varepsilon > 0$. Tyypilliselle joukolle $A_\varepsilon^{(n)}$ pätee*

(*i*) $\lim_{n\to\infty} \mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} = 1$,

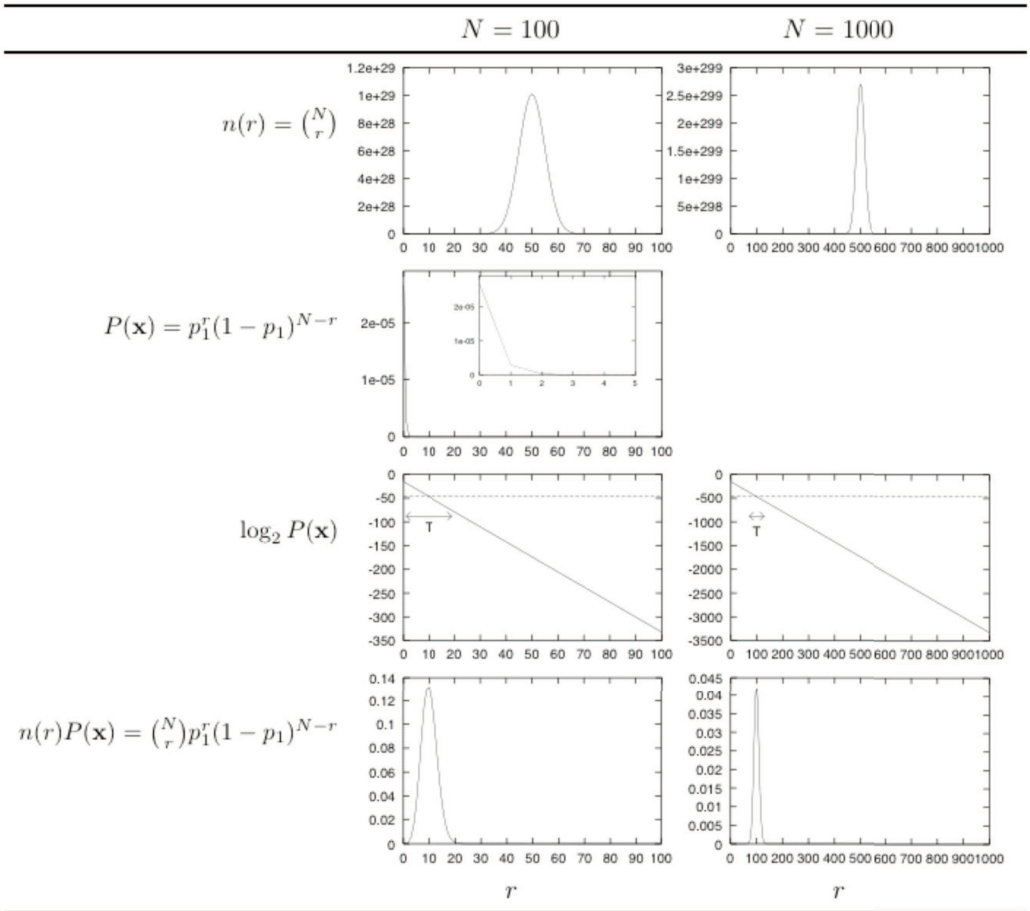
(*ii*) $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$ *kaikilla n,*

(*iii*) *kaikilla $\delta > 0$ $|A_\varepsilon^{(n)}| \geq (1-\delta)\, 2^{n(H(X)-\varepsilon)}$, kun n on riittävän suuri.*

Todistus. Kuten aikaisemmin, $X^n = (X_1, \ldots, X_n)$.

(*i*)

$$\begin{aligned}
\mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} &= \mathbb{P}\left\{2^{-n(H(X)+\varepsilon)} \leq p(X^n) \leq 2^{-n(H(X)-\varepsilon)}\right\} \\
&= \mathbb{P}\left\{H(X) - \varepsilon \leq -\frac{1}{n}\log p(X^n) \leq H(X) + \varepsilon\right\} \\
&= \mathbb{P}\left\{\left|-\frac{1}{n}\log p(X^n) - H(X)\right| \leq \varepsilon\right\}.
\end{aligned}$$



Kuva 3.2: Tyypillisen joukon havainnollistamista, kun $X_i \sim \text{Bin}(N, 0.1)$ ja $N = 100$ tai 1000. Ylhäällä $n(r)$ on niiden jonojen lukumäärä, joissa on tasan r ykköstä. Seuraavalla rivillä on sellaisen yhden jonon todennäköisyys, jossa on tasan r ykköstä ja sitä seuraavalla rivillä sama logaritmisessa skaalassa. Viimeisellä rivillä on todennäköisyys, että satunnaisessa jonossa on r ykköstä. Katkoviiva näyttää arvon $-H(X^N)$. Tyypillinen joukko on merkitty T:llä ottamalla $\varepsilon = 0.29$, kun $N = 100$ ja $\varepsilon = 0.09$, kun $N = 1000$ (esimerkki lähteestä [5]).

Tässä

$$-\frac{1}{n}\log p(X^n) = -\frac{1}{n}\log\prod_{i=1}^{n} p(X_i) = -\frac{1}{n}\sum_{i=1}^{n}\log p(X_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n}[-\log p(X_i)]$$

Toisaalta $\mathbb{E}[-\log p(X_i)] = \mathbb{E}[-\log p(X)] = H(X)$. Siten väite seuraa heikosta suurten lukujen laista, jonka mukaan komplementtitapahtumalle pätee

$$\mathbb{P}\{X^n \notin A_\varepsilon^{(n)}\} = \mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\big[-\log p(X_i)\big] - H(X)\right| > \varepsilon\right\} \to 0,$$

kun $n \to \infty$.

(ii) Tyypillisen joukon $A_\varepsilon^{(n)}$ määritelmää apuna käyttäen saadaan

$$1 = \sum_{x^n\in\mathcal{X}^n} p(x^n) \geq \sum_{x^n\in A_\varepsilon^{(n)}} p(x^n) \geq \sum_{x^n\in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = |A_\varepsilon^{(n)}|\, 2^{-n(H(X)+\varepsilon)},$$

joten

$$|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}.$$

(iii) Olkoon n niin suuri, että $\mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} \geq 1-\delta$ $\big((i)$-kohta$\big)$. Silloin

$$1-\delta \leq \mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} = \sum_{x^n\in A_\varepsilon^{(n)}} p(x^n)$$
$$\leq \sum_{x^n\in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} = |A_\varepsilon^{(n)}|\, 2^{-n(H(X)-\varepsilon)},$$

joten

$$|A_\varepsilon^{(n)}| \geq (1-\delta)\, 2^{n(H(X)-\varepsilon)}.$$

□

Huomautus. Edellä oleva lause tunnetaan nimellä AEP, Asymptotic Equipartition Principle. Sen mukaan umpimähkään valittu $x^n \in \mathcal{X}^n$ on suurella todennäköisyydellä (n suuri) joukossa $A_\varepsilon^{(n)}$ ja tässä joukossa on noin $2^{nH(X)}$ alkiota, kaikki likimain yhtä todennäköisiä.

3.2 Koodaus kompressiossa

Tarkastellaan informaatiolähdettä, jonka lähettämä viesti halutaan jostain syystä koodata.

$$\boxed{\text{informaatiolähde}} \longrightarrow X \in \mathcal{X}$$

Symbolit $X \in \mathcal{X}$ voivat olla periaatteessa mitä vain. Jos ne halutaan esimerkiksi tallentaa tietokoneen muistiin tai lähettää jonkin kanavan läpi vastaanottajalle, voi koodaus olla välttämätöntä.

Pyritään koodaamaan viesti *lohkoissa* $X^n = (X_1, \ldots, X_n)$, missä $X_1, \ldots, X_n$ ovat peräkäin lähetettyjä symboleja.

$$X^n = (X_1, \ldots, X_n) \longrightarrow \boxed{\text{kooderi}} \longrightarrow \text{koodisana}$$

Siis ensin koodataan $(X_1, \ldots, X_n)$, sitten $(X_{n+1}, \ldots, X_{2n})$ jne. Oletetaan, että $X_1, X_2, \ldots \overset{iid}{\sim} p(x)$, kuten edellisessä kappaleessa.

Nyt

$$\mathcal{X}^n = A_\varepsilon^{(n)} \cup \left[\mathcal{X}^n \setminus A_\varepsilon^{(n)}\right].$$

(Ks.kuva 3.3). Olkoon $A_\varepsilon^{(n)} = \{x^{n,1}, \ldots, x^{n,k}\}$.

Kun $t \in \mathbb{R}$, merkitään

$$\lceil t \rceil = \min\{m \in \mathbb{Z} \mid t \leq m\},$$

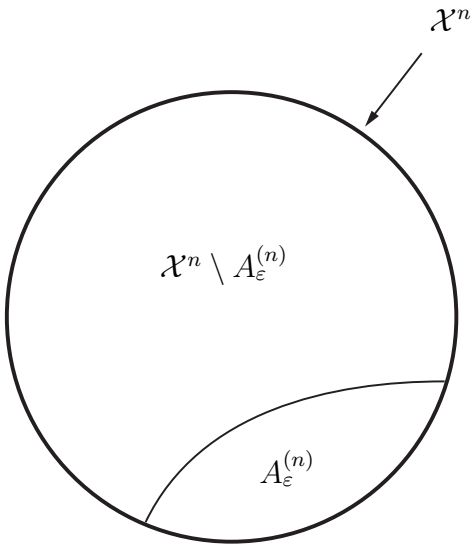
eli $\lceil t \rceil$ on ylöspäin pyöristys kokonaisluvuksi.

Nyt lauseen 3.2 kohdan (ii) nojalla $|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$, joten

$$k \leq 2^{n(H(X)+\varepsilon)} \leq 2^{\lceil n(H(X)+\varepsilon) \rceil}.$$

Siten jonot $x^n \in A_\varepsilon^{(n)}$ voidaan indeksoida $\lceil n(H(X)+\varepsilon) \rceil$ bittisillä toisistaan eroavilla binääriluvuilla (m-bittisiä binäärilukuja on 2^m kappaletta):

$$x^n \xrightarrow{\text{kooderi}} b_1 \cdots b_m, \qquad b_j \in \{0,1\},\ m = \lceil n(H(X)+\varepsilon) \rceil.$$



Kuva 3.3: Avaruuden $\mathcal{X}^n$ ositus tyypillisiin jonoihin $A_\varepsilon^{(n)}$ ja epätyypillisiin jonoihin $\mathcal{X}^n \setminus A_\varepsilon^{(n)}$.

Laitetaan vielä etunolla ilmoittamaan, että $x^n \in A_\varepsilon^{(n)}$,

$$x^n \xrightarrow{\text{kooderi}} 0\, b_1 \dots b_m \qquad \text{(koodisana)}.$$

Samoin, $|\mathcal{X}^n \setminus A_\varepsilon^{(n)}| \leq |\mathcal{X}^n| = |\mathcal{X}|^n = 2^{n \log |\mathcal{X}|}$, joten jonot $x^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)}$ voidaan koodata $\lceil n \log |\mathcal{X}| \rceil + 1$ bittisillä binääriluvuilla.

$$x^n \xrightarrow{\text{kooderi}} 1\, b_1 \cdots b_l, \qquad l = \lceil n \log |\mathcal{X}| \rceil .$$

Koodi on helppo dekoodata: ensimmäinen bitti kertoo onko $x^n \in A_\varepsilon^{(n)}$ vai $x^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)}$ ja dekoodaus voidaan sitten tehdä taulukosta.

Yleisesti on järkevää pyrkiä tietysti mahdollisimman *lyhyihin* koodisanoihin, koska se säästää esimerkiksi tilaa tallennettaessa ja nostaa nopeutta tiedonsiirrossa.

Koska yleensä $H(X) < \log |\mathcal{X}|$ (vertaa lause 2.12), tulevat tyypilliset jonot $x^n \in A_\varepsilon^{(n)}$ koodattua lyhemmin kuin epätyypilliset jonot $x^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)}$, joiden koodisanoja ei yritetty mitenkään lyhentää. Tämä on järkevää, koska

$X^n \in A_\varepsilon^{(n)}$ suurella todennäköisyydellä, ainakin kun n on suuri ($\mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} \to 1$, kun $n \to \infty$).

Suhteellisen lyhyt koodi tyypilliselle jonolle on mahdollinen, koska yleensä $A_\varepsilon^{(n)}$ on *pieni* (alkioiden lukumäärältään) verrattuna koko $\mathcal{X}^n$:ään:

$$H(X) + \varepsilon \leq \delta \log |\mathcal{X}| \qquad (0 < \delta < 1),$$

joten

$$|A_\varepsilon^{(n)}| \overset{\text{AEP}}{\leq} 2^{n(H(X)+\varepsilon)} \leq 2^{n\delta \log |\mathcal{X}|} = |\mathcal{X}|^{n\delta}.$$

Siten saadaan edelleen

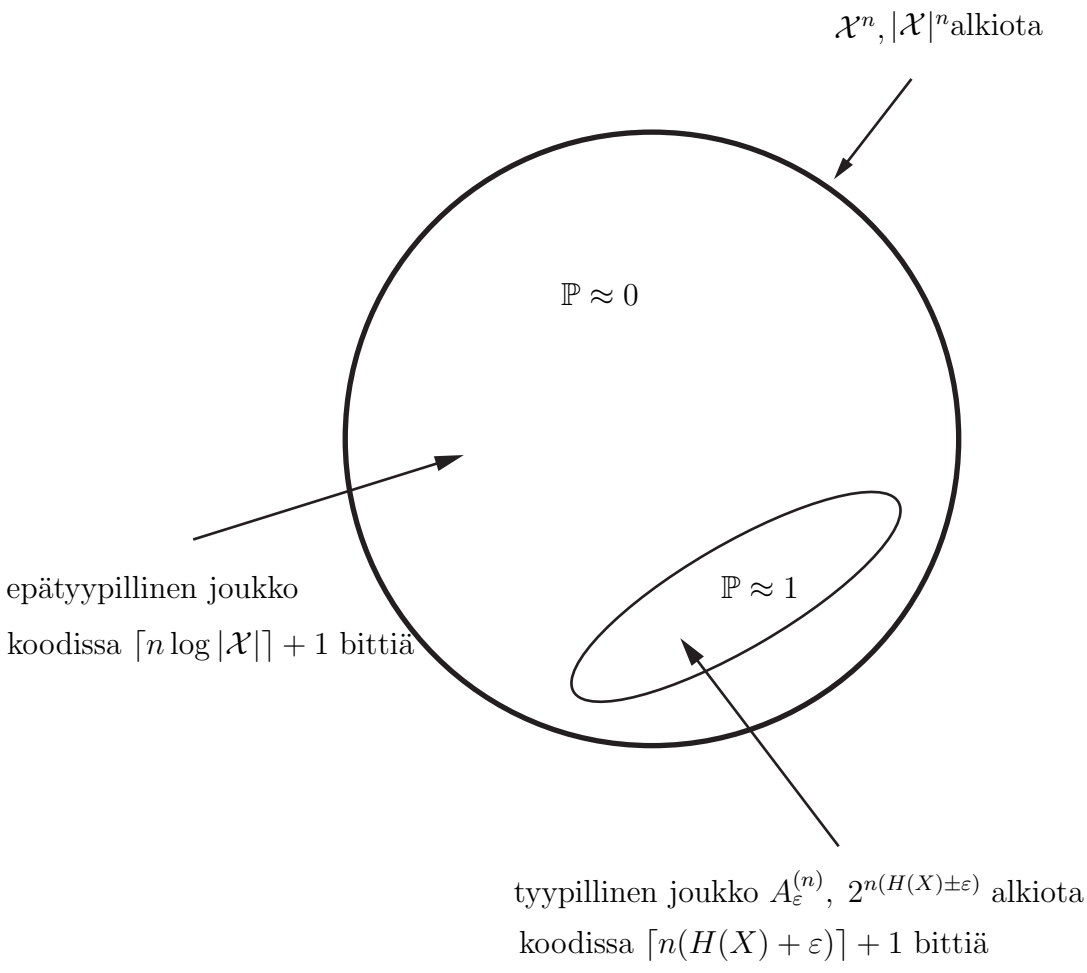
$$\frac{|A_\varepsilon^{(n)}|}{|\mathcal{X}^n|} = \frac{|A_\varepsilon^{(n)}|}{|\mathcal{X}|^n} \leq \frac{|\mathcal{X}|^{n\delta}}{|\mathcal{X}|^n} = \frac{1}{|\mathcal{X}|^{n(1-\delta)}} \longrightarrow 0,$$

kun $n \to \infty$. Tilanne on siis kuvan 3.4 kaltainen.

Lasketaan kooderin tuottamien koodisanojen keskimääräinen pituus. Olkoon $l(x^n)$ x^n:n koodisanan pituus,

$$l(x^n) = \begin{cases} \lceil n(H(X) + \varepsilon) \rceil + 1, & \text{kun } x^n \in A_\varepsilon^{(n)}, \\ \lceil n \log |\mathcal{X}| \rceil + 1, & \text{kun } x^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)}. \end{cases}$$

Olkoon n edelleen niin suuri, että $\mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} \geq 1 - \varepsilon$ (lause 3.2 (i)).



Kuva 3.4: Tyypilliseen joukkoon perustuva koodaus.

Silloin keskimääräiselle koodisanan pituudelle saadaan

$$
\begin{aligned}
&\mathbb{E}\big(l(X^n)\big) = \sum_{x^n \in \mathcal{X}^n} p(x^n) l(x^n) \\
&= \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) l(x^n) + \sum_{x^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)}} p(x^n) l(x^n) \\
&= \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) \big\{ \lceil n(H(X)+\varepsilon) \rceil + 1 \big\} + \sum_{x^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)}} p(x^n) \big\{ \lceil n \log |\mathcal{X}| \rceil + 1 \big\} \\
&\leq \{ n(H(X)+\varepsilon) + 2 \} \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) + \{ n \log |\mathcal{X}| + 2 \} \sum_{x^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)}} p(x^n) \\
&= \mathbb{P}\{ X^n \in A_\varepsilon^{(n)} \}\, n(H(X)+\varepsilon) + \mathbb{P}\{ X^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)} \}\, n \log |\mathcal{X}| \\
&\qquad\qquad + 2\big[\mathbb{P}\{ X^n \in A_\varepsilon^{(n)} \} + \mathbb{P}\{ X^n \in \mathcal{X}^n \setminus A_\varepsilon^{(n)} \} \big] \\
&\leq n(H(X)+\varepsilon) + \varepsilon n \log |\mathcal{X}| + 2 \\
&= n \left[H(X) + \varepsilon + \varepsilon \log |\mathcal{X}| + \frac{2}{n} \right].
\end{aligned}
$$

Siten, jos $\varepsilon' > 0$ on annettu, on olemassa koodi, jolle

$$\mathbb{E}\big(l(X^n)\big) \leq n \left(H(X) + \varepsilon' \right),$$

kun valitaan sellaiset ε ja n, että $\varepsilon + \varepsilon \log |\mathcal{X}| + \frac{2}{n} \leq \varepsilon'$.

Koodatuissa lohkoissa X^n on n symbolia $X_1, \ldots, X_n \in \mathcal{X}$. Siten on osoitettu:

Lause 3.3. *Olkoon $X_1, \ldots, X_n \overset{iid}{\sim} p(x)$. Jos $\varepsilon > 0$, löytyy riittävän suurella n jonojen x^n (injektiivinen) binäärikoodaus, jossa keskimääräiselle koodisanan pituudelle per symboli pätee*

$$\mathbb{E}\left[\frac{1}{n} l(X^n) \right] \leq H(X) + \varepsilon.$$

Keskimääräiselle koodisanan pituudelle siis pätee, että se on

$$
\begin{cases}
\lesssim H(X) \text{ bittiä per symboli} \\
\lesssim nH(X) \text{ bittiä per lohko } X^n
\end{cases}
$$

Huomautus. Myöhemmin osoitetaan, että *kaikille* (tietyllä tavalla yksikäsitteisille) koodeille pätee

$$H(X) \leq \mathbb{E}\left[\frac{1}{n}\, l(X^n)\right].$$

Kuten edellä todettiin, $A_\varepsilon^{(n)}$ on *pieni* joukon $\mathcal{X}^n$ osajoukko. Tämä mahdollistaa tehokkaan koodauksen, koska lisäksi $\mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} \approx 1$. Löytyisikö kuitenkin vielä parempi joukko, eli $B \subset \mathcal{X}^n$, jolle $\mathbb{P}\{X^n \in B\} \approx 1$ ja $|B| \ll |A_\varepsilon^{(n)}|$?

Lemma 3.4. *Olkoon* $0 < \varepsilon < 1/2$, $B \subset \mathcal{X}^n$ *ja* $\mathbb{P}\{X^n \in B\} \geq 1 - \varepsilon$. *Silloin*

$$\frac{1}{n} \log |B| \geq H(X) - 2\varepsilon,$$

kun n *on riittävän suuri.*

Todistus. Harjoitustehtävä. □

Jos nyt $|B| \leq |A_\varepsilon^{(n)}|$, pätee lauseen 3.2 kohdan (ii) nojalla

$$\frac{1}{n} \log |B| \leq \frac{1}{n} \log |A_\varepsilon^{(n)}| \leq H(X) + \varepsilon.$$

Lemman 3.4 mukaan suurilla n pätee

$$H(X) - 2\varepsilon \leq \frac{1}{n} \log |B| \leq H(X) + \varepsilon. \tag{3.5}$$

Toisaalta lauseen 3.2 kohdan (iii) nojalla saadaan

$$\log |A_\varepsilon^{(n)}| \geq n\big(H(X) - \varepsilon\big) + \log(1 - \delta)$$

kaikilla $0 < \delta < 1$, kun n on riittävän suuri, joten

$$H(X) - 2\varepsilon \leq \frac{1}{n} \log |A_\varepsilon^{(n)}| \leq H(X) + \varepsilon, \tag{3.6}$$

kun n on riittävän suuri. Tällöin kaavojen (3.5) ja (3.6) nojalla

$$\frac{1}{n} \log |B| \approx \frac{1}{n} \log |A_\varepsilon^{(n)}|$$

eli tässä mielessä

$$|B| \approx |A_\varepsilon^{(n)}| \approx 2^{nH(X)},$$

eikä siis joukon $A_\varepsilon^{(n)}$ kokoa voi ”oleellisesti” pienentää.

3.3 Yleisemmät informaatiolähteet

Informaatiolähteen malli edellä perustui iid satunnaismuuttujiin,

$$\boxed{\text{informaatiolähde}} \longrightarrow X_1, X_2, \ldots \overset{iid}{\sim} p(x).$$

Se on varsin yksinkertainen ja monipuolisempia malleja saadaankin yleisemmistä stokastisista prosesseista.

Määritelmä 3.5. *Stokastinen prosessi* on jono satunnaismuuttujia, $(X_n)_{n\in\mathbb{N}_+} = (X_n) = (X_1, X_2, \ldots)$.

Tässä siis $X_n : \Omega \to \mathcal{X}$, $n \in \mathbb{N}_+$, on satunnaismuuttuja kuten edellä. Prosessin saamat "arvot" ovat *realisaatioita* $(X_n(\omega)) = (x_1, x_2, \ldots)$, $X_n(\omega) = x_n$, $\omega \in \Omega$, $n \in \mathbb{N}_+$.

Prosessin (X_n) jakauma määräytyy yhteispistetodennäköisyysfunktioista

$$p(x_1, \ldots, x_n) = \mathbb{P}\{X_1 = x_1, \ldots, X_n = x_n\},$$

missä $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $n \in \mathbb{N}_+$.

Usein (kuten tällä kurssilla) on luontevaa ajatella n:ää *aikana*.

Esimerkki 3.4. Olkoon $X_1, X_2, \ldots \overset{iid}{\sim} p(x)$. Silloin (X_n) on *iid-prosessi*. Sille pätee

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i), \quad n \in \mathbb{N}_+.$$

Hyvin hyödyllinen prosessin tyyppi on stationaarinen prosessi.

Määritelmä 3.6. Stokastinen prosessi (X_n) on *stationaarinen*, jos

$$\mathbb{P}\{X_1 = x_1, \ldots, X_n = x_n\} = \mathbb{P}\{X_{1+l} = x_1, \ldots, X_{n+l} = x_n\}$$

kaikilla $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $n, l \in \mathbb{N}_+$.

Siis: Stationaarisen prosessin jakaumaominaisuudet ovat ”ajan suhteen invariantteja”; prosessi ”näyttää” samanlaiselta eri aikoina. Selvästikin iid-prosessi on aina stationaarinen.

Esimerkki 3.5. Kuvassa 3.5 on esimerkit iid-prosessin ja erään stationaarisen prosessin realisaatioista. Esimerkin stationaarisessa prosessissa näkyy selvää riippuvuutta perättäisten ajanhetkien välillä kun taas iid-prosessissa mitään riippuvuutta ei ole. Huomaa, että kyseessä on vain yksi realisaatio stationaarisesta prosessista; keskimäärin, useita realisaatioita tarkasteltaessa prosessin tilastolliset ominaisuudet näyttäisivät samanlaisilta eri aikoina. ‖

Esimerkki 3.6 (Markovin ketju)**.** Olkoon merkintöjen yksinkertaistamiseksi

$$\mathcal{X} = \{1, \ldots, m\}.$$

Kutsutaan arvoja $i \in \mathcal{X}$ *tiloiksi* ja tarkastellaan kuvan 3.6 kaltaista, tilasta toiseen siirtyvää prosessia.

Olkoon

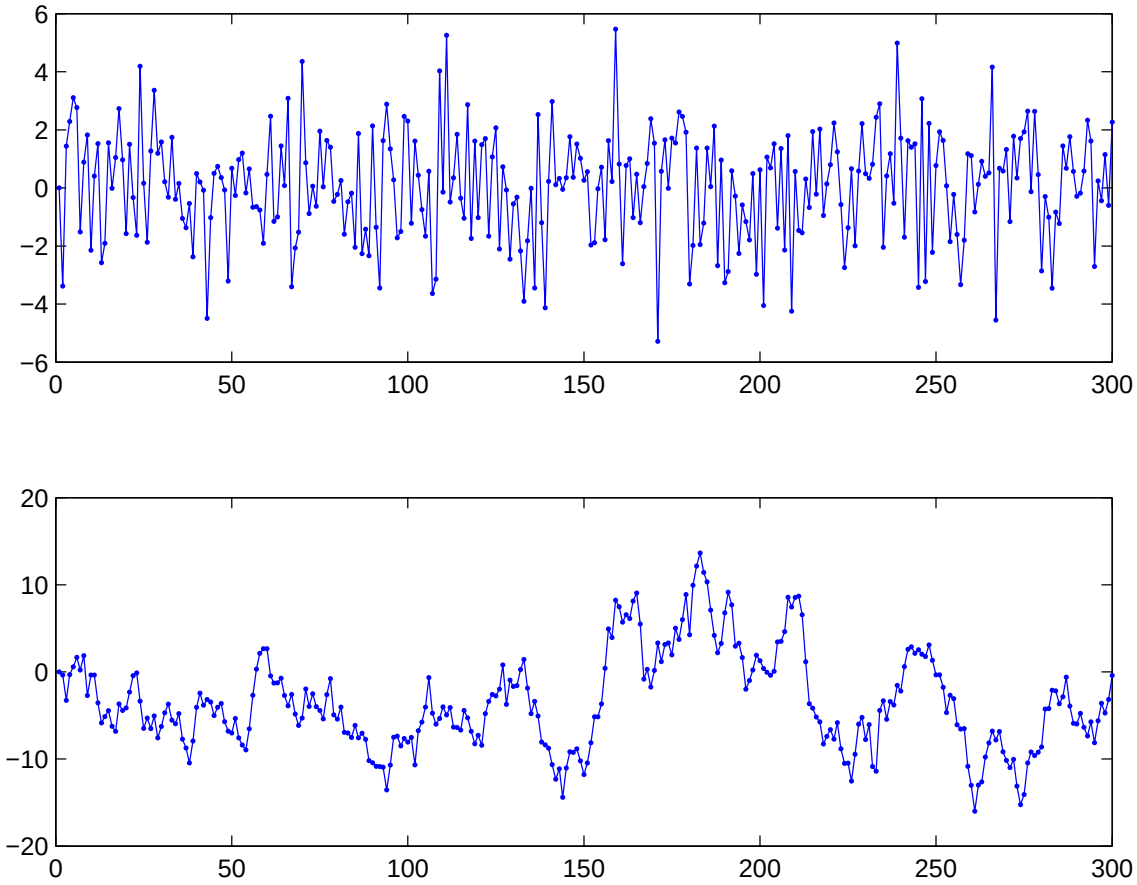
$$P = (p_{ij}) = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix} \in \mathbb{R}^{m \times m}$$

matriisi, jolle $p_{ij} \geq 0$ kaikilla $i, j \in \mathcal{X}$ ja $\sum_{j=1}^{m} p_{ij} = 1$ kaikilla $i \in \mathcal{X}$. Sanomme, että (X_n) on (aikahomogeeninen) *Markovin ketju*, jos

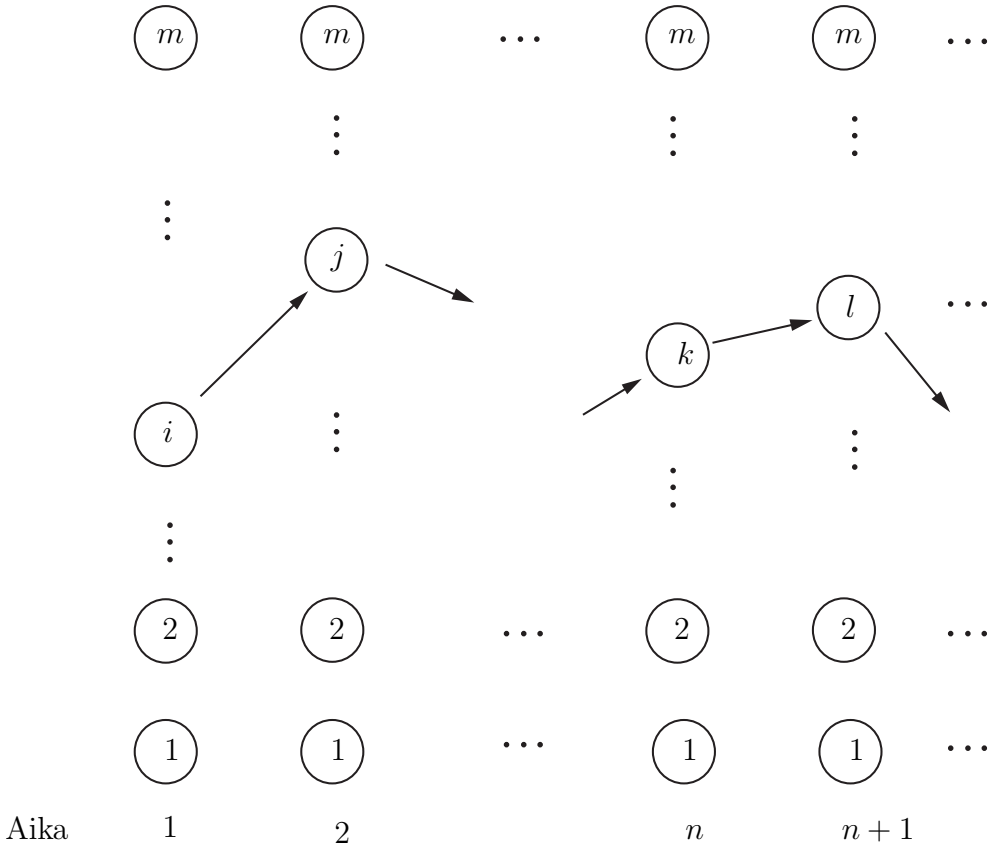
$$\begin{aligned} &\mathbb{P}\big\{X_{n+1} = j \mid X_n = i,\, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1\big\} \\ &\qquad = \mathbb{P}\big\{X_{n+1} = j \mid X_n = i\big\} = p_{ij} \end{aligned}$$

kaikilla $i, j \in \mathcal{X}$.

Siten Markovin ketjun siirtymätodennäköisyyteen tilaan $X_{n+1} = j$ vaikuttaa vain edellinen tila $X_n = i$, ei aikaisempi ”historia”. Aikahomogeenisuus tarkoittaa sitä, että siirtymätodennäköisyydet eivät riipu ajasta.



Kuva 3.5: Realisaatiot iid-prosessista (ylempi paneeli) ja eräästä sationaarisesta prosessista (alempi paneeli).



Kuva 3.6: Markovin ketju.

P on *siirtymämatriisi* ja p_{ij}:t ovat *siirtymätodennäköisyyksiä*.

Olkoon (X_n) Markovin ketju ja

$$\mathbb{P}\{X_n = i\} = p_i^{(n)}, \quad i \in \mathcal{X},\, n \in \mathbb{N}_+.$$

Siis $p^{(n)} = (p_1^{(n)}, \ldots, p_m^{(n)})$ kuvaa ketjun jakuman hetkellä n. Silloin

$$\begin{aligned} p_j^{(n+1)} &= \mathbb{P}\{X_{n+1} = j\} \\ &= \sum_{i=1}^{m} \mathbb{P}\{X_{n+1} = j \text{ ja } X_n = i\} \\ &= \sum_{i=1}^{m} \mathbb{P}\{X_{n+1} = j \mid X_n = i\}\, \mathbb{P}\{X_n = i\} \\ &= \sum_{i=1}^{m} p_{ij}\, p_i^{(n)} \end{aligned}$$

eli

$$p^{(n+1)} = p^{(n)} P.$$

Jakauma $\mu = (\mu_1, \ldots, \mu_m)$, $\mu_i \geq 0$ kaikilla i, $\sum_{i=1}^{m} \mu_i = 1$, on ketjun *tasapainojakauma*, jos

$$\mu = \mu P.$$

Siten, jos $X_1 \sim \mu$ (so. $\mathbb{P}\{X_1 = i\} = \mu_i$, $i \in \mathcal{X}$), on $X_n \sim \mu$ kaikilla n.

Voidaan osoittaa, että jos μ on Markovin ketjun tasapainojakauma ja $X_1 \sim \mu$, niin ketju on stationaarinen (harjoitustehtävä).

Huomautus. Tietyn tyyppisillä Markovin ketjuilla (pelkistymätön, aperiodinen) on yksikäsitteinen tasapainojakauma μ ja $p^{(n)} \to \mu$, kun $n \to \infty$.

$\|$

Olkoon (X_n) stokastinen prosessi. Mitä voidaan sanoa entropiasta

$$H(X_1, \ldots, X_n)?$$

Selvästikin riippuvuus vähentää entropiaa, epävarmuutta. Tätä valaisee seuraava esimerkki.

Esimerkki 3.7. Tarkastellaan kahta yksinkertaista prosessia, jotka edustavat riippuvuuden ääripäitä.

(i) (X_n) on iid-prosessi, $X_1, X_2, \ldots \overset{iid}{\sim} p(x)$. Silloin lauseen 2.16 nojalla

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) = nH(X).$$

(ii) $X_i = X$ kaikilla i. Silloin

$$H(X, \ldots, X) = H(X),$$

koska $p(x, \ldots, x) = p(x)$.

Siten prosessin (ii) entropia on vain n:s osa prosessin (i) entropiasta.

Määritelmä 3.7. Stokastisen prosessin *entropian kasvunopeus* on

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n),$$

silloin kun kyseessä oleva raja-arvo on olemassa.

Huomautus. iid-prosessille $H(\mathcal{X}) = H(X)$ ja $H(X_1, \ldots, X_n) = nH(\mathcal{X})$, $n \in \mathbb{N}_+$.

Huomautus. Voidaan myös ajatella, että $H(\mathcal{X})$ antaa (asymptoottisesti) lähteen *entropian per symboli*,

$$H(\mathcal{X}) \approx \frac{1}{n} H(X_1, \ldots, X_n).$$

Toinen tapa määritellä symbolikohtainen entropia on

$$H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n \mid X_{n-1}, \ldots, X_1),$$

kun raja-arvo on olemassa.

Osoitetaan, että stationaarisille prosesseille $H(\mathcal{X})$ ja $H'(\mathcal{X})$ ovat olemassa ja $H(\mathcal{X}) = H'(\mathcal{X})$.

Lemma 3.8. *Olkoon (X_n) stationaarinen prosessi. Silloin $H'(\mathcal{X})$ on olemassa.*

Todistus. Kaikilla $n \geq 2$ on

$$\begin{aligned} H(X_n \mid X_{n-1}, \ldots, X_1) &= H(X_{n+1} \mid X_n, \ldots, X_2) \\ &\geq H(X_{n+1} \mid X_n, \ldots, X_2, X_1), \end{aligned}$$

koska ehdollistaminen vähentää entropiaa. Tässä tarvitaan lauseen 2.11 "ehdollista" versiota

$$H(Y \mid X, Z) \leq H(Y \mid Z),$$

joka todistetaan aivan kuten lause 2.11. Koska ei-negatiivinen jono $H(X_n \mid X_{n-1}, \ldots, X_1)$ on siis vähenevä, se suppenee. □

Lemma 3.9. *Olkoon (a_n) reaalilukujono, $\lim_{n\to\infty} a_n = a \in \mathbb{R}$. Silloin*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} a_i = a.$$

Todistus. Olkoon $\varepsilon > 0$. Valitaan sellainen n_ε, että $|a_n - a| < \varepsilon$, kun $n > n_\varepsilon$. Kun $n > n_\varepsilon$, saadaan silloin

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^{n} a_i - a \right| &= \left| \frac{1}{n} \sum_{i=1}^{n} (a_i - a) \right| \\ &\leq \frac{1}{n} \sum_{i=1}^{n} |a_i - a| \\ &= \frac{1}{n} \sum_{i=1}^{n_\varepsilon} |a_i - a| + \frac{1}{n} \sum_{i=n_{\varepsilon+1}}^{n} |a_i - a| \\ &< \frac{1}{n} \sum_{i=1}^{n_\varepsilon} |a_i - a| + \frac{n - n_\varepsilon}{n} \cdot \varepsilon, \end{aligned}$$

missä $\frac{n-n_\varepsilon}{n} \cdot \varepsilon < \varepsilon$ ja $\frac{1}{n} \sum_{i=1}^{n_\varepsilon} |a_i - a| \to 0$, kun $n \to \infty$. □

Lause 3.10. *Stationaariselle prosessille* $H(\mathcal{X}) = H'(\mathcal{X})$.

Todistus. Ketjusäännön (lause 2.15) nojalla

$$\frac{1}{n}H(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1).$$

Lemman 3.8 nojalla $\lim_{n\to\infty} H(X_n \mid X_{n-1}, \ldots, X_1) = H'(\mathcal{X})$. Väite seuraa siis lemmasta 3.9 ottamalla $a_n = H(X_n \mid X_{n-1}, \ldots, X_1)$. □

Huomautus. Korvaamalla $H(X)$ suureella $H(\mathcal{X})$, voidaan AEP todistaa ns. stationaariselle ergodiselle prosessille (X_n). Lukujen 3.1 ja 3.2 tulokset saadaan näin pätemään tällaisille prosesseille.

Luku 4

Häiriöttömän lähteen koodaus, kompressio

4.1 Koodeja

Tässä kappaleessa entropian määritelmä tulee lopullisesti motivoitua datan optimaalisen kompression mittana.

Olkoon X satunnaismuuttuja, jonka arvojoukko on $\mathcal{X}$, ja $|\mathcal{X}| = m$. X kuvaa informaatiolähdettä

$$\boxed{\text{informaatiolähde}} \longrightarrow X$$

Sanomme, että $\mathcal{X}$ on *viestiaakkosto* ja jono symboleita $x_1 \cdots x_k$, $x_i \in \mathcal{X}$, $k \in \mathbb{N}_+$, on *viesti*. Kaikkien mahdollisten viestien joukko on

$$\mathcal{X}^* = \{x_1 \cdots x_k \mid x_i \in \mathcal{X},\, i = 1, \ldots, k,\, k \in \mathbb{N}_+\}.$$

Viestien siirtäminen tai tallentaminen tavallisesti vaatii koodausta (muistilaitteet, mobiili viestintä, ...). Olkoon joukko $\mathcal{D}$ *koodiaakkosto*, jossa on

$|\mathcal{D}| = D < \infty$ symbolia ja merkitään

$$\mathcal{D}^* = \left\{ d_1 \cdots d_k \mid d_i \in \mathcal{D},\, i = 1, \ldots, k,\, k \in \mathbb{N}_+ \right\}.$$

Määritelmä 4.1. Satunnaismuuttujan X *lähdekoodi* tai lyhyesti *koodi* on kuvaus $C : \mathcal{X} \to \mathcal{D}^*$.

Joskus kuvausta C sanotaan myös *kooderiksi*. Koodaus toimii seuraavan kaavion mukaisesesti:

$$\boxed{\text{informaatiolähde}} \xrightarrow{X} \boxed{\text{kooderi}} \longrightarrow C(X)$$

$C(x)$ on x:n *koodisana*. Merkitsemme $l(x)$:llä $C(x)$:n pituutta. Terminologia on tässä yhteydessä hieman horjuvaa ja joskus määritellään, että koodi on itse asiassa kuva (koodisanojen joukko) $C(\mathcal{X})$.

Viesti $x_1 \cdots x_k \in \mathcal{X}^*$ koodataan symboli kerrallaan:

$$x_1 \cdots x_k \longrightarrow C(x_1) \cdots C(x_k).$$

Koodin C *laajennus* on kuvaus $\widetilde{C} : \mathcal{X}^* \to \mathcal{D}^*$,

$$\widetilde{C}(x_1 \cdots x_k) = C(x_1) \cdots C(x_k),$$

$x_1, \ldots, x_k \in \mathcal{X}$, $k \in \mathbb{N}_+$.

Esimerkki 4.1. Olkoon $\mathcal{X} = \{a, b, c, d\}$, $\mathcal{D} = \{0, 1\}$. Alla oleva taulukko määrittelee koodin:

x	$C(x)$	$l(x)$
a	00	2
b	10	2
c	01	2
d	11	2

Nyt $\widetilde{C}(abba) = 00101000$. Toinen koodi saataisiin taulukosta

x	$C(x)$	$l(x)$
a	0	1
b	1	1
c	10	2
d	01	2

Silloin $\widetilde{C}(abba) = 0110$. ‖

Esimerkki 4.2. Kuvassa 4.1 on esitetty osa (51 ensimmäistä merkkiä) tekstitiedostojen koodauksessa käytettävää ASCII-koodia. Kunkin merkin koodisana on esitetty desimaali-, oktaali-, heksadesimaali- ja binäärimuodossa. Näitä koodeja vastaavat aakkostot ovat

Desimaalikoodi	$\mathcal{D} = \{0, 1, \ldots, 9\}$,	$\lvert\mathcal{D}\rvert = 10$
Oktaalikoodi	$\mathcal{D} = \{0, 1, \ldots, 7\}$,	$\lvert\mathcal{D}\rvert = 8$
Heksadesimaalikoodi	$\mathcal{D} = \{0, 1, \ldots, 9, A, B, \ldots, F\}$,	$\lvert\mathcal{D}\rvert = 16$
Binäärikoodi	$\mathcal{D} = \{0, 1\}$,	$\lvert\mathcal{D}\rvert = 2$

‖

Olkoon sitten $X \sim p(x)$.

Määritelmä 4.2. Satunnaismuuttujan X koodin C *keskimääräinen pituus* on

$$L(C) = \sum_{x \in \mathcal{X}} p(x) l(x).$$

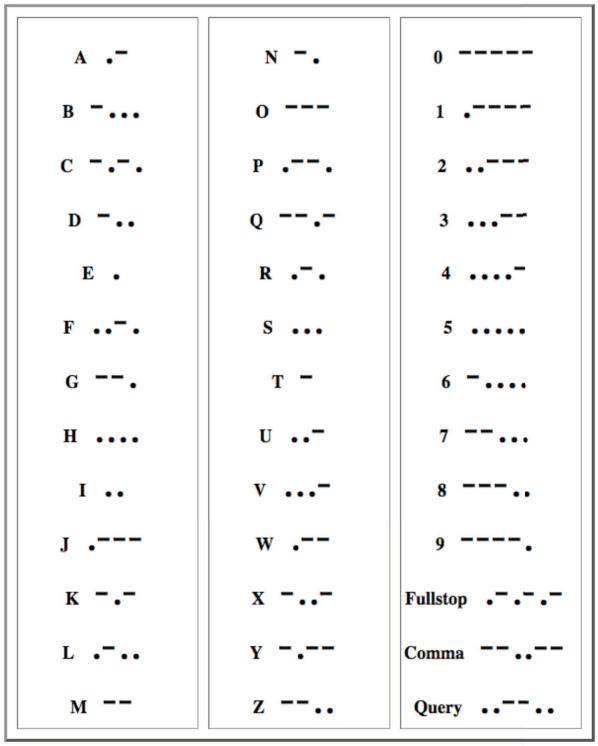
Huomautus. Selvästi kyseessä on odotusarvo

$$L(C) = \mathbb{E}\, l(X).$$

Decimal	Octal	Hex	Binary	Value	
000	000	000	00000000	NUL	(Null char.)
001	001	001	00000001	SOH	(Start of Header)
002	002	002	00000010	STX	(Start of Text)
003	003	003	00000011	ETX	(End of Text)
004	004	004	00000100	EOT	(End of Transmission)
005	005	005	00000101	ENQ	(Enquiry)
006	006	006	00000110	ACK	(Acknowledgment)
007	007	007	00000111	BEL	(Bell)
008	010	008	00001000	BS	(Backspace)
009	011	009	00001001	HT	(Horizontal Tab)
010	012	00A	00001010	LF	(Line Feed)
011	013	00B	00001011	VT	(Vertical Tab)
012	014	00C	00001100	FF	(Form Feed)
013	015	00D	00001101	CR	(Carriage Return)
014	016	00E	00001110	SO	(Shift Out)
015	017	00F	00001111	SI	(Shift In)
016	020	010	00010000	DLE	(Data Link Escape)
017	021	011	00010001	DC1	(XON) (Device Control 1)
018	022	012	00010010	DC2	(Device Control 2)
019	023	013	00010011	DC3	(XOFF)(Device Control 3)
020	024	014	00010100	DC4	(Device Control 4)
021	025	015	00010101	NAK	(Negative Acknowledgement)
022	026	016	00010110	SYN	(Synchronous Idle)
023	027	017	00010111	ETB	(End of Trans. Block)
024	030	018	00011000	CAN	(Cancel)
025	031	019	00011001	EM	(End of Medium)
026	032	01A	00011010	SUB	(Substitute)
027	033	01B	00011011	ESC	(Escape)
028	034	01C	00011100	FS	(File Separator)
029	035	01D	00011101	GS	(Group Separator)
030	036	01E	00011110	RS	(Request to Send)(Record Separator)
031	037	01F	00011111	US	(Unit Separator)
032	040	020	00100000	SP	(Space)
033	041	021	00100001	!	(exclamation mark)
034	042	022	00100010	"	(double quote)
035	043	023	00100011	#	(number sign)
036	044	024	00100100	$	(dollar sign)
037	045	025	00100101	%	(percent)
038	046	026	00100110	&	(ampersand)
039	047	027	00100111	'	(single quote)
040	050	028	00101000	(	(left/opening parenthesis)
041	051	029	00101001	)	(right/closing parenthesis)
042	052	02A	00101010	*	(asterisk)
043	053	02B	00101011	+	(plus)
044	054	02C	00101100	,	(comma)
045	055	02D	00101101	-	(minus or dash)
046	056	02E	00101110	.	(dot)
047	057	02F	00101111	/	(forward slash)
048	060	030	00110000	0	
049	061	031	00110001	1	
050	062	032	00110010	2	
051	063	033	00110011	3	

Kuva 4.1: Osa ASCII-koodia.

Jos häiriötä ei esiinny, on tiedon siirrossa ja tallentamisessa järkevää pyrkiä minimoimaan $L(C)$, koska se säästää resursseja (aikaa, tilaa). Palaamme keskimääräisen pituuden $L(C)$ minimointiin seuraavassa kappaleessa.



Kuva 4.2: Morse-koodi.

Esimerkki 4.3. Sähkötyksessä käytetyn Morse-koodin aakkosto on

$$\mathcal{D} = \{\cdot, -, \text{kirjainten väli}, \text{sanojen väli}\}.$$

Koodi on esitetty kuvassa 4.2. Huomaa, että lyhimmät koodisanat on varattu (englanninkielessä) useimmin esiintyville kirjaimille e,t,.... Koodisanat ovat pisimmät harvinaisemmille kirjaimille: z, q, ... Katso myös kuvassa 4.3 olevaa englanninkielen kirjainten frekvenssitaulukkoa. ‖

Letter	Frequency	Letter	Frequency
a	0.08167	n	0.06749
b	0.01492	o	0.07507
c	0.02782	p	0.01929
d	0.04253	q	0.00095
e	0.12702	r	0.05987
f	0.02228	s	0.06327
g	0.02015	t	0.09056
h	0.06094	u	0.02758
i	0.06966	v	0.00978
j	0.00153	w	0.02360
k	0.00772	x	0.00150
l	0.04025	y	0.01974
m	0.02406	z	0.00074

Kuva 4.3: Englannin kielen kirjainten suhteelliset frekvenssit (Robert Edward Lewand: *Cryptographical Mathematics*, Mathematical Association of America Press, (2000)).

Määritelmä 4.3. Koodi C on *ei-singulaarinen*, jos se on injektio, eli ehdosta $x_1 \neq x_2$ seuraa, että $C(x_1) \neq C(x_2)$, $x_1, x_2 \in \mathcal{X}$.

Selvästikin ei-singulaarisuus on minimivaatimus dekoodauksen täydelliselle onnistumiselle.

Määritelmä 4.4. Koodi C on *yksikäsitteisesti dekoodattavissa* (yd), jos sen laajennus $\widetilde{C}$ on ei-singulaarinen.

Huomautus. Yksikäsitteisesti dekoodattava koodi on tietysti ei-singulaarinen.

Esimerkki 4.4. Olkoon $\mathcal{X} = \{a, b, c, d\}$, $\mathcal{D} = \{0, 1\}$ ja tarkastellaan koodia C,

x	$C(x)$
a	0
b	010
c	01
d	10

C on selvästi ei-singulaarinen, mutta ei yksikäsitteisesti dekoodattava: $C(b) = C(ca) = C(ad) = 010$. ‖

Olkoon $d_1 \cdots d_k \in \mathcal{D}^*$. Jos $1 \leq l \leq k$, niin $d_1 \cdots d_l$ on $d_1 \cdots d_k$:n *etuliite*.

Määritelmä 4.5. Koodi C on *välitön*, jos minkään symbolin $x \in \mathcal{X}$ koodisana ei ole jonkin toisen symbolin $x' \in \mathcal{X}$ koodisanan etuliite.

Siis: välittömässä koodissa ei ole symboleja $x, x' \in \mathcal{X}$, $x \neq x'$, joille $C(x) = C(x')\, d_1 \cdots d_k$ joillekin $d_1, \ldots, d_k \in \mathcal{D}$.

Esimerkki 4.5. Koodi C_1:

x	$C_1(x)$
a	0
b	100
c	101
d	11

on välitön.

Koodi C_2:

x	$C_2(x)$
a	0
b	01

ei ole välitön, koska 0 on 01:n etuliite. ‖

Huomautus. Välitön koodi on ei-singulaarinen, koska jos $C(x) = C(x')$ ja $x \neq x'$, on $C(x)$ $C(x')$:n etuliite (ja päinvastoin).

Pätee enemmänkin: välitön koodi on yksikäsitteisesti dekoodattavissa. Sillä, olkoon

$$\underbrace{\widetilde{C}(x_1 \cdots x_n)}_{C(x_1)\cdots C(x_n)} = \underbrace{\widetilde{C}(x'_1 \cdots x'_{n'})}_{C(x'_1)\cdots C(x'_{n'})} = d_1 \cdots d_k.$$

Luetaan jonoa $d_1 \cdots d_k$ vasemmalta ja olkoon

$$\begin{cases} C(x_1) = d_1 \cdots d_l \\ C(x'_1) = d_1 \cdots d_{l'}, \end{cases}$$

$1 \leq l, l' \leq k$. Jos $l \neq l'$, on $C(x_1)$ $C(x'_1)$:n etuliite tai päinvastoin. Siis $l = l'$ ja $C(x_1) = C(x'_1)$. C on ei-singulaarinen, joten $x_1 = x'_1$. Näin jatketaan loppuun.

Välitön koodi on tulkittavissa (dekoodattavissa) välittömästi, sitä mukaan kun viestiä luetaan. Yksikäsitteisesti dekodaattavalla koodilla ei näin välttämättä ole.

Esimerkki 4.6. Tarkastellaan koodia C,

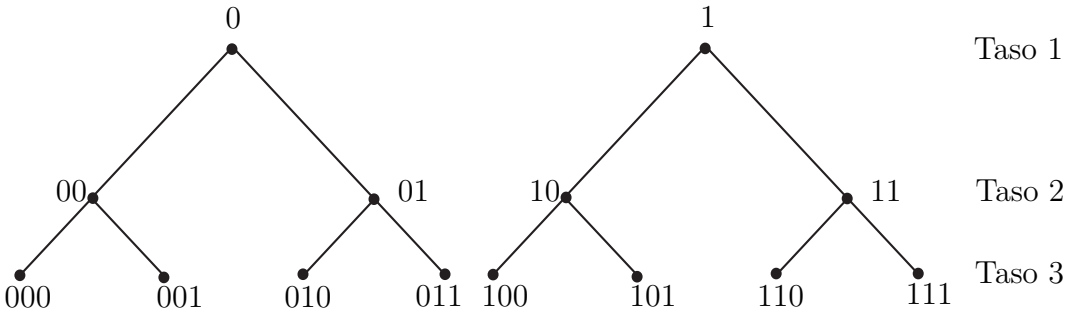
x	$C(x)$
a	0
b	$\underbrace{00\cdots00}_{n}1$

Koodi on selvästi yksikäsitteisesti dekoodattava. Oletetaan, että vastaanotetaan viesti $\underbrace{00\cdots001}_{n+1}$. Selvästi tämä voidaan tulkita vasta, kun viimeinenkin bitti on nähty. ‖

Koodille C on siis voimassa:

$$C \text{ välitön } \Rightarrow C \text{ yksikäsitteisesti dekoodattavissa } \Rightarrow C \text{ ei-singulaarinen.}$$

Implikaatiot *eivät* päde toiseen suuntaan.



Kuva 4.4: Kahteen symboliin perustuvan (binäärisen) koodin tulkinta puurakenteena.

4.2 Kraftin epäyhtälö

Tässä luvussa todistettavan epäyhtälön esitti MIT:n maisterin tutkielmassaan L.G. Kraft vuonna 1949.

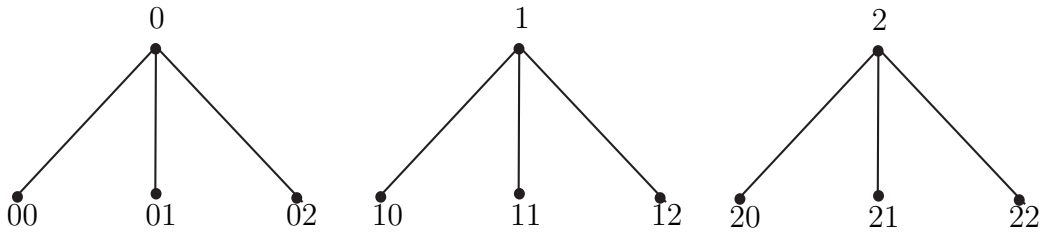
Olkoon $C : \mathcal{X} \to \mathcal{D}^*$ koodi, $|\mathcal{X}| = m$, $|\mathcal{D}| = D$. Merkitään symbolien $x \in \mathcal{X}$ koodisanojen $C(x)$ pituuksia $l_1, \ldots, l_m$.

Jotta koodi olisi välitön, eivät kaikki koodisanat voi olla lyhyitä, sillä lyhyillä koodisanoilla on vaara esiintyä etuliitteinä.

Lause 4.6 (Kraftin epäyhtälö)**.** *On olemassa välitön koodi koodisanojen pituuksilla $l_1, \ldots, l_m$, jos ja vain jos*

$$\sum_{i=1}^{m} D^{-l_i} \leq 1. \tag{4.1}$$

Todistus. Tarkastellaan välitöntä koodia, jolle yleisyyttä rajoittamatta voidaan olettaa, että $\mathcal{D} = \{0, 1, \ldots, D-1\}$ ja $l_1 \leq \cdots \leq l_m$. Koodisanojen voidaan ajatella olevan solmuja puurakenteessa, jonka kertaluku on D ja syvyys l_m (=tasojen lukumäärä). Esimerkiksi, jos $\mathcal{D} = \{0, 1\}$ ja $l_m = 3$, on tilanne kuvan 4.4 mukainen. Vastaavasti, jos $\mathcal{D} = \{0, 1, 2\}$, $l_m = 2$ on tilanne kuvan 4.5 mukainen.



Kuva 4.5: Kolmeen symboliin perustuvan koodin tulkinta puurakenteena.

Koodisanat ovat puun solmuja ja kun koodi on välitön, leikkaantuu koodisanan alipuu pois, koska kyseessä oleva koodisana on alipuun solmujen etuliite. Esimerkiksi koodi $\{0, 10, 111\}$ on välitön ja sitä vastaava binääripuu on esitetty kuvassa 4.6.

Nyt i:s koodisana (pituus, taso l_i) leikkaa pois $D^{l_m - l_i}$ alimman tason "päätesolmua" (lehteä). Päätesolmuja on täydellisessä puussa yhteensä D^{l_m} kappaletta. Kaikki m koodisanaa leikkaavat pois yhteensä $\sum_{i=1}^{m} D^{l_m - l_i}$ päätesolmua. Siten

$$\sum_{i=1}^{m} D^{l_m - l_i} \leq D^{l_m},$$

eli

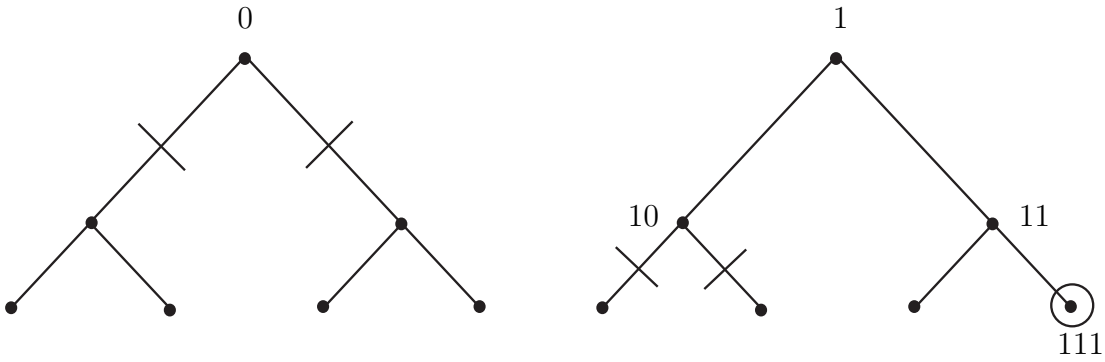
$$\sum_{i=1}^{m} D^{-l_i} \leq 1.$$

Näin tulos (4.1) pätee.

Kääntäen, olkoot $l_1, \ldots, l_m \in \mathbb{N}_+$ sellaisia, että (4.1) pätee. Oletetaan edelleen, että $l_1 \leq \cdots \leq l_m$. Välitön koodi konstruoidaan valitsemalla m solmua kertalukua D olevasta puusta, jonka syvyys on l_m.

Valitaan ensin mikä hyvänsä solmu tasolta l_1 koodisanaksi. Tällöin $D^{l_m - l_1}$ päätesolmua putoaa pois. Jos $m = 1$, lopetetaan. Jos $m \geq 2$, on oletuksen mukaan

$$\sum_{i=1}^{m} D^{-l_i} \leq 1,$$



Kuva 4.6: Välitöntä koodia $\{0, 10, 111\}$ vastaava binääripuu.

joten

$$D^{l_m} \sum_{i=1}^{m} D^{-l_i} \leq D^{l_m},$$

eli

$$\sum_{i=1}^{m} D^{l_m - l_i} \leq D^{l_m},$$

missä

$$\sum_{i=1}^{m} D^{l_m - l_i} = D^{l_m - l_1} + \sum_{i=2}^{m} D^{l_m - l_i}.$$

Siten $D^{l_m - l_1} < D^{l_m}$ ja jäljellä täytyy olla vielä päätesolmuja. Siten tasolla l_2 on jäljellä solmuja (muuten ei ole päätesolmujakaan). Valitaan sieltä solmu koodisanaksi. Jos $m = 2$, lopetetaan.

Jos $m \geq 3$, on nyt

$$D^{l_m - l_1} + D^{l_m - l_2} + \sum_{i=3}^{m} D^{l_m - l_i} \leq D^{l_m},$$

joten $D^{l_m - l_1} + D^{l_m - l_2} < D^{l_m}$ ja jäljellä on vielä päätesolmuja. Siten on tasolla l_3 solmuja. Valitaan niistä yksi koodisanaksi ja jatketaan kuten edellä, kunnes kaikki koodisanat on valittu. □

B. McMillan esitti vuonna 1956 seuraavan Kraftin epäyhtälön yleistyksen.

Lause 4.7 (McMillanin epäyhtälö)**.** *On olemassa yksikäsitteisesti dekoodattava koodi koodisanan pituuksilla $l_1, \ldots, l_m$, jos ja vain jos (4.1) pätee, eli*

$$\sum_{i=1}^{m} D^{-l_i} \leq 1.$$

Todistus. Jos (4.1) pätee, löytyy välitön koodi koodisanojen pituuksilla $l_1, \ldots, l_m$ eli löytyy myös yksikäsitteisesti dekoodattava koodi. Olkoon koodi C sitten yksikäsitteisesti dekoodattavissa. Olkoon edelleen

$$\sum_{i=1}^{m} D^{-l_i} = \sum_{j=1}^{r} m_j D^{-j}, \tag{4.2}$$

missä $r = \max\{l_1, \ldots, l_m\}$ ja

$$m_j = \big|\{i \mid l_i = j\}\big|$$

(j:n pituisten koodisanojen lukumäärä). Kun $n \in \mathbb{N}_+$, saadaan

$$\begin{aligned}\left(\sum_{j=1}^{r} m_j D^{-j}\right)^n &= \sum_{i_1,\ldots,i_n=1}^{r} m_{i_1} D^{-i_1} \cdots m_{i_n} D^{-i_n} \\ &= \sum_{k=n}^{nr} \left(\sum_{i_1+\cdots+i_n=k} m_{i_1} \cdots m_{i_n} \right) D^{-k}.\end{aligned} \tag{4.3}$$

Huomaa, että $1 \le i_1, \ldots, i_n \le r$, joten $n \le i_1 + \cdots + i_n \le nr$. Merkitään

$$N_k = \sum_{i_1+\cdots+i_n=k} m_{i_1} \cdots m_{i_n}.$$

Nyt pätee:

$$N_k = \left| \left\{ x_1 \cdots x_n \mid \sum_{i=1}^{n} l(x_i) = k \right\} \right|,$$

eli N_k on sellaisten n:n pituisten viestien lukumäärä, jotka koodattuna ovat k:n pituisia. Siten $N_k \le D^k$, koska koodi on yksikäsitteisesti dekoodattavissa. Nyt ehdon (4.3) nojalla

$$\left(\sum_{j=1}^{r} m_j D^{-j}\right)^n = \sum_{k=n}^{nr} N_k D^{-k} \le \sum_{k=n}^{nr} 1 = nr - n + 1 \le nr.$$

Siten ehdosta (4.2) saadaan

$$\sum_{i=1}^{m} D^{-l_i} \le (nr)^{\frac{1}{n}} = r^{\frac{1}{n}} n^{\frac{1}{n}}.$$

Väite seuraa, koska $n \in \mathbb{N}_+$ oli mielivaltainen ja

$$\lim_{n\to\infty} r^{\frac{1}{n}} n^{\frac{1}{n}} = \lim_{n\to\infty} r^{\frac{1}{n}} \lim_{n\to\infty} n^{\frac{1}{n}} = 1,$$

sillä $n^{\frac{1}{n}} = e^{\frac{1}{n}\ln n}$ ja $\frac{1}{n}\ln n = -\frac{1}{n}\ln\frac{1}{n} \to 0$, kun $n \to \infty$. □

4.3 Shannonin ensimmäinen lause

Kuten edellä, olkoon $X \sim p(x)$ ja olkoon X:n arvojoukko $\mathcal{X}$, $|\mathcal{X}| = m$. Tarkastellaan koodiaakkostoa $\mathcal{D}$, $|\mathcal{D}| = D$ ja koodeja $C : \mathcal{X} \to \mathcal{D}^*$. Edelleen, merkitään symbolien $x \in \mathcal{X}$ koodisanojen pituuksia $l(x)$ ja todennäköisyyksiä $p(x)$ hieman yksinkertaisemmin

$$l_1, \ldots, l_m, \qquad p_1, \ldots, p_m.$$

Keskimääräinen koodisanan pituus on

$$L = L(C) = \sum_{x \in \mathcal{X}} p(x) l(x) = \sum_{i=1}^{m} p_i l_i.$$

Lause 4.8. *Yksikäsitteisesti dekoodattavissa olevalle koodille pätee*

$$L \geq \frac{H(X)}{\log D} \tag{4.4}$$

ja yhtäsuuruus pätee jos ja vain jos $p_i = D^{-l_i}$, $i = 1, \ldots, m$.

Todistus. Olkoon

$$q_i = \frac{D^{-l_i}}{\sum_{j=1}^{m} D^{-l_j}}, \quad i = 1, \ldots, m.$$

Silloin $\sum_{i=1}^{m} q_i = 1$ ja lauseen 2.8 nojalla saadaan

$$-\sum_{i=1}^{m} p_i \log p_i \leq -\sum_{i=1}^{m} p_i \log q_i$$

eli

$$\begin{aligned} H(X) &\leq -\sum_{i=1}^{m} p_i \log \left[\frac{D^{-l_i}}{\sum_{j=1}^{m} D^{-l_j}} \right] \\ &= -\sum_{i=1}^{m} p_i (-l_i) \log D + \sum_{i=1}^{m} p_i \log \left(\sum_{j=1}^{m} D^{-l_j} \right) \\ &= \log D \sum_{i=1}^{m} l_i p_i + \log \left(\sum_{j=1}^{m} D^{-l_j} \right). \end{aligned}$$

Tässä $\sum_{i=1}^{m} l_i p_i = L$ ja McMillanin epäyhtälön nojalla $\sum_{j=1}^{m} D^{-l_j} \leq 1$, joten $H(X) \leq L \log D$ eli (4.4) pätee.

Yhtälö pätee jos ja vain jos $p_i = q_i$ kaikilla $i = 1, \ldots, m$ (lause 2.8) ja $\sum_{j=1}^{m} D^{-l_j} = 1$. Siten yhtälö pätee jos ja vain jos $p_i = D^{-l_i}$ kaikilla $i = 1, \ldots, m$.

□

Huomautus. Näemme, että optimaalisessa (keskimäärin lyhimmässä) koodissa:

$$p_i \text{ suuri } \leftrightarrow l_i \text{ pieni }, \quad p_i \text{ pieni } \leftrightarrow l_i \text{ suuri }.$$

Huomautus. Voitaisiin määritellä "D-kantainen entropia",

$$H_D(X) = -\sum_{x \in \mathcal{X}} p(x) \log_D p(x),$$

jolloin

$$\log_D t = \frac{\log t}{\log D}$$

ja edelleen

$$H_D(X) = \frac{H(X)}{\log D}.$$

Silloin ehdon (4.4) mukaan $L \geq H_D(X)$.

Oletetaan seuraavassa, että $p_i > 0$, $i = 1, \ldots, m$. Optimaaliset koodisanojen pituudet olisivat edellisen lauseen mukaan sellaiset l_i, joille

$$D^{-l_i} = p_i$$

eli

$$-l_i \log D = \log p_i$$

eli

$$l_i = \frac{\log \frac{1}{p_i}}{\log D}, \quad i = 1, \ldots, m.$$

Ongelma on siinä, että oikean puolen luku ei yleensä ole kokonaisluku. Voidaan kuitenkin ottaa

$$l_i = \left\lceil \frac{\log \frac{1}{p_i}}{\log D} \right\rceil, \quad i = 1, \ldots, m, \tag{4.5}$$

jolloin

$$\frac{\log \frac{1}{p_i}}{\log D} \leq l_i < \frac{\log \frac{1}{p_i}}{\log D} + 1, \quad i = 1, \ldots, m.$$

Tällöin

$$\sum_{i=1}^{m} \frac{p_i \log \frac{1}{p_i}}{\log D} \leq \sum_{i=1}^{m} l_i p_i < \sum_{i=1}^{m} \frac{p_i \log \frac{1}{p_i}}{\log D} + \sum_{i=1}^{m} p_i$$

eli

$$\frac{H(X)}{\log D} \leq L < \frac{H(X)}{\log D} + 1.$$

Kraftin epäyhtälöstä seuraa, että on olemassa välitön koodi, jossa koodisanojen pituudet ovat (4.5), sillä

$$\sum_{i=1}^{m} D^{-\left\lceil \frac{\log \frac{1}{p_i}}{\log D} \right\rceil} \leq \sum_{i=1}^{m} D^{-\frac{\log \frac{1}{p_i}}{\log D}} = \sum_{i=1}^{m} D^{\log_D p_i} = \sum_{i=1}^{m} p_i = 1.$$

Tällaista koodia sanotaan *Shannonin koodiksi*. Sillä siis pääsee yhden päähän keskimäärin lyhimmästä mahdollisesta koodisanan pituudesta.

Harjoitustehtävänä osoitetaan, että on olemassa välitön koodi C^*, joka minimoi keskimääräisen pituuden $L(C)$. Siten on todistettu:

Lause 4.9. *Keskimäärin lyhimmän välittömän koodin keskimääräiselle pituudelle L^* pätee*

$$\frac{H(X)}{\log D} \leq L^* < \frac{H(X)}{\log D} + 1.$$

Lauseiden 4.8 ja 4.9 tuloksia sanotaan usein Shannonin ensimmäiseksi lauseeksi. Muita nimityksiä ovat ”Source Coding Theorem” ja ”Noisless Coding Theorem”.

Huomautus. Jos joillain i on $p_i = 0$, voidaan menetellä seuraavasti:

1. Valitaan positiivisia p_i vastaamaan sellaiset l_i, että
$$\frac{\log \frac{1}{p_i}}{\log D} < l_i \leq \frac{\log \frac{1}{p_i}}{\log D} + 1.$$

2. Silloin $\sum_{p_i>0} D^{-l_i} < 1$, joten voidaan valita l_i:t myös niille i, joilla $p_i = 0$ ja edelleen pätee $\sum_{i=1}^{m} D^{-l_i} \leq 1$.

3. Kraftin epäyhtälöstä saadaan välitön koodi, jolle
$$\frac{H(X)}{\log D} < L \leq \frac{H(X)}{\log D} + 1,$$
koska $p_i = 0$ termit eivät vaikuta mitään.

4. Selvästi C^* löytyy myös nyt: todennäköisyyksiä $p_i = 0$ vastaavien koodisanojen pituuksilla ei ole merkitystä minimoinnissa.

Siten yleiselle $p(x)$ pätee
$$\frac{H(X)}{\log D} \leq L^* \leq \frac{H(X)}{\log D} + 1.$$

Edellä informaatiolähteen tuottamien perättäisten symbolien riippuvuudesta ei oletettu mitään; vain pistetodennäköisyydet $p(x)$ olivat käytössä. Siten teoria sopii vaikka luonnollisen kielen koodaukseen.

Jos perättäiset symbolit X_i ovat riippumattomia, voidaan päästä tehokkaampaan koodaukseen tarkastelemalla symboli*lohkoja*.

Esimerkki 4.7. Olkoon $X \sim p(x)$, $X_1, X_2 \overset{iid}{\sim} p(x)$ ja $Y = (X_1, X_2)$. Koodataan symboleja nyt kahden pituisissa lohkoissa:

$$\boxed{\text{informaatiolähde}} \xrightarrow[Y=(X_1,X_2)]{X_1,X_2} \boxed{\text{kooderi}} \longrightarrow C(Y)$$

Olkoon viestiaakkostona $\mathcal{X} = \{a, b\}$ ja olkoot symbolitodennäköisyydet ja koodaus määritelty seuraavien talukoiden mukaisesti:

x	$p(x)$	$C(x)$
a	$3/4$	0
b	$1/4$	1

y	$p(y)$	$C(y)$
aa	$9/16$	0
ab	$3/16$	10
ba	$3/16$	110
bb	$1/16$	111

Silloin X:n keskimääräinen koodisanan pituus on

$$L = 1 \cdot \frac{3}{4} + 1 \cdot \frac{1}{4} = 1,$$

ja Y:lle se on

$$L = 1 \cdot \frac{9}{16} + 2 \cdot \frac{3}{16} + 3 \cdot \frac{3}{16} + 3 \cdot \frac{1}{16} = \frac{27}{16}.$$

Kuitenkin Y:lle keskimääräinen koodisanan pituus *per alkuperäinen symboli* on vain

$$\frac{27/16}{2} = \frac{27}{32} < 1.$$

‖

Yleisesti, olkoon $X_1, \ldots, X_n \overset{iid}{\sim} p(x)$ ja olkoon $C : \mathcal{X}^n \to \mathcal{D}^*$ koodi satunnaisvektorille $(X_1, \ldots, X_n)$. Olkoon $l(x_1, \ldots, x_n)$ koodisanan $C(x_1, \ldots, x_n)$ pituus ja

$$L_n = L_n(C) = \frac{1}{n} \mathbb{E}\, l(X_1, \ldots, X_n) = \frac{1}{n} \sum_{(x_1,\ldots,x_n) \in \mathcal{X}^n} l(x_1, \ldots, x_n) p(x_1, \ldots, x_n)$$

keskimääräinen koodin pituus *per symboli*. Edellä esitetyn perusteella löytyy välitön koodi, jolle

$$\frac{1}{n} \cdot \frac{H(X_1, \ldots, X_n)}{\log D} \leq L_n < \frac{1}{n} \cdot \frac{H(X_1, \ldots, X_n)}{\log D} + \frac{1}{n}.$$

Tässä lauseen 2.16 nojalla

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) = nH(X),$$

koska $X_1, \ldots, X_n \overset{iid}{\sim} p(x)$. Siten

$$\frac{H(X)}{\log D} \leq L_n < \frac{H(X)}{\log D} + \frac{1}{n}$$

ja keskimääräinen koodin pituus saadaan mielivaltaisen lähelle (D-kantaista) entropiaa, kun $n \to \infty$, eli lohkon koko kasvaa rajatta.

Tässä yhden koodisymbolin tehottomuus esimerkiksi Shannonin koodissa siis oleellisesti jaetaan n:n symbolin kesken tasan.

Yleisemmin pätee:

Lause 4.10. *Olkoon (X_n) stokastinen prosessi. Keskimäärin lyhimmän koodin keskimääräiselle pituudelle per symboli L_n^* pätee*

$$\frac{1}{n} \cdot \frac{H(X_1, \ldots, X_n)}{\log D} \leq L_n^* < \frac{1}{n} \cdot \frac{H(X_1, \ldots, X_n)}{\log D} + \frac{1}{n}.$$

Jos (X_n) on stationaarinen, on

$$\lim_{n \to \infty} L_n^* = \frac{H(\mathcal{X})}{\log D}.$$

Todistus. Seuraa edellä sanotusta ja lauseesta 3.10. □

Esimerkiksi kun $D = 2$, $H(\mathcal{X}) \approx$ pienin keskimäärin tarvittavien bittien lukumäärä per symboli, kun n on suuri.

4.4 Optimaalinen koodaus

Olkoon $p(x) > 0$ kaikilla $x \in \mathcal{X}$. Optimaalista eli keskimäärin lyhintä yksikäsitteisesti dekoodattavaa koodia haettaessa riittää rajoittua välittömiin koodeihin:

Lause 4.11. *Pätee*

$$\min\{L(C) \mid C \text{ yksikäsitteisisti dekoodattava}\} = \min\{L(C) \mid C \text{ välitön}\}.$$

Todistus. Olkoon C^* välitön koodi, jolle

$$L(C^*) = \min\{L(C) \mid C \text{ välitön}\}.$$

Koska välitön koodi on yksikäsitteisesti dekoodattavissa, pätee tietysti

$$\min\{L(C) \mid C \text{ yksikäsitteisesti dekoodattava}\} \leq L(C^*). \tag{4.6}$$

Oletetaan, että löytyy yksikäsitteisesti dekoodattava koodi C', jolle $L(C') < L(C^*)$. Olkoot $l'_1, \ldots, l'_m$ koodisanojen pituudet koodissa C'. Lauseen 4.7 (McMillanin epäyhtälö) nojalla

$$\sum_{i=1}^{m} D^{-l'_i} \leq 1.$$

Mutta silloin lauseen 4.6 (Kraftin epäyhtälö) mukaan on olemassa välitön koodi C'', jossa koodisanojen pituudet ovat $l'_1, \ldots, l'_m$, joten

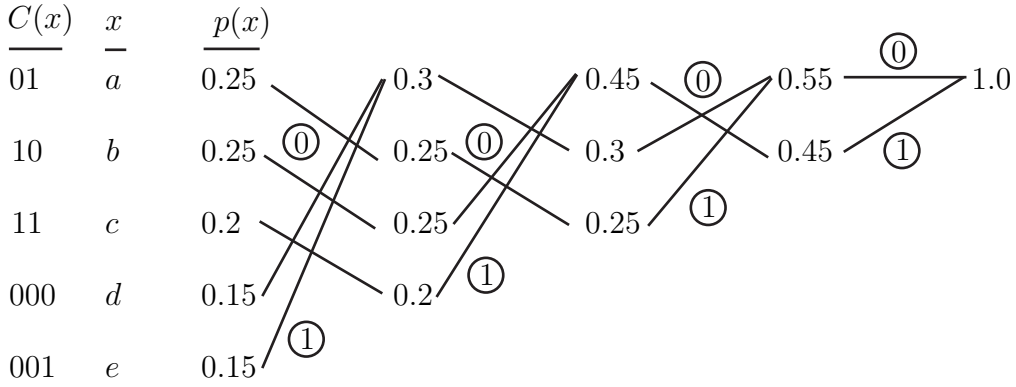
$$L(C'') = L(C') < L(C^*),$$

mikä on ristiriita, koska $L(C^*)$ oli pienin keskimääräinen koodin pituus välittömälle koodille. Siten $L(C) \geq L(C^*)$ kaikille yksikäsitteisesti dekoodattaville koodeille C, jolloin siis

$$\min\{L(C) \mid C \text{ yksikäsitteisesti dekoodattava}\} \geq L(C^*). \tag{4.7}$$

Nyt väite seuraa tulosten (4.6) ja (4.7) nojalla. □

Shannonin koodin generointi on hieman vaivalloista. Lisäksi koodi ei ole välttämättä optimaalinen. Yksinkertaisen optimaalisen välittömän koodin konstruoi ensimmäisenä David Huffman 1951. Koodia käytetään monissa sovelluksissa, mm. JPEG, MP3, ZIP jne. Koodin optimaalisuuden todistus on ”alkeellinen” mutta pitkähkö. Katso esimerkiksi Coverin and Thomasin kirjan [3] lukua 5.8 tai Ashin kirjan [2] lukua 2.6. Huffmanin koodin idea selviää parhaiten muutamalla esimerkillä.



Kuva 4.7: Esimerkki binäärisestä Huffmanin koodista.

Esimerkki 4.8. Olkoon $D = 2$, $\mathcal{D} = \{0, 1\}$ ja $\mathcal{X} = \{a, b, c, d, e\}$. Koodauksen periaate on esitetty kuvassa 4.7. Selvyyden vuoksi on hyvä järjestää symbolit koko ajan todennäköisyyksien mukaan, millä ei kuitenkaan ole vaikutusta lopputulokseen.

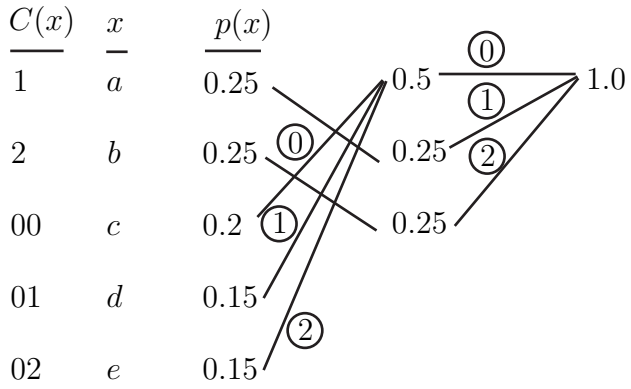
Koodin periaate on seuraava:

1. Yhdistä pienimmät todennäköisyydet kunnes jäljellä on todennäköisyys 1.
2. Koodaa viimeinen pari symboleilla 0,1.
3. Etene lopusta alkuun lisäten haarautumiskohdassa aina 0 ja 1 koodin loppuun.

Esimerkiksi:

$$\begin{array}{lccccccc} e\colon & 0 & \to & 00 & \to & 00 & \to & 001 \\ c\colon & 1 & \to & 1 & \to & 11 & \to & 11 \end{array}$$

Helpompaa on ehkä lähteä *alusta* ja lisätä 0 ja 1 koodin *alkuun* yhdistämiskohdassa:



Kuva 4.8: Esimerkki Huffmanin koodista kolmea koodiaakkosta käyttäen.

$$\begin{array}{lccccccc} e\colon & 1 & \to & 1 & \to & 01 & \to & 001 \\ c\colon & 1 & \to & 1 & \to & 11 & & \end{array}$$

Selvästikin ne x:t, joilla suuri $p(x)$ saavat lyhimmät koodit. Keskimääräinen koodin pituus on

$$L = 2 \cdot 0.25 + 2 \cdot 0.25 + 2 \cdot 0.2 + 3 \cdot 0.15 + 3 \cdot 0.15 = 2.3$$

$\|$

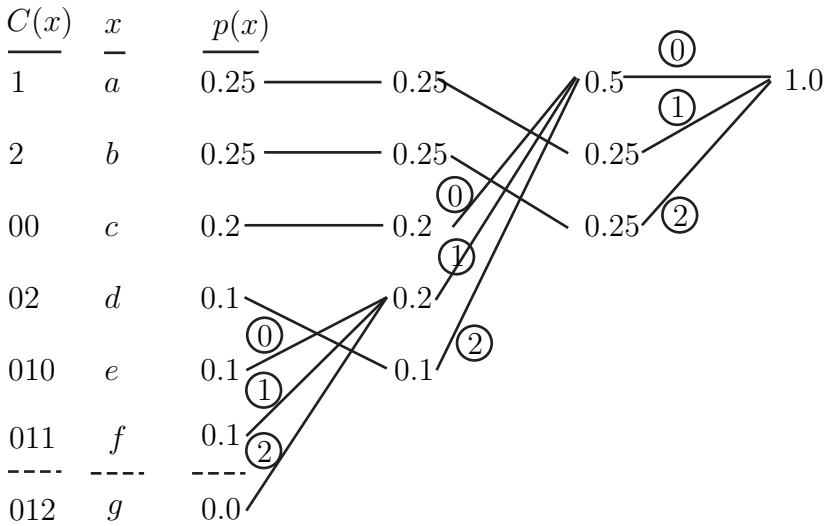
Huffmanin koodin optimaalisuuden todistuksen idea on seuraava:

1. Osoitetaan ensin, että jos koodi on optimaalinen vaiheessa k (k:n yhdistämisen jälkeen), saadaan siitä 0,1 lisäyksillä optimaalinen koodi vaiheessa $k-1$.

2. Viimeisen parin koodi 0,1 on optimaalinen.

3. Suoritetaan induktio lopusta alkuun.

Esimerkki 4.9. Olkoon $D = 3$, $\mathcal{D} = \{0, 1, 2\}$ ja $\mathcal{X}$, $p(x)$ kuten edellä. Koodaus on esitetty kuvassa 4.8.



Kuva 4.9

Nyt yhdistetään aina *kolme* pienintä todennäköisyyttä. Viimeinen kolmikko koodataan 0,1,2. Keskimääräinen koodin pituus on $L = 1.5$ (joukon $\{0, 1, 2\}$ symbolia) ‖

Entä jos lopussa ei ole ”oikeaa” määrää (D kappaletta) todennäköisyyksiä? Silloin lisätään apusymboleja, joiden todennäköisyys on nolla.

Esimerkki 4.10. Olkoon $D = 3$ ja $\mathcal{X} = \{a, b, c, d, e, f\}$. Apusymbolia g käyttäen tehty koodaus on esitetty kuvassa 4.9. Keskimääräinen koodin pituus on nyt $L = 1.7$ ‖

Huffmanin koodin eräs käytännön ongelma on se, että (kun $D = 2$), päästään vain yhden bitin päähän alarajasta $H(X)$:

$$H(X) \leq L^* < H(X) + 1.$$

Esimerkki 4.11. Kun $\mathcal{X} = \{a, b\}$, $\mathcal{D} = \{0, 1\}$, niin optimaalisessa koodissa on tietysti 1 bitti/symboli ja siis $L^* = 1$. Kuitenkin voi hyvin olla, että $H(X) \ll 1$. ‖

Ratkaisu on Huffman-koodata *lohkoja* $(X_1, \ldots, X_n)$, jolloin lauseen 4.10 mukaan

$$\frac{1}{n} \cdot \frac{H(X_1, \ldots, X_n)}{\log D} \leq L_n^* < \frac{1}{n} \cdot \frac{H(X_1, \ldots, X_n)}{\log D} + \frac{1}{n}.$$

Ongelmana on nyt se, että pitää laskea valtava määrä todennäköisyyksiä $p(x_1, \ldots, x_n)$ ja konstruoida iso taulukko. Ja, jos n vaihtuu, kaikki menee uusiksi. Erään ratkaisun tarjoaa ns. *aritmeettinen koodaus* (katso [3] luku 5.10, [5] luku 6.2). Aritmeettista koodausta voidaan myös mukauttaa (adaptoida) dynaamisesti todennäköisyyksien $p(x_1, \ldots, x_n)$ muuttuessa ajan mukana.

Luku 5

Koodaus tiedonsiirrossa

5.1 Kapasiteetti

Tarkastellaan ns. *diskreettiä kanavaa*

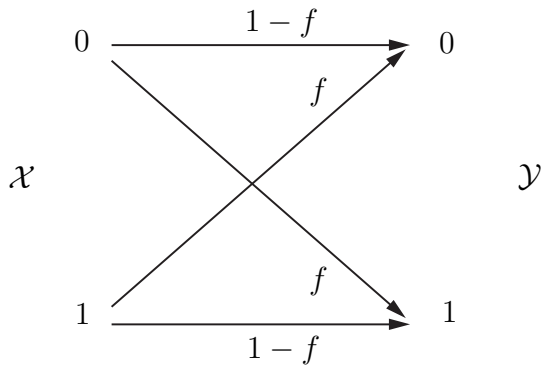
$$x \quad \longrightarrow \quad \boxed{\begin{array}{c}\text{kanava} \\ p(y \mid x)\end{array}} \quad \longrightarrow \quad y$$

Tässä x on nyt syöte *syöteaakkostosta* $\mathcal{X}$, $|\mathcal{X}| = m$ ja y on tuloste *tulosteaakkostosta* $\mathcal{Y}$, $|\mathcal{Y}| = l$. Luvut $p(y \mid x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, toteuttavat ehdot

(*i*) $p(y \mid x) \geq 0$ kaikilla $x \in \mathcal{X}$, $y \in \mathcal{Y}$,

(*ii*) $\sum_{y \in \mathcal{Y}} p(y \mid x) = 1$ kaikilla $x \in \mathcal{X}$.

Tulkinta on se, että $p(y \mid x)$ on tulosteen $y \in \mathcal{Y}$ ehdollinen todennäköisyys, kun syöte on $x \in \mathcal{X}$.



Kuva 5.1: Binäärinen symmetrinen kanava häiriötasolla f.

Esimerkki 5.1. BSK (Binäärinen symmetrinen kanava. Vrt. luku 1.2.1). Nyt $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ ja olkoon kohinataso $0 \leq f \leq 1$ (kuva 5.1).

Ehdollisten todennäköisyyksien muodostama matriisi on nyt

$$\big(p(y \mid x)\big) = \begin{pmatrix} p(0 \mid 0) & p(1 \mid 0) \\ p(0 \mid 1) & p(1 \mid 1) \end{pmatrix} = \begin{pmatrix} 1-f & f \\ f & 1-f \end{pmatrix}.$$

‖

Yleisesti matriisia $(p(y \mid x))$ sanotaan *kanavamatriisiksi*:

$$\begin{array}{cc} & \mathcal{Y} \\ & y \\ \mathcal{X} \quad x & \begin{pmatrix} & \vdots & \\ \cdots & p(y \mid x) & \cdots \\ & \vdots & \end{pmatrix} = \big(p(y \mid x)\big). \end{array}$$

Matriisissa ajatellaan, että $\mathcal{X}$:n ja $\mathcal{Y}$:n alkiot on numeroitu jollain tavalla, $\mathcal{X} = \{x_1, \ldots, x_m\}$ ja $\mathcal{Y} = \{y_1, \ldots, y_l\}$ jolloin kanavamatriisin alkio (i, j) on $p(y_j|x_i)$.

Olkoon sitten syöte satunnaismuuttuja $X \sim p(x)$ ja olkoon X:n arvojoukko $\mathcal{X}$. Määritellään

$$p(x, y) = p(x)p(y \mid x).$$

Silloin $p(x, y)$ on pistetodennäköisyysfunktio, sillä

$$\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)p(y \mid x) = \sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(y \mid x) = \sum_{x\in\mathcal{X}} p(x) = 1.$$

Tämä pistetodennäköisyysfunktio määrää jakauman parille (X, Y) ja silloin

$$\mathbb{P}\{Y = y \mid X = x\} = \frac{p(x,y)}{p(x)} = \frac{p(x)p(y \mid x)}{p(x)} = p(y \mid x).$$

Edelleen, tuloste on nyt myös satunnaismuuttuja Y ja

$$p(y) = \mathbb{P}\{Y = y\} = \sum_{x\in\mathcal{X}} p(x,y) = \sum_{x\in\mathcal{X}} p(x)p(y \mid x),$$

kun $y \in \mathcal{Y}$.

Kanavan läpi kulkevan informaation määrää kuvaa satunnaismuuttujien X ja Y keskinäisinformaatio,

$$I(X;Y) = H(X) - H(X \mid Y),$$

joka kertoo miten paljon Y antaa informaatiota X:stä.

Esimerkki 5.2. Tarkastellaan X:n ja Y:n välisen riippuvuuden ääripäitä.

(*i*) $X = g(Y)$ (X on Y:n funktio). Tällöin harjoitustehtävänä todistetun tuloksen nojalla $H(X \mid Y) = 0$ ja $I(X;Y) = H(X)$. Siten koko X:ssä oleva informaatio siirtyy kanavan läpi. Y:n arvo kertoo tarkalleen mikä X oli.

(*ii*) $X \perp\!\!\!\perp Y$, jolloin $H(X \mid Y) = H(X)$ ja edelleen $I(X;Y) = 0$. Siten Y ei kerro mitään X:stä eikä siis yhtään informaatiota kulje kanavan läpi.

‖

Yleisesti $I(X;Y)$ tietysti riippuu X:n pistetodennäköisyysfunktiosta $p(x)$. Kanavan kapasiteetti määritellään sen läpi kulkevan informaation suurimpana arvona, kun $p(x)$ voi vaihdella.

Määritelmä 5.1. Kanavan *kapasiteetti* on

$$C = \sup_{p(x)} I(X;Y).$$

Tässä siis sup otetaan yli kaikkien pistetodennäköisyysfunktioiden $p(x)$.

Lauseen 2.12 perusteella

$$0 \leq I(X;Y) \leq H(X) \leq \log |\mathcal{X}|,$$

joten

$$0 \leq C \leq \log |\mathcal{X}|.$$

Samoin

$$I(X;Y) = I(Y;X) \leq H(Y) \leq \log |\mathcal{Y}|,$$

joten

$$0 \leq C \leq \log |\mathcal{Y}|.$$

Lisäksi edellä olevassa määritelmässä itseasiassa sup = max, kuten seuraava lause osoittaa.

Lause 5.2. *On olemassa sellainen pistetodennäköisyysfunktio $p^*(x)$, että*

$$C = \sup_{p(x)} I(X;Y) = \max_{p(x)} I(X;Y) = I^*(X;Y),$$

missä $I^*(X;Y)$ on X:n ja Y:n keskinäisinformaatio, kun $X \sim p^*(x)$.

Todistus. Saadaan

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y \mid X) \\
&= -\sum_{y\in\mathcal{Y}} p(y)\log p(y) - \sum_{x\in\mathcal{X}} p(x) H(Y \mid X = x) \\
&= -\sum_{y\in\mathcal{Y}} \left[\sum_{x\in\mathcal{X}} p(x,y)\right] \log \left[\sum_{x'\in\mathcal{X}} p(x',y)\right] \\
&\qquad\qquad - \sum_{x\in\mathcal{X}} p(x)\left[-\sum_{y\in\mathcal{Y}} p(y \mid x)\log p(y \mid x)\right] \\
&= -\sum_{y\in\mathcal{Y}} \left\{ \left[\sum_{x\in\mathcal{X}} p(x)p(y \mid x)\right] \log \left[\sum_{x'\in\mathcal{X}} p(x')p(y \mid x')\right]\right\} \\
&\qquad\qquad + \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x)p(y \mid x)\log p(y \mid x).
\end{aligned}$$

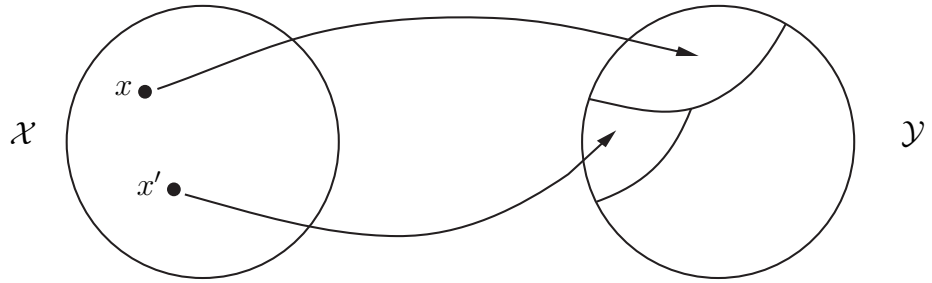
Tämä lauseke pitää maksimoida, kun $(p(x))_{x\in\mathcal{X}} = (p_1,\ldots,p_m) = \mathbf{p} \in S$,

$$S = \left\{ \mathbf{t} = (t_1,\ldots,t_m) \in \mathbb{R}^m \,\middle|\, 0 \le t_1,\ldots,t_m \le 1,\ \sum_{i=1}^{m} t_i = 1 \right\}. \tag{5.1}$$

Merkinnässä $\mathbf{p}$ siis ajatellaan, että symboleille $x \in \mathcal{X}$ on sovittu jokin kiinteä numerointi $\mathcal{X} = \{x_1,\ldots,x_m\}$ ja että $p_i = p(x_i)$. Huomaa, että luvut $p(y \mid x)$ ovat vakioita. Koska $t \mapsto t\log t$ on jatkuva kuvaus $\mathbb{R}_+ \to \mathbb{R}$, on siten $I(X;Y)$ jakauman todennäköisyyksien $\mathbf{p} = (p(x))_{x\in\mathcal{X}}$ jatkuva funktio. S on kompakti (suljettu ja rajoitettu), joten $I(X;Y)$ saavuttaa suurimman arvonsa jossain $\mathbf{p}^* = (p^*(x))_{x\in\mathcal{X}} \in S$. □

Huomautus. Yleensä kapasiteetin C laskeminen annetulle kanavalle on epätriviaali optimointitehtävä.

Huomautus. Myöhemmin tullaan osoittamaan, että tietyille kanaville C on itseasiassa *tiedonsiirtonopeuden* yläraja (bittiä/lähetetty symboli, bittiä/kanavan käyttö), jos halutaan mielivaltaisen pieni virhetodennäköisyys.



Kuva 5.2: Häviötön kanava.

5.2 Esimerkkejä kanavista

5.2.1 Häviötön kanava

Tässä $H(X \mid Y) = 0$ kaikilla $p(x)$. Nyt X määräytyy Y:stä, koska X on Y:n funktio todennäköisyydellä 1 (kuva 5.2). Harjoitustehtävänä on nimittäin osoitettu, että kaikilla y joilla $p(y) > 0$ on tasan yksi x siten, että $p(x, y) > 0$.

Nyt

$$I(X;Y) = H(X) - H(X \mid Y) = H(X),$$

joten

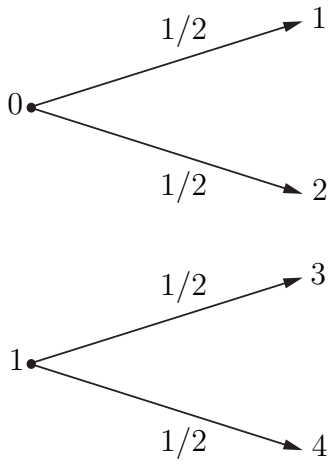
$$C = \log |\mathcal{X}|,$$

ja se saavutetaan, kun $p(x) \sim 1/|\mathcal{X}|$, $x \in \mathcal{X}$.

Esimerkki 5.3. Olkoon $\mathcal{X} = \{0, 1\}$ja $\mathcal{Y} = \{1, 2, 3, 4\}$ ja

$$\big(p(y \mid x)\big) = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

Kapasiteetti on nyt $C = \log |\mathcal{X}| = \log 2 = 1$ ja se saavutetaan, kun $(p(x)) = (1/2, 1/2)$ (ks. kuva 5.3). ‖



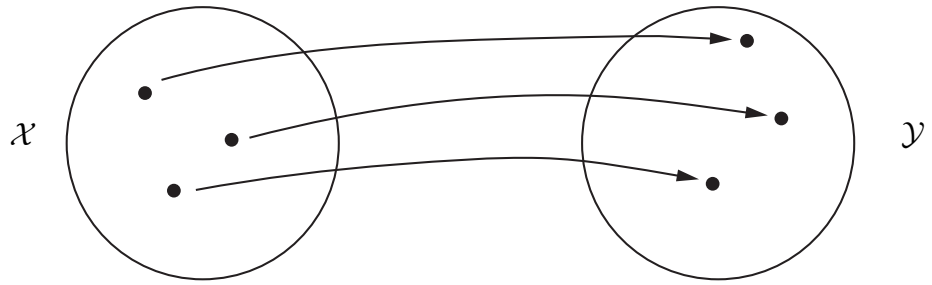
Kuva 5.3: Esimerkki häviöttömästä kanavasta.

5.2.2 Deterministinen kanava

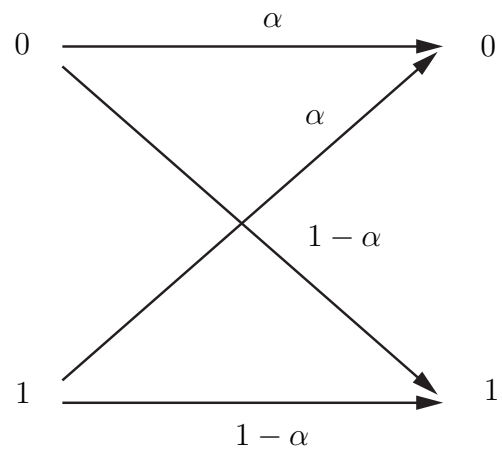
Tässä $p(y \mid x) \in \{0, 1\}$ kaikilla x, y eli y määräytyy x:stä yksikäsitteisesti (todennäköisyydellä yksi). Nyt $H(Y \mid X) = 0$ kaikilla $p(x)$, joten $I(X;Y) = H(Y) \leq \log |\mathcal{Y}|$,

$$C = \log |\mathcal{Y}|$$

ja se saavutetaan, kun $p(x)$ on sellainen, että $p(y) = 1/|\mathcal{Y}|$ kaikilla y.



Kuva 5.4: Häiriötön kanava.



Kuva 5.5: Esimerkki hyödyttömästä kanavasta.

5.2.3 Häiriötön kanava

Tämä kanava on häviötön ja deterministinen (kuva 5.4). Siten $H(X \mid Y) = H(Y \mid X) = 0$ kaikilla $p(x)$. Siis X on Y:n funktio todennäköisyydellä yksi ja Y on X:n funktio todennäköisyydellä yksi. Nyt $I(X;Y) = H(X) - H(X \mid Y) = H(X)$, joten

$$C = \log |\mathcal{X}|$$

ja se saavutetaan, kun $p(x) \sim 1/|\mathcal{X}|$, $x \in \mathcal{X}$.

5.2.4 Hyödytön kanava

Tässä $I(X;Y) = 0$ kaikilla $p(x)$ eli $X \perp\!\!\!\perp Y$ kaikilla $p(x)$, joten $C = 0$. Siten tuloste Y ei kerro mitään syötteestä X.

Esimerkki 5.4. Olkoon $0 < \alpha < 1$ ja tarkastellaan kuvassa 5.5 olevaa kanavaa. Nyt

$$\big(p(y \mid x)\big) = \begin{pmatrix} \alpha & 1-\alpha \\ \alpha & 1-\alpha \end{pmatrix}$$

ja selvästi $p(y \mid x) = p(y)$ kaikilla x, y, joten $X \perp\!\!\!\perp Y$ ja kanava on hyödytön
$\|$

5.2.5 Symmetrinen kanava

Tässä kanavamatriisille pätee:

- matriisin rivit ovat toistensa permutaatioita,
- matriisin sarakkeet ovat toistensa permutaatioita.

Esimerkki 5.5. Binäärisen symmetrisen kanavan (BSK) kanavamatriisi on

$$\begin{pmatrix} 1-f & f \\ f & 1-f \end{pmatrix}$$

eli se toteuttaa yo. symmetrisyyden vatimuksen. Samoin nämä vaatimukset toteuttaa vaikkapa matriisi

$$\begin{pmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{pmatrix}$$

$\|$

Lause 5.3. *Olkoot symmetrisen kanavan kanavamatriisin rivin alkiot $r_1, \ldots, r_l$. Silloin*

$$C = \log |\mathcal{Y}| - H(r_1, \ldots, r_l).$$

Todistus.

$$I(X;Y) = H(Y) - H(Y \mid X) = H(Y) - \sum_{x \in \mathcal{X}} p(x)\, H(Y \mid X = x).$$

Tässä

$$H(Y \mid X = x) = -\sum_{y \in \mathcal{Y}} p(y \mid x) \log p(y \mid x) = -\sum_{j=1}^{l} r_j \log r_j = H(r_1, \ldots, r_l)$$

kaikilla x, joten

$$I(X;Y) \leq \log|\mathcal{Y}| - H(r_1,\ldots,r_l)\sum_{x\in\mathcal{X}} p(x)$$
$$= \log|\mathcal{Y}| - H(r_1,\ldots,r_l).$$

Toisaalta, jos $p(x) = 1/|\mathcal{X}|$ kaikilla x, on

$$p(y) = \sum_{x\in\mathcal{X}} p(x)p(y \mid x) = \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}} p(y \mid x) = \frac{c}{|\mathcal{X}|},$$

missä c on kanavamatriisin sarakkeen summa (sama kaikilla y). Ja, koska

$$1 = \sum_{y\in\mathcal{Y}} p(y) = |\mathcal{Y}| \cdot \frac{c}{|\mathcal{X}|},$$

on $p(y) = 1/|\mathcal{Y}|$ kaikilla $y \in \mathcal{Y}$. Tällöin

$$I(X;Y) = \log|\mathcal{Y}| - H(r_1,\ldots,r_l).$$

□

Huomautus. Edellä olevassa lauseessa riitti, että kanavamatriisin rivit ovat toistensa permutaatioita ja sarakkeiden summat ovat vakio. Tällaista kanavaa sanotaan *heikosti symmetriseksi* ja lause pätee siis myös sille.

Esimerkki 5.6. Binääriselle symmetriselle kanavalle $\mathcal{X} = \mathcal{Y} = \{0,1\}$ ja

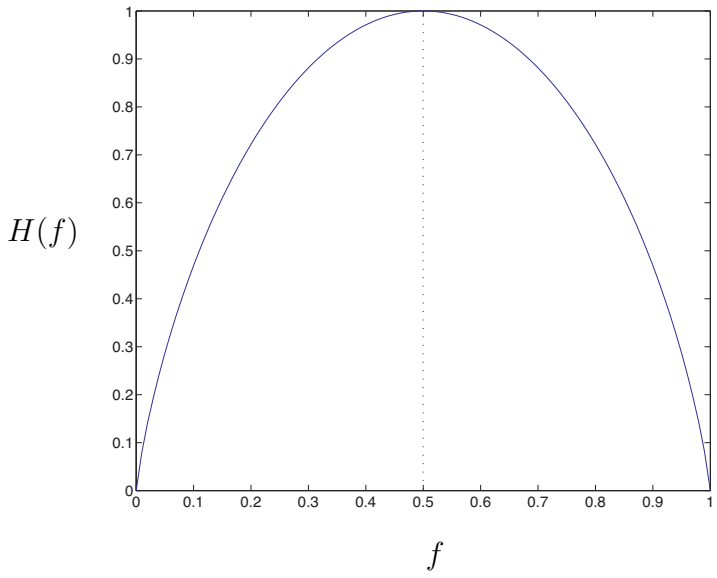
$$C = \log|\mathcal{Y}| - H(1-f,f) = 1 - H(f).$$

(ks. kuva 5.6). ‖

5.3 Kapasiteetin laskeminen

Kapasiteetin laskeminen vaatii yleensä numeerista optimointia. Seuraavassa tätä ongelmaa tarkastellaan vain teoreettisesti.

Palautetaan mieleen, että joukko $B \subset \mathbb{R}^m$ on *konveksi*, jos ehdoista $\mathbf{t}, \mathbf{t}' \in B$, $0 \leq a \leq 1$ seuraa, että $a\mathbf{t} + (1-a)\mathbf{t}' \in B$ (kuva 5.7).



Kuva 5.6: Funktio $H(f)$.

Määritelmä 5.4. Olkoon $B \subset \mathbb{R}^m$ konveksi. Funktio $g : B \to \mathbb{R}$ on *konkaavi*, jos ehdoista $\mathbf{t}, \mathbf{t}' \in B$, $0 \leq a \leq 1$ seuraa, että

$$g\big(a\mathbf{t} + (1-a)\mathbf{t}'\big) \geq ag(\mathbf{t}) + (1-a)g(\mathbf{t}').$$

(Vrt. kuva 5.8).

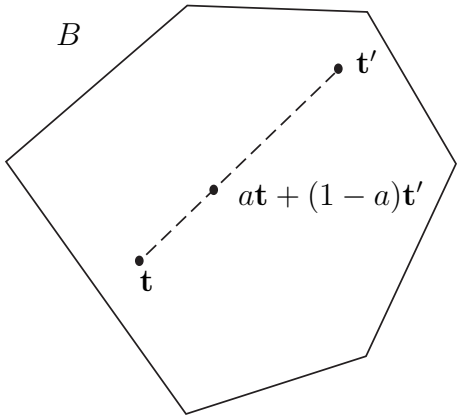
Selvästi joukko (5.1),

$$S = \left\{ \mathbf{t} = (t_1, \ldots, t_m) \in \mathbb{R}^m \,\middle|\, 0 \leq t_1, \ldots, t_m \leq 1, \sum_{i=1}^{m} t_i = 1 \right\}$$

on konveksi. Kuten lauseen 5.2 todistuksessa, ajatellaan $I(X;Y)$:tä vektorin $\mathbf{p} = \big(p(x)\big)_{x \in \mathcal{X}} \in S$ funktiona $S \to \mathbb{R}$.

Lause 5.5. *$I(X;Y)$ on konkaavi.*

Todistus. Olkoot $p_1(x)$ ja $p_2(x)$ pistetodennäköisyysfunktioita, $0 \leq a \leq 1$ ja $p(x) = ap_1(x) + (1-a)p_2(x)$. Merkitään näitä vastaavia suureita $I_1(X;Y)$, $I_2(X;Y)$, $I(X;Y)$ jne.



Kuva 5.7: Konveksi joukko B.

Nyt tulee osoittaa, että

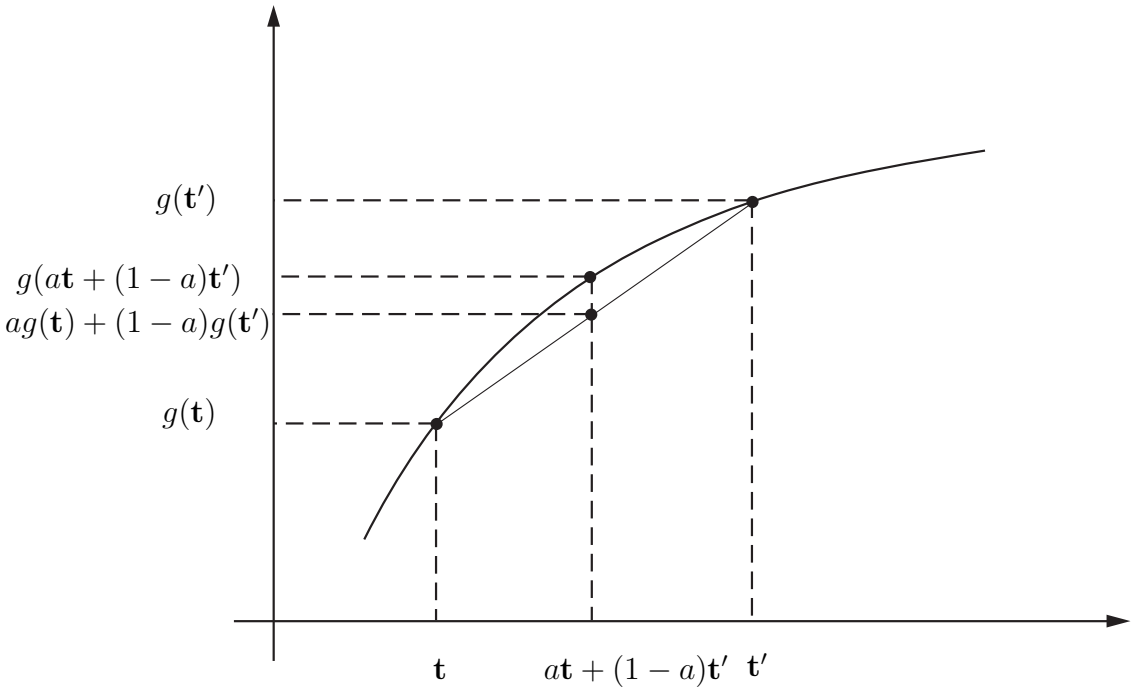
$$I(X;Y) \geq a\, I_1(X;Y) + (1-a)\, I_2(X;Y).$$

Olkoon

$$\begin{aligned}\Delta &= I(X;Y) - a\, I_1(X;Y) - (1-a)\, I_2(X;Y) \\ &= H(Y) - H(Y \mid X) - a[H_1(Y) - H_1(Y \mid X)] \\ &\qquad - (1-a)[H_2(Y) - H_2(Y \mid X)].\end{aligned}$$

Tässä

$$\begin{aligned}H(Y \mid X) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y \mid x) \\ &= -a\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_1(x)p(y \mid x)\log p(y \mid x) \\ &\qquad - (1-a)\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_2(x)p(y \mid x)\log p(y \mid x) \\ &= a\, H_1(Y \mid X) + (1-a)\, H_2(Y \mid X),\end{aligned}$$



Kuva 5.8: Konkaavi funktio g $\mathbb{R}^1$:ssä.

joten

$$\begin{aligned}\Delta &= H(Y) - a\,H_1(Y) - (1-a)\,H_2(Y) \\ &= a\left[-\sum_{y\in\mathcal{Y}} p_1(y)\log p(y) + \sum_{y\in\mathcal{Y}} p_1(y)\log p_1(y)\right] \\ &\quad + (1-a)\left[-\sum_{y\in\mathcal{Y}} p_2(y)\log p(y) + \sum_{y\in\mathcal{Y}} p_2(y)\log p_2(y)\right].\end{aligned}$$

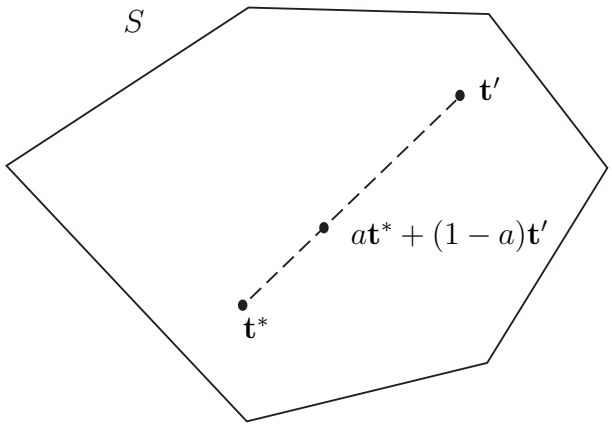
Tässä

$$\left[-\sum_{y\in\mathcal{Y}} p_1(y)\log p(y) + \sum_{y\in\mathcal{Y}} p_1(y)\log p_1(y)\right] \geq 0$$

ja

$$\left[-\sum_{y\in\mathcal{Y}} p_2(y)\log p(y) + \sum_{y\in\mathcal{Y}} p_2(y)\log p_2(y)\right] \geq 0$$

lauseen 2.8 nojalla, joten $\Delta \geq 0$. □



Kuva 5.9: Lemman 5.6 todistuksen havainnollistus.

Olkoon S edellä oleva konveksi joukko (5.1) ja

$$\mathbb{R}^m_+ = \{\mathbf{t} = (t_1, \ldots, t_m) \mid t_1, \ldots, t_m \geq 0\},$$

$$T = \{\mathbf{t} = (t_1, \ldots, t_m) \mid t_1, \ldots, t_m > 0\}. \tag{5.2}$$

Lemma 5.6. *Olkoon* $g : \mathbb{R}^m_+ \to \mathbb{R}$, $g|S$ *konkaavi ja* $g|T$ *differentioituva. Oletetaan, että* $\mathbf{t}^* \in S \cap T$ *ja* $\nabla g(\mathbf{t}^*) = 0$. *Silloin* $g|S$ *saa suurimman arvonsa pisteessä* $\mathbf{t}^*$:

$$g(\mathbf{t}) \leq g(\mathbf{t}^*) \quad \textit{kaikilla } \mathbf{t} \in S.$$

Todistus. Oletetaan, että $\mathbf{t}' \in S$, $\mathbf{t}' \neq \mathbf{t}^*$ ja $g(\mathbf{t}') > g(\mathbf{t}^*)$ (kuva 5.9). Nyt $g|S$ on konkaavi, joten kaikilla $0 \leq a < 1$ pätee

$$\begin{aligned}
&\frac{g\big(\mathbf{t}^* + (1-a)(\mathbf{t}' - \mathbf{t}^*)\big) - g(\mathbf{t}^*)}{1-a} = \frac{g\big(a\mathbf{t}^* + (1-a)\mathbf{t}'\big) - g(\mathbf{t}^*)}{1-a} \\
&\geq \frac{ag(\mathbf{t}^*) + (1-a)g(\mathbf{t}') - g(\mathbf{t}^*)}{1-a} = \frac{(1-a)[g(\mathbf{t}') - g(\mathbf{t}^*)]}{1-a} \\
&= g(\mathbf{t}') - g(\mathbf{t}^*) > 0.
\end{aligned}$$

Kun $a \to 1$, niin $1-a \to 0$ ja nähdään, että g:n suunnattu derivaatta pisteessä $\mathbf{t}^*$ suuntaan $\mathbf{t}' - \mathbf{t}^*$ on aidosti suurempi kuin nolla. Tämä on ristiriita, koska $\nabla g(\mathbf{t}^*) = 0$ ja siis kyseessä oleva suunnattu derivaatta on $(\mathbf{t}' - \mathbf{t}^*) \cdot \nabla g(\mathbf{t}^*) = 0$. □

Lause 5.7. *Olkoon* $|\mathcal{X}| = |\mathcal{Y}| = m$ *ja oletetaan, että kanavamatriisi* $\big(p(y \mid x)\big)$ *on ei-singulaarinen,*

$$\big(p(y \mid x)\big)^{-1} = \big(q(y,x)\big).$$

Olkoon

$$d(x) = \sum_{y\in\mathcal{Y}} q(y,x)\, 2^{\left[-\sum\limits_{x'\in\mathcal{X}} q(y,x')H(Y|X=x')\right]} > 0$$

kaikilla $x \in \mathcal{X}$*. Silloin kanavan kapasiteetti on*

$$C = \log\left\{\sum_{y\in\mathcal{Y}} 2^{\left[-\sum\limits_{x\in\mathcal{X}} q(y,x)H(Y|X=x)\right]}\right\}$$

ja se saavutetaan, kun $p(x) = 2^{-C}\, d(x)$*,* $x \in \mathcal{X}$*.*

Huomautus. Olkoon $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^m$. Silloin $\sum\limits_{y\in\mathcal{Y}} p(y \mid x) = 1$ kaikilla x, joten $\big(p(y \mid x)\big)\mathbf{1} = \mathbf{1}$ ja edelleen $\mathbf{1} = \big(q(y,x)\big)\mathbf{1}$. Siten

$$\sum_{x\in\mathcal{X}} q(y,x) = 1 \quad \text{kaikilla } y,$$

joten

$$\begin{aligned}
&\sum_{x\in\mathcal{X}} p(x) = 2^{-C} \sum_{x\in\mathcal{X}} d(x) \\
&= 2^{-C} \sum_{y\in\mathcal{Y}} \left[\sum_{x\in\mathcal{X}} q(y,x)\right] 2^{\left[-\sum\limits_{x\in\mathcal{X}} q(y,x')H(Y|X=x')\right]} \\
&= 2^{-C} 2^{C} = 1.
\end{aligned}$$

Siten $p(x) = 2^{-C}\, d(x)$ on todella pistetodennäköisyysfunktio.

Todistus. Meidän tulee maksimoida

$$\begin{aligned}
I(X;Y) = &-\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x)p(y \mid x) \log\left[\sum_{x'\in\mathcal{X}} p(x')p(y \mid x')\right] \\
&-\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x)p(y \mid x) \log p(y \mid x)
\end{aligned}$$

(vertaa lauseen 5.2 todistus) vektorin $\mathbf{p} = (p(x)) \in S$ suhteen (vertaa (5.1)). Kaavan oikea puoli on kuitenkin määritelty kaikilla $\mathbf{p} \in \mathbb{R}_+^m$ ja siis määrittelee funktion $g : \mathbb{R}_+^m \to \mathbb{R}$. Edelleen, $g|T$ (vertaa (5.2)) on differentioituva, joten voidaan soveltaa lemmaa 5.6. Koska lopulta halutaan, että kuitenkin $\sum_{x\in\mathcal{X}} p(x) = 1$, tarkastellaan Lagrangen funktiota

$$I(X;Y) + \lambda\Big(\sum_{x\in\mathcal{X}} p(x) - 1\Big) \tag{5.3}$$

ja sen kriittistä pistettä $(p^*(x))$, jolle siis tulee olla

$$\frac{\partial}{\partial p(x)}\left[I(X;Y) + \lambda\Big(\sum_{x'\in\mathcal{X}} p(x') - 1\Big)\right]_{(p(x))=(p^*(x))} = 0 \tag{5.4}$$

kaikilla $x \in \mathcal{X}$. Nyt (5.4):ssä

$$\begin{aligned}
&\frac{\partial}{\partial p(x)}\Big[H(Y) - H(Y \mid X)\Big] + \lambda\frac{\partial}{\partial p(x)}\sum_{x'\in\mathcal{X}} p(x') \\
&= \sum_{y\in\mathcal{Y}} \frac{\partial H(Y)}{\partial p(y)} \cdot \frac{\partial p(y)}{\partial p(x)} + \sum_{y\in\mathcal{Y}} p(y \mid x)\log p(y \mid x) + \lambda \\
&= -\sum_{y\in\mathcal{Y}} p(y \mid x)\log p(y) - \frac{1}{\ln 2}\sum_{y\in\mathcal{Y}} p(y \mid x) - H(Y \mid X = x) + \lambda \\
&= -\frac{1}{\ln 2} - \sum_{y\in\mathcal{Y}} p(y \mid x)\log p(y) - H(Y \mid X = x) + \lambda.
\end{aligned} \tag{5.5}$$

Siten vaaditaan (vrt. (5.4)), että (5.5) häviää, eli

$$\frac{1}{\ln 2} - \lambda + \sum_{y\in\mathcal{Y}} p(y \mid x)\log p(y) = -H(Y \mid X = x) \tag{5.6}$$

kaikilla $x \in \mathcal{X}$. Koska $\sum_{y\in\mathcal{Y}} p(y \mid x) = 1$ kaikilla x, on tämä

$$\sum_{y\in\mathcal{Y}} p(y \mid x)\left[\frac{1}{\ln 2} - \lambda + \log p(y)\right] = -H(Y \mid X = x)$$

kaikilla $x \in \mathcal{X}$. Ratkaisemalla yhtälöryhmä kerroinmatriisin käännöllä saadaan siten

$$\frac{1}{\ln 2} - \lambda + \log p(y) = -\sum_{x\in\mathcal{X}} q(y,x)H(Y \mid X = x)$$

kaikilla $y \in \mathcal{Y}$, missä siis $\big(q(y,x)\big) = \big(p(y \mid x)\big)^{-1}$. Siten

$$2^{\left(\frac{1}{\ln 2}-\lambda\right)} p(y) = 2^{\left[-\sum_{x\in\mathcal{X}} q(y,x) H(Y|X=x)\right]}. \tag{5.7}$$

Edelleen,

$$p(y) = \sum_{x\in\mathcal{X}} p(x,y) = \sum_{x\in\mathcal{X}} p(y \mid x)p(x) \quad \text{kaikilla } y,$$

joten

$$p(x) = \sum_{y\in\mathcal{Y}} q(y,x)p(y) \quad \text{kaikilla } x,$$

ja siis edelleen

$$p(x) = 2^{\lambda-\frac{1}{\ln 2}} \underbrace{\sum_{y\in\mathcal{Y}} q(y,x)\, 2^{\left[-\sum_{x'\in\mathcal{X}} q(y,x') H(Y|X=x')\right]}}_{d(x)>0}.$$

Erityisesti $p(x) > 0$ kaikilla x joten $(p(x)) \in T$ (vrt. (5.2)).

Valitaan nyt sellainen λ, että $\sum\limits_{x\in\mathcal{X}} p(x) = 1$, eli

$$2^{\lambda-\frac{1}{\ln 2}} \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} q(y,x)\, 2^{\left[-\sum_{x'\in\mathcal{X}} q(y,x') H(Y|X=x')\right]} = 1,$$

eli

$$\lambda = -\log \sum_{y\in\mathcal{Y}} 2^{\left[-\sum_{x'\in\mathcal{X}} q(y,x') H(Y|X=x')\right]} + \frac{1}{\ln 2}. \tag{5.8}$$

Olkoon $\big(p^*(x)\big)$ näin saatu pistetodennäköisyysfunktio. Koska funktio (5.3) on selvästi konkaavi, on lemman 5.6 nojalla $\big(p^*(x)\big)$ (5.3):n maksimikohta joukossa S. Selvästi $\big(p^*(x)\big)$ myös maksimoi $I(X;Y)$:n joukossa S, koska $\sum\limits_{x\in\mathcal{X}} p(x) = 1$, kun $\big(p(x)\big) \in S$.

Lasketaan vielä kapasiteetti, eli $I(X;Y)$, kun $p(x) = p^*(x)$. Kaavasta (5.6) saadaan

$$\begin{aligned}\sum_{x\in\mathcal{X}} p^*(x)\big(\frac{1}{\ln 2} - \lambda\big) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p^*(x)p(y \mid x)\log p^*(y) - \sum_{x\in\mathcal{X}} p^*(x)H(Y \mid X = x) \\ &= I^*(X;Y) = C\end{aligned}$$

eli

$$C = \frac{1}{\ln 2} - \lambda.$$

Siten tuloksen (5.8) nojalla

$$C = \log \sum_{y\in\mathcal{Y}} 2^{\left[-\sum\limits_{x\in\mathcal{X}} q(y,x)H(Y|X=x)\right]}$$

ja se saavutetaan, kun $p^*(x) = 2^{\lambda - \frac{1}{\ln 2}}\, d(x) = 2^{-C}\, d(x)$. □

5.4 Muistiton diskreetti kanava

Syötetään nyt diskreettiin kanavaan symboleita $x \in \mathcal{X}$ *lohkoissa*:

$$\mathcal{X}^n \ni (x_1, \ldots, x_n) \longrightarrow \boxed{\text{informaatiokanava}} \longrightarrow (y_1, \ldots, y_n) \in \mathcal{Y}^n$$

Kuten aikaisemmin, tällaista kanavaa sanotaan diskreetiksi, koska symboleja x_i syötetään tietyssä tahdissa, yksi kerrallaan.

Kanavan ominaisuudet määrittelee funktiojoukko

$$p(y_1, \ldots, y_n \mid x_1, \ldots, x_n), \quad n \in \mathbb{N}_+,$$

joille

$$p(y_1, \ldots, y_n \mid x_1, \ldots, x_n) \geq 0$$

kaikilla $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $(y_1, \ldots, y_n) \in \mathcal{Y}^n$ ja

$$\sum_{(y_1,\ldots,y_n)\in\mathcal{Y}^n} p(y_1, \ldots, y_n \mid x_1, \ldots, x_n) = 1$$

kaikilla $(x_1, \ldots, x_n) \in \mathcal{X}^n$. Tässä

$$p(y_1, \ldots, y_n \mid x_1, \ldots, x_n) = p_n(y_1, \ldots, y_n \mid x_1, \ldots, x_n)$$

eli n:n pituiselle syötteelle on oma, indeksistä n riippuva funktionsa, $n \in \mathbb{N}_+$.

Tulkinta on se, että $p(y_1, \ldots, y_n \mid x_1, \ldots, x_n)$ on tulosteiden $(y_1, \ldots, y_n)$ ehdollinen todennäköisyys, kun syöte on $(x_1, \ldots, x_n)$.

Kuten aikaisemmin, merkitään

$$\begin{cases} x^n = (x_1, \ldots, x_n), \\ y^n = (y_1, \ldots, y_n), \quad n \in \mathbb{N}_+. \end{cases}$$

Määritelmä 5.8. Diskreetti kanava on *muistiton*, jos

$$p(y^n \mid x^n) = p(y_1, \ldots, y_n \mid x_1, \ldots, x_n) = \prod_{i=1}^{n} p(y_i \mid x_i)$$

kaikilla $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$ ja $n \in \mathbb{N}_+$.

Siis "p" kussakin $p(y_i \mid x_i)$ on nyt *sama* $p_1(y_i \mid x_i)$. On vain yksi kanavamatriisi $\big(p(y \mid x)\big)$, josta kaikki $p(y^n \mid x^n)$:t voidaan laskea.

Määritellään sitten

$$\begin{aligned} p(y^{n-k} \mid x^n) &= p(y_1, \ldots, y_{n-k} \mid x_1, \ldots, x_n) \\ &\equiv \sum_{y_{n-k+1},\ldots,y_n\in\mathcal{Y}} p(y_1, \ldots, y_n \mid x_1, \ldots, x_n), \end{aligned}$$

$1 \le k \le n-1$ ja

$$\begin{aligned} p(y_n \mid x^n, y^{n-1}) &= p(y_n \mid x_1, \ldots, x_n, y_1, \ldots, y_{n-1}) \\ &\equiv \frac{p(y_1, \ldots, y_{n-1}, y_n \mid x_1, \ldots, x_n)}{p(y_1, \ldots, y_{n-1} \mid x_1, \ldots, x_n)}. \end{aligned} \tag{5.9}$$

Tulkinta on, että

- $p(y^{n-k} \mid x^n)$ = todennäköisyys, että tulosteen $n-k$ ensimmäistä komponenttia ovat $y_1, \ldots, y_{n-k}$, kun syöte on $(x_1, \ldots, x_n)$,
- $p(y_n \mid x^n, y^{n-1})$ = todennäköisyys, että tulosteen n:s komponentti on y_n, kun $n-1$ ensimmäistä ovat $y_1, \ldots, y_{n-1}$ ja syöte on $(x_1, \ldots, x_n)$.

Muistiton diskreetti kanava voidaan nyt luonnehtia seuraavasti:

Lause 5.9. *Diskreetti kanava on muistiton jos ja vain jos*

(i) $p(y_n \mid x^n, y^{n-1}) = p(y_n \mid x_n)$,

(ii) $p(y^{n-k} \mid x^n) = p(y^{n-k} \mid x^{n-k})$

kaikilla $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$, $n \in \mathbb{N}_+$, $1 \leq k \leq n-1$.

Todistus. Oletetaan, että (i) ja (ii) pätevät. Silloin

$$p(y^n \mid x^n) \overset{(5.9)}{=} p(y^{n-1} \mid x^n) p(y_n \mid x^n,\, y^{n-1}) \overset{(i),(ii)}{=} p(y^{n-1} \mid x^{n-1}) p(y_n \mid x_n).$$

Induktiolla saadaan, että $p(y^n \mid x^n) = \prod_{i=1}^{n} p(y_i \mid x_i)$, joten kanava on muistiton.

Käänteinen puoli on harjoitustehtävä. □

Huomautus. Nähdään, että

(i) y_n riippuu vain vastaavasta syötteestä x_n, *ei* aikaisemmista syötteistä tai tulosteista. Kyseessä on siis ”muistiton kanava”.

(ii) Tulosteet eivät riipu *myöhemmin* tulevista syötteistä. Tämä on luonteva ”kausaalisuusehto”.

5.5 Koodaus ja dekoodaus

Luvussa 4 tarkasteltiin viestejä $x_1 \cdots x_k$, $x_i \in \mathcal{X}$, $i = 1, \ldots, k$, joita koodattiin koodisanoilla $C(x_1 \cdots x_k) = d_1 \cdots d_l$, $d_i \in \mathcal{D}$, $i = 1, \ldots, l$.

Tässä luvussa tarkastellaan viestejä, jotka on koodattu n:n pituisilla koodisanoilla. Koodisanat syötetään kanavaan ja siksi käytetään koodiaakkostona jatkossa joukkoa $\mathcal{X}$. Alkuperäisten viestien sisällöillä ei ole merkitystä. Oletetaan vain, että niitä on M kappaletta ja numeroidaan ne $1, 2, \ldots, M$.

Kooderi on kuvaus $C : \{1, \ldots, M\} \to \mathcal{X}^n$,

$$C(j) = x^n(j) = \big(x_1(j), \ldots, x_n(j)\big) \in \mathcal{X}^n,\ j = 1, \ldots, M,$$

$x^n(j) =$ viestin j koodisana.

Kaaviona:

$$\text{viesti } j \longrightarrow \boxed{\text{kooderi } C} \xrightarrow{x^n(j)} \boxed{\begin{matrix}\text{inf.kanava} \\ p(y^n \mid x^n)\end{matrix}} \xrightarrow{y^n} \boxed{\text{dekooderi } g} \longrightarrow \hat{j}$$

Tässä dekoodattu viesti $\hat{j}$ on siis arvaus alkuperäisen viestin j sisällöstä.

Dekoodaus perustuu siihen, että vastaanottaja *tietää* mitä koodisanoja $x^n(j)$ lähetyksessä käytetään.

Dekooderi on jokin deterministinen funktio $g : \mathcal{Y}^n \to \{1, \ldots, M\}$, $\hat{j} = g(y^n)$.

Määritelmä 5.10. Olkoon $M, n \in \mathbb{N}_+$. (M, n)-*koodi* on pari (C, g), missä $C : \{1, \ldots, M\} \to \mathcal{X}^n$ on kooderi ja $g : \mathcal{Y}^n \to \{1, \ldots, M\}$ on dekooderi.

Koodisanaa $C(j) = x^n(j)$ vastaavan tulosteen y^n jakaumaa kuvaa $p\big(y^n \mid x^n(j)\big)$. Dekooderi g tekee *virheen*, jos $g(y^n) \neq j$, kun syöte oli $x^n(j)$.

Määritelmä 5.11. (M, n)-koodin (C, g) *viestin* j *virhetodennäköisyys* on

$$\lambda_j = \sum_{\substack{y^n \\ g(y^n) \neq j}} p\big(y^n \mid x^n(j)\big). \tag{5.10}$$

Siis λ_j on todennäköisyys, että viesti j dekoodataan väärin.

Määritelmä 5.12. (M, n)-koodin (C, g) *keskimääräinen virhetodennäköisyys* on

$$\bar{\lambda} = \frac{1}{M} \sum_{j=1}^{M} \lambda_j.$$

Määritelmä 5.13. (M, n)-koodin (C, g) *maksimaalinen virhetodennäköisyys* on

$$\lambda^{(n)} = \max\{\lambda_1, \ldots, \lambda_M\}.$$

Huomautus. Selvästi $\bar{\lambda} \leq \lambda^{(n)}$.

Jos viestit j valitaan satunnaisesti (ajatusta tullaan jatkossa käyttämään todistuksissa) jonkin jakauman mukaan,

$$J \sim p(j),$$

voidaan myös määritellä virhetodennäköisyys

$$P_e = \mathbb{P}\{g(Y^n) \neq J\},$$

eli todennäköisyys, että satunnaisesti valittu viesti dekoodataan väärin. Parilla (J, Y^n) on nyt määritelmän mukaan pistetodennäköisyysfunktio

$$p(j, y^n) \equiv p(j)p(y^n \mid j) \equiv p(j)p(y^n \mid C(j)) = p(j)p(y^n \mid x^{(n)}(j)).$$

Nyt voidaan laskea myös satunnaismuuttujan (J, Z) jakauma, kun $Z = g(Y^n)$:

$$\begin{aligned} &p(j, z) = \mathbb{P}\{J = j \text{ ja } g(Y^n) = z\} = \\ &\mathbb{P}\{J = j \text{ ja } Y^n \in g^{-1}(z)\} = \sum_{y^n \in g^{-1}(z)} p(j, y^n). \end{aligned}$$

Virhetodennäköisyydelle saadaan:

$$\begin{aligned} P_e &= \sum_{j=1}^{M} \mathbb{P}\{g(Y^n) \neq J \text{ ja } J = j\} \\ &= \sum_{j=1}^{M} \mathbb{P}\{J = j\}\, \mathbb{P}\{g(Y^n) \neq j \mid J = j\} \\ &= \sum_{j=1}^{M} p(j)\lambda_j. \end{aligned}$$

Huomautus. Jos $p(j) = 1/M$, $j = 1, \ldots, M$ (J:llä tasainen jakauma), niin

$$P_e = \frac{1}{M} \sum_{j=1}^{M} \lambda_j = \bar{\lambda}.$$

Kun vastaanotetaan y^n, optimaalinen (pienin virhetodennäköisyys P_e) dekooderi $g^* : \mathcal{Y}^n \to \{1, \ldots, M\}$ on $g^*(y^n) = \hat{j}$, jossa

$$\max_j \mathbb{P}\{J = j \mid Y^n = y^n\} = \mathbb{P}\{J = \hat{j} \mid Y^n = y^n\}.$$

Perustellaan tämän dekooderin optimaalisuus. Ensinnäkin Bayesin kaavan nojalla

$$\mathbb{P}\{J = j \mid Y^n = y^n\} = \frac{\mathbb{P}\{J = j\}\, \mathbb{P}\{Y^n = y^n \mid J = j\}}{\mathbb{P}\{Y^n = y^n\}} = \frac{p(j)\, p\big(y^n \mid x^n(j)\big)}{p(y^n)},$$

eli $\hat{j} = g^*(y^n)$ maksimoi tulon $p(j)\, p\big(y^n \mid x^n(j)\big)$. Siten, jos g on mikä hyvänsä dekooderi,

$$p\big(g^*(y^n)\big)\, p\big(y^n \mid x^n(g^*(y^n))\big) \geq p\big(g(y^n)\big)\, p\big(y^n \mid x^n(g(y^n))\big), \tag{5.11}$$

kun $y^n \in \mathcal{Y}$, $j = 1, \ldots, M$. Siis kaikilla j,

$$\begin{aligned}
1 - P_e &= \sum_{j=1}^{M} p(j) - \sum_{j=1}^{M} p(j)\lambda_j \\
&= \sum_{j=1}^{M} p(j)(1 - \lambda_j) \overset{(5.10)}{=} \sum_{j=1}^{M} p(j) \sum_{\substack{y^n \\ g(y^n)=j}} p\big(y^n \mid x^n(j)\big) \\
&= \sum_{j=1}^{M} \sum_{g(y^n)=j} p(j)\, p\big(y^n \mid x^n(j)\big) \\
&= \sum_{y^n \in \mathcal{Y}^n} p\big(g(y^n)\big)\, p\big(y^n \mid x^n(g(y^n))\big) \\
&\overset{(5.11)}{\leq} \sum_{y^n \in \mathcal{Y}^n} p\big(g^*(y^n)\big)\, p\big(y^n \mid x^n(g^*(y^n))\big) \\
&= 1 - P_e^*,
\end{aligned}$$

missä P_e^* on dekooderin g^* virhe. Siten

$$P_e^* \leq P_e$$

ja g^* on todellakin optimaalinen.

Tätä dekooderia on kuitenkin vaikea analysoida ja siksi tulemme käyttämään dekoodauksessa tyypillisiä joukkoja. Tällainen dekooderi on kuitenkin asymptoottisesti optimaalinen. Ideana on asettaa $g(y^n) = j$, kun $(x^n(j), y^n)$ on ”yhteistyypillinen”.

5.6 Yhteistyypillisyys

Olkoon $(X^n, Y^n) \sim p(x^n, y^n)$ ja

$$p(x^n) = \sum_{y^n \in \mathcal{Y}^n} p(x^n, y^n), \quad p(y^n) = \sum_{x^n \in \mathcal{X}^n} p(x^n, y^n).$$

Määritelmä 5.14. Olkoon $n \in \mathbb{N}_+$ ja $\varepsilon > 0$. *Yhteistyypillisten jonojen* joukko $A_\varepsilon^{(n)}$ on

$$\begin{aligned} A_\varepsilon^{(n)} = \big\{(x^n, y^n) \,|&2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}, \\ &2^{-n(H(Y)+\varepsilon)} \leq p(y^n) \leq 2^{-n(H(Y)-\varepsilon)}, \\ &2^{-n(H(X,Y)+\varepsilon)} \leq p(x^n, y^n) \leq 2^{-n(H(X,Y)-\varepsilon)}\big\}. \end{aligned}$$

Huomautus. Kun (x^n, y^n) on yhteistyypillinen, ovat x^n ja y^n tyypillisiä määritelmän 3.1 mielessä.

Seuraava tulos on AEP parille (X, Y).

Lause 5.15. *Olkoon* $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} p(x, y)$ *ja* $\varepsilon > 0$. *Silloin yhteistyypilliselle joukolle* $A_\varepsilon^{(n)}$ *pätee:*

(*i*) $\lim\limits_{n\to\infty} \mathbb{P}\big\{(X^n, Y^n) \in A_\varepsilon^{(n)}\big\} = 1,$

(*ii*) $|A_\varepsilon^{(n)}| \leq 2^{n(H(X,Y)+\varepsilon)}$ *kaikilla* n *ja jos* $\delta > 0$, *on*

$$|A_\varepsilon^{(n)}| \geq (1-\delta)\, 2^{n(H(X,Y)-\varepsilon)},$$

kun n *on riittävän suuri.*

(*iii*) *Jos* $\widetilde{X}^n \sim p(x^n)$, $\widetilde{Y}^n \sim p(y^n)$ (*siis* $\widetilde{X}^n \sim X^n$, $\widetilde{Y}^n \sim Y^n$) *ja* $\widetilde{X}^n \perp\!\!\!\perp \widetilde{Y}^n$, *on*

$$\mathbb{P}\big\{(\widetilde{X}^n, \widetilde{Y}^n) \in A_\varepsilon^{(n)}\big\} \leq 2^{-n(I(X;Y)-3\varepsilon)}$$

kaikilla n *ja jos* $\delta > 0$, *on*

$$\mathbb{P}\big\{(\widetilde{X}^n, \widetilde{Y}^n) \in A_\varepsilon^{(n)}\big\} \geq (1-\delta)\, 2^{-n(I(X;Y)+3\varepsilon)},$$

kun n *on riittävän suuri.*

Huomautus. Olkoon n suuri. Silloin kohdan (i) mukaan tyypillinen syöte X^n tuottaa tavallisesti tyypillisen tulosteen Y^n ja yhteistyypillisen parin

(X^n, Y^n):

$$\begin{aligned} 1 &\geq \mathbb{P}\{Y^n \text{ tyypillinen } \mid X^n \text{ tyypillinen } \} \\ &= \frac{\mathbb{P}\{Y^n \text{ tyypillinen ja } X^n \text{ tyypillinen } \}}{\mathbb{P}\{X^n \text{ tyypillinen}\}} \geq \frac{\mathbb{P}\big\{(X^n, Y^n) \in A_\varepsilon^{(n)}\big\}}{\mathbb{P}\{X^n \text{ tyypillinen}\}} \approx \frac{1}{1} = 1, \end{aligned}$$

kun apuna käytetään lauseita 3.2 ja 5.15. Siten

$$\mathbb{P}\{Y^n \text{ tyypillinen } \mid X^n \text{ tyypillinen } \} \approx 1.$$

Lisäksi epäyhtälöketjun oikeanpuoleinen suhde on

$$\mathbb{P}\{(X^n, Y^n) \text{ tyypillinen} \mid X^n \text{ tyypillinen}\},$$

joten myös se on likimain 1. Samoin tyypillinen Y^n vastaa tavallisesti tyypillistä X^n:ää ja yhteistyypillistä paria (X^n, Y^n). Toisaalta kohdan (iii) mukaan riippumattomasti valitut jonot X^n ja Y^n *eivät* tavallisesti ole yhteistyypillisiä.

Todistus.

(i) Olkoon

$$\begin{aligned} B_1^{(n)} &= \big\{\big|-\tfrac{1}{n}\log p(X^n) - H(X)\big| > \varepsilon\big\}, \\ B_2^{(n)} &= \big\{\big|-\tfrac{1}{n}\log p(Y^n) - H(Y)\big| > \varepsilon\big\}, \\ B_3^{(n)} &= \big\{\big|-\tfrac{1}{n}\log p(X^n, Y^n) - H(X,Y)\big| > \varepsilon\big\}. \end{aligned}$$

Nyt iid-oletuksen nojalla

$$p(x^n) = \prod_{i=1}^n p(x_i),\ p(y^n) = \prod_{i=1}^n p(y_i),\ p(x^n, y^n) = \prod_{i=1}^n p(x_i, y_i),$$

joten

$$-\frac{1}{n}\log p(X^n) = -\frac{1}{n}\sum_{i=1}^n \log p(X_i) = \frac{1}{n}\sum_{i=1}^n \big[-\log p(X_i)\big].$$

Heikon suurten lukujen lain nojalla

$$-\frac{1}{n}\log p(X^n) \longrightarrow \mathbb{E}\big[-\log p(X)\big] = H(X) \ \text{ stokastisesti.}$$

Samoin

$$-\frac{1}{n}\log p(Y^n) \longrightarrow H(Y),\ -\frac{1}{n}\log p(X^n, Y^n) \longrightarrow H(X,Y) \text{ stokastisesti.}$$

Siten $\lim_{n\to\infty} \mathbb{P}\big(B_i^{(n)}\big) = 0$, $i = 1, 2, 3$.

Toisaalta

$$\big\{(X^n, Y^n) \in A_\varepsilon^{(n)}\big\} = \big(B_1^{(n)} \cup B_2^{(n)} \cup B_3^{(n)}\big)^c = \big(B_1^{(n)}\big)^c \cap \big(B_2^{(n)}\big)^c \cap \big(B_3^{(n)}\big)^c,$$

(tässä "c" merkitsee joukon komplementtia) koska esimerkiksi $\big(B_1^{(n)}\big)^c$:ssä

$$\left|-\frac{1}{n}\log p(X^n) - H(X)\right| \le \varepsilon,$$

eli

$$2^{-n(H(X)+\varepsilon)} \le p(X^n) \le 2^{-n(H(X)-\varepsilon)}$$

ja vastaavasti $p(Y^n)$:lle ja $p(X^n, Y^n)$:lle. Siten

$$\begin{aligned} 1 \ge \mathbb{P}\big\{(X^n, Y^n) \in A_\varepsilon^{(n)}\big\} &= 1 - \mathbb{P}\big(B_1^{(n)} \cup B_2^{(n)} \cup B_3^{(n)}\big) \\ &\ge 1 - \Big[\mathbb{P}\big(B_1^{(n)}\big) + \mathbb{P}\big(B_2^{(n)}\big) + \mathbb{P}\big(B_3^{(n)}\big)\Big] \longrightarrow 1, \end{aligned}$$

kun $n \to \infty$, koska $\mathbb{P}\big(B_i^{(n)}\big) \to 0$, kun $n \to \infty$, $i = 1, 2, 3$.

(*ii*) Käyttämällä apuna joukon $A_\varepsilon^{(n)}$ määritelmää saadaan

$$\begin{aligned} 1 &= \sum_{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n} p(x^n, y^n) \ge \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n, y^n) \\ &\ge \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} 2^{-n(H(X,Y)+\varepsilon)} = |A_\varepsilon^{(n)}|\, 2^{-n(H(X,Y)+\varepsilon)}, \end{aligned}$$

joten

$$|A_\varepsilon^{(n)}| \le 2^{n(H(X,Y)+\varepsilon)}.$$

Ja, jos $\delta > 0$ ja n on niin suuri, että $\mathbb{P}\big\{(X^n, Y^n) \in A_\varepsilon^{(n)}\big\} \ge 1 - \delta$ (vrt. (*i*)), on

$$1 - \delta \le \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n, y^n) \le |A_\varepsilon^{(n)}|\, 2^{-n(H(X,Y)-\varepsilon)},$$

joten

$$|A_\varepsilon^{(n)}| \geq (1-\delta)\, 2^{n(H(X,Y)-\varepsilon)}.$$

(*iii*) Nyt siis $\widetilde{X}^n \sim p(x^n)$, $\widetilde{Y}^n \sim p(y^n)$, $\widetilde{X}^n \perp\!\!\!\perp \widetilde{Y}^n$, joten $(\widetilde{X}^n, \widetilde{Y}^n) \sim p(x^n)p(y^n)$. Siten

$$\begin{aligned}
\mathbb{P}\big\{(\widetilde{X}^n, \widetilde{Y}^n) \in A_\varepsilon^{(n)}\big\} &= \sum_{(x^n,y^n)\in A_\varepsilon^{(n)}} p(x^n)p(y^n) \\
&\leq \sum_{(x^n,y^n)\in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)}\, 2^{-n(H(Y)-\varepsilon)} \\
&\overset{(ii)}{\leq} 2^{n(H(X,Y)+\varepsilon)}\, 2^{-n(H(X)-\varepsilon)}\, 2^{-n(H(Y)-\varepsilon)} \\
&= 2^{-n(I(X;Y)-3\varepsilon)},
\end{aligned}$$

koska $-H(X,Y) + H(X) + H(Y) = I(X;Y)$.

Viimein, kun $\delta > 0$ ja n on riittävän suuri, on $A_\varepsilon^{(n)}$:n määritelmän ja kohdan (*ii*) nojalla

$$\begin{aligned}
\mathbb{P}\big\{(\widetilde{X}^n, \widetilde{Y}^n) \in A_\varepsilon^{(n)}\big\} &= \sum_{(x^n,y^n)\in A_\varepsilon^{(n)}} p(x^n)p(y^n) \\
&\geq (1-\delta)\, 2^{n(H(X,Y)-\varepsilon)}\, 2^{-n(H(X)+\varepsilon)}\, 2^{-n(H(Y)+\varepsilon)} \\
&= (1-\delta)\, 2^{-n(I(X;Y)+3\varepsilon)}.
\end{aligned}$$

□

Määritelmä 5.16. (M,n)-koodin (*tiedonsiirto*)*nopeus* on

$$R = \frac{\log M}{n} \quad \text{(bittiä per symboli/kanavan käyttö)}.$$

Tämä määritelmä voidaan tulkita seuraavasti. Jos viestejä on $M = 2^l$ kappaletta, on $R = l/n$. Siten virheettömässä tiedonsiirrossa koodilla voidaan lähettää R bittiä per syötesymboli kun ajatellaan, että koodisanan $x^{(n)}(j) \in \mathcal{X}^n$ mukana kulkee tieto indeksistä $j \in \{1, \ldots, 2^l\}$, jonka arvot voidaan esittää l:n bitin binääriluvuilla.

Määritelmä 5.17. Kanavan (tiedonsiirto)nopeus R on *saavutettavissa*, jos on olemassa $\big(\lceil 2^{nR} \rceil, n\big)$-koodit (C_n, g_n), $n \in \mathbb{N}_+$, joille $\lim\limits_{n\to\infty} \lambda^{(n)} = 0$.

Tässä siis $\lambda^{(n)}$ on koodin (C_n, g_n) maksimivirhe.

Huomautus. Virheettömässä tiedonsiirrossa $\big(\lceil 2^{nR} \rceil, n\big)$-koodin koodisanan avulla lähetettävissä olevien bittien määrä on välillä $[nR, nR+1[$, koska

$$2^{nR} \leq \lceil 2^{nR} \rceil < 2^{nR} + 1 < 2^{nR} + 2^{nR} = 2 \cdot 2^{nR} = 2^{nR+1}.$$

Siten per symboli siirtyvien bittien lukumäärä on välillä $\big[R, R + \frac{1}{n}\big[$, eli $\approx R$, kun n on suuri.

Tulemme osoittamaan, että

$$\sup\{R \mid \text{ kanavan tiedonsiirtonopeus } R \text{ saavutettavissa}\} = C,$$

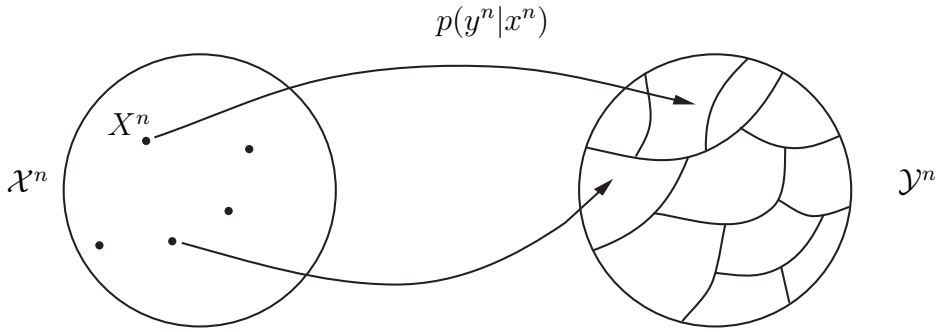
missä C on kanavan kapasiteetti.

Yhteistyypillisyyten perustuva idea hyvälle koodille on seuraava (kuva 5.10). Syöte ja tuloste ovat yhteistyypillisiä suurella todennäköisyydellä (AEP:n kohta (i)). Toisaalta, olkoon syöte X^n ja $Y^n \sim p(y^n)$ satunnainen. Silloin AEP:n kohdan (iii) mukaan todennäköisyys, että (X^n, Y^n) on yhteistyypillinen on noin $2^{-nI(X;Y)}$. Yksi X^n siis "varaa" noin $100 \cdot 2^{-nI(X;Y)}$ prosenttia Y^n:ien todennäköisyysmassasta (X^n:n kanssa yhteistyypilliset Y^n:t). Siten $2^{nI(X;Y)}$ tuntuisi olevan maksimaalinen X^n:in lukumäärä, kun halutaan, että varaukset "eivät mene päällekkäin" eli että X^n voidaan dekoodata tulosteesta Y^n yhteistyypillisyyten perustuen.

Käyttämällä näitä syötteitä X^n koodisanoina on

$$R_n = \frac{\log 2^{nI(X;Y)}}{n} = I(X;Y) \leq C,$$

joten $R_n \leq C$. Itseasiassa käy niin, että $R_n \to C$ ja $\lambda^{(n)} \to 0$, kun $n \to \infty$.



Kuva 5.10: Tehokkaan koodauksen idea.

5.7 Shannonin toinen lause

Lause tunnetaan myös nimellä ”informaatioteorian peruslause” (The Fundamental Theorem of Information Theory), ”Channel Capacity Theorem”, ”Channel Coding Theorem” ja ”Noisy Channel Coding Theorem”.

Lause 5.18. *Tarkastellaan diskreettiä muistitonta kanavaa jonka kapasiteetti on C ja olkoon $0 \leq R < C$. Silloin on olemassa $(\lceil 2^{nR} \rceil, n)$-koodit (C_n, g_n), joille maksimivirhe $\lambda^{(n)} \to 0$, kun $n \to \infty$. Kääntäen, jos $R \geq 0$ ja $(\lceil 2^{nR} \rceil, n)$-koodeille (C_n, g_n) pätee $\lambda^{(n)} \to 0$, kun $n \to \infty$, on välttämättä $R \leq C$.*

Seuraus 5.19. *Diskreetille muistittomalle kanavalle pätee*

$$\sup\{R \mid \text{ kanavan tiedonsiirtonopeus } R \text{ saavutettavissa}\} = C.$$

Todistus. Seuraa suoraan lauseesta 5.18. □

Todistus. Todistetaan sitten itse lause 5.18.

Selvästi voidaan olettaa, että $R > 0$. Osoitetaan ensin, että jokainen nopeus $0 < R < C$ on saavutettavissa. Olkoon $n \in \mathbb{N}_+$ ja $\varepsilon > 0$. Merkitään $M_n = \lceil 2^{nR} \rceil$.

Olkoon $C_n : \{1, \ldots, M_n\} \to \mathcal{X}^n$ kooderi,

$$C_n(j) = x^n(j), \quad j = 1, \ldots, M_n.$$

Kooderiin C_n liitetään dekooderi $g_n : \mathcal{Y}^n \to \{1, \ldots, M_n\}$, jolle

$$g_n(y^n) = \hat{j},$$

jos $\big(x^n(\hat{j}), y^n\big) \in A_\varepsilon^{(n)}$ (yhteistyypillisiä) ja $\big(x^n(k), y^n\big) \notin A_\varepsilon^{(n)}$ kaikilla $k \neq \hat{j}$ (mikään muu $x^n(k)$ ei ole y^n:n kanssa yhteistyypillinen). Jos

$$\big|\big\{k \,\big|\, \big(x^n(k), y^n\big) \in A_\varepsilon^{(n)}\big\}\big| \neq 1, \quad (0 \text{ tai } \geq 2)$$

asetetaan (mielivaltaisesti) $g_n(y^n) = 1$.

Valitaan jokin pistetodennäköisyysfunktio $p(x)$ ja asetetaan

$$p(x^n) = \prod_{i=1}^{n} p(x_i),$$

$x^n = (x_1, \ldots, x_n)$. Silloin syöte-tuloste parille (X^n, Y^n) saadaan pistetodennäköisyysfunktio

$$p(x^n, y^n) = p(x^n)\, p(y^n \mid x^n),$$
$$p(y^n \mid x^n) = \prod_{i=1}^{n} p(y_i \mid x_i)$$

ja luvut $p(y_i \mid x_i)$ saadaan kanavamatriisista.

Olkoon $X^n(1), \ldots, X^n(M_n) \overset{iid}{\sim} p(x^n)$ ja käytetään satunnaisvektoreita $X^{(n)}(j)$ kooderin $\mathcal{C}_n$ koodisanoina. Koska koodisanat ovat satunnaisia, on kooderi itsekin satunnainen. Merkitään jatkossa

$$\mathcal{C}_n = \big(X^n(1), \ldots, X^n(M_n)\big)$$

samaistaen kooderi sen kuvan kanssa. Vastaavasti merkitään kooderin $\mathcal{C}_n$ arvoa (”realisaatiota”)

$$C_n = \big(x^n(1), \ldots, x^n(M_n)\big),$$

joka siis on jokin konkreettinen, tietty kooderi. Tällaisia koodereita on vain äärellisen monta. Huomaa, että nyt jotkut $x^n(j)$:t voivat hyvin olla myös

samoja. Kooderin $\mathcal{C}_n$ keskimääräinen virhetodennäköisyys on (vertaa määritelmä 5.12)

$$\bar{\lambda}(\mathcal{C}_n) = \frac{1}{M_n} \sum_{j=1}^{M_n} \lambda_j(\mathcal{C}_n),$$

missä

$$\lambda_j(\mathcal{C}_n) = \sum_{\substack{y^n \\ g_n(y^n;\mathcal{C}_n)\neq j}} p\big(y^n \mid X^n(j)\big),$$

ja $g_n(y^n;\mathcal{C}_n)$ on kooderiin $\mathcal{C}^n$ liittyvä (yhteistyypillisyyteen perustuva) dekooderi. Sekä $\bar{\lambda}(\mathcal{C}_n)$ että $\lambda_j(\mathcal{C}_n)$:t ovat satunnaismuuttujia. Suure $p\big(y^n \mid X^n(j)\big)$ on määritelty kanavamatriisin alkiona jokaiselle toteutuneelle satunnaisvektorin $X^n(j)$ arvolle.

Tarkastellaan nyt odotusarvoa

$$\mathbb{E}\,\bar{\lambda}(\mathcal{C}_n) = \sum_{C_n} \mathbb{P}\{\mathcal{C}_n = C_n\}\bar{\lambda}(C_n), \tag{5.12}$$

missä

$$\bar{\lambda}(C_n) = \frac{1}{M_n} \sum_{j=1}^{M_n} \lambda_j(C_n), \tag{5.13}$$

$$\lambda_j(C_n) = \sum_{\substack{y^n \\ g(y^n;C_n)\neq j}} p\big(y^n \mid x^n(j)\big),$$

ja

$$\mathbb{P}\{\mathcal{C}_n = C_n\} \equiv p(C_n) = \prod_{j=1}^{M_n} p\big(x^n(j)\big) = \prod_{j=1}^{M_n}\prod_{i=1}^{n} p\big(x_i(j)\big),$$

koska $X^n(1), \ldots, X^n(M_n) \overset{iid}{\sim} p(x^n)$. Osoitetaan, että

$$\lim_{n\to\infty} \mathbb{E}\,\bar{\lambda}(\mathcal{C}_n) = 0. \tag{5.14}$$

Nyt kaavojen (5.12) ja (5.13) nojalla saadaan

$$\mathbb{E}\,\bar{\lambda}(\mathcal{C}_n) = \frac{1}{M_n} \sum_{j=1}^{M_n} \sum_{C_n} p(C_n)\,\lambda_j(C_n).$$

Olkoon j kiinteä. Yllä

$$\lambda_j(C_n) = \mathbb{P}\{\text{"viesti } j \text{ dekoodataan väärin"} \mid \text{"kooderi on } C_n\text{"}\},$$

$$\begin{aligned}
&\sum_{C_n} p(C_n)\lambda_j(C_n) \\
&\quad = \mathbb{P}\{\text{"viesti } j \text{ dekoodataan väärin, kun kooderi } C_n \text{ valitaan satunnaisesti}\} \\
&\quad \equiv \mathbb{P}(B_{j,n}),
\end{aligned}$$

missä merkittiin $B_{j,n}$:llä yllä olevassa kaavassa esiintyvää tapahtumaa "viesti $j \cdots$ satunnaisesti".

Arvioidaan todennäköisyyttä $\mathbb{P}(B_{j,n})$. Selvästi

$$B_{j,n} \subset D_{j,n} \cup E_{j,n},$$

missä

$$\begin{aligned}
D_{j,n} &= \big\{\big(X^n(j), Y^n\big) \notin A_\varepsilon^{(n)}\big\}, \\
E_{j,n} &= \big\{\big(X^n(k), Y^n\big) \in A_\varepsilon^{(n)} \text{ jollain } k \neq j\big\}.
\end{aligned}$$

Nyt

$$\mathbb{P}(D_{j,n}) = \mathbb{P}\big\{\big(X^n(j), Y^n\big) \notin A_\varepsilon^{(n)}\big\} = \nu_n$$

ei riipu j:stä, koska $\big(X^n(j), Y^n\big) \sim p(x^n, y^n)$. Lauseen 5.15 (AEP) kohdan (i) nojalla

$$\mathbb{P}(D_{j,n}) = \nu_n \to 0, \text{ kun } n \to \infty. \tag{5.15}$$

(On helppo osoittaa, että $\big(X_1(j), Y_1\big), \ldots, \big(X_n(j), Y_n\big) \overset{iid}{\sim} p(x, y)$, joten AEP pätee. Tämä on harjoitustehtävä.)

Olkoon $k \neq j$. Syöte-tuloste parille on voimassa $\big(X^n(j), Y^n\big) \perp\!\!\!\perp X^n(k)$, joten $Y^n \perp\!\!\!\perp X^n(k)$. Mutta $X^n(k) \sim p(x^n)$, joten lauseen 5.15 kohdan (iii) nojalla

$$\mathbb{P}\big\{\big(X^n(k), Y^n\big) \in A_\varepsilon^{(n)}\big\} \leq 2^{-n(I(X;Y)-3\varepsilon)}. \tag{5.16}$$

Siten $E_{j,n}$:n todennäköisyydelle saadaan

$$\mathbb{P}(E_{j,n}) = \mathbb{P}\Big(\bigcup_{k\neq j}\big\{\big(X^n(k), Y^n\big) \in A_\varepsilon^{(n)}\big\}\Big) \leq \sum_{k\neq j}\mathbb{P}\big\{\big(X^n(k), Y^n\big) \in A_\varepsilon^{(n)}\big\}$$
$$\overset{(5.16)}{\leq} (M_n - 1)\, 2^{-n(I(X;Y)-3\varepsilon)} \leq 2^{-n(I(X;Y)-R-3\varepsilon)}, \tag{5.17}$$

koska $M_n = \lceil 2^{nR} \rceil \leq 2^{nR} + 1$.

Olkoon nyt alussa valittu pistetodennäköisyysfunktio $p(x) = p^*(x)$ sellainen, jolla kanavan kapasiteeti saavutetaan. Silloin

$$I(X;Y) - R = C - R > 0.$$

Olkoon $\varepsilon < \big(I(X;Y) - R\big)/3$. Silloin

$$\mathbb{P}(E_{j,n}) \leq 2^{-n(I(X;Y)-R-3\varepsilon)} \equiv \eta_n \to 0,$$

kun $n \to \infty$. Siten tästä ja tuloksesta (5.15) saadaan

$$\begin{aligned}
\mathbb{E}\,\bar{\lambda}(\mathcal{C}_n) &= \frac{1}{M_n}\sum_{j=1}^{M_n}\sum_{C_n} p(C_n)\,\lambda_j(C_n) = \frac{1}{M_n}\sum_{j=1}^{M_n}\mathbb{P}(B_{j,n}) \\
&\leq \frac{1}{M_n}\sum_{j=1}^{M_n}\mathbb{P}(D_{j,n}\cup E_{j,n}) \\
&\leq \frac{1}{M_n}\sum_{j=1}^{M_n}\big[\mathbb{P}(D_{j,n}) + \mathbb{P}(E_{j,n})\big] \leq \frac{1}{M_n}\cdot M_n\,(\nu_n + \eta_n) \\
&= \nu_n + \eta_n \to 0,
\end{aligned}$$

kun $n \to \infty$ ja (5.14) siis pätee.

Olkoon C_n^* (jokin) kooderi, joka minimoi $\bar{\lambda}(C_n)$:n. Silloin

$$\lim_{n\to\infty}\bar{\lambda}(C_n^*) = 0,$$

koska jos $\bar{\lambda}(C_n^*) \geq \delta > 0$ äärettömän monella n, on näillä n:n arvoilla

$$\mathbb{E}\,\bar{\lambda}(\mathcal{C}_n) = \sum_{C_n} p(C_n)\,\bar{\lambda}(C_n) \geq \sum_{C_n} p(C_n)\delta = \delta > 0,$$

jolloin (5.14) ei pätisi.

Haluamme kuitenkin, että *maksimivirhe* $\lambda^{(n)} \to 0$. Olkoon kooderissa C_n^*

$$\lambda_{j_{n,1}}(C_n^*) \leq \lambda_{j_{n,2}}(C_n^*) \leq \cdots \leq \lambda_{j_{n,M_n}}(C_n^*)$$

viestien j virhetodennäköisyydet suuruusjärjestyksessä. Olkoon vielä $l_n = \lceil M_n/2 \rceil$. Sillon

$$\lim_{n\to\infty} \lambda_{j_{n,l_n}}(C_n^*) = 0,$$

koska jos

$$\lambda_{j_{n,l_n}}(C_n^*) \geq \delta > 0$$

äärettömän monella n, on näillä n

$$\begin{aligned}
\bar{\lambda}(C_n^*) &= \frac{1}{M_n} \sum_{k=1}^{M_n} \lambda_{j_{n,k}}(C_n^*) \geq \frac{1}{M_n} \sum_{k=l_n}^{M_n} \lambda_{j_{n,k}}(C_n^*) \\
&\geq \frac{1}{M_n} \left(M_n - l_n + 1\right) \delta = \frac{1}{M_n} \left(M_n - \lceil M_n/2 \rceil + 1\right) \delta \\
&\geq \frac{1}{M_n} \left(M_n - \frac{M_n}{2}\right) \delta = \frac{\delta}{2},
\end{aligned}$$

mikä on ristiriita, koska $\lim_{n\to\infty} \bar{\lambda}(C_n^*) = 0$.

Siten $C_n^* | \{j_{n,1}, \ldots, j_{n,l_n}\}$ on kooderi, jolla maksimivirhe $\lambda^{(n)} \to 0$. Numeroidaan viestit uudestaan $1, \ldots, l_n$ ja haetaan sellainen R_n, että $l_n = 2^{nR_n}$. Silloin on siis olemassa $(\lceil 2^{nR_n} \rceil, n)$-koodit, joille $\lambda^{(n)} \to 0$, kun $n \to \infty$. On helppo nähdä, että

$$2^{n(R-\frac{1}{n})} \leq l_n < 2^{n(R+\frac{1}{n})},$$

joten $R - \frac{1}{n} \leq R_n < R + \frac{1}{n}$. Siten, jos $R < C$ on annettu, voidaan ensin valita $R < R' < C$ ja tehdä koko edellä oleva konstruktio R':lle. Koska $R'_n \geq R' - \frac{1}{n} > R$ suurilla n, voidaan löydetyistä $(\lceil 2^{nR'_n} \rceil, n)$-koodeista pudottaa $\lceil 2^{nR'_n} \rceil - \lceil 2^{nR} \rceil$ huonointa (suurimman virheen antavaa) koodisanaa pois ja saada näin $(\lceil 2^{nR} \rceil, n)$-koodit, joilla $\lambda^{(n)} \to 0$.

Todistetaan sitten lauseen toinen puoli. Oletetaan, että $(\lceil 2^{nR} \rceil, n)$-koodeille (C_n, g_n), $n \in \mathbb{N}_+$, pätee $\lim_{n\to\infty} \lambda^{(n)} = 0$. Merkitään taas $M_n = \lceil 2^{nR} \rceil$ ja olkoot $x^n(1), \ldots, x^n(M_n)$ koodin (C_n, g_n) koodisanat. Oletetaan, että J_n on tasaisesti jakautunut viestien $\{1, \ldots, M_n\}$ joukossa, eli

$$p(j) = \mathbb{P}\{J_n = j\} = \frac{1}{M_n}, \quad j = 1, \ldots, M_n.$$

Olkoon $X^n = C_n(J_n)$ J_n:ää vastaava koodisana. Silloin X^n:llä on pistetodennäköisyysfunktio

$$\begin{aligned} p(x^n) &= \mathbb{P}\{X^n = x^n\} = \mathbb{P}\{C_n(J_n) = x^n\} \\ &= \mathbb{P}\left\{J_n \in C_n^{-1}(\{x^n\})\right\} = \sum_{j, C_n(j) = x^n} p(j) \\ &= \frac{|\{j \mid C_n(j) = x^n\}|}{M_n}, \quad x^n \in \mathcal{X}^n. \end{aligned}$$

Erityisesti $p(x^n) = 0$, kun $x^n \in \mathcal{X}^n$ ei ole koodisana. Nyt kanavamatriisin avulla määräytyy pistetodennäköisyysfunktio satunnaismuuttujaparille (X^n, Y^n),

$$p(x^n, y^n) = p(x^n)\, p(y^n \mid x^n) = p(x^n) \prod_{i=1}^{n} p(y_i \mid x_i).$$

Tässä siis Y^n kuvaa tulostetta, kun syöte on X^n.

Selvästi $H(J_n) = \log M_n$ (lause 2.12). Toisaalta

$$I(J_n; Y^n) = H(J_n) - H(J_n \mid Y^n).$$

Siten

$$\log M_n = H(J_n \mid Y^n) + I(J_n; Y^n). \tag{5.18}$$

Arvioidaan oikean puolen termejä erikseen. Ei ole vaikea nähdä (harjoitustehtävä), että

$$I(J_n; Y^n) \leq I(X^n; Y^n), \tag{5.19}$$

koska $X^n = C_n(J_n)$ on J_n:n funktio. Edelleen,

$$I(X^n; Y^n) = H(Y^n) - H(Y^n \mid X^n) \tag{5.20}$$

ja, koska kanava on diskreetti ja muistiton,

$$\begin{aligned} H(Y^n \mid X^n) &= -\sum_{x^n \in \mathcal{X}^n} \sum_{y^n \in \mathcal{Y}^n} p(x^n, y^n) \log p(y^n \mid x^n) \\ &= -\sum_{x^n \in \mathcal{X}^n} \sum_{y^n \in \mathcal{Y}^n} p(x^n, y^n) \Big[\sum_{i=1}^{n} \log p(y_i \mid x_i) \Big] \\ &= \sum_{i=1}^{n} \Big[-\sum_{x^n \in \mathcal{X}^n} \sum_{y^n \in \mathcal{Y}^n} p(x^n, y^n) \log p(y_i \mid x_i) \Big] \\ &= \sum_{i=1}^{n} \Big[-\sum_{x_i \in \mathcal{X}} \sum_{y_i \in \mathcal{Y}} p(x_i, y_i) \log p(y_i \mid x_i) \Big] \\ &= \sum_{i=1}^{n} H(Y_i \mid X_i). \end{aligned} \tag{5.21}$$

Toisaalta lauseen 2.16 nojalla

$$H(Y^n) \leq \sum_{i=1}^{n} H(Y_i). \tag{5.22}$$

Siten kaavojen (5.19), (5.20), (5.21) ja (5.22) nojalla

$$\begin{aligned} I(J_n; Y^n) &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i) \\ &= \sum_{i=1}^{n} \big[H(Y_i) - H(Y_i \mid X_i) \big] \\ &= \sum_{i=1}^{n} I(X_i; Y_i) \leq nC, \end{aligned} \tag{5.23}$$

koska $I(X_i; Y_i) \leq C$ kaikilla $i = 1, \ldots, n$.

Arvioidaan sitten (5.18):n toista termiä $H(J_n \mid Y^n)$. Olkoon

$$P_e(C_n) = \mathbb{P}\{g_n(Y^n) \neq J_n\}, \quad n \in \mathbb{N}_+.$$

Silloin Fanon epäyhtälön (lause 2.17) nojalla

$$H(J_n \mid Y^n) \leq H\big(P_e(C_n)\big) + P_e(C_n)\log(M_n - 1), \tag{5.24}$$

missä $H(t) = -t\log t - (1-t)\log(1-t)$, $0 \leq t \leq 1$. Tässä kokonaistodennäköisyyden kaavan avulla

$$\begin{aligned} P_e(C_n) &= \sum_{j=1}^{M_n} \mathbb{P}\{J_n = j\}\, \mathbb{P}\{g_n(Y^n) \neq J_n \mid J_n = j\} \\ &= \frac{1}{M_n}\sum_{j=1}^{M_n} \lambda_j(C_n) \\ &\leq \frac{1}{M_n}\sum_{j=1}^{M_n} \lambda^{(n)} = \lambda^{(n)}, \end{aligned} \tag{5.25}$$

missä siis $\lambda_j(C_n)$ on viestin j virhetodennäköisyys ja $\lambda^{(n)}$ on koodin (C_n, g_n) maksimaalinen virhetodennäköisyys.

Kokoamalla tulokset (5.18), (5.23), (5.24) ja (5.25) saadaan

$$\begin{aligned} \log M_n &\leq H\big(P_e(C_n)\big) + P_e(C_n)\log(M_n - 1) + nC \\ &\leq 1 + \lambda^{(n)}\log(M_n - 1) + nC, \end{aligned} \tag{5.26}$$

koska $H(t) \leq 1$ kaikilla $t \in [0,1]$. Edelleen,

$$M_n = \lceil 2^{nR} \rceil \in [2^{nR}, 2^{nR} + 1[,$$

joten tuloksen (5.26) nojalla

$$\begin{aligned} nR = \log 2^{nR} &\leq \log M_n \leq 1 + \lambda^{(n)}\log(2^{nR} + 1 - 1) + nC \\ &= 1 + \lambda^{(n)} nR + nC \end{aligned}$$

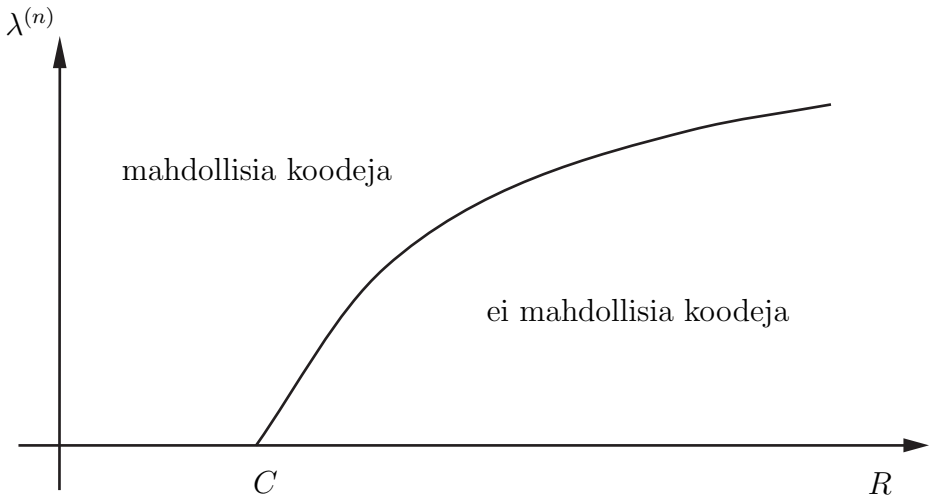
ja siten

$$R \leq \frac{1}{n} + R\lambda^{(n)} + C. \tag{5.27}$$

Koska $\lim\limits_{n\to\infty} \lambda^{(n)} = 0$, on

$$\lim_{n\to\infty}\left(\frac{1}{n} + R\lambda^{(n)} + C\right) = C.$$

Siten täytyy olla $R \leq C$. □



Kuva 5.11: Shannonin toisen lauseen havainnollistus.

Huomautus. Todistuksessa käytetyt satunnaismuuttujat $X^n(j)$, X^n, J^n, Y^n ovat vain todistuksen aputyökaluja. Itse lauseen tulos puhuu vain koodeista ja niiden maksimaalisesta virheestä. Nämä ovat puhtaasti kooderiin, dekooderiin ja kanavaan (=kanavamatriisiin) liittyviä käsitteitä.

Huomautus. Hyvä koodi C_n^* löytyisi *periaatteessa* käymällä läpi kaikki mahdolliset koodisanojen kokoelmat $C_n = \{x^{(n)}(1), \ldots, x^{(n)}(M_n)\}$. Tämä on kuitenkin käytännössä liian raskasta ja myös dekoodaamisen kompleksisuus (taulukon koko 2^{nR}) kasvaa eksponentiaalisesti R:n mukana. Koodausteoriassa pyritään helpommin dekoodattaviin koodeihin.

Huomautus. Tuloksen (5.27) nojalla

$$\lambda^{(n)} \geq \frac{R - \frac{1}{n} - C}{R} = 1 - \frac{1}{nR} - \frac{C}{R}.$$

Siten, kun $R > C$, on $\lambda^{(n)} \geq \delta > 0$ suurilla n. Itseasiassa tällöin $\lambda^{(n)} \geq \delta' > 0$ kaikilla n, sillä jos $\lambda^{(n_0)} = 0$ jollain n_0, voidaan koodisanoja yhdistämällä (konkatenoimalla) konstruoida mielivaltaisen suurella n koodeja, joilla $\lambda^{(n)} = 0$. Kuva 5.11 havainnollistaa Shannonin toisen lauseen tulosta.

Huomautus. Voidaan osoittaa, että itseasiassa $\lim_{n\to\infty} \lambda^{(n)} = 1$, kun $R > C$ (Wolfowitz 1961).

Luku 6

Jatkuvat satunnaismuuttujat ja informaatio

6.1 Differentiaalientropia

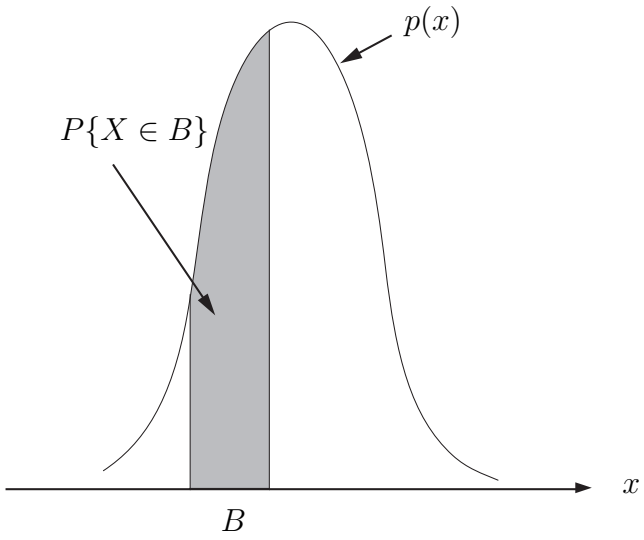
Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus. Kuvaus $X : \Omega \to \mathbb{R}$ on satunnaismuuttuja, jos kaikilla $B \subset \mathbb{R}$

$$\{X \in B\} = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F}.$$

Satunnaismuuttuja X on *jatkuva* tai satunnaismuuttujalla X on *jatkuva jakauma*, jos on olemassa integroituva $p : \mathbb{R} \to [0, \infty[$, jolle

$$\mathbb{P}\{X \in B\} = \int_B p(x)\, dx$$

kaikilla $B \subset \mathbb{R}$. Sanomme, että p on X:n (tai X:n jakauman) *tiheysfunktio* (tf) (kuva 6.1). Merkitseme jatkossa $p = p(x)$ ja vastaavasti satunnaismuuttujan Y tiheysfunktiota merkitään $p(y)$:llä. Samat asiat ilmaistaan myös merkinnöillä $X \sim p(x)$, $Y \sim p(y)$.



Kuva 6.1: Satunnaismuuttujan jakauman tiheysfunktio.

Huomautus. Edellä annetuissa määritelmissä B:n tulee tarkkaan ottaen olla ns. Borelin joukko. Borelin joukot määritellään $\mathbb{R}$:n pienimpänä σ-algebrana, joka sisältää avoimet joukot. Edelleen, tiheysfunktion tulee olla ns. Borelin funktio eli $p^{-1}(U)$ on Borelin joukko kaikille avoimille $U \subset \mathbb{R}$. Integraalit ovat yleisesti ottaen Lebesguen integraaleja. Tavallisimmat funktiot ovat Borelin funktiota ja niille Lebesguen integraalit ovat samoja kuin Riemannin integraalit. Emme jatkossa juuri kiinnitä enää huomioita tämän tyyppisiin yksityiskohtiin.

Huomautus. Pätee

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

Esimerkki 6.1 (Normaalijakauma)**.** Olkoon $\mu \in \mathbb{R}$ ja $\sigma > 0$ ja

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad x \in \mathbb{R}.$$

Sanomme, että $X \sim p(x)$ on normaalijakautunut parametrein μ ja σ^2,

$$X \sim \mathrm{N}(\mu, \sigma^2).$$

Tunnetusti

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xp(x)\,dx = \mu \quad (X\text{:n odotusarvo}),$$

$$D^2(X) = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)\,dx = \sigma^2 \quad (X\text{:n varianssi}).$$

Varianssin neliöjuuri $\sqrt{D^2(X)} = \sigma$ on X:n keskihajonta. ‖

Yleisemmin, olkoon $n \in \mathbb{N}_+$. Kuvaus

$$X^n = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$$

on satunnaisvektori (sv), jos kaikilla $B \subset \mathbb{R}^n$

$$\{X^n \in B\} = \{\omega \in \Omega \mid X^n(\omega) \in B\} \in \mathcal{F}.$$

Vastaavasti, satunnaisvektori X^n on jatkuva ja integroituva $p : \mathbb{R}^n \to [0, \infty[$ on X^n:n tiheysfunktio, jos kaikilla $B \subset \mathbb{R}^n$

$$\mathbb{P}\{X^n \in B\} = \int_B p(x^n)\,dx^n,$$

missä $x^n = (x_1, \ldots, x_n) \in \mathbb{R}^n$ ja $dx^n = dx_1 \cdots dx_n$.

On helppo nähdä, että jatkuvan satunnaisvektorin X^n komponentit ovat jatkuvia satunnaismuuttujia ja $X_i \sim p(x_i)$, missä

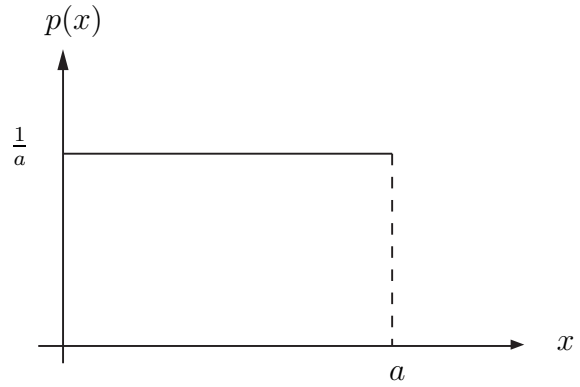
$$p(x_i) = \int_{\mathbb{R}^{n-1}} p(x^n)\,dx_1 \cdots dx_{i-1}\,dx_{i+1} \cdots dx_n$$

on ns. X_i:n reuna- eli marginaalitiheysfunktio.

Määritelmä 6.1. Olkoon $X \sim p(x)$ jatkuva satunnaismuuttuja. Silloin X:n *differentiaalientropia* on

$$H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x)\,dx,$$

edellyttäen, että kyseessä oleva integraali suppenee.



Kuva 6.2: Jakauman $\mathrm{Tas}(0, a)$ tiheysfunktio.

Huomautus. Tässä käytetään sopimusta $0 \log 0 = 0$. Integraalin suppeneminen tarkoittaa suppenemista Lebesguen mielessä eli

$$\int_{-\infty}^{\infty} p(x) \big| \log p(x) \big| \, dx < \infty.$$

Huomautus. Voidaan myös tulkita, että

$$H(X) = \mathbb{E}\big[-\log p(X) \big],$$

jos $-\log p\big(X(\omega)\big)$ määritellään sopivasti, kun $p\big(X(\omega)\big) = 0$.

Differentiaalientropian yksikkö on bitti.

Esimerkki 6.2. Olkoon $X \sim \mathrm{Tas}(0, a)$, $a > 0$, eli $X \sim p(x)$, missä

$$p(x) = \begin{cases} \frac{1}{a}\,, & x \in]0, a[, \\ 0\,, & x \notin]0, a[. \end{cases}$$

(Kuva 6.2). Funktiolle $p(x)$ selvästi pätee

$$\int_{-\infty}^{\infty} p(x) \, dx = \int_{0}^{a} \frac{1}{a} \, dx = 1.$$

Edelleen,

$$H(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx = -\int_{0}^{a} \frac{1}{a}\log\frac{1}{a}\,dx = -\log\frac{1}{a} = \log a.$$

Siten, kun $0 < a < \infty$, voi $H(X)$ saada minkä tahansa reaaliarvon. Tilanne on siten erilainen kuin diskreettien satunnaismuuttujien tapauksessa, jolloin entropia on aina ei-negatiivinen. ‖

Huomautus. Joillain $p(x)$ on myös $\int_{-\infty}^{\infty} p(x)\log p(x)dx = \pm\infty$ (harjoitustehtävä).

Esimerkki 6.3. Lasketaan satunnaismuuttujan $X \sim \mathrm{N}(\mu, \sigma^2)$ differentiaalientoropia. Käytetään laskussa apuna kaavaa $\log t = \ln t/\ln 2$.

$$\begin{aligned}
H(X) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \log\left[\frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\right] dx \\
&= -\frac{1}{\ln 2}\cdot\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\left[-\ln(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}(x-\mu)^2\right] dx \\
&= \frac{\ln(\sqrt{2\pi}\sigma)}{\ln 2}\,\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx \\
&\qquad + \frac{1}{\ln 2}\cdot\frac{1}{2\sigma^2}\cdot\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx \\
&= \log\left(\sqrt{2\pi}\sigma\right) + \frac{1}{2\ln 2} = \log\left(\sqrt{2\pi\sigma^2 e}\right) = \frac{1}{2}\log\left(2\pi e\sigma^2\right).
\end{aligned}$$

‖

Lauseen 2.8 vastine on nyt

Lause 6.2. *Olkoot $p(x) > 0$ ja $q(x) > 0$ positiivisia tiheysfunktioita. Silloin*

$$-\int_{-\infty}^{\infty} p(x)\log p(x)\,dx \le -\int_{-\infty}^{\infty} p(x)\log q(x)\,dx \tag{6.1}$$

edellyttäen, että kyseessä olevat integraalit suppenevat. Yhtäsuuruus pätee epäyhtälössä (6.1) *jos ja vain jos* $p(x) = q(x)$ *m.k.* $x \in \mathbb{R}$.

Huomautus. "m.k. $x \in \mathbb{R}$" tarkoittaa melkein kaikilla $x \in \mathbb{R}$ eli kaikilla $x \in \mathbb{R} \setminus A$, jossa joukon A ns. Lebesguen mitta $m(A) = 0$. Itseasiassa (minkä tahansa joukon A) Lebesguen mitta voidaan laskea myös integraalina

$$m(A) = \int_A dx = \int_{-\infty}^{\infty} 1_A(x)\, dx; \quad 1_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

Todistus. Kuten lauseen 2.8 todistuksessa, käytetään tässäkin luonnollista logaritmia. Kuten luvussa 2 (kuva 2.7) todettiin,

$$\ln x \leq x - 1, \quad 0 < x < \infty$$

ja yhtäsuuruus on voimassa jos ja vain jos $x = 1$. Siten kaikilla $x \in \mathbb{R}$

$$\ln \left[\frac{q(x)}{p(x)} \right] \leq \frac{q(x)}{p(x)} - 1, \tag{6.2}$$

eli

$$p(x)\, \ln \left[\frac{q(x)}{p(x)} \right] \leq q(x) - p(x),$$

ja niin

$$-p(x) \ln p(x) \leq -p(x) \ln q(x) + q(x) - p(x).$$

Kun integraalit suppenevat, on siis

$$\begin{aligned} -\int_{-\infty}^{\infty} p(x) \ln p(x)\, dx &\leq -\int_{-\infty}^{\infty} p(x) \ln q(x)\, dx + \int_{-\infty}^{\infty} q(x)\, dx - \int_{-\infty}^{\infty} p(x)\, dx \\ &= -\int_{-\infty}^{\infty} p(x) \ln q(x)\, dx, \end{aligned}$$

koska $\int_{-\infty}^{\infty} q(x)\, dx = \int_{-\infty}^{\infty} p(x)\, dx = 1$. Siten (6.1) pätee.

Jos $p(x) = q(x)$ m.k. x, pätee epäyhtälössä (6.1) yhtäsuuruus Lebesguen integraalin ominaisuuksien perusteella.

Kääntäen, olkoon

$$g(x) = p(x)\ln p(x) - p(x)\ln q(x) + q(x) - p(x).$$

Silloin edellisen perusteella $g(x) \geq 0$ kaikilla x ja jos epäyhtälössä (6.1) pätee yhtäsuuruus, on $\int_{-\infty}^{\infty} g(x)\,dx = 0$. Silloin $g(x) = 0$ m.k. x eli (6.2):ssa pätee yhtälö m.k. x, joten $\frac{q(x)}{p(x)} = 1$ m.k. x, eli $p(x) = q(x)$ m.k. x. □

Huomautus. Jos sopimuksen $0\log 0 = 0$ lisäksi sovitaan, että $0\cdot(-\infty) = 0$, pätee lause 6.2 kaikille tiheysfunktioille $p(x)$ ja $q(x)$.

Millä jakaumalla sitten mahtaa olla suurin differentiaalientropia? Kysymys ei ole tässä muodossa aivan mielekäs, koska esimerkiksi jos $X \sim \mathrm{Tas}(0,a)$, on

$$H(X) = \log a \to \infty,$$

kun $a \to \infty$. Jakauma pitää ensin ”normeerata” jotenkin. Kiinnitetään seuraavassa varianssi $D^2(X)$. Vaikka ehkä tasainen jakauma tuntuisi lupaavalta ehdokkaalta differentiaalientropian maksimoivaksi jakaumaksi, optimijakauma on kuitenkin normaali.

Lause 6.3. *Olkoon $X \sim p(x)$ jatkuva satunnaismuuttuja, jolla on olemassa differentiaalientropia $H(X)$ ja olkoon $D^2(X) = \sigma^2 > 0$. Silloin*

$$H(X) \leq \frac{1}{2}\log(2\pi e\sigma^2) \tag{6.3}$$

ja yhtäsuuruus pätee jos ja vain jos $X \sim \mathrm{N}(\mu,\sigma^2)$, missä $\mu = E(X)$.

Todistus. Otetaan edellisessä lauseessa $q(x)$:ksi jakauman $\mathrm{N}(\mu,\sigma^2)$ tiheysfunktio, missä $\mu = E(X)$. Silloin (vertaa lauseen 6.2 todistuksen jälkeinen

huomautus)

$$
\begin{aligned}
H(X) &= -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx \le -\int_{-\infty}^{\infty} p(x)\log q(x)\,dx \\
&= -\frac{1}{\ln 2}\int_{-\infty}^{\infty} p(x)\,\ln\left[\frac{1}{\sqrt{2\pi}\sigma}\,e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\right]\,dx \\
&= \frac{1}{\ln 2}\left\{\ln(\sqrt{2\pi}\sigma)\int_{-\infty}^{\infty} p(x)\,dx + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty}(x-\mu)^2 p(x)\,dx\right\} \\
&= \frac{1}{\ln 2}\left\{\ln(\sqrt{2\pi}\sigma)+\frac{1}{2}\right\} = \frac{\ln\left(\sqrt{2\pi\sigma^2 e}\right)}{\ln 2} \\
&= \log\sqrt{2\pi\sigma^2 e} = \frac{1}{2}\log(2\pi e\sigma^2),
\end{aligned}
$$

joten (6.3) pätee. Lauseen 6.2 mukaan epäyhtälössä (6.3) pätee yhtäsuuruus jos ja vain jos $p(x) = q(x)$ m.k. x eli jos ja vain jos $X \sim \mathrm{N}(\mu, \sigma^2)$. □

Siis: kiinteällä $D^2(X) = \sigma^2$, on normaalijakaumalla suurin entropia.

Määritelmä 6.4. Olkoot $p(x)$ ja $q(x)$ tiheysfunktioita. Silloin $p(x)$:n ja $q(x)$:n *Kullbackin-Leiblerin etäisyys* on

$$
\begin{aligned}
D(p \parallel q) &= \int_{-\infty}^{\infty} p(x)\log\left[\frac{p(x)}{q(x)}\right]\,dx \\
&\equiv \int_{-\infty}^{\infty} p(x)\log p(x)\,dx - \int_{-\infty}^{\infty} p(x)\log q(x)\,dx,
\end{aligned}
$$

edellyttäen, että kyseessä olevat oikean puolen integraalit suppenevat.

Lauseen 6.2 mukaan

$$
D(p \parallel q) \ge 0 \tag{6.4}
$$

kaikilla $p(x)$, $q(x)$, kun integraalit suppenevat ja yhtäsuuruus pätee jos ja vain jos $p(x) = q(x)$ m.k. x.

Määritelmä 6.5. Jatkuvan satunnaisvektorin $(X, Y) \sim p(x, y)$ *yhteisdifferentiaalientropia* on

$$H(X, Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y) \log p(x, y)\, dx\, dy,$$

edellyttäen, että kyseessä oleva integraali suppenee.

Tarkastellaan jatkuvaa satunnaisvektoria $(X, Y) \sim p(x, y)$ ja olkoon $X \sim p(x)$. Satunnaismuuttujan Y ehdollinen tiheysfunktio ehdolla $X = x$ on

$$p(y \mid x) = \begin{cases} \frac{p(x,y)}{p(x)}, & \text{kun } p(x) > 0 \\ 0, & \text{kun } p(x) = 0, \end{cases}$$

kun $y \in \mathbb{R}$. Jatkossa otetaan (X, Y):n tiheysfunktioksi aina

$$p(x, y) = p(x)\, p(y \mid x).$$

Erityisesti siis $p(x, y) = 0$, kun $p(x) = 0$. On helppo nähdä, että näin modifioitu $p(x, y)$ on edelleen (X, Y):n tiheysfunktio.

Määritelmä 6.6. Jatkuvan satunnaismuuttujan Y *differentiaalientropia ehdolla* $X = x$ on

$$H(Y \mid X = x) = -\int_{-\infty}^{\infty} p(y \mid x) \log p(y \mid x)\, dy,$$

edellyttäen, että kyseessä oleva integraali suppenee.

Määritelmä 6.7. Jatkuvan satunnaismuuttujan Y *differentiaalientropia ehdolla* X on

$$\begin{aligned} H(Y \mid X) &= \int_{-\infty}^{\infty} p(x) H(Y \mid X = x)\, dx \\ &= \int_{-\infty}^{\infty} p(x) \left[-\int_{-\infty}^{\infty} p(y \mid x) \log p(y \mid x)\, dy \right] dx \end{aligned}$$

edellyttäen, että kyseessä oleva integraali suppenee.

Siten, Lebesquen integraalia koskevan Fubinin lauseen nojalla,

$$\begin{aligned} H(Y \mid X) &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x)p(y \mid x)\log p(y \mid x)\,dy\,dx \\ &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(y \mid x)\,dy\,dx, \end{aligned}$$

ainakin kun

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\big|\log p(y \mid x)\big|\,dx\,dy < \infty.$$

Lause 6.8 (Ketjusääntö)**.** *Pätee*

$$H(X,Y) = H(X) + H(Y \mid X),$$

edellyttäen, että kyseessä olevat suureet ovat olemassa.

Todistus.

$$\begin{aligned} H(X,Y) &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x,y)\,dx\,dy \\ &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x)\,dx\,dy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(y \mid x)\,dx\,dy \\ &= -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx + \int_{-\infty}^{\infty} p(x)\Big[-\int_{-\infty}^{\infty} p(y \mid x)\log p(y \mid x)\,dy\Big]\,dx \\ &= H(X) + H(Y \mid X), \end{aligned}$$

missä tarkkaan ottaen kolmannessa yhtälössä käytettiin taas Fubinin lausetta kaksinkertaisen integraalin laskemisessa. □

Olkoon (X,Y):n tiheysfunktio f. Silloin (Y,X):llä on tiheysfunktio g, $g(x,y) = f(y,x)$. Siten tällä kurssilla käytetyin merkinnöin $p(x,y) = p(y,x)$ ja pätee

$$H(X,Y) = H(Y,X).$$

Ketjusäännön mukaan on myös voimassa

$$H(X,Y) = H(Y,X) = H(Y) + H(X \mid Y),$$

jos kyseessä olevat suureet ovat olemassa.

Lause 6.9. *Ehdollistaminen vähentää entropiaa, eli*

$$H(Y \mid X) \leq H(Y),$$

kun kyseessä olevat suureet ja lisäksi $H(X,Y)$ sekä $H(X)$ ovat olemassa. Yhtäsuuruus on voimassa jos ja vain jos $X \perp\!\!\!\perp Y$.

Todistus. Epäyhtälön (6.4) (pätee myös $\mathbb{R}^n$:ssä) nojalla

$$\begin{aligned}
0 &\leq D\big(p(x,y) \parallel p(x)p(y)\big) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x,y)\,dx\,dy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\big(p(x)p(y)\big)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(y \mid x)\,dx\,dy \\
&\quad - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(x)\,dx\,dy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log p(y)\,dx\,dy \\
&= -H(Y \mid X) + H(Y),
\end{aligned}$$

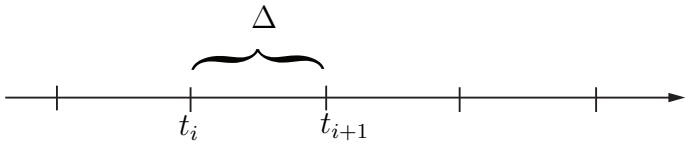
joten

$$H(Y \mid X) \leq H(Y).$$

Yhtäsuuruus on voimassa jos ja vain jos $p(x,y) = p(x)p(y)$ m.k. x, y eli jos ja vain jos $X \perp\!\!\!\perp Y$. □

Yhdistämällä lauseet 6.8 ja 6.9 nähdään, että

$$H(X,Y) = H(X) + H(Y \mid X) \leq H(X) + H(Y).$$



Kuva 6.3: Reaaliakselin osiinjako.

Määritelmä 6.10. Jatkuvien satunnaismuuttujien *keskinäisinformaatio* on

$$I(X;Y) = H(X) - H(X \mid Y),$$

edellyttäen, että oikean puolen suureet ovat olemassa.

Yhdistelemällä edellä olevia tuloksia todetaan esimerkiksi tutut kaavat:

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) = I(Y;X),$$
$$I(X;Y) = H(X) - H(X \mid Y) = H(X) + H(Y) - H(X,Y).$$

Edellä olevat määritelmät ja tulokset yleistyvät helposti satunnaisvektoreille X^n, Y^n. Esimerkiksi satunnaisvektorin $X^n \sim p(x^n)$ differentiaalientropia määritellään kaavalla

$$H(X^n) = -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x^n) \log p(x^n)\, dx^n,$$

missä $x^n = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Vastaavasti myös muut määritelmät.

Tarkastellaan vielä lopuksi lyhyesti diskreetin entropian ja differentiaalientropian välistä suhdetta. Oletetaan, että $X \sim p(x)$ on jatkuva satunnaismuuttuja ja tiheysfunktio $p(x)$ on jatkuva. Olkoon $\Delta > 0$ ja $t_i = i\Delta$, $i \in \mathbb{Z}$ (kuva 6.3). Integraalilaskennan väliarvolauseen mukaan jokaista $i \in \mathbb{Z}$ kohti on olemassa sellainen $x_i \in [t_i, t_{i+1}]$, että

$$\int_{t_i}^{t_{i+1}} p(x)\, dx = p(x_i)\Delta.$$

Määritellään sellainen diskreetti satunnaismuuttuja X^Δ, että

$$X^\Delta = x_i, \text{ kun } X \in [t_i, t_{i+1}[.$$

X^Δ on siis satunnaismuuttujan X "kvantisointi" (diskretointi) tarkkuudella Δ. Nyt

$$p_i \equiv \mathbb{P}\{X^\Delta = x_i\} = \mathbb{P}\big\{X \in [t_i, t_{i+1}[\big\} = \int_{t_i}^{t_{i+1}} p(x)\,dx = p(x_i)\Delta.$$

Siten X^Δ:n entropia on (vrt. harjoitus 2, tehtävä 1).

$$\begin{aligned} H(X^\Delta) &= -\sum_i p_i \log p_i = -\sum_i p(x_i)\Delta \log\big(p(x_i)\Delta\big) \\ &= -\sum_i p(x_i)\Delta \log p(x_i) - \sum_i p(x_i)\Delta \log \Delta \\ &= -\sum_i \big[p(x_i)\log p(x_i)\big]\Delta - \log\Delta \sum_i p(x_i)\Delta. \end{aligned}$$

Tässä

$$\sum_i p(x_i)\Delta = \sum_{i=-\infty}^{\infty} \int_{t_i}^{t_{i+1}} p(x)\,dx = \int_{-\infty}^{\infty} p(x)\,dx = 1.$$

Edelleen, kun $p(x)$ on riittävän säännöllinen, on

$$\lim_{\Delta\to 0+} -\sum_{i=-\infty}^{\infty} \big[p(x_i)\log p(x_i)\big]\Delta = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx = H(X).$$

(Ks. kuva 6.4). Tällöin siis

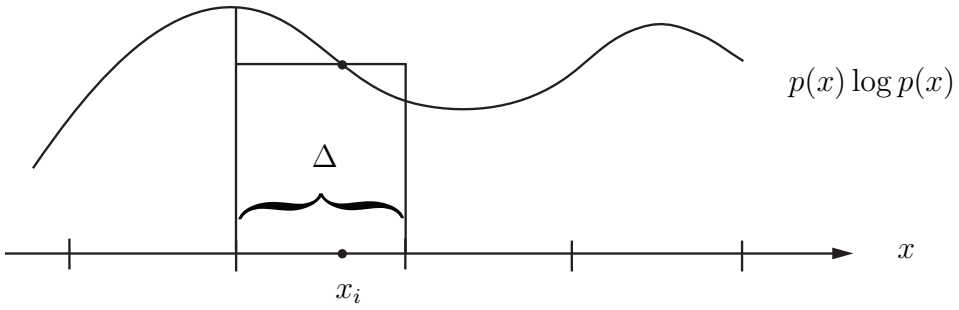
$$\lim_{\Delta\to 0+} [H(X^\Delta) + \log\Delta] = H(X)$$

ja pienillä $\Delta > 0$ pätee

$$H(X^\Delta) \approx H(X) - \log\Delta.$$

Erityisesti, jos $\Delta = 2^{-l}$, on X^Δ X:n "l:n bitin kvantisointi". Sille siis

$$H(X^\Delta) \approx H(X) + l.$$



Kuva 6.4: Integraalin laskeminen Riemannin summan avulla.

Esimerkki 6.4. Olkoon $X \sim \text{Tas}(0, 1/8)$. Tällöin

$$H(X) = \log \frac{1}{8} = -3$$

ja jos $\Delta = 2^{-l}$, niin

$$H(X^{\Delta}) \approx l - 3. \tag{6.5}$$

Kuten tiedämme, entropia likimain kertoo kuinka monta bittiä (esimerkiksi 0/1 kysymystä) tarvitaan satunnaismuuttujan arvojen kuvaamiseen keskimäärin. Nyt $X \in]0, \frac{1}{8}[$, joten satunnaismuuttujan X kolme ensimmäistä bittiä ovat 0 ja X:n arvon binääriesitys on

$$0.000 b_4 b_5 \cdots b_l b_{l+1} \cdots .$$

Siten l:n bitin tarkkuuteen riittää itseasiassa vain $l - 3$ bittiä ja (6.5) on luonteva tulos. ‖

6.2 AEP

Olkoot $X_1, \ldots, X_n \overset{iid}{\sim} p(x)$ eli $X_1, \ldots, X_n \perp\!\!\!\perp$ ja

$$\mathbb{P}\{X_i \in B\} = \int_B p(x)\, dx, \quad B \subset \mathbb{R}^n,$$

$i = 1, \ldots, n$. Silloin satunnaisvektorilla $X^n = (X_1, \ldots, X_n)$ on tiheysfunktio

$$p(x^n) = \prod_{i=1}^{n} p(x_i), \quad x^n = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Määritelmä 6.11. Olkoon $H(X)$ olemassa ja $n \in \mathbb{N}_+$, $\varepsilon > 0$. *Tyypillinen joukko* $A_\varepsilon^{(n)}$ on

$$A_\varepsilon^{(n)} = \left\{ x^n = (x_1, \ldots, x_n) \mid 2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)} \right\}.$$

Huomautus. Määritelmä on siis sama kuin diskreetissä tapauksessa (paitsi, että $H(X)$:n olemassaolo pitää erikseen olettaa).

Kuten edellä, merkitään joukon B Lebesquen mittaa $m(B)$:llä. Yksinkertaiselle B (pallo, suorakulmainen särmiö,...) $m(B)$ on itse asiassa joukon B *tilavuus*.

Lause 6.12 (AEP)**.** *Olkoon $H(X)$ olemassa ja $\varepsilon > 0$. Tyypilliselle joukolle pätee*

(*i*) $\lim_{n\to\infty} \mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} = 1$,

(*ii*) $m(A_\varepsilon^{(n)}) \leq 2^{n(H(X)+\varepsilon)}$ *kaikilla* n,

(*iii*) *kaikilla* $\delta > 0$ *on* $m(A_\varepsilon^{(n)}) \geq (1-\delta)\, 2^{n(H(X)-\varepsilon)}$, *kun* n *on riittävän suuri.*

Huomautus. Nähdään, että diskreetin AEP:n (lause 3.2) $|A_\varepsilon^{(n)}|$ korvataan jatkuvassa tapauksessa ”tilavuudella” $m(A_\varepsilon^{(n)})$.

Todistus.

(*i*) Tyypillisen joukon jonolle x^n

$$2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}$$

eli

$$\left| -\frac{1}{n} \log p(x^n) - H(X) \right| \leq \varepsilon,$$

joten

$$\mathbb{P}\{X^n \notin A_\varepsilon^{(n)}\} = \mathbb{P}\left\{ \left| -\frac{1}{n} \log p(X^n) - H(X) \right| > \varepsilon \right\}.$$

Edelleen,

$$-\frac{1}{n}\log p(X^n) = -\frac{1}{n}\log\prod_{i=1}^{n} p(X_i) = \frac{1}{n}\sum_{i=1}^{n}\big[-\log p(X_i)\big].$$

Koska $\mathbb{E}\big[-\log p(X_i)\big] = H(X)$, seuraa suurten lukujen laista, että

$$\lim_{n\to\infty}\mathbb{P}\{X^n \notin A_\varepsilon^{(n)}\} = 0,$$

josta saadaan (i).

(ii) Käyttämällä joukon $A_\varepsilon^{(n)}$ määritelmää saadaan

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(x^n)\,dx^n \geq \int_{A_\varepsilon^{(n)}} p(x^n)\,dx^n \\
&\geq \int_{A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)}\,dx^n = 2^{-n(H(X)+\varepsilon)}\int_{A_\varepsilon^{(n)}} dx^n \\
&= 2^{-n(H(X)+\varepsilon)}\,m(A_\varepsilon^{(n)}),
\end{aligned}$$

joten

$$m(A_\varepsilon^{(n)}) \leq 2^{n(H(X)+\varepsilon)}.$$

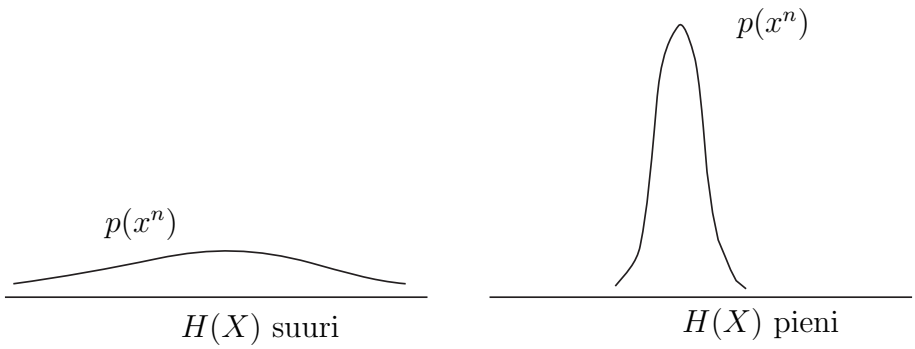
(iii) Olkoon n niin suuri, että $\mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} \geq 1-\delta$ ((i)-kohta). Silloin

$$\begin{aligned}
1-\delta \leq \mathbb{P}\{X^n \in A_\varepsilon^{(n)}\} &= \int_{A_\varepsilon^{(n)}} p(x^n)\,dx^n \\
&\leq \int_{A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)}\,dx^n = 2^{-n(H(X)-\varepsilon)}\,m(A_\varepsilon^{(n)}),
\end{aligned}$$

joten

$$m(A_\varepsilon^{(n)}) \geq (1-\delta)\,2^{n(H(X)-\varepsilon)}.$$

□



Kuva 6.5: Differentiaalientropia ja jakauman keskittyneisyys.

Yllä oleva todistus oli siis täysin analoginen diskreetin tapauksen kanssa. Aivan vastaavasti kuin diskreetissä tapauksessa, voidaan myös määritellä *yhteistyypillisyys*, kun

$$(X^n, Y^n) \sim p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$$

(vertaa määritelmä 5.14). AEP parille (X, Y) todistetaan kuten lauseessa 5.15, kunhan vaan $|A_\varepsilon^{(n)}|$ korvataan mitalla $m(A_\varepsilon^{(n)})$ ja oletetaan, että $H(X)$, $H(Y)$, $H(X, Y)$ ja $I(X; Y)$ ovat kaikki olemassa.

Lauseen 6.12 tulkinta on, että suurin osa $p(x^n)$:n todennäköisyysmassasta on joukossa, jonka "tilavuus" on $\approx 2^{nH(X)}$. Siten (vrt. kuva 6.5):

$$\begin{cases} H(X) \text{ suuri } \Leftrightarrow p(x^n) \text{ laajalle levinnyt} \\ H(X) \text{ pieni } \Leftrightarrow p(x^n) \text{ pienelle alueelle keskittynyt.} \end{cases}$$

6.3 Multinormaalijakauma

Vektorien $x^n, y^n \in \mathbb{R}^n$ *sisätulo* on

$$\langle x^n, y^n \rangle = \sum_{i=1}^{n} x_i y_i,$$

kun $x^n = (x_1, \ldots, x_n)$, $y^n = (y_1, \ldots, y_n)$. Vektorin x^n *normi* on

$$\|x^n\| = \sqrt{\langle x^n, x^n \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Satunnaisvektorilla $X^n = (X_1, \ldots, X_n)$ on *standardimultinormaalijakauma*, jos sillä on tiheysfunktio

$$p(x^n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\|x^n\|^2}, \quad x^n \in \mathbb{R}^n.$$

Tällöin

$$p(x^n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2} = \prod_{i=1}^{n} p(x_i),$$

missä $p(x_i)$ on $\mathrm{N}(0,1)$-jakauman tiheysfunktio, joten $X_1, \ldots, X_n \overset{iid}{\sim} N(0,1)$.

Matriisi $K \in \mathbb{R}^{n \times n}$ on *symmetrinen*, jos $K^T = K$. Symmetrinen matriisi K on *positiivisesti definiitti*, jos $\langle x^n, Kx^n \rangle > 0$ kaikilla $x^n \neq 0$. Tässä K:ta ajatellaan lineaarikuvauksena $\mathbb{R}^n \to \mathbb{R}^n$,

$$Kx^n \equiv \left[\begin{pmatrix} K_{11} & \ldots & K_{1n} \\ \vdots & & \vdots \\ K_{n1} & \ldots & K_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right]^T = x^n K^T = x^n K.$$

Voidaan osoittaa, että positiivisesti definiitti matriisi on aina säännöllinen ja $\det(K) > 0$.

Olkoon $\mu^n \in \mathbb{R}^n$ ja $K \in \mathbb{R}^{n \times n}$ positiivisesti definiitti. Satunnaisvektorilla X^n on *multinormaalijakauma parametrein μ^n ja K*, jos sillä on tiheysfunktio

$$p(x^n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(K)}} e^{-\frac{1}{2}\left\langle x^n - \mu^n, K^{-1}(x^n - \mu^n) \right\rangle}.$$

Tällöin merkitään $X^n \sim \mathrm{N}(\mu^n, K)$. Selvästikin satunnaisvektorilla X^n on standardimultinormaalijakauma, jos

$$X^n \sim \mathrm{N}(0, I_n),$$

missä $I_n \in \mathbb{R}^{n\times n}$ on yksikkömatriisi.

Edelleen, olkoon $N > 0$ positiivinen ja

$$K = NI_n = \begin{pmatrix} N & & 0 \\ & \ddots & \\ 0 & & N \end{pmatrix},$$

jolloin

$$K^{-1} = \frac{1}{N}\, I_n, \quad \det(K) = N^n.$$

Silloin $X^n \sim \mathrm{N}(\mu^n, K)$ jos ja vain jos X^n:llä on tiheysfunktio

$$p(x^n) = \frac{1}{(2\pi N)^{n/2}}\, e^{-\frac{1}{2N}\|x^n-\mu^n\|^2} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi N}}\, e^{-\frac{1}{2N}(x_i-\mu_i)^2},$$

jolloin siis $X_1, \ldots, X_n \perp\!\!\!\perp$, $X_i \sim \mathrm{N}(\mu_i, N)$.

Olkoon sitten yleisesti $X^n \sim \mathrm{N}(\mu^n, K)$, jossa K on positiivisesti definiitti. Voidaan osoittaa, että

$$\begin{cases} \mu^n = (\mu_1, \ldots, \mu_n),\ \mu_i = \mathbb{E}(X_i),\ i = 1, \ldots, n, \\ K = (K_{ij}),\ K_{ij} = \mathbb{E}\big[(X_i - \mu_i)(X_j - \mu_j)\big],\ i, j = 1, \ldots, n, \end{cases}$$

eli μ^n on X^n:n odotusarvo(vektori) ja K on X^n:n kovarianssimatriisi.

Matriisiteorian nojalla tiedetään, että symmetrisellä positiivisesti definiitillä matriisilla K on ortonormaalit ominaisvektorit $u_1^n, \ldots, u_n^n$ ja niitä vastaavat positiiviset ominaisarvot $a_1, \ldots, a_n$,

$$\begin{cases} Ku_i^n = a_i u_i^n, \\ \langle u_i^n, u_j^n \rangle = \delta_{ij}, \end{cases}$$

$i, j = 1, \ldots, n$. Kirjoittamalla

$$x^n - y^n = \sum_{i=1}^{n} \langle x^n - y^n, u_i^n \rangle\, u_i^n$$

ja huomaamalla, että $K^{-1}u_i^n = \frac{1}{a_i}\,u_i^n$, saadaan helposti

$$\left\langle x^n - \mu^n, K^{-1}(x^n - \mu^n)\right\rangle = \sum_{i=1}^{n} \frac{1}{a_i}\left\langle x^n - \mu^n, u_i^n\right\rangle^2 .$$

Siten $p(x)$:n tasa-arvopinnat $\mathbb{R}^n$:ssä ovat muotoa

$$\sum_{i=1}^{n} \frac{1}{a_i}\left\langle x^n - \mu^n, u_i^n\right\rangle^2 = c, \qquad c > 0$$

eli

$$\sum_{i=1}^{n} \frac{\left\langle x^n - \mu^n, u_i^n\right\rangle^2}{(\sqrt{ca_i})^2} = 1.$$

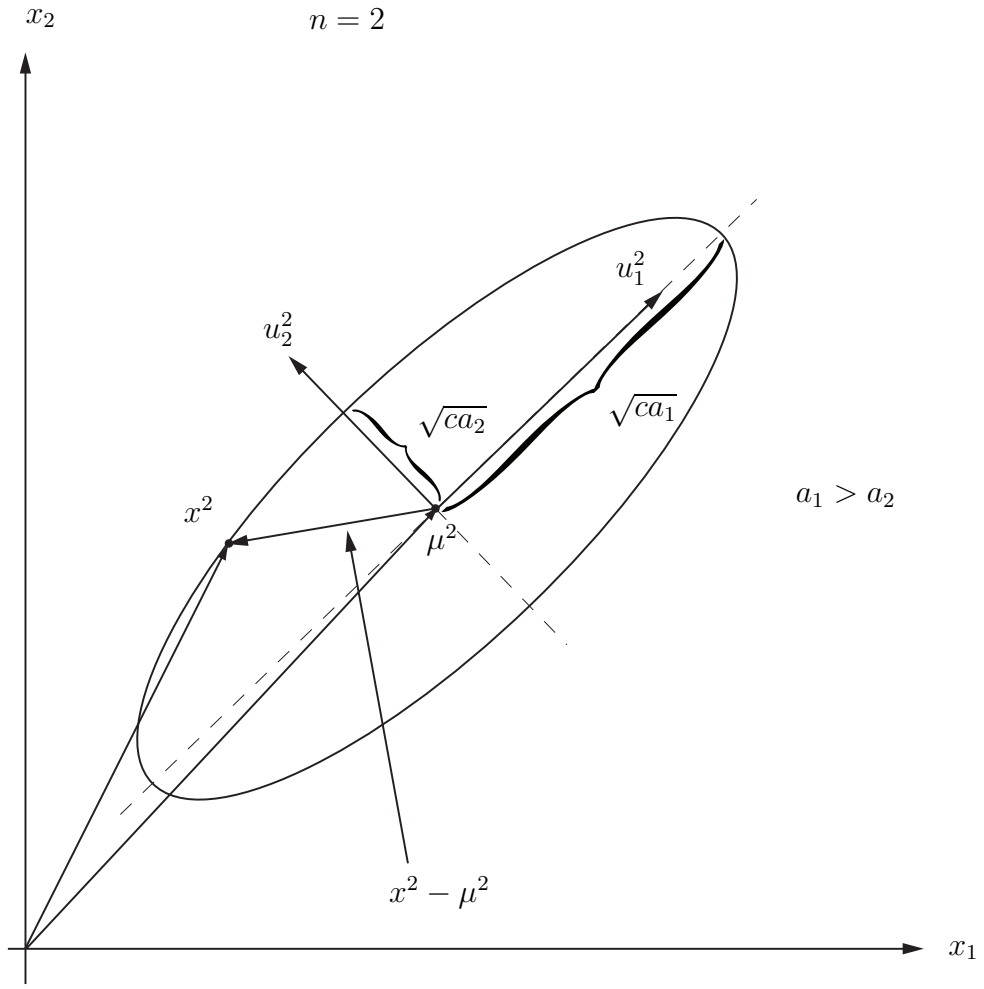
Nämä ovat yleisessä tapauksessa ($n > 2$) *ellipsoideja*, joiden keskipiste on μ^n, puoliakselien suunnat u_i^n ja puoliakselien pituudet $\sqrt{ca_i}$ (kuva 6.6).

Erityisesti tiheysfunktion $p(x^n) \sim \mathrm{N}(\mu^n, NI_n)$ tasa-arvopinnat ovat *palloja*, koska NI_n:n ominaisarvot ovat $N, \ldots, N$. Tämä on tietysti muutenkin selvää, koska $\|x^n - \mu^n\|^2 = c$ on pallon yhtälö.

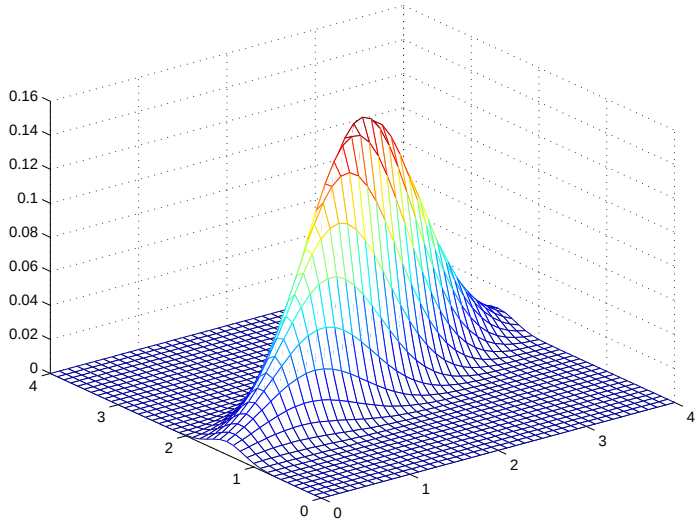
Esimerkki 6.5. Kuvissa 6.7 ja 6.8 on perspektiivikuva ja tasa-arvokäyrät kaksiulotteisesta normaalijakaumasta, kun $\mu^2 = (2, 2)$ ja

$$K = \begin{pmatrix} 0.66 & 0.2 \\ 0.2 & 0.12 \end{pmatrix}.$$

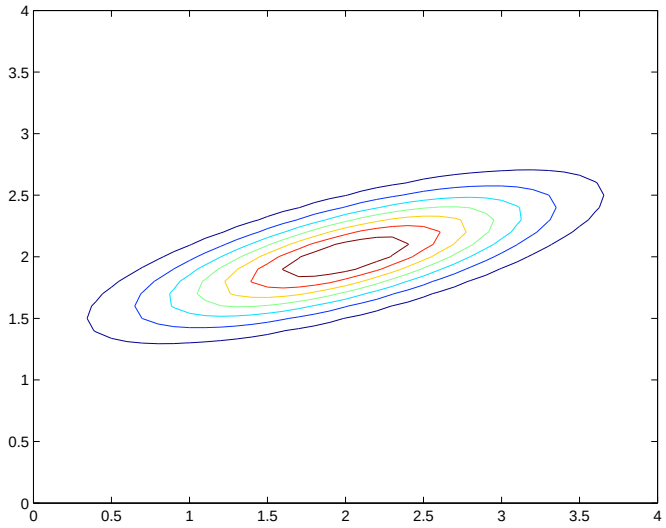
‖



Kuva 6.6: Kaksiulotteisen normaalijakauman tiheysfunktion tasa-arvokäyrä.



Kuva 6.7: Kaksiulotteisen normaalijakauman tiheysfunktion kuvaaja.



Kuva 6.8: Kaksiulotteisen normaalijakauman tiheysfunktion tasa-arvokäyrät.

Luku 7

Diskreettiaikainen Gaussin kanava

7.1 Kanavamalli

Tarkastellaan kanavaa, jossa syöte- ja tulosteaakkostot ovat *jatkuvia*, $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, mutta symbolien syöttö tapahtuu edelleen yksi kerrallaan (diskreetti aika). Tämä malli kuvaa hyvin joitain tiedonsiirtotilanteita.

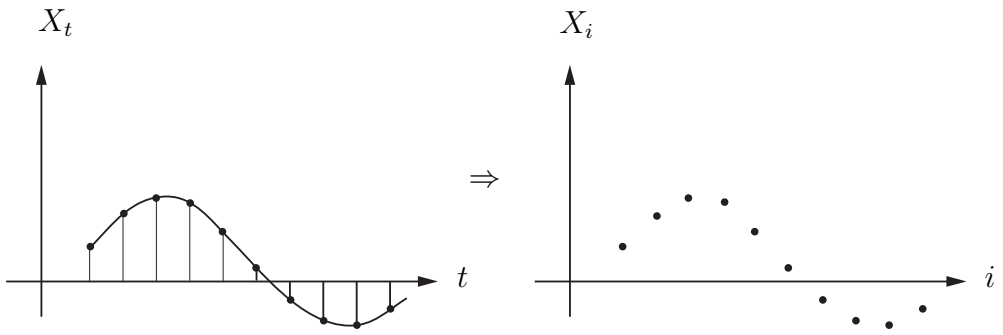
Esimerkki 7.1. Tällainen tilanne syntyy esimerkiksi, kun jatkuva signaali X_t diskretoidaan ajan suhteen tiedonsiirtoa varten (kuva 7.1). ‖

Syötetään symbolit kanavaan taas *lohkoissa*:

$$\mathbb{R}^n \ni (x_1, \ldots, x_n) \longrightarrow \boxed{\text{informaatiokanava}} \longrightarrow (y_1, \ldots, y_n) \in \mathbb{R}^n.$$

Kanavan ominaisuudet kuvaa ei-negatiiviset funktiot

$$p(y_1, \ldots, y_n \mid x_1, \ldots, x_n), \quad n \in \mathbb{N}_+.$$



Kuva 7.1: Jatkuva-aikainen signaali diskretoidaan.

Täsmällisesti ottaen $p = p_n$ on muuttujien $x_1, \ldots, x_n, y_1, \ldots, y_n$ Borelin funktio. Lisäksi vaaditaan, että

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(y_1, \ldots, y_n \mid x_1, \ldots, x_n)\, dy_1 \cdots dy_n = 1$$

kaikilla $(x_1, \ldots, x_n) \in \mathbb{R}^n$. Tulkinta on se, että $p(y_1, \ldots, y_n \mid x_1, \ldots, x_n)$ on tulosteiden $(y_1, \ldots, y_n)$ ehdollinen tiheysfunktio, kun syöte on $(x_1, \ldots, x_n)$.

Merkitään seuraavassa $x^n = (x_1, \ldots, x_n)$, $y^n = (y_1, \ldots, y_n)$, kuten ennenkin. Olkoon $N > 0$. Kyseessä on *Gaussin kanava*, jos

$$p(y^n \mid x^n) = \frac{1}{(2\pi N)^{n/2}}\, e^{-\frac{1}{2N}\|y^n - x^n\|^2}. \tag{7.1}$$

Siis $p(y^n \mid x^n)$ on $\mathrm{N}(x^n, NI_n)$-jakauman tiheysfunktio (vrt. luku 6.3).

Olkoon $X^n \sim p(x^n)$ jatkuva satunnaismuuttuja. Silloin saadaan satunnainen syöte-tulostepari $(X^n, Y^n) \sim p(x^n, y^n)$ määrittelemällä

$$p(x^n, y^n) = p(x^n)\, p(y^n \mid x^n).$$

Jos $Z^n = Y^n - X^n$, niin tekemällä muuttujan vaihto tiheysfunktiossa $p(x^n, y^n)$ nähdään helposti (vrt. harjoitustehtävät), että $(X^n, Z^n) \sim p(x^n, z^n)$, missä

$$p(x^n, z^n) = p(x^n) \frac{1}{(2\pi N)^{n/2}}\, e^{-\frac{1}{2N}\|z^n\|^2}.$$

Siten $X^n \perp\!\!\!\perp Z^n$ ja $Z^n \sim p(z^n)$,

$$p(z^n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x^n, z^n)\, dx^n = \frac{1}{(2\pi N)^{n/2}}\, e^{-\frac{1}{2N}\|z^n\|^2} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi N}}\, e^{-\frac{1}{2N} z_i^2}.$$

Siis $Z^n = (Z_1, \ldots, Z_n) \sim \mathrm{N}(0, NI_n)$ ja $Z_1, \ldots, Z_n \overset{iid}{\sim} \mathrm{N}(0, N)$.

Kääntäen, jos $Y^n = X^n + Z^n$ ja $X^n \perp\!\!\!\perp Z^n$, $Z^n \sim \mathrm{N}(0, NI_n)$, on helppo nähdä, että $p(y^n \mid x^n)$ on muotoa (7.1).

Näin diskreettiaikainen Gaussin kanava (satunnaisella syötteellä X^n) voitaisiin määritellä myös kaavalla

$$Y^n = X^n + Z^n, \qquad X^n \perp\!\!\!\perp Z^n, \qquad Z^n \sim \mathrm{N}(0, NI_n),$$

missä $N =$*kohinavarianssi* (kuva 7.2). Kanavassa syötteeseen X^n siis lisätään normaalijakautunutta riippumatonta kohinaa (häiriötä). Komponenteittain tämä merkitsee, että

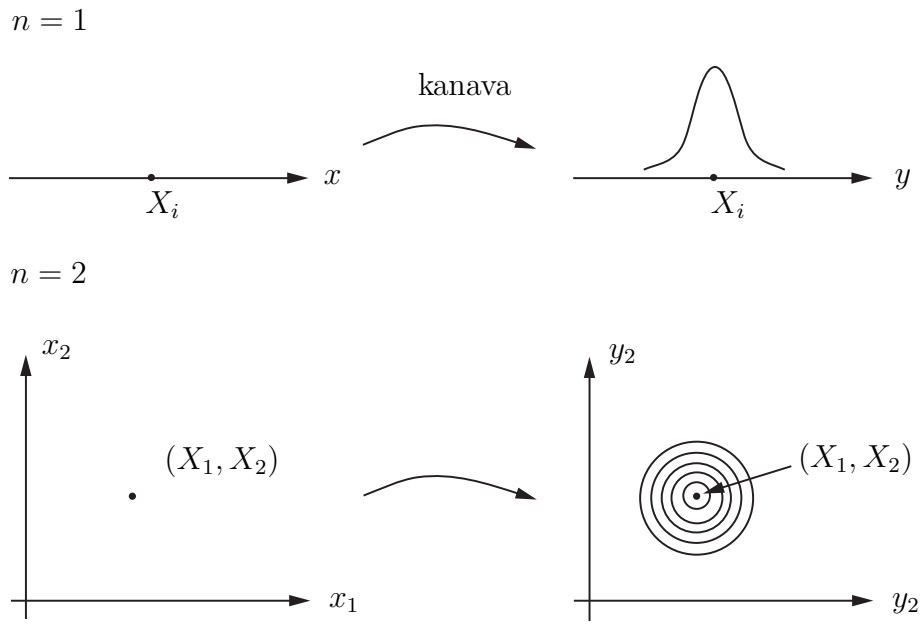
$$Y_i = X_i + Z_i,$$

missä $Z_i \perp\!\!\!\perp X_i$, $Z_i \sim \mathrm{N}(0, N)$. Tällainen malli on usein perusteltu todennäköisyyslaskennan keskeisen raja-arvolauseen nojalla.

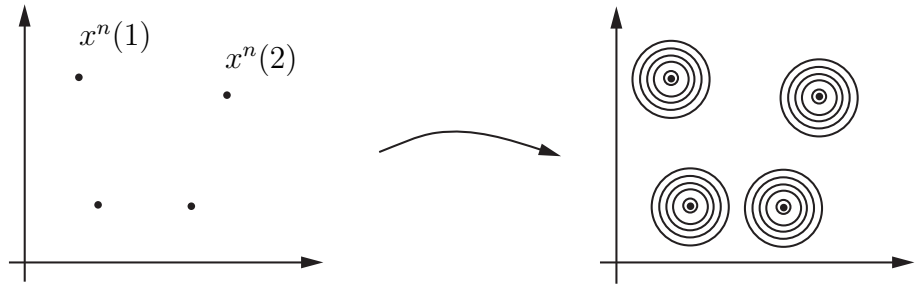
Käytännössä $|x_i|$ ei voi olla mielivaltaisen suuri silloin kun se esimerkiksi on jonkin fysikaalisen signaalin amplitudi. Jatkossa asetammekin rajoituksen "keskimääräiselle teholle" vaatimalla, että

$$\frac{1}{n}\|x^n\|^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \le P,$$

jollain kiinteällä $P > 0$. Tässä P on *tehoraja*. Ilman tällaista (realistista) rajoitusta voitaisiin kanavan läpi siirtää informaatiota miten suurella nopeudella tahansa liki virheettömästi, koska koodisanat $x^n(1), \ldots, x^n(M)$ voitaisiin valita niin että ne ovat hyvin kauakana toisistaan ja tällöin dekoodaus liki virheettömästi olisi helppoa. Koodisanojen määrälle M:llä ei tarvitsisi asettaa mitään rajoitusta (kuva 7.3).



Kuva 7.2: Diskreettiaikainen Gaussin kanava, kun lohkon pituus on 1 (ylempi kuva) tai 2 (alempi kuva).



Kuva 7.3: Tehoa lisäämällä voitaisiin koodisanat periaatteessa valita siten, että ne ovat kaukana toisistaan ja näin päästä liki virheettömään tiedonsiirtoon.

7.2 Koodaus ja dekoodaus

Koodauksesa ja dekoodauksessa tarvittavat käsitteet määritellään melkein kuten luvussa 5. Tarkastellaan viestejä $1, \ldots, M$ ja olkoon $P > 0$. Kun $x^n \in \mathbb{R}^n$, $r > 0$, merkitään

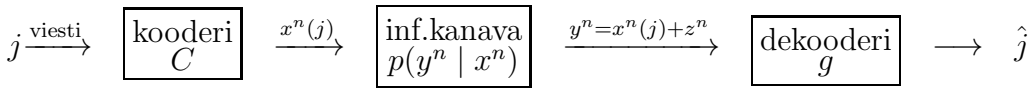
$$\bar{B}(x^n, r) = \{y^n \in \mathbb{R}^n \mid \|y^n - x^n\|^2 \leq r^2\}.$$

Kooderi on kuvaus $C : \{1, \ldots, M\} \to \bar{B}\big(0, \sqrt{nP}\big)$.

Koodisanoille $C(j) = x^n(j) = \big(x_1(j), \ldots, x_n(j)\big)$ pätee

$$\|x^n(j)\|^2 \leq nP \qquad \text{eli} \qquad \frac{1}{n}\sum_{i=1}^{n} x_i(j)^2 \leq P.$$

Dekooderi on funktio $g : \mathbb{R}^n \to \{1, \ldots, M\}$. Tiedonssiirtoa kanavan läpi kuvaa kaavio



Tässä $\hat{j} = g(y^n)$ on arvaus lähetetystä viestistä j.

(M, n)-*koodi* on pari (C, g), kuten ennen.

Määritelmä 7.1. (M, n)-koodin (C, g) *viestin* j *virhetodennäköisyys* on

$$\lambda_j = \int\limits_{\{y^n \mid g(y^n) \neq j\}} p\big(y^n \mid x^n(j)\big)\, dy^n,$$

missä $\{y^n \mid g(y^n) \neq j\}$ on niiden y^n:ien joukko, jotka dekoodataan väärin, kun lähetetään viesti j ja λ_j on tällaisen y^n:n sattumistodennäköisyys.

(M,n)-koodin (C,g) *keskimääräinen* ja *maksimaalinen virhetodennäköisyys* määritellään kuten ennen,

$$\bar{\lambda} = \frac{1}{M}\sum_{j=1}^{M} \lambda_j, \qquad \lambda^{(n)} = \max\{\lambda_1, \ldots, \lambda_M\}.$$

Edelleen, jos viesti on satunnainen, $J \sim p(j)$, voidaan määritellä virhetodennäköisyys

$$P_e = \mathbb{P}\big\{g(Y^n) \neq J\big\},$$

ja, kuten ennen, $P_e = \sum_{j=1}^{M} p(j)\,\lambda_j$.

(M,n)-koodin (*tiedonsiirto*)*nopeus* on $R = \log(M/n)$. Nopeus R on *saavutettavissa*, jos on olemassa $(\lceil 2^{nR} \rceil, n)$-koodit (C_n, g_n), joille $\lambda^{(n)} \to 0$, kun $n \to \infty$.

Mikä voisi olla aikadiskreetin Gaussin kanavan maksimaalinen tiedonsiirtonopeus R, jos syötteeltä vaaditaan tehorajoitus $\frac{1}{n}\|x^n\|^2 \leq P$? Aikaisemmin osoitettiin (luku 5.7), että maksimaalisen R:n määräsi kapasiteetti,

$$C = \max_{p(x)} I(X;Y).$$

Tarkastellaan siis syötettä X, $X^2 \leq P$ ja tulostetta

$$Y = X + Z,$$

$X \perp\!\!\!\perp Z$, $Z \sim \mathrm{N}(0,N)$, $N > 0$. Oletetaan, että X on lisäksi sellainen, että kaikki suureet seuraavissa laskuissa on määritelty. Saadaan

$$I(X;Y) = H(Y) - H(Y \mid X) = H(Y) - H(Z) \overset{\text{luku 6.1}}{=} H(Y) - \frac{1}{2}\log(2\pi e N),$$

missä toisessa yhtälössä käytettiin harjoitustehtävänä johdettua tulosta $H(Y|X) = H(Z)$. Edelleen,

$$\begin{aligned} D^2(Y) = D^2(X+Z) &\overset{X \perp\!\!\!\perp Z}{=} D^2(X) + D^2(Z) \\ &= E(X^2) - \big[E(X)\big]^2 + N \leq P + N. \end{aligned} \tag{7.2}$$

Siten lauseen 6.3 nojalla

$$H(Y) \leq \frac{1}{2}\log[2\pi e D^2(Y)] \leq \frac{1}{2}\log\big[2\pi e(P+N)\big], \tag{7.3}$$

joten

$$\begin{aligned} I(X;Y) &\leq \frac{1}{2}\log\big[2\pi e(P+N)\big] - \frac{1}{2}\log(2\pi eN) = \frac{1}{2}\log\left(\frac{P+N}{N}\right) \\ &= \frac{1}{2}\log\left(1+\frac{P}{N}\right). \end{aligned}$$

Yhtäsuuruus pätee jos ja vain jos $X \sim \mathrm{N}(0,P)$, koska tällöin $Y \sim \mathrm{N}(0,N+P)$ ja epäyhtälöissä (7.3) pätevät yhtäsuuruudet (lause 6.3) ja käänteen. Kun $X \sim \mathrm{N}(0,P)$, pätevät kaikki edellä olevat laskut myös.

Aikadiskreetin Gaussin kanavan *kapasiteetti* on

$$C = \sup\{R \mid \text{kanavan tiedonsiirtonopeus } R \text{ saavutettavissa}\}.$$

Tulemme osoittamaan:

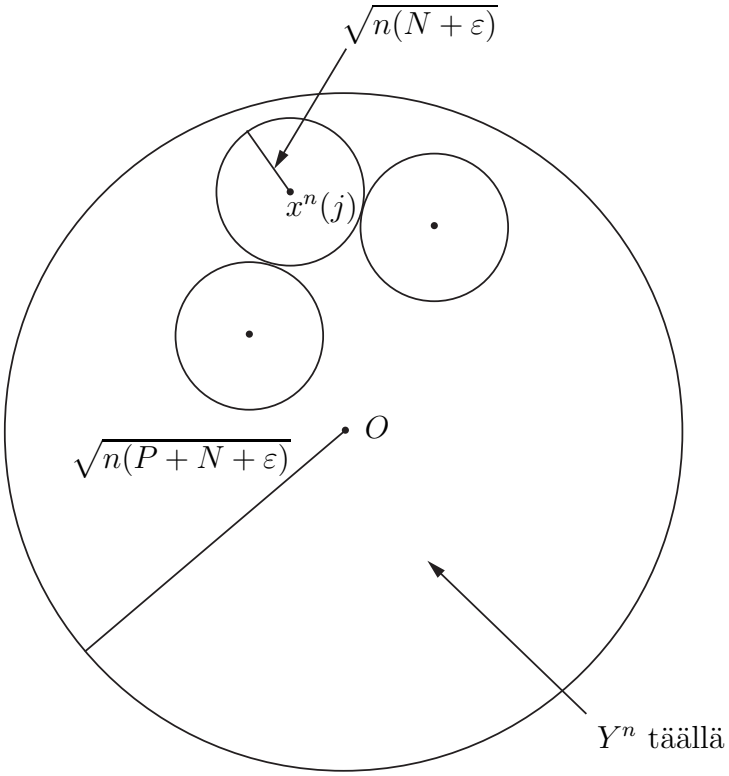
$$C = \frac{1}{2}\log\left(1+\frac{P}{N}\right).$$

Esitetään vielä heuristinen perustelu nopeuden $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$ saavuttamiselle. Tarkastellaan koodisanaa $x^n(j)$ ja sitä vastaavaa tulostetta $Y^n = x^n(j) + Z^n$. Olkoon $\varepsilon > 0$. Silloin

$$\begin{aligned} &\mathbb{P}\left\{\|Y^n - x^n(j)\|^2 > n(N+\varepsilon) \mid X^n = x^n(j)\right\} \\ &= \int\limits_{\|y^n - x^n(j)\|^2 > n(N+\varepsilon)} p\big(y^n \mid x^n(j)\big)\, dy^n = \frac{1}{(2\pi N)^{n/2}} \int\limits_{\|z^n\|^2 > n(N+\varepsilon)} e^{-\frac{1}{2N}\|z^n\|^2}\, dz^n \\ &= \mathbb{P}\left\{\|Z^n\|^2 > n(N+\varepsilon)\right\} = \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i^2 > N+\varepsilon\right\}. \end{aligned}$$

Mutta $E(Z_i^2) = D^2(Z_i) = N$, joten suurten lukujen lain nojalla

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i^2 > N+\varepsilon\right\} = 0.$$



Kuva 7.4: Heuristinen perustelu nopeuden $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$ saavuttamiselle.

Siten

$$\mathbb{P}\left\{Y^n \in \bar{B}\big(x^n(j), \sqrt{n(N+\varepsilon)}\,\big) \mid X^n = x^n(j)\right\} \approx 1,$$

kun n on suuri. Jos nyt dekooderi g määritellään siten, että

$$g(y^n) = j \quad \text{kaikilla } y^n \in \bar{B}\big(x^n(j), \sqrt{n(N+\varepsilon)}\,\big),$$

on virheen todennäköisyys pieni. Pallojen $\bar{B}\big(x^n(j), \sqrt{n(N+\varepsilon)}\,\big)$ täytyy tietysti olla pistevieraita, jotta g olisi hyvin määritelty.

Samoin, koska $Y^n = x^n(j)+Z^n$ ja $\|x^n(j)\|^2 \le nP$ kaikilla j, voidaan osoittaa, että

$$\mathbb{P}\left\{Y^n \in B\big(0, \sqrt{n(P+N+\varepsilon)}\,\big)\right\} \approx 1,$$

kun n on suuri. Erottuvia koodisanoja on siis niin monta, kuin $\sqrt{n(N+\varepsilon)}$-säteisiä palloja mahtuu toisiaan leikkaamatta $\sqrt{n(P+N+\varepsilon)}$-säteiseen palloon (kuva 7.4).

Yleisesti on n-ulotteisen pallon tilavuudelle on voimassa $m\big(\bar{B}(x^n, r)\big) = a_n r^n$

eräällä n:stä riippuvalla vakiolla $a_n > 0$. Siten erottuvia koodisanoja on noin

$$M = \frac{a_n\big(\sqrt{n(P+N+\varepsilon)}\,\big)^n}{a_n\big(\sqrt{n(N+\varepsilon)}\,\big)^n}$$

kappaletta. Nyt

$$M = \left(\frac{P+N+\varepsilon}{N+\varepsilon}\right)^{n/2} \approx \left(\frac{P+N}{N}\right)^{n/2} = 2^{\frac{n}{2}\log\left(1+\frac{P}{N}\right)},$$

eli

$$R \approx \frac{\log M}{n} \approx \frac{1}{2}\log\left(1+\frac{P}{N}\right).$$

Näin siis väitetty tiedonsiirtonopeus todella saavutetaan.

Diskreetti syöte, jatkuva tuloste

Shannonin toisen lauseen käsittelemässä tilanteessa äärellisen monesta koodisanasta yksi kerrallaan lähetetään Gaussin kanavaan:

$$X^n(j) \longrightarrow \boxed{\text{Gaussin kanava}} \longrightarrow Y^n$$

Siten syöte on *diskreetti* ja tuloste on *jatkuva*. Tähän asti kuitenkin olemme käsitelleet vain diskreetti/diskreetti ja jatkuva/jatkuva yhdistelmät syötteelle ja tulosteelle. Nyt pitää siis tarkastella erikseen myös diskreetti/jatkuva tilannetta.

Olkoon $X : \Omega \to \mathcal{X}$ diskreetti satunnaismuuttuja, $|\mathcal{X}| = m$, ja $Y : \Omega \to \mathbb{R}$ jatkuva satunnaismuuttuja. Oletetaan, että kaikilla $x \in \mathcal{X}$ on määritelty ehdollinen tiheysfunktio $p(y \mid x)$. Asetamme

$$p(x,y) = p(x)\,p(y \mid x).$$

Tämä tarkoittaa, että kaikilla $B \subset \mathbb{R}$ (Borelin joukko) ja $x \in \mathcal{X}$,

$$\mathbb{P}\{X = x,\, Y \in B\} = p(x) \int_B p(y \mid x)\,dy.$$

Tällöin erityisesti

$$\mathbb{P}\{Y \in B \mid X = x\} = \frac{\mathbb{P}\{X = x, Y \in B\}}{p(x)} = \int_B p(y \mid x)\, dy,$$

kuten pitääkin. Siis ehdollisen tiheysfunktion avulla voidaan laskea Y:n ehdollinen todennäköisyys, kun $X = x$. Satunnaismuuttujan Y tiheysfunktio on nyt

$$p(y) = \sum_{x \in \mathcal{X}} p(x)\, p(y \mid x),$$

minkä osoittaminen on harjoitustehtävänä.

Oletetaan seuraavassa, että kaikki määriteltävät suureet ovat olemassa. Näin tulee olemaan jatkossa kiinnostavassa tilanteessa, jossa $Y = X + Z$, $X \perp\!\!\!\perp Z$ ja $Z \sim \mathrm{N}(0, N)$, jolloin $p(y \mid x) \sim \mathrm{N}(x, N)$.

Määritellään satunnaismuuttujan Y *differentiaalientropia ehdolla* $X = x$,

$$H(Y \mid X = x) = -\int_{-\infty}^{\infty} p(y \mid x) \log p(y \mid x)\, dy$$

ja satunnaismuuttujan Y *differentiaalientropia ehdolla* X,

$$H(Y \mid X) = \sum_{x \in \mathcal{X}} p(x)\, H(Y \mid X = x).$$

Asetetaan

$$p(x \mid y) = \frac{p(x)\, p(y \mid x)}{p(y)}, \qquad \text{(”Bayesin kaava”)}$$

missä $p(x \mid y)$ on x:n *posterioritodennäköisyys* ehdolla $Y = y$. Määritellään satunnaismuuttujan X *entropia ehdolla* $Y = y$,

$$H(X \mid Y = y) = -\sum_{x \in \mathcal{X}} p(x \mid y) \log p(x \mid y)$$

ja satunnaismuuttujan X *entropia ehdolla* Y,

$$H(X \mid Y) = \int_{-\infty}^{\infty} p(y)\, H(X \mid Y = y)\, dy.$$

Nyt voidaan osoittaa (harjoitustehtävä), että

$$H(Y \mid X) \leq H(Y) \tag{7.4}$$

(ehdollistaminen vähentää entropiaa) ja yhtäsuuruus on voimassa jos ja vain jos $X \perp\!\!\!\perp Y$. Edelleen (todistus harjoitustehtävänä),

$$H(Y) + H(X \mid Y) = H(X) + H(Y \mid X). \tag{7.5}$$

Määritellään satunnaismuuttujien X *ja* Y *keskinäisinformaatio*

$$I(X;Y) = H(X) - H(X \mid Y).$$

Kaavan (7.5) nojalla

$$I(X;Y) = H(Y) - H(Y \mid X).$$

Voidaan myös todistaa *Fanon epäyhtälö*: Jos $g : \mathbb{R} \to \mathcal{X}$ (Borelin funktio) ja $P_e = \mathbb{P}\{g(Y) \neq X\}$, niin

$$H(P_e) + P_e \log\big(|\mathcal{X}| - 1\big) \geq H(X \mid Y),$$

missä $H(t) = -t \log t - (1-t)\log(1-t)$, $t \in [0,1]$. Todistus on samanlainen kuin (täysin) diskreetissä tapauksessa ja sivuutetaan.

Huomautus. Edellä voisi joka paikassa olla satunnaismuuttujan Y paikalla myös n-ulotteinen satunnaisvektori Y^n.

7.3 Shannonin toinen lause diskreettiaikaiselle Gaussin kanavalle

Lause 7.2. *Olkoon diskreettiaikaisen Gaussin kanavan kohinavarianssi* $N > 0$ *ja tehoraja* $P > 0$. *Silloin kanavan kapasiteetti on*

$$C = \frac{1}{2} \log\left(1 + \frac{P}{N}\right).$$

Todistus. Pitää siis osoittaa, että

$$\sup\{R \mid R \text{ saavutettavissa}\} = C.$$

Olkoon $R < C$. Osoitetaan ensin, että löytyy sellaiset $(\lceil 2^{nR} \rceil, n)$-koodit (C_n, g_n), että $\lambda^{(n)} \to 0$, kun $n \to \infty$. Voidaan olettaa, että $R > 0$. Olkoon $\varepsilon > 0$ ja merkitään $M_n = \lceil 2^{nR} \rceil$. Olkoon $p(x)$ jakauman $\mathrm{N}(0, P - \varepsilon)$ tiheysfunktio ja asetetaan

$$p(x^n) = \prod_{i=1}^{n} p(x_i),$$
$$p(x^n, y^n) = p(x^n)\, p(y^n \mid x^n),$$

$x^n, y^n \in \mathbb{R}^n$. Erityisesti siis $X^n \sim \mathrm{N}\big(0, (P - \varepsilon) I_n\big)$.

Olkoon $X^n(1), \ldots, X^n(M_n) \overset{iid}{\sim} p(x^n)$ ja käytetään satunnaisvektoreita $X^n(j)$ "kooderin" $\mathcal{C}_n$ koodisanoina. On syytä huomauttaa, että $\mathcal{C}_n$ ei välttämättä ole "oikea" kooderi (vertaa luvun 7.2 määritelmä), koska ei ehkä ole $\|X^n(j)\|^2 \leq nP$ kaikilla j. Lopuksi kuitenkin koodisanoja karsitaan siten, että tehorajoitus tulee olemaan voimassa.

Merkitään

$$\mathcal{C}_n = \big(X^n(1), \ldots, X^n(M_n)\big)$$

samaistaen kooderi $\mathcal{C}_n$ satunnaisten koodisanojensa kanssa. Merkitään vastaavasti

$$C_n = \big(x^n(1), \ldots, x^n(M_n)\big),$$

kun $(x^n(1), \ldots, x^n(M_n))$ on satunnaisvektorin $(X^n(1), \ldots, X^n(M_n))$ arvo (eli "realisaatio"). Siten $\mathcal{C}_n$:ää ajatellaan $(\mathbb{R}^n)^{M_n} \equiv \mathbb{R}^{nM_n}$-arvoisena satunnaisvektorina ja $C_n \in \mathbb{R}^{nM_n}$ on sen saama arvo, eräs konkreettinen kooderi.

Käytetään yhteistyypillisyysdekooderia g_n, kuten lauseessa 5.18, mutta lisätään tehorajoitus: $g_n(y^n) = \hat{j}$, jos $\big(x^n(\hat{j}), y^n\big) \in A_\varepsilon^{(n)}$, $\big(x^n(k), y^n\big) \notin A_\varepsilon^{(n)}$ kaikilla $k \neq \hat{j}$ ja lisäksi $||x^n(\hat{j})||^2 \leq nP$. Muutoin asetamme (mielivaltaisesti) $g_n(y^n) = 1$.

Kooderin $\mathcal{C}_n$ keskimääräinen virhe määritellään kaavalla

$$\bar{\lambda}(\mathcal{C}_n) = \frac{1}{M_n} \sum_{j=1}^{M_n} \lambda_j(\mathcal{C}_n),$$
$$\lambda_j(\mathcal{C}_n) = \int\limits_{\{y^n | g_n(y^n;\mathcal{C}_n) \neq j\}} p\big(y^n \mid X^n(j)\big)\, dy^n.$$

Vastaavasti määritellään $\bar{\lambda}(C_n)$ ja $\lambda_j(C_n)$ korvaamalla $\mathcal{C}_n$ C_n:llä ja $X^n(j)$ $x^n(j)$:llä. Silloin

$$\mathbb{E}\big[\bar{\lambda}(\mathcal{C}_n)\big] = \int\limits_{\mathbb{R}^{nM_n}} \bar{\lambda}(C_n)\, p(C_n)\, dC_n,$$

missä $p(C_n)$ on jakauman $\mathrm{N}\big(0, (P-\varepsilon) I_{nM_n}\big)$ tiheysfunktio ja siis $C_n \in \mathbb{R}^{nM_n}$.

Osoitetaan seuraavaksi, että

$$\lim_{n\to\infty} \mathbb{E}\big[\bar{\lambda}(\mathcal{C}_n)\big] = 0. \tag{7.6}$$

Nyt

$$\mathbb{E}\big[\bar{\lambda}(\mathcal{C}_n)\big] = \frac{1}{M_n} \sum_{j=1}^{M_n} \int\limits_{\mathbb{R}^{nM_n}} \lambda_j(C_n)\, p(C_n)\, dC_n.$$

Olkoon j kiinteä. Edellä olevassa kaavassa

$$\lambda_j(C_n) = \mathbb{P}\{\text{"viesti } j \text{ dekoodataan väärin"} \mid \text{"kooderi on } C_n\text{"}\},$$

ja kokonaistodennäköisyyden kaavan nojalla

$$\int\limits_{\mathbb{R}^{nM_n}} \lambda_j(C_n)\, p(C_n)\, dC_n$$
$$= \mathbb{P}\{\text{"viesti } j \text{ dekoodataan väärin, kun kooderi } C_n \text{ valitaan satunnaisesti"}\}$$
$$\equiv \mathbb{P}(B_{j,n}),$$

missä vielä merkittiin $B_{j,n}$:llä yo. kaavassa olevaa tapahtumaa "viesti ... satunnaisesti".

Nyt $B_{j,n} \subset D_{j,n} \cup E_{j,n} \cup F_{j,n}$, missä

$$\begin{aligned} D_{j,n} &= \left\{ \left(X^n(j), Y^n\right) \notin A_\varepsilon^{(n)} \right\}, \\ E_{j,n} &= \left\{ \left(X^n(k), Y^n\right) \in A_\varepsilon^{(n)} \text{ jollain } k \neq j \right\}, \\ F_{j,n} &= \left\{ \|X^n(j)\|^2 > nP \right\}, \end{aligned}$$

koska selvästi $D_{j,n}^c \cap E_{j,n}^c \cap F_{j,n}^c \subset B_{j,n}^c$. Aivan kuten lauseessa 5.18, AEP:stä saadaan (lause 6.12 ja vastaava yhteistyypillisyys AEP),

$$\mathbb{P}(D_{j,n}) = \nu_n,$$

jossa ν_n ei riipu j:stä ja $\nu_n \to 0$, kun $n \to \infty$ ja myös että

$$\mathbb{P}(E_{j,n}) \leq 2^{-n(I(X;Y)-R-3\varepsilon)}.$$

Nyt $X \sim \mathrm{N}(0, P-\varepsilon)$, joten luvun 7.2 nojalla

$$I(X;Y) = \frac{1}{2} \log \left(1 + \frac{P-\varepsilon}{N}\right).$$

Koska $R < C = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$, on silloin

$$I(X;Y) - R - 3\varepsilon > 0,$$

kun $\varepsilon > 0$ on tarpeeksi pieni. Siten $\mathbb{P}(E_{j,n}) = \mu_n$, missä μ_n ei riipu j:stä ja $\lim_{n\to\infty} \mu_n = 0$. Edelleen, $X^n(j) \sim \mathrm{N}\big(0, (P-\varepsilon)I_n\big)$, joten $\mathbb{E}\big[X_i(j)^2\big] = P - \varepsilon$. Siten heikon suurten lukujen lain nojalla

$$\frac{1}{n}\|X^n(j)\|^2 = \frac{1}{n}\sum_{i=1}^{n} X_i(j)^2 \longrightarrow P - \varepsilon$$

stokastisesti, kun $n \to \infty$. Näin

$$\mathbb{P}(F_{j,n}) = \mathbb{P}\left\{\frac{1}{n}\|X^n(j)\|^2 > P\right\} = \rho_n$$

ei riipu j:stä ja $\rho_n \to 0$, kun $n \to \infty$. Siten $\lim\limits_{n\to\infty} \mathbb{P}(B_{j,n}) = 0$ ja kuten lauseessa 5.18, $\lim\limits_{n\to\infty} \mathbb{E}\big[\bar{\lambda}(\mathcal{C}_n)\big] = 0$, eli (7.6) pätee.

Jokaiselle kooderille C_n on $0 \leq \bar{\lambda}(C_n) \leq 1$. Olkoon

$$\tau_n = \inf_{C_n} \bar{\lambda}(C_n).$$

Silloin $\lim_{n\to\infty} \tau_n = 0$, koska jos $\tau_n \geq \delta > 0$ äärettömän monella n, on näille n

$$\mathbb{E}\big[\bar{\lambda}(\mathcal{C}_n)\big] = \int_{\mathbb{R}^{nM_n}} \bar{\lambda}(C_n)\, p(C_n)\, dC_n \geq \tau_n \int_{\mathbb{R}^{nM_n}} p(C_n)\, dC_n = \tau_n \geq \delta > 0,$$

mikä on ristiriidassa tuloksen 7.6 kanssa. Valitaan nyt kaikilla $n \in \mathbb{N}_+$ kooderi C_n^*, jolle

$$\tau_n \leq \bar{\lambda}(C_n^*) < \tau_n + \frac{1}{n}.$$

Silloin

$$\lim_{n\to\infty} \bar{\lambda}(C_n^*) = 0.$$

Pitämällä $l_n = \lceil M_n/2 \rceil$ parasta koodisanaa $x^n(j_{n,1}), \ldots, x^n(j_{n,l_n})$ saadaan koodit $C_n^*|\{j_{n,1}, \ldots, j_{n,l_n}\}$, joilla maksimaalinen virhe $\lambda^{(n)} \to 0$, kun $n \to \infty$.

Kun n on riittävän suuri, on $||x^n(j_{n,k})||^2 \leq nP$ kaikilla $j_{n,k} \neq 1$, koska muutoin $\lambda_{j_{n,k}}(C_n^*) = 1$ eikä voisi olla $\lambda^{(n)} \to 0$. Siten, pudottamalla tarvittaessa pois vielä $x^n(1)$ (jos se on mukana) , saadaan kooderit $C_n^*|\{j'_{n,1}, \ldots, j'_{n,k_n}\}$, $\lceil M_n/2 \rceil \geq k_n \geq \lceil M_n/2 \rceil - 1$, joilla $\lambda^{(n)} \to 0$. Nyt löytyy sellainen R_n, että $k_n = 2^{nR_n}$,

$$R - \frac{2}{n} \leq R_n < R + \frac{1}{n}$$

ja loppu menee kuten lauseessa 5.18.

Kääntäen, olkoon $(\lceil 2^{nR} \rceil, n)$-koodeille voimassa $\lambda^{(n)} \to 0$. Osoitetaan, että silloin välttämättä on $R \leq \frac{1}{2} \log\big(1 + \frac{P}{N}\big)$.

Olkoot koodin (C_n, g_n) koodisanat $x^n(1), \ldots, x^n(M_n)$, $M_n = \lceil 2^{nR} \rceil$, ja

$$||x^n(j)||^2 \leq nP, j = 1, \ldots, M_n.$$

Olkoon viesti J_n tasaisesti jakautunut joukossa $\{1, \ldots, M_n\}$,

$$p(j) = \mathbb{P}\{J_n = j\} = \frac{1}{M_n}, \quad j = 1, \ldots, M_n.$$

Olkoon $X^n = C_n(J_n)$ viestiä J_n vastaava koodisana. Silloin X^n on diskreetti satunnaismuuttuja, jonka arvojoukko on $\mathcal{X} = \{x^n(1), \ldots, x^n(M_n)\}$ ja jolle (vrt. lause 5.18)

$$p(x^n) = \mathbb{P}\{X^n = x^n\} = \frac{|\{j \mid C_n(j) = x^n\}|}{M_n}, \quad x^n \in \mathcal{X}.$$

Näin saadaan syöte-tulostepari $(X^n, Y^n) \sim p(x^n, y^n)$, missä $p(x^n, y^n) \equiv p(x^n)\, p(y^n \mid x^n)$, jolloin

$$\mathbb{P}\{X = x^n, y^n \in B\} = p(x^n) \int_B p(y^n \mid x^n)\, dy^n,$$

kun $x^n \in \mathcal{X}$, $B \subset \mathbb{R}^n$ (ks. luku 7.2).

Nyt

$$I(J_n; Y^n) = H(J_n) - H(J_n \mid Y^n)$$

ja $H(J_n) = \log M_n$ (lause 2.12). Siis

$$\log M_n = H(J_n \mid Y^n) + I(J_n; Y^n). \tag{7.7}$$

Arvioidaan nyt (7.7):n oikean puolen termejä erikseen. Ensiksikin

$$I(J_n; Y^n) \leq I(X^n; Y^n), \tag{7.8}$$

koska $X^n = C_n(J_n)$ on J_n:n funktio. Toisaalta,

$$I(X^n; Y^n) = H(Y^n) - H(Y^n \mid X^n) = H(Y^n) - H(Z^n),$$

koska $Y^n = X^n + Z^n$, $Z^n \sim \mathrm{N}(0, NI_n)$, $X^n \perp\!\!\!\perp Z^n$. Edelleen,

$$H(Y^n) \leq \sum_{i=1}^{n} H(Y_i)$$

differentiaalientropian perusominaisuuksien perusteella. Koska $Z_1, \ldots, Z_n \perp\!\!\!\perp$, on lisäksi

$$H(Z^n) = \sum_{i=1}^{n} H(Z_i).$$

Siten

$$I(X^n;Y^n) \le \sum_{i=1}^{n} \big[H(Y_i) - H(Z_i)\big]. \tag{7.9}$$

Edelleen,

$$Y_i = X_i + Z_i,$$

missä X_i on koodisanan i:s komponentti, $Z_i \sim \mathrm{N}(0,N)$ ja $X_i \perp\!\!\!\perp Z_i$. Merkitään

$$P_i = \mathbb{E}(X_i^2) = \frac{1}{M_n}\sum_{j=1}^{M_n} x_i(j)^2.$$

Silloin

$$\begin{aligned} D^2(Y_i) &= \mathbb{E}(Y_i^2) - \big[\mathbb{E}(Y_i)\big]^2 \le \mathbb{E}(Y_i^2) \\ &= \mathbb{E}\big(X_i^2 + 2X_iZ_i + Z_i^2\big) \\ &= \mathbb{E}(X_i^2) + 2\mathbb{E}(X_i)\,\mathbb{E}(Z_i) + \mathbb{E}(Z_i^2) \\ &= P_i + N, \end{aligned}$$

koska $\mathbb{E}(Z_i) = 0$. Siten lauseesta 6.3 saadaan

$$H(Y_i) \le \frac{1}{2}\log\big[2\pi e D^2(Y_i)\big] \le \frac{1}{2}\log\big[2\pi e(P_i+N)\big].$$

Toisaalta $H(Z_i) = \frac{1}{2}\log(2\pi eN)$, joten kaavasta (7.9) saadaan

$$\begin{aligned} I(X^n;Y^n) &\le \sum_{i=1}^{n}\left\{\frac{1}{2}\log\big[2\pi e(P_i+N)\big] - \frac{1}{2}\log(2\pi eN)\right\} \\ &= \sum_{i=1}^{n}\frac{1}{2}\log\left[\frac{2\pi e(P_i+N)}{2\pi eN}\right] = \sum_{i=1}^{n}\frac{1}{2}\log\left(1+\frac{P_i}{N}\right). \end{aligned} \tag{7.10}$$

Kaavan (7.7) oikean puolen ensimmäiselle termille saadaan arvio taas Fanon epäyhtälöstä:

$$H(J_n \mid Y^n) \le H\big(P_e(C_n)\big) + P_e(C_n)\log(M_n - 1), \tag{7.11}$$

missä $P_e(C_n) = \mathbb{P}\{g_n(Y^n) \ne J_n\}$. Mutta $H\big(P_e(C_n)\big) \le 1$, $M_n = \lceil 2^{nR}\rceil$ ja, kuten lauseessa 5.18 (vrt. kaava (5.25)), on $P_e(C_n) \le \lambda^{(n)}$.

Siten tulosten (7.7), (7.8), (7.10) ja (7.11) nojalla

$$\begin{aligned} nR &\leq \log M_n \\ &\leq 1 + \lambda^{(n)} nR + \sum_{i=1}^{n} \frac{1}{2} \log\left(1 + \frac{P_i}{N}\right), \end{aligned}$$

joten

$$R \leq \frac{1}{n} + \lambda^{(n)} R + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} \log\left(1 + \frac{P_i}{N}\right). \tag{7.12}$$

Lisäksi tiedetään (harjoitustehtävä), että

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} \log\left(1 + \frac{P_i}{N}\right) \leq \frac{1}{2} \log\left(1 + \frac{1}{n} \sum_{i=1}^{n} \frac{P_i}{N}\right). \tag{7.13}$$

Vielä $||x^n(j)||^2 \leq nP$ kaikilla j, joten

$$\frac{1}{n} \sum_{i=1}^{n} P_i = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{M_n} \sum_{j=1}^{M_n} x_i(j)^2 = \frac{1}{M_n} \sum_{j=1}^{M_n} \frac{1}{n} \sum_{i=1}^{n} x_i(j)^2 \leq P.$$

Siten epäyhtälöiden (7.12) ja (7.13) nojalla

$$R \leq \frac{1}{n} + \lambda^{(n)} R + \frac{1}{2} \log\left(1 + \frac{P}{N}\right).$$

Kun $n \to \infty$, saadaan tästä väite. □

Luku 8

Jatkuva-aikainen Gaussin kanava

8.1 Hilbertin avaruuksista

Olkoon V (reaalikertoiminen) vektoriavaruus. Kuvaus $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ on *sisätulo*, jos kaikilla $x, y, z \in V$ ja $a \in \mathbb{R}$ pätee

(i) $\langle x, y \rangle = \langle y, x \rangle$,

(ii) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$,

(iii) $\langle ax, y \rangle = a \langle x, y \rangle$,

(iv) $\langle x, x \rangle > 0$, kun $x \neq 0$.

Huomautus. Kohdan (ii) (tai (iii)) nojalla $\langle x, x \rangle = 0$, kun $x = 0$.

Pari $\big(V, \langle \cdot, \cdot \rangle\big)$ on *sisätuloavaruus*.

Esimerkki 8.1. Olkoon $x^n = (x_1, \ldots, x_n)$, $y^n = (y_1, \ldots, y_n) \in \mathbb{R}^n$ ja

$$\langle x^n, y^n \rangle = \sum_{i=1}^{n} x_i y_i.$$

Silloin $\big(\mathbb{R}^n, \langle \cdot, \cdot \rangle\big)$ on sisätuloavaruus. ‖

Esimerkki 8.2. Olkoon $a, b \in \mathbb{R}$, $a < b$ ja

$$C\big([a,b]\big) = \{x \mid x : [a,b] \to \mathbb{R} \text{ jatkuva funktio}\}.$$

Asetetaan

$$\langle x, y \rangle = \int_a^b x(t)y(t)\,dt,$$

kun $x, y \in C([a,b])$. Silloin $\big(C([a,b]), \langle \cdot, \cdot \rangle\big)$ on sisätuloavaruus. ‖

Sisätulo $\langle \cdot, \cdot \rangle$ määrää *normin* $\|\cdot\|$,

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Huomautus. $\big(V, \|\cdot\|\big)$ on normiavaruus. Jos asetetaan $d(x,y) = \|x - y\|$, niin $d : V \times V \to [0, \infty[$ on metriikka ja (V, d) on metrinen avaruus.

Esimerkki 8.3. Avaruudessa $\mathbb{R}^n$

$$\|x^n\| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

‖

Sisätuloavaruuden V jonolla (x_n) on *raja-arvo* $x \in V$, jos $\lim\limits_{n\to\infty} \|x_n - x\| = 0$ eli jos jokaista $\varepsilon > 0$ kohti on olemassa sellainen $n_\varepsilon \in \mathbb{N}_+$, että ehdosta $n > n_\varepsilon$ seuraa, että $\|x_n - x\| < \varepsilon$. Merkitään $\lim\limits_{n\to\infty} x_n = x$.

Jono (x_n) on *Cauchyn jono*, jos jokaista $\varepsilon > 0$ kohti on olemassa sellainen $n_\varepsilon \in \mathbb{N}_+$, että $\|x_n - x_m\| < \varepsilon$ aina, kun $n, m > n_\varepsilon$.

Sisätuloavaruus on *täydellinen*, jos sen jokaisella Cauchyn jonolla on raja-arvo. Täydellinen sisätuloavaruus on *Hilbertin avaruus*.

Esimerkki 8.4. Avaruus $\mathbb{R}^n$ edellä olevalla sisätulolla varustettuna on Hilbertin avaruus. ‖

Esimerkki 8.5. Olkoon

$$l^2 = \left\{ (x_n) \,\middle|\, x_n \in \mathbb{R},\ n \in \mathbb{N}_+ \text{ ja } \sum_{n=1}^{\infty} x_n^2 < \infty \right\}.$$

Asetetaan

$$\langle (x_n), (y_n) \rangle = \sum_{n=1}^{\infty} x_n y_n.$$

Silloin $\big(l^2, \langle \cdot, \cdot \rangle\big)$ on Hilbertin avaruus. Todistus sivuutetaan. ‖

Esimerkki 8.6. $C\big([a,b]\big)$ edellä olevalla sisätulolla *ei* ole Hilbertin avaruus (harjoitustehtävä). ‖

Esimerkki 8.7. Olkoon

$$L^2\big([a,b]\big) = \left\{ x \,\middle|\, x : [a,b] \to \mathbb{R} \text{ Borelin funktio ja } \int_a^b x(t)^2\, dt < \infty \right\}.$$

Asetetaan

$$\langle x, y \rangle = \int_a^b x(t) y(t)\, dt,$$

kun $x, y \in L^2\big([a,b]\big)$. Silloin $\big(L^2\big([a,b]\big), \langle \cdot, \cdot \rangle\big)$ on Hilbertin avaruus. Todistus sivuutetaan. ‖

Huomautus. Jotta ehto (iv) sisätulolle olisi voimassa esimerkissä 8.7, pitää kuitenkin samaistaa funktiot, jotka yhtyvät m.k. $t \in [a,b]$. Tämä tarkoittaa, että avaruuden $L^2\big([a,b]\big)$ alkiot ovat itseasiassa funktioiden x määräämiä (ekvivalenssi)luokkia

$$[x] = \big\{ y \mid y(t) = x(t) \text{ m.k. } t \in [a,b] \big\}.$$

Esimerkki 8.8. Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus ja asetetaan

$$L^2(\Omega) = \{ X \mid X : \Omega \to \mathbb{R} \text{ on satunnaismuuttuja, jolle } \mathbb{E}(X^2) < \infty \}.$$

Määritellään

$$\langle X, Y \rangle = \mathbb{E}(XY),$$

kun $X, Y \in L^2(\Omega)$. Silloin $\big(L^2(\Omega), \langle \cdot, \cdot \rangle\big)$ on Hilbertin avaruus. Todistus sivuutetaan. ‖

Huomautus. Abstraktissa todennäköisyyslaskennassa odotusarvo on integraali,

$$\mathbb{E}(X) = \int_\Omega X(\omega)\, d\mathbb{P},$$

missä $\mathbb{P}$ on todennäköisyysmitta. Tällöin

$$\langle X, Y \rangle = \int_\Omega X(\omega) Y(\omega)\, d\mathbb{P},$$

joten $L^2(\Omega)$ on analoginen $L^2\big([a,b]\big)$:n kanssa.

Avaruudessa $\mathbb{R}^n$ on standardi kanta $\{e_1^n, \ldots, e_n^n\}$,

$$e_i^n = (\underbrace{0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0}_{n \text{ kpl}}).$$

Se on *ortonormaali* (o.n.), eli $\big\langle e_i^n, e_j^n \big\rangle = \delta_{ij}$, $i, j = 1, \ldots, n$. Jos $x^n \in \mathbb{R}^n$, on

$$x^n = \sum_{i=1}^n \langle x^n, e_i^n \rangle\, e_i^n$$

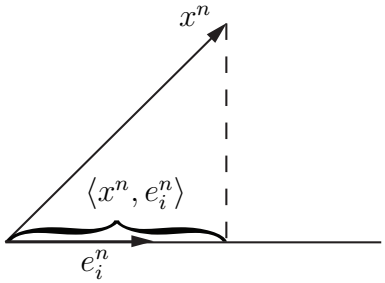
(kuva 8.1).

Vastaavasti Hilbertin avaruuden V jono (e_n) on *ortonormaali kanta*, jos $\langle e_n, e_m \rangle = \delta_{nm}$ kaikilla $n, m \in \mathbb{N}_+$ ja

$$x = \sum_{n=1}^\infty \langle x, e_n \rangle\, e_n,$$

kun $x \in V$. Sarjan suppeneminen takoittaa, että

$$\lim_{m \to \infty} \sum_{n=1}^m \langle x, e_n \rangle\, e_n = x,$$



Kuva 8.1

eli

$$\lim_{m\to\infty} \left\| \sum_{n=1}^{m} \langle x, e_n \rangle \, e_n - x \right\| = 0.$$

Voidaan osoittaa, että Hilbertin avaruudella on ortonormaali kanta jos ja vain jos se on *separoituva* eli on olemassa numeroituva ja tiheä avaruuden V osajoukko.

Esimerkki 8.9. Avaruudella l^2 on ortonormaali kanta (e_n), $e_n = (\delta_{in})_{i\in\mathbb{N}_+}$, eli

$$e_n = (0, \ldots, 0, \overset{n}{1}, 0, 0, \ldots).$$

‖

Esimerkki 8.10. Avaruudella $L^2\big([0,1]\big)$ on esimerkiksi ortonormaali kanta (e_n),

$$e_n(t) = \begin{cases} 1, & n = 1 \\ \sqrt{2}\cos(2\pi k t), & n = 2k \\ \sqrt{2}\sin(2\pi k t), & n = 2k+1, \end{cases}$$

$k = 1, 2, \ldots$ (vrt. harjoitustehtävä). ‖

Olkoon V Hilbertin avaruus. Kuvaus $A : V \to V$ on *operaattori*, jos se on lineaarinen ja jatkuva. Lineaarisuus tarkoittaa, että

$$A(ax + by) = aA(x) + bA(y)$$

kaikilla $x, y \in V$, $a, b \in \mathbb{R}$. Jatkuvuus puolestaan tarkoittaa, että kaikilla $x \in V$ ja $\varepsilon > 0$ on olemassa sellainen $\delta > 0$, että

$$\|A(x) - A(y)\| < \varepsilon \text{ aina, kun } \|y - x\| < \delta.$$

Lineaariselle kuvaukselle A merkitään tavallisesti $A(x) \equiv Ax$, kun $x \in V$.

Olkoon $A : V \to V$ lineaarinen. On helppo nähdä, että seuraavat ehdot ovat yhtäpitävät:

(i) A on jatkuva

(ii) A on jatkuva jossain pisteessä $x \in V$

(iii) A on jatkuva origossa $0 \in V$

(iv) on olemassa sellainen $M \geq 0$, että $\|Ax\| \leq M \|x\|$ kaikilla $x \in V$.

Operaattori A on *symmetrinen*, jos

$$\langle Ax, y \rangle = \langle x, Ay \rangle$$

kaikilla $x, y \in V$.

Luku $a \in \mathbb{R}$ on operaattorin A *ominaisarvo*, jos on olemassa sellainen $x \neq 0$, että

$$Ax = ax.$$

Vektori x on tällöin (eräs) ominaisarvoon a liittyvä *ominaisvektori*.

Esimerkki 8.11. Symmetrinen matriisi $K \in \mathbb{R}^{n \times n}$ määrittelee symmetrisen operaattorin $K : \mathbb{R}^n \to \mathbb{R}^n$, koska

$$\langle Kx^n, y^n \rangle = \sum_{i,j=1}^{n} K_{ij}\, x_j\, y_i \overset{K^T = K}{=} \sum_{i,j=1}^{n} K_{ji}\, y_i\, x_j = \langle x^n, Ky^n \rangle ,$$

kun $x^n, y^n \in \mathbb{R}^n$. ‖

Lineaarialgebrasta tiedetään, että symmetrisellä matriisilla on ortonormaalit ominaisvektorit $u_1^n, \ldots, u_n^n$. Siis, jos A on symmetrinen,

$$Au_i^n = a_i u_i^n,\ \langle u_i^n, u_j^n \rangle = \delta_{ij},\ x^n = \sum_{i=1}^{n} \langle x^n, u_i^n \rangle\, u_i^n,$$

kun $i, j = 1, \ldots, n$, $x^n \in \mathbb{R}^n$.

Hilbertin avaruudessa analoginen ominaisuus on kompaktilla symmetrisellä operaattorilla. Operaattori $A : V \to V$ on *kompakti*, jos jokaiselle rajoitetulle jonolle (x_n) (siis $\|x_n\| \leq M$, $n \in \mathbb{N}_+$, jollain $M \geq 0$) jonolla (Ax_n) on suppeneva osajono.

Olkoon $A : V \to V$ kompakti ja symmetrinen operaattori. Voidaan osoittaa, että operaattorilla A on sellaiset ortonormaalit ominaisvektorit e_n, että (e_n) on avaruuden V kanta. Siis,

$$Ae_n = a_n e_n,\ \langle e_n, e_m \rangle = \delta_{mn},\ x = \sum_{n=1}^{\infty} \langle x, e_n \rangle\, e_n,$$

kun $n, m \in \mathbb{N}_+$, $x \in V$. Summa on tietysti äärellinen, jos V on äärellisdimensioinen.

Esimerkki 8.12. Tarkastellaan Hilbertin avaruutta $L^2\big([a,b]\big)$ ja olkoon $R : [a,b] \times [a,b] \to \mathbb{R}$ jatkuva. Merkitään

$$M = \max\Big\{ |R(t,\tau)| \,\Big|\, t, \tau \in [a,b] \Big\}.$$

Olkoon $x \in L^2\big([a,b]\big)$ ja määritellään

$$y(t) = \int_a^b R(t,\tau)x(\tau)\, d\tau, \qquad t \in [a,b]. \tag{8.1}$$

Silloin

$$|y(t)|^2 \leq \left[\int_a^b |R(t,\tau)|\,|x(\tau)|\,d\tau \right]^2 = \left[\, \langle |R(t,\cdot)|, |x| \rangle \,\right]^2$$

$$\leq \|R(t,\cdot)\|^2 \|x\|^2 = \int_a^b R(t,\tau)^2\,d\tau \cdot \|x\|^2$$

$$\leq (b-a)\,M^2 \|x\|^2 < \infty$$

missä toinen epäyhtälö seuraa Schwarzin epäyhtälöstä (harjoitustehtävä). Erityisesti siis integraali (8.1) suppenee. Edelleen,

$$\int_a^b |y(t)|^2\,dt \leq (b-a)^2 M^2 \|x\|^2 < \infty, \tag{8.2}$$

joten $y \in L^2\big([a,b]\big)$. Voidaan osoittaa, että y on itseasiassa myös *jatkuva*.

Edelleen, määritellään kuvaus $A : L^2\big([a,b]\big) \to L^2\big([a,b]\big)$ kaavalla $Ax = y$. Selvästi A on lineaarinen ja tuloksen (8.2) nojalla se on myös jatkuva, sillä

$$\|Ax\|^2 = \|y\|^2 \leq (b-a)^2 M^2 \|x\|^2,$$

joten

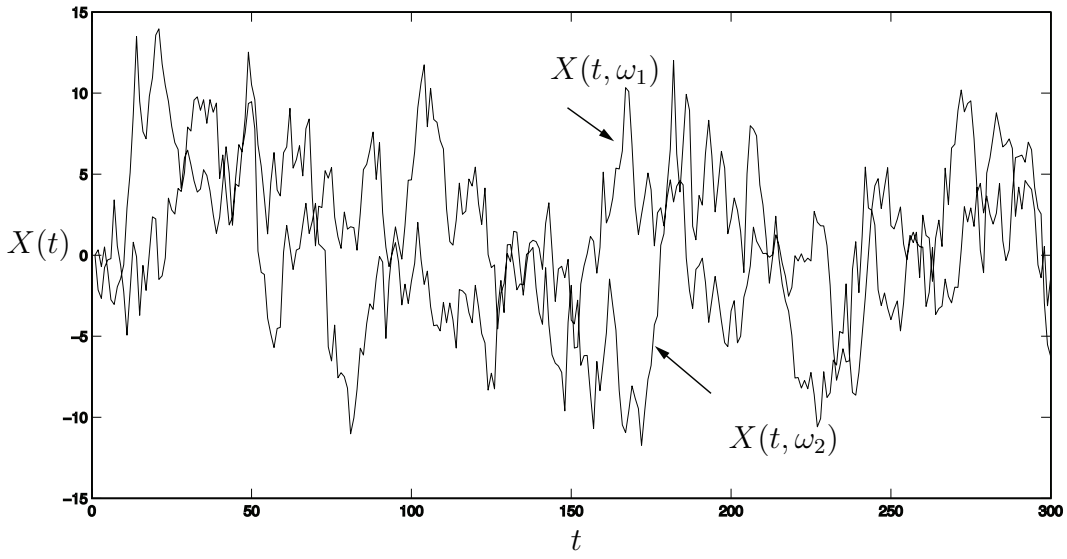
$$\|Ax\| \leq (b-a)\,M \|x\|$$

(vertaa kohta (*iv*) edellä). A on siis operaattori, ns. funktioon R liittyvä *integraalioperaattori*. Voidaan osoittaa, että A on myös kompakti.

Jos R on symmetrinen, $R(t,\tau) = R(\tau,t)$ kaikilla $t, \tau \in [a,b]$, silloin A on myös symmetrinen ja siis sen ominaisvektoreista voidaan muodostaa avaruuden $L^2\big([a,b]\big)$ ortonormaali kanta. ‖

8.2 Karhusen-Loèven kehitelmä

Otsikon Karhunen viittaa suomalaiseen matemaatikkoon, Kari Karhuseen.



Kuva 8.2: Stokastisen prosessin kaksi realisaatiota.

Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus. Aikaisemmin stokastinen prosessi määriteltiin Ω:lla määriteltyjen satunnaismuuttujien X_n jonona (X_n). Yleisemmin, *stokastinen prosessi* on satunnaismuuttujaperhe $(X_t)_{t\in I} = (X_t)$, missä I on indeksijoukko ja $X_t : \Omega \to \mathbb{R}$ on satunnaismuuttuja, $t \in I$. Seuraavassa tavallisesti $I = [a, b] \subset \mathbb{R}$ ja indeksiä $t \in [a, b]$ ajatellaan aikana.

Jatkossa merkitään $X_t = X(t)$. Satunnaismuuttujan X_t arvoa pisteessä $\omega \in \Omega$ voitaisiin tällöin merkitä $X_t(\omega) = X(t, \omega)$. Olkoon $\omega \in \Omega$. Funktio $t \mapsto X(t, \omega)$ on prosessin $\big(X(t)\big)$ (eräs) *realisaatio* (vastaa satunnaismuuttujan realisaatiota, arvoa). Kuvassa 8.2 on esimerkki erään välillä $[0, 300]$ määritellyn prosessin kahdesta realisaatiosta.

Stokastisen prosessin $\big(X(t)\big)$ jakaumaa voidaan kuvata sen äärellisulotteisilla jakaumilla eli satunnaisvektoreiden $\big(X(t_1), \ldots, X(t_k)\big)$ jakaumilla, $t_1, \ldots, t_k \in [a, b]$, $k \in \mathbb{N}_+$.

Esimerkki 8.13. *Gaussin prosessissa* satunnaisvektoreilla $\big(X(t_1), \ldots, X(t_k)\big)$ on multinormaalijakaumat. (Tarkkaan ottaen sallitaan myös singulaariset multinormaalijakaumat, joilla ei ole tiheysfunktiota; ne ovat keskittyneet vek-

torialiavaruuksille, jolloin kovarianssimatriisi on singulaarinen.)

Olkoon erityisesti $\mathbb{E}[X(t)] = 0$ kaikilla $t \in [a, b]$. Silloin Gaussin prosessin äärellisulotteiset jakaumat määräytyvät kovariansseista

$$\mathbb{E}\big[X(t_i)\, X(t_j)\big]$$

(vertaa luku 6.3). ‖

Yleisesti, olkoon $\big(X(t)\big)$ prosessi, jolla $\mathbb{E}[X(t)] = 0$ kaikilla $t \in [a, b]$ ja oletetaan, että odotusarvot $\mathbb{E}[X(t)X(\tau)]$ ovat olemassa, kun $t, \tau \in [a, b]$. Silloin funktio $R : [a, b] \times [a, b] \to \mathbb{R}$,

$$R(t, \tau) = \mathbb{E}\big[X(t)\, X(\tau)\big]$$

on prosessin $\big(X(t)\big)$ *kovarianssifunktio*. Selvästi R on symmetrinen, $R(t, \tau) = R(\tau, t)$.

Prosessi $\big(X(t)\big)$ on *stationaarinen*, jos kaikilla $t_1, \ldots, t_k \in \mathbb{R}$, $k \in \mathbb{N}_+$ ja $r \in \mathbb{R}$ on satunnaisvektoreilla $\big(X(t_1), \ldots, X(t_k)\big)$ ja $\big(X(t_1+r), \ldots, X(t_k+r)\big)$ sama jakauma (vertaa määritelmä 3.6). Silloin, jos kovarianssifunktio on olemassa, on siis

$$\begin{aligned} R(t, \tau) &= \mathbb{E}\big[X(t)\, X(\tau)\big] \\ &= \mathbb{E}\big[X\big((t-\tau)+\tau\big)\, X(0+\tau)\big] \\ &= \mathbb{E}\big[X(t-\tau)\, X(0)\big] \\ &\equiv R(t-\tau). \end{aligned}$$

Funktiota $R(\tau) = \mathbb{E}\big[X(\tau)\, X(0)\big]$ kutsutaan myös prosessin kovarianssifunktioksi.

Erityisesti stationaarisen Gaussin prosessin äärellisulotteiset jakaumat määrää täysin kovarianssifunktio $R(\tau)$.

Olkoon nyt X^n satunnaisvektori, jolla $\mathbb{E}(X^n) = 0$ ja olkoon $R = (R_{ij}) \in \mathbb{R}^{n\times n}$ X^n:n kovarianssimatriisi,

$$R_{ij} = \mathbb{E}(X_i\, X_j), \qquad i, j = 1, \ldots, n.$$

R on symmetrinen, joten sillä on ortonormaalit ominaisvektorit $e_1^n, \ldots, e_n^n$ (ei välttämättä tietenkään $\mathbb{R}^n$:n standardikanta),

$$R\, e_i^n = a_i\, e_i^n, \tag{8.3}$$

$i = 1, \ldots, n$. Siten, kun $\omega \in \Omega$, $X^n(\omega) \in \mathbb{R}^n$ ja

$$X^n(\omega) = \sum_{i=1}^{n} \langle X^n(\omega), e_i^n \rangle\, e_i^n.$$

Voidaan siis kirjoittaa

$$X^n = \sum_{i=1}^{n} \langle X^n, e_i^n \rangle\, e_i^n. \tag{8.4}$$

Tässä

$$Z_i = \langle X^n, e_i^n \rangle \tag{8.5}$$

on satunnaismuuttuja, jolle

$$\mathbb{E}(Z_i) = \mathbb{E}\big(\, \langle X^n, e_i^n \rangle\, \big) = \mathbb{E}\Big(\sum_{k=1}^{n} X_k\, e_{ik} \Big) = \sum_{k=1}^{n} \mathbb{E}(X_k)\, e_{ik} = 0,$$

koska $\mathbb{E}(X_k) = 0$ kaikilla k. Siten siis

$$\mathbb{E}(Z_i) = 0, \qquad i = 1, \ldots, n. \tag{8.6}$$

Lisäksi

$$\begin{aligned}
\mathbb{E}(Z_i\, Z_j) &= \mathbb{E}\big(\, \langle X^n, e_i^n \rangle \, \langle X^n, e_j^n \rangle\, \big) = \mathbb{E}\Big(\sum_{k=1}^{n} X_k\, e_{ik} \sum_{l=1}^{n} X_l\, e_{jl} \Big) \\
&= \sum_{k=1}^{n} \sum_{l=1}^{n} \mathbb{E}(X_k\, X_l)\, e_{ik}\, e_{jl} = \sum_{k=1}^{n} \sum_{l=1}^{n} R_{kl}\, e_{ik}\, e_{jl} \\
&= \big\langle e_i^n, R\, e_j^n \big\rangle = \big\langle e_i^n, a_j\, e_j^n \big\rangle = a_j \, \big\langle e_i^n, e_j^n \big\rangle \\
&= a_j\, \delta_{ij},
\end{aligned}$$

kun $i, j = 1, \ldots, n$. Siten

$$\mathbb{E}(Z_i\, Z_j) = 0, \text{ kun } i \neq j \tag{8.7}$$

ja

$$\mathbb{E}(Z_i^2) = a_i, \tag{8.8}$$

$i, j = 1, \ldots, n$. Siten on voimassa *Karhusen-Loèven kehitelmä*

$$X^n = \sum_{i=1}^{n} Z_i\, e_i^n, \tag{8.9}$$

$$\mathbb{E}(Z_i) = 0, \quad \mathbb{E}(Z_i\, Z_j) = 0, \quad \mathbb{E}(Z_i^2) = a_i,$$

kun $i, j = 1, \ldots, n$, $i \neq j$. Edelleen, kun $a_i = 0$, on $\mathbb{E}(Z_i^2) = 0$ eli $Z_i = 0$ todennäköisyydellä 1. Siten kaavassa (8.9) voidaan kehitelmään ottaa mukaan vain termit, joissa $a_i > 0$ ja se edelleen pätee todennäköisyydellä 1.

Analoginen tulos pätee yleisemmin myös tietyille stokastisille prosesseille; stokastista prosessia voidaan ajatella eräänlaisena ääretönulotteisena satunnaisvektorina ja siis äärellisulotteisen vektorin X^n yleistyksenä.

Olkoon $\big(X(t)\big)_{t\in[a,b]} = \big(X(t)\big)$ stokastinen prosessi, jolle $E\big(X(t)\big) = 0$ kaikilla t ja jolla kovarianssifunktio $R(t,\tau) = \mathbb{E}[X(t)\, X(\tau)]$ on jatkuva. Olkoot (e_k) kompaktin operaattorin (8.1) ortonormaalien ominaisvektoreiden jono,

$$\int_a^b R(t,\tau)\, e_k(\tau)\, d\tau = a_k\, e_k(t),$$

$k \in \mathbb{N}_+$ (vrt. (8.3)). Ominaisfunktiot e_k ovat itseasiassa jatkuvia (vertaa esimerkki luvussa 8.1). Tästä johtuen ei seuraavassa ole mitään ”m.k. epämääräisyyttä” tai määrittelemättömyyttä. Avaruuden $L^2([a,b])$ ekvivalenssiluokan jatkuva edustaja on yksikäsitteinen ja voidaan puhua tavallisten funktioiden arvoista pisteissä t. Nyt voidaan osoittaa, että

$$X(t) = \sum_{k=1}^{\infty} Z_k\, e_k(t), \tag{8.10}$$

missä satunnaismuuttujat Z_k toteuttavat ehdot

$$\mathbb{E}(Z_k) = 0, \quad \mathbb{E}(Z_k\, Z_l) = 0, \quad \mathbb{E}(Z_k^2) = a_k, \tag{8.11}$$

kun $k, l \in \mathbb{N}_+$, $k \neq l$. Suppeneminen tapahtuu tasaisesti avaruudessa $L^2(\Omega)$, eli

$$\sup_{t\in[a,b]} \mathbb{E}\Big\{ \Big[X(t) - \sum_{k=1}^{n} Z_k\, e_k(t)\Big]^2 \Big\} \longrightarrow 0,$$

kun $n \to \infty$. Kehitelmää (8.10) sanotaan prosessin $\big(X(t)\big)$ *Karhusen-Loèven kehitelmäksi.*

Satunnaismuuttujat Z_k määritellään analogisesti kaavan (8.5) kanssa,

$$Z_k = \int_a^b X(t)\, e_k(t)\, dt. \tag{8.12}$$

Integraali (8.12) voidaan määritellä "Riemannin summien" avulla seuraavasti. Tarkastellaan välin $[a, b]$ jakoja $D : a = t_0 < t_1 < \cdots < t_n = b$ ja olkoon

$$|D| = \max\big\{|t_i - t_{i-1}| \mid i = 1, \ldots, n\big\}.$$

Olkoon

$$I(D) = \sum_{i=1}^{n} X(t_i)\, e_k(t_i)(t_i - t_{i-1}).$$

Silloin (8.12) tarkoittaa, että Z_k on satunnaismuuttuja, jolle jokaista $\varepsilon > 0$ kohti on olemassa sellainen $\delta > 0$, että $\mathbb{E}\big[(I(D) - Z_k)^2\big] < \varepsilon$ aina, kun $|D| < \delta$.

Edelleen, ehdon (8.11) nojalla $a_k \geq 0$ ja summa (8.10) voidaan ottaa vain niiden k yli, joilla $a_k > 0$, koska ehdosta $a_k = 0$ seuraa, että $Z_k = 0$ todennäköisyydellä 1.

Jos $\big(X(t)\big)$ on Gaussin prosessi, ovat satunnaismuuttujat Z_k itseasiassa *normaalisti jakautuneita* ja *riippumattomia.* Lisäksi kiinteällä $t \in [a, b]$ sarja (8.10) suppenee todennäköisyydellä 1.

Hyvä esitys kaikkine yksityiskohtineen lukujen 8.1 ja 8.2 asioista löytyy Ashin kirjasta [2].

8.3 Shannonin toinen lause jatkuva-aikaiselle Gaussin kanavalle

Syötteenä kanavaan on nyt funktio $s(t)$ ja kanavan aiheuttamaa häiriötä mallinnetaan häiriö/kohinaprosessilla $\big(n(t)\big)$. Oletamme, että $\big(n(t)\big)$ on stationaarinen Gaussin prosessi.

$$\begin{array}{cccc} s(t) & \longrightarrow & \boxed{\text{Informaatiokanava}} & \longrightarrow & s(t)+n(t) \\ \text{syötesignaali} & & & & \text{tulostesignaali} \end{array} \tag{8.13}$$

Eri syöttökerroilla syötesignaaliin lisätään eri realisaatio $n(t,\omega)$ häiriöprosessista,

$$s(t) \longrightarrow s(t)+n(t,\omega),$$

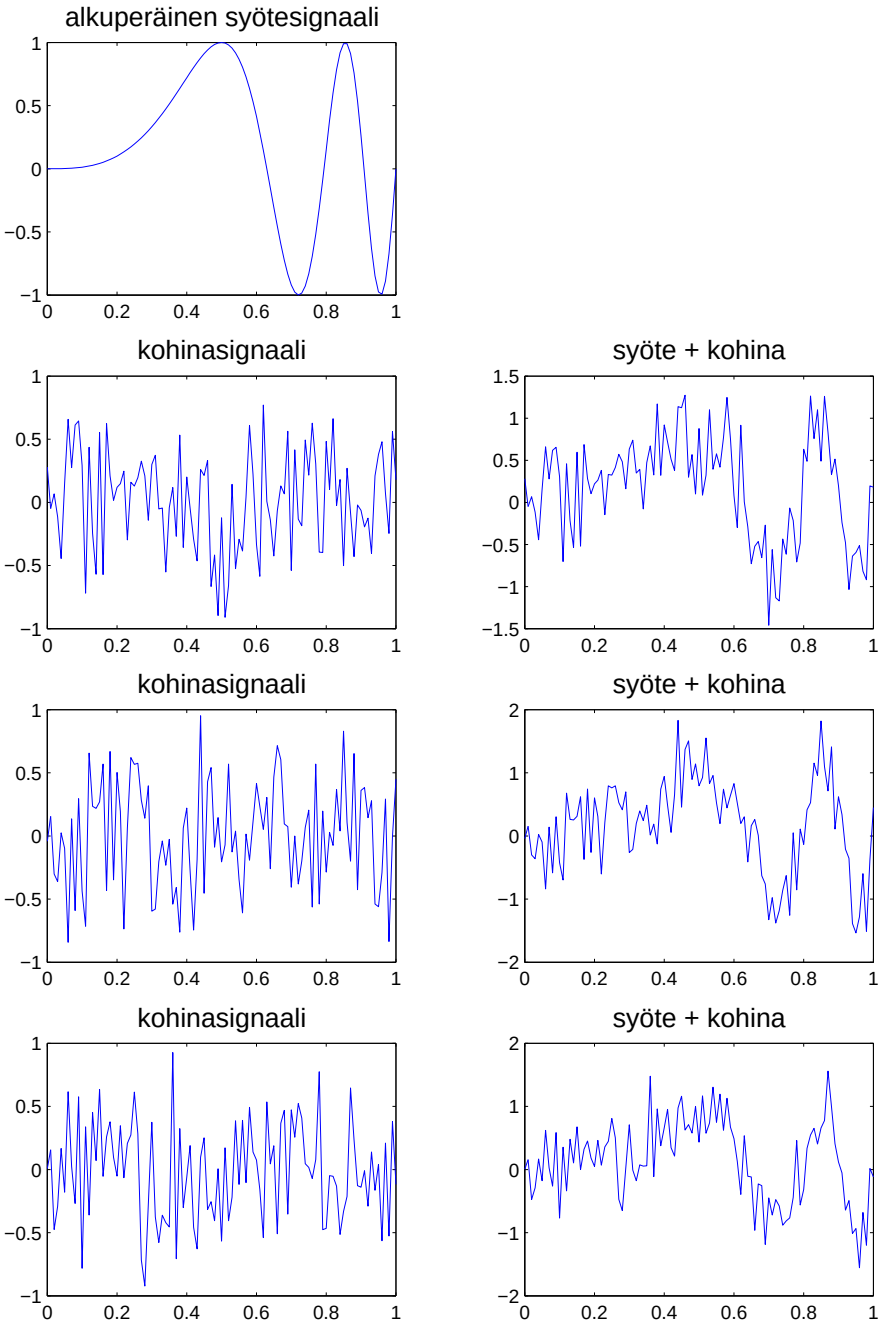
$\omega \in \Omega$. Siten tuloste on satunnainen.

Esimerkki 8.14. Kuvassa 8.3 on esimerkki syötesignaalista ja siihen lisättävästä kohinaprosessin kolmesta eri realisaatiosta. Lopputulos on aina hieman erilainen riippuen lisätyn häiriön tarkasta muodosta. ‖

Oletetaan, että $\mathbb{E}[n(t)] = 0$ kaikilla t ja olkoon prosessin $\big(n(t)\big)$ kovarianssifunktio $R(\tau)$ jatkuva. Olkoon $T > 0$ ja tarkastellaan Hilbertin avaruutta $L^2\big([-T/2, T/2]\big)$.

Luvun 8.2 nojalla kaava

$$(A_T\, x)(t) = \int_{-T/2}^{T/2} R(t-\tau)\, x(\tau)\, d\tau, \qquad t \in [-T/2, T/2]$$



Kuva 8.3: Syötesignaali ja siihen lisättäviä kohinaprosessin realisaatioita.

määrittelee kompaktin symmetrisen operaattorin $A_T : L^2\big([-T/2,T/2]\big) \to L^2\big([-T/2,T/2]\big)$. Olkoot $\big(e_k(\,\cdot\,;T)\big)_{k\in\mathbb{N}_+}$ operaattorin A_T ortonormaalien ominaisvektorien muodostama avaruuden $L^2\big([-T/2,T/2]\big)$ kanta ja $a_k(T)$, $k \in \mathbb{N}_+$, vastaavat ominaisarvot. Oletetaan, että $a_k(T) > 0$ kaikilla k.

Tarkastellaan syötettä $s_T \in L^2\big([-T/2,T/2]\big)$ ja olkoon $(n_T(t))_{t\in[-T/2,T/2]}$ kohinaprosessin $\big(n(t)\big)_{t\in\mathbb{R}}$ rajoittuma väliin $[-T/2,T/2]$. Lukujen 8.1 ja 8.2 nojalla

$$s_T(t) = \sum_{k=1}^{\infty} s_k(T)\, e_k(t;T), \tag{8.14}$$

missä $s_k(T) = \langle s_T, e_k(\,\cdot\,;T)\rangle$ ja

$$n_T(t) = \sum_{k=1}^{\infty} Z_k\, e_k(t;T), \tag{8.15}$$

missä satunnaismuuttujat $Z_k : \Omega \to \mathbb{R}$ ovat riippumattomia ja

$$Z_k \sim \mathrm{N}\big(0, a_k(T)\big),\ k \in \mathbb{N}_+.$$

Siten saadaan vastaavuudet

$$s_T(t) \longleftrightarrow \big(s_k(T)\big)_{k\in\mathbb{N}_+}$$
$$n_T(t) \longleftrightarrow (Z_k)_{k\in\mathbb{N}_+}$$

ja siis oleellisesti

$$s_T(t) + n_T(t) \longleftrightarrow \big(s_k(T) + Z_k\big)_{k\in\mathbb{N}_+}.$$

Näin kanavan (8.13) voi ajatella vastaavan kanavaa

$$\big(s_k(T)\big)_{k\in\mathbb{N}_+} \longrightarrow \boxed{\text{Informaatiokanava}} \longrightarrow \big(s_k(T) + Z_k\big)_{k\in\mathbb{N}_+}. \tag{8.16}$$

Edelleen, skaalaamalla luvuilla $a_k(T)^{-1/2} > 0$ voidaan yhtä hyvin tarkastella kanavaa

$$\big(a_k(T)^{-1/2} s_k(T)\big)_{k\in\mathbb{N}_+} \longrightarrow \big(a_k(T)^{-1/2} s_k(T) + a_k(T)^{-1/2} Z_k\big)_{k\in\mathbb{N}_+}.$$

Tässä $Z_k \sim \mathrm{N}\big(0, a_k(T)\big)$, joten satunnaismuuttujille $Z_k^* = a_k(T)^{-1/2} Z_k$ pätee

$$D^2(Z_k^*) = D^2\big(a_k(T)^{-1/2} Z_k\big) = [a_k(T)^{-1/2}]^2 \, D^2(Z_k) = a_k(T)^{-1} \cdot a_k(T) = 1.$$

Siten voidaan jatkossa tarkastella kanavaa

$$x \longrightarrow \boxed{\text{Informaatiokanava}} \longrightarrow Y,$$

missä $x = (x_1, x_2, \ldots)$ on reaalilukujono,

$$Y = x + Z,$$

missä $Z = (Z_1^*, Z_2^*, \ldots)$ on kohinaprosessi, jolle

$$Z_1^*, Z_2^*, \ldots \overset{iid}{\sim} \mathrm{N}(0, 1).$$

Tämä kanava on *jatkuva-aikainen Gaussin kanava*.

Merkitään

$$\mathbb{R}^\infty = \big\{ x \mid x = (x_1, x_2, \ldots),\, x_k \in \mathbb{R},\, k \in \mathbb{N}_+ \big\}.$$

Siis jatkuva-aikaisen Gaussin kanavan syöte $x \in \mathbb{R}^\infty$ ja tuloste on stokastinen prosessi $Y = (x_k + Z_k^*)_{k \in \mathbb{N}_+}$. Kun tätä vertaa luvun 7.1 diskreettiaikaiseen Gaussin kanavaan, nähdään että jatkuva-aikainen Gaussin kanava vastaa matemaattisesti ääretönulotteista diskreettiaikaista Gaussin kanavaa.

Kuten diskreettiaikaisessa Gaussin kanavassa, asetetaan nytkin syötteelle tehorajoitus. Tähän voidaan ajatella olevan selvät fysikaaliset syyt ja jos rajoitusta ei aseteta, voitaisiin helposti taas määritellä koodeja, joilla tiedonsiirtonopeus saadaan mielivaltaisen suureksi virhetodennäköisyyden pysyessä mielivaltaisen pienenä.

Syötteen tehoa hetkellä t mitataan suureella $x(t)^2$. Keskimääräinen teho välillä $[-T/2, T/2]$ on tällöin noin (kuva 8.4)

$$\frac{1}{m} \sum_{i=1}^{m} x(t_i)^2 = \frac{1}{m\Delta} \sum_{i=1}^{m} x(t_i)^2 \Delta \quad \xrightarrow[m \to \infty]{} \quad \frac{1}{T} \int_{-T/2}^{T/2} x(t)^2 \, dt.$$

Kuva 8.4: Välin $[-T/2, T/2]$ diskretointi syötteen keskimääräisen tehon laskemiseksi.

Siten syötesignaalin $x(t)$ keskimääräinen teho määritellään kaavalla

$$\frac{1}{T} \int_{-T/2}^{T/2} x(t)^2 \, dt.$$

Olkoon nyt

$$x(t) = \sum_{k=1}^{\infty} x_k \, e_k(t; T).$$

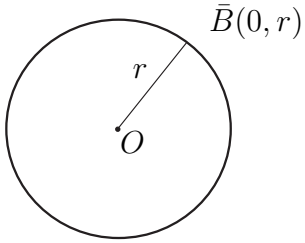
Silloin kannan $\big(e_k(\,\cdot\,; T)\big)$ ortonormaalisuuden perusteella

$$\begin{aligned}
\int_{-T/2}^{T/2} x(t)^2 \, dt &= \int_{-T/2}^{T/2} \left[\sum_{k=1}^{\infty} x_k \, e_k(t; T) \sum_{l=1}^{\infty} x_l \, e_l(t; T) \right] dt \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} x_k \, x_l \int_{-T/2}^{T/2} e_k(t; T) \, e_l(t; T) \, dt \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} x_k \, x_l \, \langle e_k(\,\cdot\,; T), e_l(\,\cdot\,; T) \rangle \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} x_k \, x_l \, \delta_{kl} = \sum_{k=1}^{\infty} x_k^2.
\end{aligned}$$

Siten keskimääräiselle teholle saadaan kaava

$$\frac{1}{T} \int_{-T/2}^{T/2} x(t)^2 \, dt = \frac{1}{T} \sum_{k=1}^{\infty} x_k^2 = \frac{1}{T} \, \|x\|^2 \,,$$

missä $\|x\|$ on jonon $x = (x_k)$ l^2-normi.



Kuva 8.5: Origokeskinen r-säteinen pallo avaruudessa l^2.

Merkitään

$$\bar{B}(0,r) = \big\{x \in l^2 \mid \|x\| \le r\big\},$$

kun $r > 0$. $\bar{B}(0,r)$ on siis origokeskinen r-säteinen (suljettu) pallo avaruudessa l^2 (kuva 8.5).

Tarkastellaan nyt viestejä $1,\dots,M$ ja olkoon $K > 0$. (M,T)-*koodi* on pari (C,g), missä C on *kooderi*,

$$C : \{1,\dots,M\} \to \bar{B}(0,\sqrt{KT}) \subset l^2$$

ja (Borelin funktio) $g : \mathbb{R}^\infty \to \{1,\dots,M\}$ on *dekooderi*. K on *tehoraja* ja *koodisanat*

$$C(j) = x(j) = (x_1(j), x_2(j), \dots)$$

toteuttavat ehdon

$$\|x(j)\|^2 \le KT \qquad \text{eli} \qquad \frac{1}{T}\sum_{k=1}^{\infty} x_k(j)^2 \le K. \tag{8.17}$$

Näin koodisanat siis vastaavat funktioita, joiden keskimääräinen teho välillä $[-T/2, T/2]$ on korkeintaan K.

Viestin j virhetodennäköisyys määritellään kaavalla

$$\lambda_j = \mathbb{P}\big\{g\big(x(j)+Z\big) \ne j\big\}.$$

Huomautus. Ei ole vaikeaa osoittaa, että $\big\{g\big(x(j)+Z\big) \ne j\big\} \in \mathcal{F}$ ja että

siis yllä oleva todennäköisyys on määritelty:

$$\begin{aligned}\big\{g\big(x(j)+Z\big)\neq j\big\} &= \bigcup_{i\neq j}\big\{g\big(x(j)+Z\big)=i\big\} = \bigcup_{i\neq j}\big\{x(j)+Z\in g^{-1}(i)\big\}\\ &= \bigcup_{i\neq j}\big\{Z\in -x(j)+g^{-1}(i)\big\}\end{aligned}$$

ja $\big\{Z\in -x(j)+g^{-1}(i)\big\}\in\mathcal{F}$, koska $-x(j)+g^{-1}(i)$ on $\mathbb{R}^\infty$:n Borelin joukko ja $Z=(Z_1^*,Z_2^*,\ldots)$, missä $Z_k^*:\Omega\to\mathbb{R}$ on satunnaismuuttuja, $k\in\mathbb{N}_+$.

Koodin *maksimaalinen virhe* on taas

$$\lambda^{(T)}=\max\{\lambda_1,\ldots,\lambda_M\}.$$

Edelleen, tiedonsiirtonopeus $R\geq 0$ on *saavutettavissa*, jos on olemassa sellaiset $\big(\lceil 2^{RT_n}\rceil,T_n\big)$-koodit, että $T_n\to\infty$, $\lambda^{(T_n)}\to 0$, kun $n\to\infty$. Tällöin siis, jos $M_n=\lceil 2^{RT_n}\rceil$, on

$$R\leq\frac{\log M_n}{T_n}\leq R+\frac{1}{T_n},$$

eli välillä $[-T_n/2,T_n/2]$ siirtyy kanavan läpi noin R bittiä per aikayksikkö lähes virheettömästi, kun $T_n\to\infty$ (eli $n\to\infty$).

Kanavan *kapasiteetti* on

$$C=\sup\{R\mid\text{kanavan tiedonsiirtonopeus }R\text{ saavutettavissa}\}.$$

Lause 8.1. *Olkoon jatkuva-aikaisen Gaussin kanavan tehoraja $K>0$. Silloin kanavan kapasiteetti on*

$$C=\frac{1}{2}K\log e.$$

Todistus. Osoitetaan ensin, että $C\geq\frac{1}{2}K\log e$. Olkoon $l\in\mathbb{N}_+$ ja

$$\frac{1}{2}\log\left(1+\frac{K}{l}\right)-\frac{1}{l^2}<R_l<\frac{1}{2}\log\left(1+\frac{K}{l}\right). \tag{8.18}$$

Tarkastellaan diskreettiaikaista Gaussin kanavaa kohinavarianssilla $N=1$ ja tehorajalla $P=K/l$. Lauseen 7.2 nojalla nyt on olemassa sellaiset $\big(\lceil 2^{nR_l}\rceil,n\big)$-koodit (C_n,g_n), että maksimaalinen virhe $\lambda^{(n)}\to 0$, kun $n\to\infty$.

Olkoon $T \in \mathbb{N}_+$ ja tarkastellaan koodien osajonoa $(C_{lT}, g_{lT})_{T\in\mathbb{N}_+}$. Koodin (C_{lT}, g_{lT}) koodisanoille $x^{lT}(j) = \big(x_1(j), \ldots, x_{lT}(j)\big)$ pätee tehorajoituksen nojalla

$$\frac{1}{lT}\left\|x^{lT}(j)\right\|^2 \leq \frac{K}{l}$$

eli

$$\left\|x^{lT}(j)\right\|^2 \leq KT,$$

missä $\left\|x^{lT}(j)\right\|^2 = \sum_{k=1}^{lT} x_k(j)^2$.

Määritellään nyt $x(j) \in \mathbb{R}^\infty$ koodisanan $x^{lT}(j)$ "nollajatkona",

$$x(j) = \big(x_1(j), \ldots, x_{lT}(j), 0, 0, \ldots\big).$$

Silloin

$$\|x(j)\|^2 = \sum_{k=1}^{\infty} x_k(j)^2 = \left\|x^{lT}(j)\right\|^2 \leq KT$$

(vrt. kaava (8.17)), joten voidaan määritellä jatkuva-aikaiselle Gaussin kanavalle kooderi

$$C_T : \big\{1, \ldots, \lceil 2^{lTR_l}\rceil\big\} \to \bar{B}\big(0, \sqrt{KT}\big) \subset l^2, \quad C_T(j) = x(j).$$

Vastaava dekooderi $g_T : \mathbb{R}^\infty \to \big\{1, \ldots, \lceil 2^{lTR_l}\rceil\big\}$ määritellään sitten kaavalla

$$g_T(y) = g_{lT}(y_1, \ldots, y_{lT}),$$

kun $y = (y_1, \ldots, y_{lT}, y_{lT+1}, \ldots) \in \mathbb{R}^\infty$. Silloin selvästi koodin (C_T, g_T) maksimivirhe on sama kuin diskreetin kanavan koodilla (C_{lT}, g_{lT}), jolle maksimivirhe $\lambda^{(lT)} \to 0$, kun $T \to \infty$. Siten myös jatkuva-aikaisessa kanavassa maksimivirhe $\lambda^{(T)} \to 0$.

Nyt $lTR_l = (lR_l)T$, joten nopeus lR_l on saavutettavissa jatkuva-aikaisessa kanavassa ($T_n = T = n$ saavutettavuuden määritelmässä). Siten $C \geq lR_l$. Mutta $l \in \mathbb{N}_+$ oli mielivaltainen ja kaavan (8.18) nojalla

$$\frac{l}{2}\log\left(1 + \frac{K}{l}\right) - \frac{1}{l} < lR_l < \frac{l}{2}\log\left(1 + \frac{K}{l}\right).$$

Lisäksi

$$\lim_{l\to\infty} \frac{l}{2} \log\left(1+\frac{K}{l}\right) = \frac{1}{2} \lim_{l\to\infty} \log\left(1+\frac{K}{l}\right)^l$$
$$= \frac{1}{2} \log \lim_{l\to\infty} \left(1+\frac{K}{l}\right)^l = \frac{1}{2} \log e^K = \frac{1}{2} K \log e.$$

Siten

$$C \geq \lim_{l\to\infty} l R_l = \frac{1}{2} K \log e.$$

Osoitetaan sitten toiseksi, että $C \leq \frac{1}{2} K \log e$. Olkoon siis $\big(\lceil 2^{RT_n} \rceil, T_n\big)$-koodeille (C_{T_n}, g_{T_n}) voimassa $T_n \to \infty$, $\lambda^{(T_n)} \to 0$, kun $n \to \infty$. Osoitetaan, että tällöin $R \leq \frac{1}{2} K \log e$. Olkoon $M_n = \lceil 2^{RT_n} \rceil$. Tarkastellaan kiinteätä $n \in \mathbb{N}_+$ ja merkitään

$$B_j = g_{T_n}^{-1}(j) = \big\{ y \in \mathbb{R}^\infty \mid g_{T_n}(y) = j \big\},$$

$j = 1, \ldots, M_n$. Siis B_j koostuu niistä $y \in \mathbb{R}^\infty$, jotka dekoodataan j:ksi. Pyrimme seuraavassa approksimoimaan jatkuva-aikaista kanavaa diskreettiaikaisella kanavalla.

Olkoon $x(j) \in \mathbb{R}^\infty$ viestin j koodisana ja

$$Y_j = x(j) + Z$$

tuloste, kun syöte on $x(j)$, $j = 1, \ldots, M_n$.

Mittateorian tuloksista seuraa (esimerkiksi Halmos: Measure Theory, s. 56), että kaikilla $\varepsilon > 0$ on olemassa $n_j \in \mathbb{N}_+$ ja $B'_{jn_j} \subset \mathbb{R}^{n_j}$ siten, että jos $B'_j = B'_{jn_j} \times \mathbb{R}^\infty$, niin

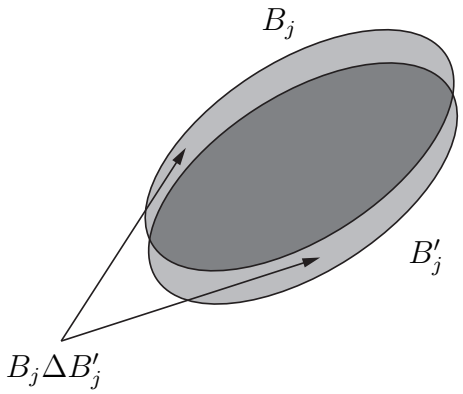
$$\mathbb{P}\{Y_j \in B_j \,\Delta\, B'_j\} \leq \varepsilon,$$

$j = 1, \ldots, M_n$. Tässä $B_j \,\Delta\, B'_j$ on joukkojen B_j ja B'_j symmetrinen erotus,

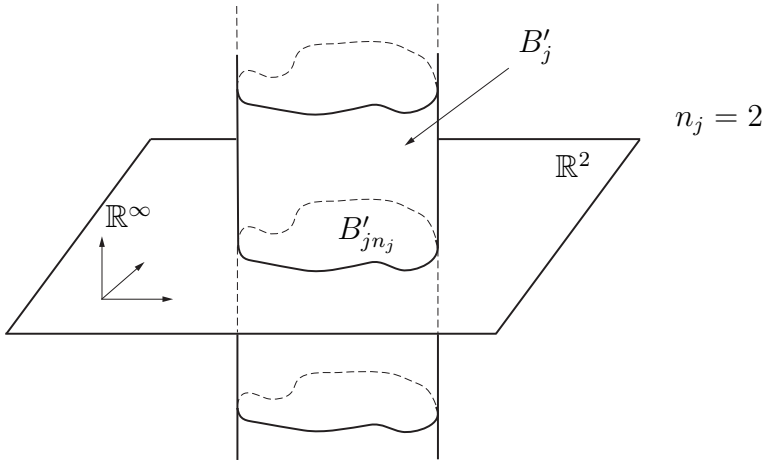
$$B_j \,\Delta\, B'_j = (B_j \setminus B'_j) \cup (B'_j \setminus B_j)$$

(kuva 8.6). Joukko B'_j on *n_j-ulotteinen sylinteri*, jonka *pohja* on B'_{jn_j},

$$B'_j = B'_{jn_j} \times \mathbb{R}^\infty = \big\{ x \in \mathbb{R}^\infty \mid (x_1, \ldots, x_{n_j}) \in B'_{jn_j} \big\} \tag{8.19}$$



Kuva 8.6: Joukkojen B_j ja B'_j symmetrinen erotus $B_j \,\Delta\, B'_j$.



Kuva 8.7: Kaavan (8.19) n_j-ulotteinen sylinteri, jonka pohja on B_{jn_j}.

(kuva 8.7).

Otetaan nyt $\varepsilon = \frac{\lambda^{(T_n)}}{2M_n}$, jolloin siis

$$\mathbb{P}\big\{Y_j \in B_j \,\Delta\, B'_j\big\} \leq \frac{\lambda^{(T_n)}}{2M_n}, \tag{8.20}$$

$j = 1, \ldots, M_n$. Korvataan sitten joukot B'_j pistevierailla joukoilla

$$A_j = B'_j \setminus \bigcup_{k \neq j} B'_k.$$

Joukot A_j ovat myös äärellisulotteisia sylintereitä. Koska halutaan, että $\bigcup_{j=1}^{M_n} A_j =$

$\mathbb{R}^\infty$, vaihdetaan vielä joukoksi A_1

$$\mathbb{R}^\infty \setminus \bigcup_{j=2}^{M_n} A_j,$$

joka on myös äärellisulotteinen sylinteri. Merkitään tätä edelleen A_1:llä.

Nyt

$$\begin{aligned}\mathbb{P}\{Y_j \in A_j\} &\geq \mathbb{P}\{Y_j \in B_j'\} - \mathbb{P}\Big\{Y_j \in B_j' \cap \big(\bigcup_{k\neq j} B_k'\big)\Big\} \\ &\geq \mathbb{P}\{Y_j \in B_j'\} - \sum_{k\neq j} \mathbb{P}\{Y_j \in B_j' \cap B_k'\}.\end{aligned}$$

Koska $B_j' \supset B_j \setminus (B_j \,\Delta\, B_j')$, saadaan edelleen

$$\begin{aligned}\mathbb{P}\{Y_j \in A_j\} &\geq \mathbb{P}\{Y_j \in B_j\} - \mathbb{P}\{Y_j \in B_j \,\Delta\, B_j'\} - \sum_{k\neq j} \mathbb{P}\{Y_j \in B_j' \cap B_k'\} \\ &\overset{(8.20)}{\geq} \mathbb{P}\{Y_j \in B_j\} - \frac{\lambda^{(T_n)}}{2M_n} - \sum_{k\neq j} \mathbb{P}\{Y_j \in B_j' \cap B_k'\}.\end{aligned}$$

On helppo nähdä, että

$$B_j' \cap B_k' \subset (B_j \,\Delta\, B_j') \cup (B_k \,\Delta\, B_k'),$$

kun $j \neq k$. Siten kaavan (8.20) nojalla

$$\mathbb{P}\{Y_j \in A_j\} \geq \mathbb{P}\{Y_j \in B_j\} - \frac{\lambda^{(T_n)}}{2M_n} - (M_n - 1)\frac{\lambda^{(T_n)}}{M_n}.$$

Tässä

$$\begin{aligned}-\frac{\lambda^{(T_n)}}{2M_n} - (M_n-1)\frac{\lambda^{(T_n)}}{M_n} &= -\lambda^{(T_n)}\left[\frac{1}{2M_n} + \frac{M_n-1}{M_n}\right] = -\lambda^{(T_n)}\left[\frac{2M_n-1}{2M_n}\right] \\ &= -\lambda^{(T_n)}\left[1 - \frac{1}{2M_n}\right] \geq -\lambda^{(T_n)}.\end{aligned}$$

Siten

$$\begin{aligned}\mathbb{P}\{Y_j \in A_j\} &\geq \mathbb{P}\{Y_j \in B_j\} - \lambda^{(T_n)} \\ &= 1 - \mathbb{P}\{Y_j \notin B_j\} - \lambda^{(T_n)} \\ &\geq 1 - 2\lambda^{(T_n)}.\end{aligned} \tag{8.21}$$

Valitaan nyt sellainen $m \in \mathbb{N}_+$, että jokaisen sylinterin A_j pohja $A_{jm} \subset \mathbb{R}^m$. (Esimerkiksi sylinterien $[0,1] \times \mathbb{R}^\infty = ([0,1] \times \mathbb{R}) \times \mathbb{R}^\infty$ ja $[1,2] \times [3,4] \times \mathbb{R}^\infty$ pohjiksi voidaan ottaa $[0,1] \times \mathbb{R} \subset \mathbb{R}^2$ ja $[1,2] \times [3,4] \subset \mathbb{R}^2$; tässä siis $m = 2$.)

Tarkastellaan sitten diskreetti-aikaista Gaussin kanavaa kohinavarianssilla $N = 1$ ja tehorajalla $P = \frac{KT_n}{m}$. Olkoon

$$x^m(j) = \big(x_1(j), \ldots, x_m(j)\big),$$

$j = 1, \ldots, M_n$, missä $x_1(j), \ldots, x_m(j)$ ovat koodisanan $x(j) \in \mathbb{R}^\infty$ m ensimmäistä komponenttia. Silloin

$$\frac{1}{m} \left\| x^m(j) \right\|^2 = \frac{1}{m} \sum_{k=1}^{m} x_k(j)^2 \leq \frac{1}{m} \sum_{k=1}^{\infty} x_k(j)^2 = \frac{1}{m} \left\| x(j) \right\|^2 \leq \frac{1}{m} KT_n$$

tehorajoituksen (8.17) nojalla. Siten vektorit $x^m(j)$ kelpaavat tämän diskreettiaikaisen kanavan koodisanoiksi.

Määritellään kanavan dekooderi $g : \mathbb{R}^m \to \{1, \ldots, M_n\}$ kaavalla

$$g(y^m) = j, \text{ kun } y^m \in A_{jm},$$

missä $y^m = (y_1, \ldots, y_m) \in \mathbb{R}^m$ ja A_{jm} on sylinterin A_j pohja. Huomaa, että $\mathbb{R}^m = \bigcup_{j=1}^{M_n} A_{jm}$ ja joukot A_{jm} ovat pistevieraita, joten g on siis hyvin määritelty.

Tässä kanavassa viestiä j vastaava tuloste on $Y_j^m = x^m(j) + Z^m$, missä $Z^m = (Z_1^*, \ldots, Z_m^*) \sim \mathrm{N}(0, I_m)$. Siten

$$\mathbb{P}\{Y_j^m \in A_{jm}\} = \mathbb{P}\{Y_j \in A_j\} \overset{(8.21)}{\geq} 1 - 2\lambda^{(T_n)},$$

joten viestin j virhetodennäköisyydelle saadaan

$$\mathbb{P}\{Y_j^m \notin A_{jm}\} = 1 - \mathbb{P}\{Y_j^m \in A_{jm}\} \leq 1 - \big(1 - 2\lambda^{(T_n)}\big) = 2\lambda^{(T_n)}.$$

Lauseen 7.2 todistuksen tapaan saadaan sitten

$$RT_n \leq 1 + RT_n \cdot 2\lambda^{(T_n)} + \frac{m}{2} \log\left(1 + \frac{KT_n}{m}\right),$$

joten

$$R \leq \frac{1}{T_n} + 2R\lambda^{(T_n)} + \frac{m}{2T_n} \log \left(1 + \frac{KT_n}{m} \right).$$

Mutta m voidaan valita mielivaltaisen suureksi ja, kuten edellä,

$$\lim_{m \to \infty} m \log \left(1 + \frac{KT_n}{m} \right) = KT_n \log e.$$

Siten

$$R \leq \frac{1}{T_n} + 2R\lambda^{(T_n)} + \frac{1}{2T_n} KT_n \log e = \frac{1}{T_n} + 2R\lambda^{(T_n)} + \frac{1}{2} K \log e.$$

Mutta $T_n \to \infty$ ja $\lambda^{(T_n)} \to 0$, kun $n \to \infty$, joten täytyy olla $R \leq \frac{1}{2} K \log e$. Siten $C \leq \frac{1}{2} K \log e$. □

Kirjallisuutta

[1] Norman Abramson. *Information Theory and Coding*. McGraw-Hill, 1963.

[2] Robert B. Ash. *Information Theory*. Dover, 1990. Vuonna 1965 julkaistun kirjan uusintapainos.

[3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[4] Robert G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

[5] David J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2004.

[6] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1998.