Tero Tilus

Jatkuvan kanavan kapasiteetti

Matematiikan (yleinen linja)
pro gradu -tutkielma
30. heinäkuuta 2007

JYVÄSKYLÄN YLIOPISTO
MATEMATIIKAN JA TILASTOTIETEEN LAITOS

Jyväskylä

Tekijä: Tero Tilus
Yhteystiedot: tero@tilus.net
Työn nimi: Jatkuvan kanavan kapasiteetti
Title in English: Capacity of Continuous Channel
Työ: Matematiikan (yleinen linja) pro gradu -tutkielma
Sivumäärä: 67
Tiivistelmä: Informaatioteorian merkittävät edistysaskeleet 1940- ja 50-luvuilla loivat matemaattisen perustan modernille tietoliikennetekniikalle. Tässä opinnäytetyössä esitellään informaatioteorian ja siihen tiiviisti kytkeytyvän koodausteorian keskeisimmät käsitteet ja tulokset. Tavoitteena on auttaa lukijaa hahmottamaan näiden käsitteiden sisältö, intuitiiviset tulkintamahdollisuudet, keskinäiset suhteet, sekä joitain kytkentöjä muille aloille.
English abstract: Remarkable advances in information theory taken on 1940's and 50's laid the mathematical foundation of modern telecommunication technology. In this thesis the core concepts of information theory and closely related coding theory are presented. The goal is to provide reader with an understanding of the content of the concepts, possibilities of intuitive interpretation, mutual relationships and a few connection to other fields.
Avainsanat: informaatioteoria, koodausteoria, kanava, jatkuva kanava, kapasiteetti
Keywords: information theory, coding theory, channel, continuous channel, capacity

Esipuhe

Tätä opinnäytetyötäni ohjasi professori Tapani Kuusalo Jyväskylän yliopiston Matematiikan ja tilastotieteen laitokselta. Toisena tarkastajana toimi Ari Lehtonen. Ohjaajan ja toisen tarkastajan lisäksi kiitokset Seija Sirkiälle ja Anni Toivolalle arvokkaista lisätiedoista ja kommenteista, Antti-Juhani Kaijanaholle ja Matthieu Weberille tämän opinnäytetyön taittoon käyttämästäni LaTeX-dokumenttiluokasta, sekä rakkaalle vaimolle runsaasta kannustuksesta ja pitkästä pinnasta.

Sisältö

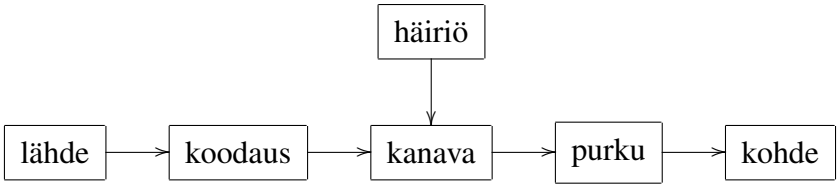
1 Johdanto

Tämä opinnäytetyö käsittelee informaatioteoriaksi kutsuttua viestinnän matemaattista formalisointia. Informaatioteoria henkilöityy vahvasti Claude E. Shannoniin. Häntä voi syystä pitää modernin informaatioteorian isänä. Hänen tarkastelunsa edusti uutta lähestymistapaa, jossa viestin ominaisuudet ja informaation välittämiseen käytetty kanava kuvattiin tilastollisin käsittein.

Shannonin esittämä informaation välitysprosessin malli [17] on esitetty kuvassa 1.1. Norbert Wienerin malli [21] ei erotellut informaation välitysprosessin ”tasoja” kuten Shannon teki. Lähteestä otettu viesti välitettiin kanavan kautta sellaisenaan. Shannonin mallissa kaikille mahdollisille viesteille, joita informaatiolähde voi tuottaa, valitaan kanavan välittämiä signaaleja käyttävä ”koodattu” esitys, joka lähetetään. Vastaanottavassa päässä signaali pyritään ”purkamaan” takaisin viestiksi. Koodaus, välitys kanavan kautta ja purku muodostavat lähettäjän ja vastaanottajan näkökulmasta läpinäkyvän tiedonsiirtolaitteen, joka kapseloi kanavan erilleen viesteistä. Shannonin esittämä malli on vakiintunut ja siihen tukeudutaan myös tässä opinnäytetyössä.

Shannonin voidaan klassikkoartikkelillaan katsoa käytännössä perustaneen kolme uutta tutkimusalaa: informaatioteorian, koodausteorian ja informaatiolähteiden (tilastollisten ominaisuuksien) tutkimuksen. [9] Näistä kolmas sivuutetaan tässä opinnäytetyössä, koska sen käsitteistö ja tulokset eivät ole välttämättömiä tiedonsiirtokanavan kapasiteettia koskevien tulosten muotoilemiseen ja ymmärtämiseen. Samasta syystä sivuutetaan koodausteorian osalta virheitä korjaavien koodien tarkastelu.

Informaatioteoria analysoi kanavan ominaisuuksia. Koodausteoria analysoi tapoja manipuloida (eli koodata) viestejä. Informaatioteoria työskentelee todennäköisyys- ja mittateoreettisella käsitteistöllä kun taas koodausteoriassa operoidaan pääosin algebrallisin käsittein mm. merkkijonoilla ja graafeilla.



Kuva 1.1: Informaation välitysprosessin malli Shannonin mukaan

Informaatioteorian tutkimus on lähtökohdiltaan hyvinkin pragmaattinen ala. Raa'at käytännön ongelmat ovat motivoineet matemaattisen formuloinnin ja tutkimuksen. Shannonin ongelma, johon hän informaatioteorialla pureutui, oli tietoliikennetekniikka ja siihen liittyvät laskelmat.

Pääasiallisina lähteinä tässä opinnäytetyössä ovat kirjat [8, 2, 13], Ari Lehtosen informaatioteorian luennot [11] ja luonnollisesti Shannonin klassikkotyö [17]. Luvussa 2 johdetaan epävarmuuden mitta, intuitiivisesti perustellusta joukosta aksioomia, osoitetaan joitakin sen perusominaisuuksia, johdetaan informaatioteoreettinen informaation mitta, analysoidaan sen suhdetta muutamiin läheisiin aloihin ja informaation käsitettä yleisesti. Luvussa 3 käydään läpi jatkuvan kanavan kapasiteetin määrittämiseksi tarpeelliset koodausteoreettiset työkalut. Luvussa 4 esitellään diskreetissä tapauksessa Shannonin mallin mukaisen viestinnän tarkastelun peruskäsitteet, jotka sitten luvussa 5 yleistetään jatkuvaan tapaukseen ja viedään koodauslauseen (engl. *coding theorem*) ja käänteisen (engl. *converse*) koodauslauseen todistus loppuun.

2 Informaation mittaamisesta

Informaatioteoreettisessa tarkastelussa on Shannonin mukaan [18] kyse tiedonsiirtoprosessin ja itse viestin ominaisuuksista, ei semantiikasta tai informaation arvosta.

> *Frequently the messages have meaning... these semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages.*

Tätä ei pidä kuitenkaan ymmärtää niin, että muut viestinnän piirteet olisivat merkityksettömiä. Shannon vain rajaa informaatioteorian koskemaan viestien välittämisen teknistä näkökulmaa.

2.1 Aksioomat epävarmuuden mitalle

Informaatioteoria perustuu epävarmuuden kvantifiointiin. Tässä luvussa käydään läpi ne vaatimukset, jotka epävarmuuden mitalle halutaan sen sovelluksia ajatellen asettaa. Lisäksi osoitetaan, että vaatimukset määrittelevät epävarmuuden mitan vakiokerrointa vaille yksikäsitteisesti.

Informaatioteoria hyödyntää tilastollisia käsitteitä. Siksi on aluksi on syytä kerrata lyhyesti todennäköisyysteorian keskeiset käsitteet (lisätietoa esim. [19]).

Satunnaiskokeen *näyteavaruus* on joukko Ω, joka sisältää alkioina kokeen kaikki mahdolliset tulokset. Esimerkiksi arpakuution heitolle näyteavaruus voisi olla $\Omega = \{1, 2, 3, 4, 5, 6\}$ tai $\Omega = \mathbb{N}$. Näyteavaruuden osajoukkoja kutsutaan *tapahtumiksi*. Esimerkiksi ”parillinen silmäluku” on em. näyteavaruuden tapahtuma. Vain yhden näyteavaruuden alkion sisältäviä tapahtumia kutsutaan *alkeistapahtumiksi*. Kun satunnaiskoe toteutetaan, tietty tapahtuma $A \subset \Omega$ joko tapahtuu, jos kokeen tulos $x \in A$ tai ei tapahdu, jos $x \notin A$. Tapahtumien joukon $\mathcal{F}$ edellytetään muodostavan näyteavaruuden σ-algebran, eli kokoelman johon kuuluu näyteavaruus itse ja kaikki kokoelman joukkojen komplementit (näyteavaruuden suhteen) ja numeroituvat yhdisteet. Pienin Ω:n σ-algebra on $\{\emptyset, \Omega\}$ ja suurin $\mathbb{P}(\Omega)$.

Mitta-avaruus $(\Omega, \mathcal{F}, P)$, jossa $\mathcal{F}$ on Ω:n σ-algebra ja P toteuttaa Kolmogorovin todennäköisyysaksioomat (ei-negatiivisuus, $P(\Omega) = 1$ ja σ-additiivisuus) on *todennäköisyysavaruus*.

Jos Ω on äärellinen, voidaan aina valita $\mathcal{F} = \mathbb{P}(\Omega)$, jolloin riittää merkitä todennäköisyysavaruutta (Ω, P). *Satunnaismuuttuja* X on mitallinen funktio $X\colon (\Omega, \mathcal{F}) \mapsto S$ (riittää merkitä $X\colon \Omega \mapsto S$, jos Ω on äärellinen tai σ-algebra muuten implisiittisesti selvä), jossa S on mitallinen avaruus, usein $\mathbb{R}$ euklidisella mitalla varustettuna. Arpakuutioesimerkkiä jatkaen satunnaismuuttujia voisivat olla esim. $X =$"parillisten silmälukujen määrä yhdellä heitolla" ja $X =$"kahden heiton silmälukujen summa". Tapahtuman $\{\omega \in \Omega : X(\omega) \in s \subset S\}$ *todennäköisyys* on $P(\{\omega \in \Omega : X(\omega) = x\}) = P(X = x)$. Lyhennysmerkintöinä käytetään tapahtumalle $\{X \in s\}$ ja todennäköisyydelle $P(X \in s)$ ja vastaavasti $\{X = x\}$ ja $P(X = x)$, kun $s = \{x\}$.

Ennen aksioomien tarkastelua on vielä tarpeen muistuttaa, ettei ole sovellusten kannalta lopulta erityisen tärkeää se, miten epävarmuuden mittaan päädytään. Se, että tietty yksikäsitteinen epävarmuuden mitta voidaan järjestelmällisesti johtaa perustelluista vaatimuksista luo tietysti informaatioteorialle vankan matemaattisen perustan. Lisäksi on hyvä muistaa, että epävarmuuden mitan keskeisin arvo on siinä miten sen avulla voidaan määritellä välitetyn informaation määrä ja todistaa "hyvien koodien" olemassaolo ja toisaalta koodauksen teoreettiset rajat.

Tässä luvussa läpikäytävät aksioomat ovat olennaisesti samat kuin Shannonilla [17]. Olennaisesti heikommatkin aksioomat (esim. [10]) riittävät epävarmuuden mitan yksikäsitteiseen määrittelyyn.

Oletetaan, että tilastolliseen kokeeseen liittyy satunnaismuuttuja $X\colon \Omega \mapsto \{x_1, \ldots, x_M\}$ ja X:n arvoja vastaavat todennäköisyydet ovat $\{p_1, \ldots, p_M\}$ ja $p_i > 0$. Konstruoidaan seuraavaksi satunnaistapahtuman epävarmuus $h\colon]0, 1] \mapsto [0, \infty[$ ja satunnaismuuttujan keskimääräinen epävarmuus H. $h(p_i)$ on tapahtuman $\{X = x_i\}$ epävarmuus tai toisin päin ajateltuna epävarmuus, joka poistuu (tai informaatio, joka saadaan) kun havaitaan X saaneen kokeessa arvon x_i. Kaikille M määritellään $H_M(p_1, \ldots, p_M) = \sum_{i=1}^{M} p_i h(p_i)$, eli että $H_M(p_1, \ldots, p_M)$, jatkossa lyhemmin $H(X)$, on h:n odotusarvo, eli X:n keskimääräinen epävarmuus tai lyhyemmin vain epävarmuus.

Seuraavaksi aletaan asettaa H:lle vaatimuksia. Merkitään epävarmuutta tasaisen jakauman tapauksessa $f(M) = H(M^{-1}, \ldots, M^{-1})$. Nyt siis $f(2)$ on kolikon heiton, $f(6)$ nopan heiton ja $f(5, 3 \times 10^6)$ satunnaisen suomalaisen valinnan epävarmuus. f:n tulee olla aidosti kasvava M:n funktio, eli $M < M' \Rightarrow f(M) < f(M')$.

Tarkastellaan riippumattomia tasajakautuneita satunnaismuuttujia $X\colon \Omega \mapsto \{x_1, \ldots, x_M\}$ ja $Y\colon \Omega \mapsto \{y_1, \ldots, y_N\}$. Yhdessä tarkasteltuna muuttujilla on MN yhtä todennäköistä arvoa, joten yhteisen kokeen epävarmuus on $f(MN)$. Jos X:n arvo havaitaan sen ei pitäisi vaikuttaa Y:n epävarmuuteen, koska muuttujien oletettiin olevan riippumattomia. Täsmällisesti

muotoiltuna $f(MN) - f(M) = f(N) \Leftrightarrow f(MN) = f(M) + f(N)$.

Luovutaan nyt tasaisen jakauman oletuksesta ja tarkastellaan mielivaltaisia jakaumia. Jaetaan satunnaismuuttujan X mahdolliset arvot kahteen ryhmään $A = \{x_1, \ldots, x_r\}$ ja $B = \{x_{r+1}, \ldots, x_M\}$. Muodostetaan yhdistetty tilastollinen koe seuraavasti. Valitaan aluksi toinen ryhmistä siten, että ryhmä tulee valituksi todennäköisyydellä, joka on ryhmän alkioiden todennäköisyyksien summa. Eli ryhmä A valitaan todennäköisyydellä $p_A = \sum_{i=1}^{r} p_i$ ja B todennäköisyydellä $p_B = \sum_{i=r+1}^{M} p_i$. Jos valituksi tuli A, valitaan x_i todennäköisyydellä $P\{X = x_i | x_i \in A\} = p_i/p_A$. Jos valituksi tuli B tehdään x_i:n valinta vastaavaan tapaan. Tuloksena saatu yhdistetty koe vastaa alkuperäistä muuttujaa X.

Ennen edellä kuvattua yhdistettyä koetta epävarmuus on $H(p_1, \ldots, p_M)$. Paljastamalla tuliko valituksi ryhmä A vai B, vähennämme epävarmuutta $H(p_A, p_B)$:lla. Jos valittiin A, jäljelle jäävä epävarmuus on $H(p_1/p_A, \ldots, p_r/p_A)$ ja jos B niin $H(p_{r+1}/p_B, \ldots, p_M/p_B)$. Siten jäljelle jäävä keskimääräinen epävarmuus on

$$p_A H(p_1/p_A, \ldots, p_r/p_A) + p_B H(p_{r+1}/p_B, \ldots, p_M/p_B).$$

Tästä rakentuu kolmas vaatimus. Yhdistetyn kokeen epävarmuus vähennettynä valittu ryhmä paljastamalla vähenevällä epävarmuudella on oltava ryhmän paljastumisen jälkeen jäljelle jäävä keskimääräinen epävarmuus, eli

$$H(p_1, \ldots, p_M) = H(p_A, p_B) + p_A H(p_1/p_A, \ldots, p_r/p_A) + p_B H(p_{r+1}/p_B, \ldots, p_M/p_B)$$

Lopuksi vaaditaan matemaattisen mukavuudenhalun vuoksi vielä, että $H(p, 1-p)$ on jatkuva p:n funktio. On toki luontevaa olettaa, että pieni muutos satunnaismuuttujan jakaumassa vaikuttaa epävarmuuteen vain vähän. Ja tästähän jatkuvuudessa on olennaisesti kyse.

Määritelmä 2.1 (Epävarmuuden aksioomat) *Diskreetin satunnaismuuttujan keskimääräinen epävarmuus H toteuttaa seuraavat aksioomat.*

1. $H(M^{-1}, \ldots, M^{-1}) = f(M)$ *on M:n ($M \in \mathbb{Z}_+$) funktiona aidosti kasvava.*
2. $f(MN) = f(M) + f(N)$, $M, N \in \mathbb{Z}_+$.
3. $H(p_1, \ldots, p_M) = H(p_A, p_B) + p_A H(p_1/p_A, \ldots, p_r/p_A) + p_B H(p_{r+1}/p_B, \ldots, p_M/p_B)$, *jossa $p_A = \sum_{i=1}^{r} p_i$ ja $p_B = \sum_{i=r+1}^{M} p_i$. Tätä aksioomaa kutsutaan usein nimellä ryhmittelyaksiooma.*
4. $H(p, 1-p)$ *on p:n jatkuva funktio.*

Osoittautuu, että nämä neljä aksioomaa määrittelevät funktion H vakiokerrointa vaille yksikäsitteisesti.

Lause 2.2 *Ainut määritelmän 2.1 aksioomat toteuttava funktio on*

$$H(p_1, \ldots, p_M) = C \sum_{i=1}^{M} p_i \log_b p_i^{-1}, \tag{2.1}$$

missä $C > 0$ *ja* $b > 1$.

Todistus: Voimme kiinnittää kantaluvun b, koska sen muutos voidaan vaihtaa kertoimen C muutokseen. Varmistetaan aluksi, että annettu H toteuttaa aksioomat.

1. Oletetaan, että $M, M' \in \mathbb{Z}_+$ ja $M < M'$. Funktio $\log_b$, $b > 1$, on aidosti kasvava, joten

$$\begin{aligned} MM^{-1} \log M &< M'M'^{-1} \log M' \\ \Rightarrow C \sum_{i=1}^{M} M^{-1} \log M &< C \sum_{i=1}^{M'} M'^{-1} \log M' \\ \Rightarrow f(M) &< f(M') \end{aligned}$$

2. Myös tämän aksiooman toteutuminen seuraa suoraan määritelmistä

$$\begin{aligned} f(MN) &= C \sum_{i=1}^{MN} (MN)^{-1} \log MN \\ &= C \log MN = C \log M + C \log N \\ &= C \sum_{i=1}^{M} M^{-1} \log M + C \sum_{i=1}^{N} N^{-1} \log N = f(M) + f(N) \end{aligned}$$

3. Olkoon $1 \leq r < M$ ja merkitään $p_A = \sum_{i=1}^{r} p_i$ ja $p_B = \sum_{i=r+1}^{M} p_i$.

$$\begin{aligned} H(p_1, \ldots, p_M) &= C \sum_{i=1}^{M} p_i \log p_i^{-1} \\ &= \frac{p_A}{p_A} C \sum_{i=1}^{r} p_i \log \left(p_i^{-1} \frac{p_A}{p_A} \right) + \frac{p_B}{p_B} C \sum_{i=r+1}^{M} p_i \log \left(p_i^{-1} \frac{p_B}{p_B} \right) \\ &= p_A C \sum_{i=1}^{r} \frac{p_i}{p_A} \left[\log \left(\frac{p_i}{p_A} \right)^{-1} + \log p_A^{-1} \right] + \\ &\quad p_B C \sum_{i=r+1}^{M} \frac{p_i}{p_B} \left[\log \left(\frac{p_i}{p_B} \right)^{-1} + \log p_B^{-1} \right] \\ &= \sum_{i=1}^{r} p_i \log p_A^{-1} + \sum_{i=r+1}^{M} p_i \log p_B^{-1} + \\ &\quad p_A C \sum_{i=1}^{r} \frac{p_i}{p_A} \log \left(\frac{p_i}{p_A} \right)^{-1} + p_B C \sum_{i=r+1}^{M} \frac{p_i}{p_B} \log \left(\frac{p_i}{p_B} \right)^{-1} \\ &= H(p_A, p_B) + p_A H(\frac{p_1}{p_A}, \ldots, \frac{p_r}{p_A}) + p_B H(\frac{p_{r+1}}{p_B}, \ldots, \frac{p_M}{p_B}) \end{aligned}$$

4. Logaritmi, lineaarikuvaukset ja jatkuvien kuvausten äärelliset summat ovat jatkuvia, joten $H(p, 1-p)$ on jatkuva.

Seuraavaksi osoitamme toisen suunnan, eli että jos funktio toteuttaa annetut aksioomat niin sen on muotoa 2.1. Aloitetaan osoittamalla

$$f(M^k) = kf(M) \qquad \forall M, k \in \mathbb{Z}_+ \tag{2.2}$$

Tapaus $k = 1$ pätee triviaalisti. Tehdään induktio-oletus, että yhtälö 2.2 pätee $k-1$ saakka. Nyt induktio-oletuksen ja aksiooman 2 nojalla $f(M^k) = f(M \cdot M^{k-1}) = f(M) + f(M^{k-1}) = f(M) + (k-1)f(M) = kf(M)$, joten yhtälö 2.2 pätee kaikille k. Seuraavaksi osoitamme, että

$$f(M) = C \log M \qquad \forall M \in \mathbb{Z}_+,\, C \in \mathbb{R}_+ \tag{2.3}$$

Olkoon $M = 1$. Nyt aksiooman 2 nojalla $f(1) = f(1 \cdot 1) = f(1) + f(1) \Rightarrow f(1) = 0$. Se ettei tapahtumaan, jolla on vain yksi mahdollinen lopputulos, liity lainkaan epävarmuutta, vastaa hyvin intuitiota. Kiinnitetään nyt M siten, että se on yhtä suurempi positiivinen kokonaisluku. Jos nyt valitaan mielivaltainen $r \in \mathbb{Z}_+$, on 2^r kahden M:n potenssin välissä, eli on olemassa k siten, että $M^k \leq 2^r < M^{k+1}$. Aksiooman 1 ja yhtälön 2.2 nojalla pätee $kf(M) \leq rf(2) < (k+1)f(M) \Leftrightarrow k/r \leq f(2)/f(M) < (k+1)/r$. Logaritmi, jonka kantaluku on suurempi kuin 1, on aidosti kasvava funktio, joten $\log M^k \leq \log 2^r < \log M^{k+1}$, joka saadaan muotoon $k/r \leq (\log 2)/(\log M) < (k+1)/r$. Siis

$$\left| \frac{\log 2}{\log M} - \frac{f(2)}{f(M)} \right| < \frac{1}{r} \tag{2.4}$$

Koska M oli kiinnitetty ja r mielivaltainen, voimme antaa r:n kasvaa rajatta, joten itseisarvon termit saadaan rajankäynnillä yhtäsuuriksi, jolloin $f(M) = C \log M$, jossa $C = f(2)/(\log 2)$. Huomataan, että C on aina positiivinen, koska $f(1) = 0$ ja $f(M)$ on kasvava funktio. Seuraavaksi osoitamme, että

$$H(p, 1-p) = C\left[p \log p^{-1} + (1-p)\log(1-p)^{-1}\right] \qquad \forall p \in \mathbb{Q} \cap]0,1[\tag{2.5}$$

Olkoon $p = r/s$, jossa $r, s \in \mathbb{Z}_+$. Oletusten ja aksiooman 3 nojalla

$$\begin{aligned} f(s) &= H(\underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{r \text{ kpl}}, \underbrace{\frac{1}{s}, \ldots, \frac{1}{s}}_{s-r \text{ kpl}}) \\ &= H\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{r}{s} f(r) + \frac{s-r}{s} f(s-r) \end{aligned}$$

Yhtälön 2.3 avulla saadaan $C \log s = H(p, 1-p) + Cp \log r + C(1-p)\log(s-r)$, joten

$$\begin{aligned} H(p,1-p) &= -C\left[p\log r - \log s + (1-p)\log(s-r)\right] \\ &= -C\left[p\log r - p\log s + p\log s - \log s + (1-p)\log(s-r)\right] \\ &= -C\left[p\log\frac{r}{s} + (1-p)\log\frac{s-r}{s}\right] \\ &= C\left[p\log p^{-1} + (1-p)\log(1-p)^{-1}\right] \end{aligned}$$

Nyt aksiooman 4 (jatkuvuus) nojalla yhtälö 2.5 pätee myös kaikille reaaliluvuille välillä $]0,1[$.

Päätämme todistuksen osoittamalla induktiolla, että yhtälö 2.1 pätee. Olemme jo osoittaneet, että yhtälö pätee tapauksissa $M = 1$ (ks. yhtälö 2.3) ja $M = 2$ (ks. yhtälö 2.5). Tehdään induktio-oletus, että yhtälö 2.1 pätee positiivisille kokonaisluvuille $M-1$ saakka. Merkitään $p^* = \sum_{i=1}^{M-1} p_i$. Oletetaan $M > 2$. Oletusten ja aksiooman 3 nojalla

$$\begin{aligned} H(p_1,\ldots,p_M) &= H(p^*,p_M) + p^* H(p_1/p^*,\ldots,p_{M-1}/p^*) + p_M H(1) \\ &= C[p^*\log(p^*)^{-1} + p_M \log p_M^{-1}] + Cp^* \sum_{i=1}^{M-1} \frac{p_i}{p^*}\log\left(\frac{p_i}{p^*}\right)^{-1} + p_M \cdot 0 \\ &= C\left[p^*\log(p^*)^{-1} + p_M\log p_M^{-1}\right] + C\left[\sum_{i=1}^{M-1} p_i \log p_i^{-1} - p^*\log(p^*)^{-1}\right] \\ &= C\sum_{i=1}^{M} p_i \log p_i^{-1} \end{aligned}$$

□

2.2 Entropian ominaisuuksia

Informaatioteoriassa käytetään yleisesti käsitteitä diskreetti ja äärellinen synonyymeinä. Pitäydyn samassa konventiossa, joten jäljempänä ”diskreetti” merkitsee äärellistä. Jatkossa merkitään diskreeteille satunnaismuuttujille X ja Y.

$$\begin{aligned} p_i &= P\{X = x_i\} \\ p(y) &= P\{Y = y\} \\ p(x|y) &= P\{X = x | Y = y\} \\ p(x,y) &= P\{X = x, Y = y\} \end{aligned}$$

Lisäksi muistetaan, että $p(x|y) = p(x,y)p(y)^{-1}$. Huomataan myös, että satunnaismuuttuja, jonka jakauman mukaisesta todennäköisyydestä on kyse, ilmoitetaan implisiittisesti lyhennysmerkinnän muuttujan valinnalla. Siksi $p(x) = P\{X = x\}$, eikä $p(x) = P\{Y = x\}$.

Edellisessä luvussa johdettiin mitta satunnaismuuttujan keskimääräiselle epävarmuudelle. Seuraavaksi määritellään edellisen luvun tulosten pohjalta epävarmuuden mitta eli *entropia*. Tämä tehdään vakiintuneen tavan mukaan satunnaistapahtuman entropian avulla.

Määritelmä 2.3 (Satunnaistapahtuman entropia) *Olkoon funktio* $h_b\colon]0,1] \mapsto [0,\infty]$

$$h_b(p_i) = \log_b p_i^{-1},$$

missä $b > 1$.

Funktio h kuvaa epävarmuutta, joka liittyy tapahtumaan $\{X = x_i\}$, jonka todennäköisyys on p_i tai vastaavasti epävarmuutta, joka poistuu kun satunnaistapahtuman lopputulos havainnoidaan.

Tämän avulla voidaan määritellä avainkäsite, satunnaismuuttujan entropia. Se on epävarmuuden odotusarvo, eli satunnaismuuttujan keskimääräinen entropia. Jatkossa lyhyesti entropia.

Määritelmä 2.4 (Diskreetin satunnaismuuttujan entropia) *Diskreetin satunnaismuuttujan* X *b-kantainen* entropia $H_b(X)$ *määritellään seuraavasti:*

$$H_b(X) \;=\; E(h_b(X)) = \sum_{i=1}^{M} p_i \log_b \frac{1}{p_i}$$

jossa $X\colon \Omega \mapsto \{x_1, x_2, \ldots\, x_M\}$

Huomaa, että p_i^{-1} ei ole määritelty jos $p_i = 0$, joten myöskään $\log_b p_i^{-1}$ ei ole määritelty. Koska kuitenkin $\lim_{p_i \to 0} p_i \log_b p_i^{-1} = 0$, voitaisiin määritellä $p_i \log_b p_i^{-1} = 0$, kun $p_i = 0$.

Jatkossa, ellei toisin ilmoiteta, oletetaan $b = 2$ ja merkitään $H(X) = H_2(X)$. Kantaluku merkitään jatkossa näkyville vain niissä yhteyksissä, joissa sillä on merkitystä. Kantaluvun valinta (kunhan pysytään oletuksissa ja valittu kanta $b > 1$) ei vaikuta entropian ominaisuuksiin, eikä tuleviin todistuksiin. Se kuitenkin vaikuttaa konkreettisten laskelmien lopputulokseen. Entropian yksikkö määräytyy kantaluvun valinnan mukaan. Valinnalla $b = 2$ yksikkö on bitti (engl. *bit,* ***bi****nary digi****t***).

Esimerkki 2.5 Olkoon diskreetti satunnaismuuttuja X kolikon heitto. Kolikolle $X\colon \Omega \mapsto \{0,1\}$ ja jakauma on tasainen ($p_i = \frac{1}{M} = \frac{1}{2}$). Nyt

$$\begin{aligned} H(X) &= \sum_{i=1}^{M} p_i \log_b \frac{1}{p_i} = \sum_{i=1}^{2} \frac{1}{2} \log_2 \frac{1}{\frac{1}{2}} \\ &= \frac{1}{2} + \frac{1}{2} = 1 \text{ bit} \end{aligned}$$

Esimerkki 2.6 Olkoon diskreetti satunnaismuuttuja X arpakuutio. Arpakuutiolle $X\colon \Omega \mapsto \{1,2,3,4,5,6\}$ ja jakauma on tasainen ($p_i = \frac{1}{M} = \frac{1}{6}$). Nyt

$$\begin{aligned} H(X) &= \sum_{i=1}^{M} p_i \log_b \frac{1}{p_i} = \sum_{i=1}^{6} \frac{1}{6} \log_2 \frac{1}{\frac{1}{6}} \\ &= \log_2 6 \approx 2,5849 \text{ bit} \end{aligned}$$

Huomataan, että koska $0 \leq p_i \leq 1$, niin $p_i \log p_i^{-1} \geq 0$ kaikille i. Siten $H(X)$ on aina ei-negatiivinen.

Päästäksemme käsiksi $H(X)$:n ominaisuuksiin tarvitsemme funktion konveksiuden käsitteen ja muutamia aputuloksia.

Määritelmä 2.7 (funktion konveksius) *Olkoon $I \subset \mathbb{R}$ väli. Funktio $f\colon I \mapsto \mathbb{R}$ on* konveksi *joss*

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b) \qquad \forall a, b \in I, t \in [0,1] \tag{2.6}$$

Funktio $f\colon I \mapsto \mathbb{R}$ on konkaavi*, joss $-f$ on konveksi.*
Funktio $f\colon I \mapsto \mathbb{R}$ on aidosti konveksi *(-konkaavi vastaavasti) kun epäyhtälö 2.6 edellä on aito aina kun $a \neq b$ ja $t \in]0,1[$.*

Lause 2.8 *Olkoon $i \in \{1,\ldots,M\}$, $p_i > 0$, $q_i > 0$ ja $\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} q_i = 1$. Nyt $\sum_{i=1}^{M} p_i \log_b p_i^{-1} \leq \sum_{i=1}^{M} p_i \log_b q_i^{-1}$ ja yhtäsuuruus pätee joss $p_i = q_i$ kaikille i.*

Todistus: Luonnollinen logaritmi $\ln(x)$ on derivoituva ja aidosti konkaavi funktio, joten sen pisteen $x = 1$ kautta kulkeva tangentti $x \mapsto x - 1$ on pistettä $x = 1$ lukuun ottamatta logaritmin graafin yläpuolella, eli $\ln(y) \leq \ln(1) + \mathrm{D}(\ln)(1)(y-1) \Leftrightarrow \ln(y) \leq y - 1$ yhtäsuuruudella joss $y = 1$. Siten $\ln(q_i/p_i) \leq q_i/p_i - 1$ yhtäsuuruudella joss $p_i = q_i$.

Kertomalla epäyhtälö puolittain p_i:llä ja summaamalla yli i:n saadaan

$$\begin{aligned}
\sum_{i=1}^{M} p_i \ln q_i p_i^{-1} &\leq \sum_{i=1}^{M} (q_i - p_i) = 1 - 1 = 0 \\
\Leftrightarrow \sum_{i=1}^{M} p_i \ln p_i^{-1} &\leq \sum_{i=1}^{M} p_i \ln q_i^{-1} \\
\Leftrightarrow \log_b e \left(\sum_{i=1}^{M} p_i \ln p_i^{-1} \right) &\leq \log_b e \left(\sum_{i=1}^{M} p_i \ln q_i^{-1} \right) \\
\Leftrightarrow \sum_{i=1}^{M} p_i \log_b p_i^{-1} &\leq \sum_{i=1}^{M} p_i \log_b q_i^{-1}
\end{aligned}$$

yhtäsuuruudella joss $p_i = q_i$ kaikille i. □

On luonnollista olettaa, että satunnaismuuttujan epävarmuus on suurin juuri silloin kun kaikki mahdolliset arvot ovat yhtä todennäköisiä, eli kun jakauma on tasainen. Todistetaan seuraavaksi, että entropialla $H(X)$ todella on tämä ominaisuus.

Lause 2.9 *$H(p_1, \ldots, p_M) \leq \log M$ yhtäsuuruudella joss $p_i = 1/M$ kaikille i.*

Todistus: Kun valitaan $q_i = 1/M$, saadaan lauseen 2.8 nojalla

$$H(p_1, \ldots, p_M) = \sum_{i=1}^{M} p_i \log p_i^{-1} \leq \sum_{i=1}^{M} p_i \log \left(\frac{1}{M} \right)^{-1} = \log M \sum_{i=1}^{M} p_i = \log M$$

ja yhtäsuuruus pätee joss $p_i = q_i = 1/M$ kaikille i. □

Yleistetään seuraavaksi entropian tarkastelua useamman satunnaismuuttujan tapauksiin. Olkoot nyt X ja Y samaan satunnaisprosessiin liittyviä diskreettejä satunnaismuuttujia. Olkoon muuttujilla yhteisjakauma $p_{ij} = p(x_i, y_j) = P\{X = x_i \text{ ja } Y = y_j\}$, jossa $i \in \{1, \ldots, M\}$ ja $j \in \{1, \ldots, N\}$. Näin ollen on MN mahdollisuutta ja tapahtuman $\{X = x_i \text{ ja } Y = y_j\}$ todennäköisyys on p_{ij}. Tällaisessa tilanteessa voidaan luontevasti puhua näiden kahden muuttujan yhdessä karakterisoiman tapahtuman entropiasta ja siten myös muuttujien yhteisestä entropiasta.

Määritelmä 2.10 (Yhteisentropia) *Olkoot $X : \Omega \mapsto \{x_1, \ldots, x_M\}$ ja $Y : \Omega \mapsto \{y_1, \ldots, y_N\}$ diskreettejä satunnaismuuttujia ja p_{ij} niiden yhteistodennäköisyys. Nyt muuttujien X ja Y* yhteisentropia *(engl.* joint entropy*) on*

$$H_b(X,Y) = \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log_b p(x_i,y_j)^{-1}.$$

Mikä sitten on kahden muuttujan entropioiden ja yhteisentropian keskinäinen suhde? Käsitteen kannalta on johdonmukaista ajatella, että kahden muuttujan yhteinen epävarmuus on aina yhden muuttujan epävarmuutta suurempi. Tämä pätee. Lisäksi osoitamme, että yhteisentropia on enintään entropioiden summa sekä yhtäsuuri jos ja vain jos muuttujat ovat riippumattomat.

Lause 2.11 *$H(X) \leq H(X,Y)$ ja $H(Y) \leq H(X,Y)$ Oletukset kuten määritelmässä 2.10.*
Todistus:

$$\begin{aligned} H(X) &= \sum_{i=1}^{M} p(x_i)\log p(x_i)^{-1} = \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log p(x_i)^{-1} \\ &\leq \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log p(x_i,y_j)^{-1} = H(X,Y) \end{aligned}$$

Sama argumentti soveltuu luonnollisesti myös toiselle muuttujalle. □

Lause 2.12 *$H(X,Y) \leq H(X) + H(Y)$ yhtäsuuruudella joss X ja Y ovat riippumattomat. Oletukset kuten määritelmässä 2.10.*

Todistus:

$$\begin{aligned} H(X)+H(Y) &= \sum_{i=1}^{M} p(x_i)\log p(x_i)^{-1} + \sum_{j=1}^{N} p(y_j)\log p(y_j)^{-1} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log p(x_i)^{-1} + \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log p(y_j)^{-1} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)(\log p(x_i)^{-1} + \log p(y_j)^{-1}) = \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}\log q_{ij} \\ &\geq \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}\log p_{ij} = H(X,Y) \end{aligned}$$

Epäyhtälö edellä saadaan lauseesta 2.8 kun vain tarkistetaan oletuksen pitävyys, eli

$$\sum_{i=1}^{M}\sum_{j=1}^{N} q_{ij} = \sum_{i=1}^{M} p(x_i)\sum_{j=1}^{N} p(y_j) = \sum_{i=1}^{M} p(x_i)\cdot 1 = 1.$$

Kaksoissummaus voidaan helposti indeksoida uudelleen yhdeksi summaksi eikä se siten vaikuta lauseen soveltamiseen. Lisäksi epäyhtälössä yhtäsuuruus pätee joss $p_{ij} = q_{ij} \Leftrightarrow p(x_i, y_j) = p(x_i)p(y_i)$ kaikille i, j, joka taas pätee joss X ja Y ovat riippumattomat. □

Edelliset tulokset (lauseet 2.11 ja 2.12) yleistyvät esitettyjen kanssa analogisin todistuksin myös useampien muuttujien tapaukseen.

$$\begin{aligned} H(X_1, \ldots, X_{m-1}) &\leq H(X_1, \ldots, X_m) \\ &\leq H(X_1, \ldots, X_n) + \ldots + H(X_{n+1}, \ldots, X_m) \qquad \text{jossa } 1 < n < m. \end{aligned}$$

Määritelmä 2.13 (Ehdollinen entropia) *Olkoot $X\colon \Omega \mapsto \{x_1, x_2, \ldots x_M\}$ ja $Y\colon \Omega \mapsto \{y_1, y_2, \ldots y_N\}$ diskreettejä satunnaismuuttujia. Nyt $x \mapsto p(x|y)$ on X:n ehdollinen jakauma kun $Y = y$. Merkitään*

$$H_b(X|Y = y) = \sum_{i=1}^{M} p(x_i|y) \log_b \frac{1}{p(x_i|y)}.$$

Ehdollinen b-kantainen entropia $H_b(X|Y)$ on $y \mapsto H_b(X|Y = y)$:n odotusarvo

$$\begin{aligned} H_b(X|Y) &= \sum_{i=1}^{N} p(y_j) H_b(X|Y = y_j) \\ &= \sum_{j=1}^{N} p(y_j) \sum_{i=1}^{M} p(x_i|y_j) \log_b \frac{1}{p(x_i|y_j)} \\ &= \sum_{j=1}^{N} \sum_{i=1}^{M} p(y_j) \frac{p(x_i, y_j)}{p(y_j)} \log_b \frac{1}{p(x_i|y_j)} \\ &= \sum_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} p(x_i, y_j) \log_b \frac{1}{p(x_i|y_j)} \end{aligned}$$

Ehdollinen entropia kuvaa satunnaismuuttujan X epävarmuutta kun Y on tunnettu. Myös ehdollinen entropia yleistyy useamman kuin kahden muuttujan tapaukseen aiempien yleistysten kanssa analogisella tavalla. Seuraavat lauseet kuvaavat entropian, yhteisentropian ja ehdollisen entropian keskinäiset suhteet.

Lause 2.14 $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$.

Todistus: Tulos seuraa suoraan määritelmistä.

$$\begin{aligned} H(X,Y) &= \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log p(x_i,y_j)^{-1} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log(p(x_i)p(y_j|x_j))^{-1} \\ &= \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log p(x_i)^{-1} + \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log p(y_j|x_j)^{-1} \\ &= \sum_{i=1}^{M} p(x_i\log p(x_i)^{-1} + H(Y|X) = H(X) + H(Y|X) \end{aligned}$$

Yhtälön toisen osan todistukseen käy (vaihtaen muuttujien x_i ja y_j keskinäistä asemaa) täsmälleen sama argumentti. □

Lause 2.15 *$H(Y|X) \leq H(Y)$ yhtäsuuruudella joss X ja Y ovat riippumattomat.*

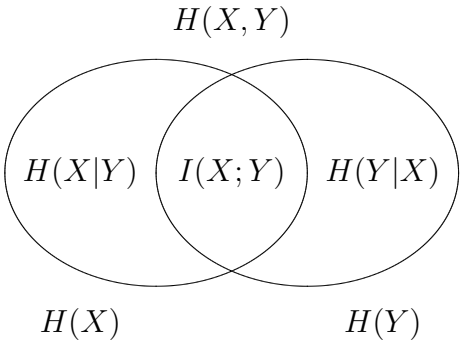
Todistus: Lauseen 2.14 nojalla $H(X,Y) = H(X) + H(Y|X)$ ja lauseen 2.12 nojalla $H(X,Y) \leq H(X) + H(Y)$ yhtäsuuruudella joss X ja Y ovat riippumattomat. Väite seuraa. □

2.3 Informaation mitta

Palataan vielä esimerkin 2.5 kolikkoon. Jos oletamme, että kolikko voi olla joko tavallinen tai sellainen, jossa on molemmilla puolilla klaava. Heitetään kolikkoa kahdesti ja kirjataan ylös montako klaavaa saatiin. Paljonko informaatiota tämän koetuloksen avulla saamme siitä kumman kaltainen kolikko on kyseessä? Jos saatiin vähemmän kuin kaksi klaavaa, kyseessä on varmasti tavallinen kolikko. Jos taas molemmilla kerroilla tuli klaava, kyseessä on todennäköisesti – muttei varmasti – kaksiklaavainen kolikko.

Informaatioteoriassa informaatioksi kutsutaan epävarmuuden vähenemistä. Jos merkitään satunnaismuuttujalla X kolikon valintaa ja Y taas on kahden kolikonheiton klaavamäärä, on $H(X)$ alkuperäinen epävarmuus siitä kumpi kolikko on valittu ja $H(X|Y)$ samainen epävarmuus kolikonheiton jälkeen. Näin muuttujan Y toisesta muuttujasta X antama informaatio, eli epävarmuuden väheneminen kun Y:n arvo paljastuu, on $H(X) - H(X|Y)$. Tähän suureeseen viitataan termillä keskinäisinformaatio (engl. *mutual information*).

Määritelmä 2.16 (keskinäisinformaatio) *Olkoot X ja Y diskreettejä satunnaismuuttujia. Muuttujien keskinäisinformaatio $I(X;Y) = H(X) - H(X|Y)$.*



Kuva 2.1: Entropian ja keskinäisinformaation suhteet

Informaation mittana toimiva keskinäisinformaatio kuvaa sitä, miten paljon informaatiota yksi satunnaismuuttuja toisesta antaa. Tulkinta on intuitiivinen kun ajattelee määritelmän merkitystä. Muuttujan X epävarmuudesta vähennetään samaisen muuttujan epävarmuus, joka jää jäljelle, kun Y:n arvo on havainnoitu. Keskinäisinformaatio on avainkäsite kun tarkastellaan viestien lähettämistä häiriöisen tiedonsiirtokanavan kautta. Tähän aiheeseen päästään aikanaan luvussa 4.

Keskinäisinformaatiolta on luontevaa edellyttää ei-negatiivisuutta.

Lause 2.17 $I(X;Y) \geq 0$ *yhtäsuuruudella joss* X *ja* Y *ovat riippumattomat.*

Todistus: Tulos seuraa suoraan lauseesta 2.15, jonka nojalla $H(X|Y) \leq H(X)$. $\square$

Keskinäisinformaatiolla on varsin yllättävä symmetriaominaisuus. Kahden muuttujan keskinäisinformaatiot toistensa suhteen ovat samat.

Lause 2.18 $I(X;Y) = I(Y;X)$

Todistus: Lauseen 2.14 ja sen, että $H(X,Y) = H(Y,X)$, avulla saadaan

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) = H(X) - (H(X,Y) - H(Y)) \\ &= H(Y) - (H(Y,X) - H(X)) = I(Y;X). \quad \square \end{aligned}$$

Myös nämä määritelmät ja tulokset yleistyvät suoraan useampien satunnaismuuttujien tapauksiin. $I(X_1,\ldots,X_n;X_{n+1},\ldots,X_m) = H(X_1,\ldots,X_n) - H(X_1,\ldots,X_n|X_{n+1},\ldots,X_m)$ vastaavin ominaisuuksin.

Kuvan 2.1 Venn-diagrammi pyrkii selventämään entropian ja keskinäisinformaation suhteita.

informaatioteoria	mittateoria
muuttujat X, Y	joukot A, B
mitat H, I	mitta μ
$H(X)$	$\mu(A)$
$H(X,Y)$	$\mu(A \cup B)$
$H(X\|Y)$	$\mu(A \setminus B)$
$I(X;Y)$	$\mu(A \cap B)$

Taulukko 2.1: Informaatioteorian ja mittateorian analogioita

Mitasta puhuminen on, vaikkakin intuitiivisesti oikeutettua, sikäli harhaanjohtavaa, että informaatioteoriassa käytetty entropian käsitteeseen perustuva tarkastelu ei ole suoranaisesti mittateoreettinen. Informaatioteorian ja mittateorian tuloksista voidaan kuitenkin jossain määrin löytää vastaavuuksia [11]. Muutamia niistä on esitetty taulukossa 2.1. Suhde ei kuitenkaan ole aivan triviaali, eikä siihen tässä opinnäytetyössä perehdytä.

2.4 Entropian tulkinnasta

Suora määritelmän mukainen tulkinta on pitää satunnaismuuttujan entropiaa satunnaistapahtuman epävarmuuden odotusarvona yli annetun satunnaismuuttujan jakauman. Tämä on tietysti tulkintana varsin pinnallinen. Se kuitenkin paljastaa epävarmuuden tilastollisen luonteen.

Satunnaismuuttujan 2-kantainen entropia voidaan myös ajatella olevan pienin määrä kyllä-ei-kysymyksiä, joka keskimäärin tarvitaan selvittämään yksi havainto satunnaismuuttujan arvosta. Yleisesti D-kantainen entropia on pienin yhden havainnon selvittämiseen keskimäärin tarvittava määrä vastauksia D-kohtaisiin monivalintakysymyksiin. Luvussa 3.4 todistettavan häiriöttömän koodauslauseen (lause 3.9) olennainen sisältö on se, ettei edellä mainittu kysymysten määrä voi alittaa entropiaa.

Satunnaismuuttujan entropia voidaan tulkita myös riippumattomien saman jakauman omaavien satunnaismuuttujien sarjan asymptoottisen käyttäytymisen kautta käyttäen luonteenomaisia jonoja (engl. *typical sequences*). Tämä on peräisin Shannonilta [17]. Luonteenomaisten jonojen käyttäminen ei ole tämän esityksen kannalta välttämätöntä, joten se on sivuutettu. Luonteenomaisia jonoja käsittelee mm. [2, s. 14] ja [11].

2.5 Informaation käsite

Sana ”informaatio” lienee yksi moniselitteisimmistä ja kuormitetuimmista. Tarkastelun kohteena olevan informaatioteorian rajojen hahmottamiseksi on tarpeen eritellä sanan sisältöä.

Suomen kielessä ei ole täsmällistä vastinetta englannin sanalle ”*information*”. Sana ”tieto” on arkikielessä moniselitteinen. Se voi viitata kontekstista riippuen ainakin englannin kielen käsitteisiin *knowledge*, *information*, *data* ja *fact*. Filosofian tietoteorian (epistemologia) mielessä vastine on *knowledge*. Sanan ”tieto” moniselitteisyyden johdosta joudutaan muiden käsitteiden yhteydessä tavanomaisesti käyttämään lainasanoja ”informaatio”, ”data” ja ”fakta”.

Claude Shannonin perustaman tutkimusalueen nimeksi on vakiintunut ”informaatioteoria”, vaikka Shannon itse kutsui sitä nimellä *”mathematical theory of communication”*. Shannonin käyttämä termi on siinä mielessä täsmällisempi, että informaatioteorian keskeiset kysymykset liittyvät nimenomaan informaation välittämiseen, eivät informaatioon sinänsä.

2.5.1 Informaatio-sanan käyttöyhteydet

Käsitteenä ”informaatio” esiintyy kirjallisuudessa ja yleisessä kielenkäytössä tässä opinnäytetyössä käsiteltävän informaatioteoreettisen merkityksen lisäksi myös lukuisissa muissa merkityksissä. Informaatio on kuitenkin aina kontekstuaalista, se on informaatiota jostain. Lisäksi se on aina riippumatonta sen esittämiseen käytetystä välineestä (engl. *media*), mutta kuitenkin viime kädessä luonteeltaan fysikaalista [6].

Sanan juuret ovat latinan sanassa *informare*, jonka kantana on käsitettä tai ajatusta merkitsevä sana *forma*. Filosofiset juuret löytyvät Platonin idealismin käsitteistä $\varepsilon\acute{\iota}\delta\omega\varsigma$ (eídos) ja $\mu\omega\rho\varphi\acute{\eta}$ (morphé), jotka viittaavat asioiden todelliseen perimmäiseen ideaan. [7]

Sana informaatio esiintyy useissa käyttöyhteyksissä, joiden erot ja suhteet on hyvä hahmottaa. Kyseessä voi olla [6, 7] informaation

hahmo/muoto (engl. *pattern/form of* $\sim$) informaation esiintymän tai osan abstrakti sisältö (konkreettisen esiintymän vastakohtana), samanhahmoisten esiintymien sanotaan olevan toistensa kopioita

esiintymä (engl. *instance of* $\sim$) informaation hahmon nimenomainen esiintymä, eri esiintymillä voi olla sama hahmo

virta (engl. *stream of* $\sim$) määrittelemättömän kokoinen informaation esiintymä, jolla on ulottuvuutta myös ajassa

kantaja (engl. *holder of* $\sim$) esiintymä, jonka muoto voi muuttua ja joka siten voi toimia informaation esiintymän tallennuspaikkana

osa (engl. *piece of* $\sim$) tietty ”fakta”, informaation esiintymän osajoukko

määrä (engl. *amount of* $\sim$) mitta, joka kuvaa informaation esiintymän, osan tai hahmon koon

Tässä opinnäytetyössä on kyse nimenomaan viimeksi mainitusta. Frank [7] mainitsee lisäksi ajatuksen (engl. *nugget of* $\sim$) ja informaation aaveen (engl. *wraith of* $\sim$), jotka eivät kuitenkaan ole yleisesti käytössä.

Informaation yhteydessä käytetään myös seuraavia käsitteitä [6, 7]

symboli ja viesti (engl. *symbol* ja engl. *message*) informaation hahmo, esiintymä tai ilmentymä, joka on luotu välittämään tietty semanttinen sisältö, viesti koostuu usein kokoelmasta symboleita

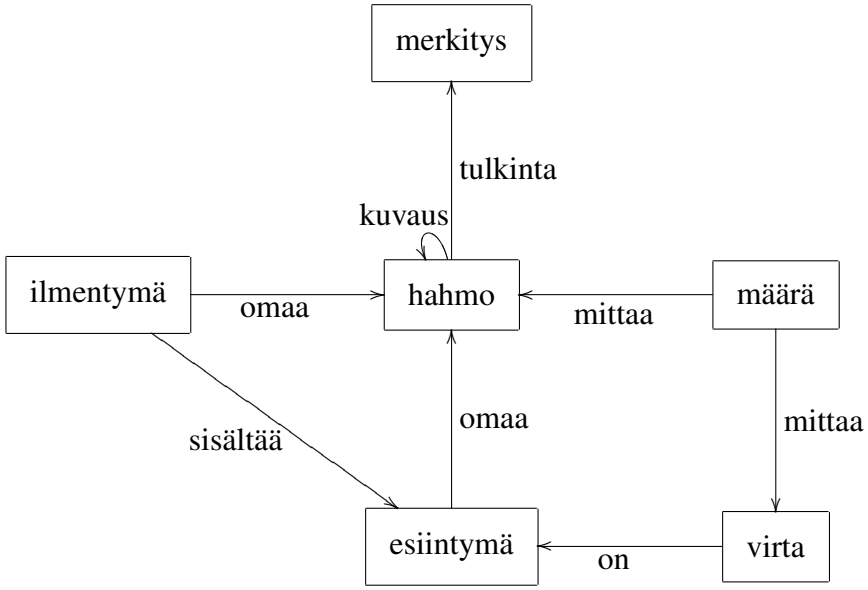
ilmentymä/säilö (engl. *embodiment/container of* $\sim$) olio (esine tai asia), jonka olennainen ydin (engl. *essence*) informaation esiintymä on

kuvaus (engl. *representation of* $\sim$) hahmo, joka yksikäsitteisesti kuvaa jonkin toisen hahmon tai tapahtuma, jossa hahmo kuvaillaan toisella hahmolla

tulkinta (engl. *interpretation of* $\sim$) symbolin tai viestin yhdistäminen sen välittämään semanttiseen sisältöön

aihe (engl. *subject of* $\sim$) olio (abstrakti olio tai konkreettinen esine), jonka informaation esiintymä tai osa kuvailee tai yksilöi

Sanan informaatio käyttöyhteyksien ja niihin liittyvien käsitteiden keskinäisiä yhteyksiä hahmottaa kuvan 2.2 käsitekartta. Esimerkiksi bittijono ”00001001” on informaation *hahmo*. Tämän hahmon *esiintymä* voi olla vaikkapa samainen bittijono tietokoneen prosessorin rekisterissä. Tässä tapauksessa prosessori fysikaalisena järjestelmänä on tämän esiintymän *ilmentymä*. Tuossa rekisterissä voi eri aikoina olla tallennettuna eri esiintymä, joten rekisteri toimii *kantaja*na. Mainittu rekisteriin tallennettu bittijono *kuvaa* lukua 9, joka sekin on hahmo ja myös *symboli*. Tämän nimenomaisen esiintymän *tulkinta* voisi olla vaikka se, että Tero Tiluksen lukion päättötodistuksen liikunnan arvosana on 9. Esiintymän aihe on tässä tapauksessa Tero Tiluksen lukion päättötodistuksen liikunnan arvosana. Rekisterin sisältämän informaation *määrä* voisi tavanomaisessa kotitietokoneessa olla $f_2(2^{32}) = \log_2 2^{32} = 32$ bittiä (engl. *bit,* ***bi****nary digi****t***). Rekisteriin tallennetun esiintymän (tai tarkalleen ottaen sen



Kuva 2.2: Informaation käsitekartta

hahmon) kuvaama hahmo, 9, sitä vastoin voi sisältää vain $f_2(7) = \log_2 7 \approx 2,807354922$ bittiä informaatiota (koska arvosanoja $4, \ldots, 10$ on 7 kpl). Funktio f_b löytyy määritelmästä 2.1 sivulta 5.

2.5.2 Informaatio eri tieteenaloilla

Shannonin informaatioteorian erityinen vahvuus aiempiin informaation välitystä abstrahoineisiin kehittelyihin nähden on lähtökohta, jossa informaatiolähteitä ja kanavia käsitellään tilastollisesti. Tämä lähtökohta on paradoksaalisesti myös sen heikkous. Monissa yhteyksissä (mm. laskettavuuden teoriassa) on tarpeen tarkastella tiettyjen yksittäisten olioiden satunnaisuuden, mutkikkuuden tai informaatiosisällön määrää. Tähän Shannonin tarkastelu ei anna mahdollisuutta.

Yksittäisiin merkkijonoihin (tai muihin tietorakenteisiin) sisältyvään informaatioon liittyvää tutkimusta kutsutaan algoritmiseksi informaatioteoriaksi. Siinä keskeinen käsite – informaation mitta – on ns. Kolmogorov-kompleksisuus, johon saatetaan eri yhteyksissä viitata myös termeillä algoritminen informaatio/entropia/satunnaisuus, Kolmogorov-Chaitin-kompleksisuus tai pienin ohjelmapituus. Algoritmisen informaatioteorian mielessä informaatio merkitsee kompressoitumattomuutta. Merkkijonossa on informaatiota täsmälleen lyhimmän alkuperäisen merkkijonon generoivan Turingin koneen ohjelman pituuden verran.

Informaatioteoria ja algoritminen informaatioteoria eivät kummatkaan puutu semantiikkaan eivätkä informaation arvoon. Paksussa tietosanakirjassa on siis sekä algoritmista informaa-

tiota, että entropiaa merkittävästi vähemmän kuin saman kokoisessa täysin satunnaisesti tuotetussa merkkijonossa, vaikka tietosanakirja epäilemättä on olennaisesti hyödyllisempi. Entropian ja algoritmisen informaation eron tuovat parhaiten esiin ns. näennäissatunnaiset (engl. *pseudorandom*) prosessit. Katkelmassa piin desimaaleja on entropiaa osapuilleen katkelman pituuden verran, mutta algoritmista informaatiota hyvin pieni olennaisesti jonon pituudesta riippumaton määrä. Piin desimaaleja kun voidaan tuottaa mielivaltaisen pitkä jono varsin yksinkertaisella algoritmilla. Kolmogorov-kompleksisuus on siinä mielessä epäkäytännöllinen mitta, että se ei yleisessä tapauksessa ole laskettavissa.

Algoritmisen informaatioteorian juuret ovat Shannonin klassikkoartikkelissa. Neuvostoliittolainen matemaatikko Andrei Kolmogorov kiinnostui Shannonin tuloksista. Hän onnistui järjestämään aiheen tiimoilta seminaarin vuonna 1954, vaikka Shannonin työ laskettiin kuuluvaksi "kybernetiikan" alalle, jota pidettiin Neuvostoliitossa tuohon aikaan pseudotieteenä. Tuo seminaari ja Solomonoffin ja Chaitinin itsenäiset julkaisut samoihin aikoihin käynnistivät tämän tutkimusalan. [9]

Tilastotieteessä informaation käsite tulee vastaan Fisher-informaation yhteydessä. Fisher-informaatio $\mathcal{I}(\theta)$ on tilastotieteellisin termein ilmaistuna pistelukufunktion varianssi tai vaihtoehtoisesti satunnaisprosessin logaritmin toisen derivaatan odotusarvo.

$$\mathcal{I}(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right],$$

jossa X on satunnaismuuttuja, jonka jakauma riippuu parametrista θ ja f on todennäköisyysjakauman tiheysfunktio. Fisher-informaatio kertoo kuinka paljon havainnoitavissa oleva satunnaismuuttuja X sisältää informaatiota tuntemattomasta parametrista θ. Fisher-informaatiota käytetään esimerkiksi vertailtaessa eri tapoja havainnoida satunnaisprosessia.

Todennäköisyysteoriassa (ja osin myös informaatioteoriassa) käytetyllä Kullback-Leibler-divergenssillä on yhteys informaatioteoreettiseen entropiaan. Kullback-Leibler-divergenssi

$$D_{KL}(P||Q) = \sum_i p_i \log \frac{p_i}{q_i}$$

on eräänlainen satunnaismuuttujien P ja Q jakaumien etäisyyden mitta. Kyseessä ei kuitenkaan ole aito metriikka, koska se ei ole symmetrinen eikä toteuta kolmioepäyhtälöä. Entropian ja divergenssin yhteys on jo lausekkeen perusteella ilmeinen. Divergenssiä kutsutaankin joskus informaatiosaannoksi (engl. *information gain*) käytettäessä P:tä Q:n sijaan tai suhteelliseksi entropiaksi (engl. *relative entropy*) käytettäessä Q:tä P:n sijaan. Jakaumien korvaamiseen liittyvä lause 2.8 (s. 10) voidaan itseasiassa muotoilla Kullback-Leibler-divergenssin kautta seuraavasti: $D_{KL}(P||Q) \geq 0$ yhtäsuuruudella joss $P = Q$.

Informaatioteoreettinen entropian käsite on sukua termodynaamiselle entropialle, mutta näitä ei kuitenkaan pidä sekoittaa toisiinsa. Shannonin määritelmä entropialle yhtyy vakiota vaille termodynamiikassa (erityisesti tilastollisessa mekaniikassa) käytettyyn ns. Gibbsin entropiaan.

$$S = -k_B \sum_i p_i \ln p_i$$

kertoimena oleva Boltzmannin vakiokin on lähinnä historian painolastia. Se takaa, että tilastollinen entropia yhtyy termodynamiikan klassiseen Clausiuksen entropiaan ja Boltzmannin entropiaan

$$S_B = k_B \Omega$$

kun systeemin kaikkien mikrotilojen (joiden määrä Ω on) todennäköisyydet ovat yhtä suuret.

Termodynaamisen ja informaatioteoreettisen entropian suhde on yleisellä tasolla suoraviivainen. Gibbsin entropian määrä on sama kuin se Shannonin informaation määrä, joka tarvitaan kuvaamaan systeemin mikrotila kun sen makrotila on tiedossa.

Termodynamiikan entropiassa käytetään luonnollista logaritmia. Informaatioteorian entropiassa käytetty kanta vaihtelee, joskin tavallisimmin käytetään kantalukua 2.

Termodynamiikan entropian yläkäsite on fysikaalinen informaatio, tai vain lyhyesti informaatio, koska kaikki käsiteltävissä oleva informaatio on fysikaalista. Systeemin fysikaalinen informaatio on termodynaamisen entropian maksimi, jonka antaa suoraan Boltzmannin entropian lauseke. Entropia on se osuus systeemin fysikaalisesta informaatiosta, jota (määrätty havaitsija) ei tunne. Loppuosaan viitataan yksinkertaisesti tunnettuna informaationa. Hämmästyttävä yhteys fysiikkaan on se, että informaatiolle voidaan antaa fysikaalinen yksikkö, energia lämpötilayksikköä kohden. Tämä johtuu siitä, että lämpötila voidaan ilmaista energiamääränä, joka tarvitaan kasvattamaan systeemin mahdollisten tilojen määrän logaritmia tietyllä vakiokertoimella (joka on itseasiassa sama kuin käytetyn informaation yksikön kantaluku). Esimerkiksi yksi bitti (tilamäärän logaritmin tuplaus) on $9,57 \times 10^{-24}$ Joulea Kelviniä kohden. [6]

Entropia ja tunnettu informaatio ovat fysikaalisen informaation eri muotoja. Fysikaalinen informaatio – kuten energiakin – säilyy suljetussa systeemissä. Informaatiota ei siis voi tuhota. Ainut tapa hankkiutua eroon tunnetusta informaatiosta on muuttaa se entropiaksi. Eräs hämmästyttävä seuraus tästä on se, että transistorien muodostaman loogisen JA-portin yhdessä operaatiossa vapautuvan termisen energian (yksi bitti entropiaa) määrän teoreettinen alaraja, ns. von Neumann-Landauer-raja on suurempi kuin EI-portin. Edellinen ”unohtaa” operaatiossa yhden bitin ja jälkimmäinen ei. [6]

Kvanttimekaniikan ja informaatioteorian keskinäisten yhteyksien tutkimus on vielä suhteellisen tuore aihe. Shannon tekee informaation käsitteensä määrittelyssä kaksi lausumatonta

oletusta, joiden yhteensopivuus kvanttimekaniikan kanssa ei ole ilmeistä eikä ongelmatonta.

- oletus, että on olemassa asian ”havaittavissa oleva tila” ennen havainnointia
- oletus kommutatiivisuudesta, eli että tieto tilasta ei riipu tehtyjen havaintojen keskinäisestä järjestyksestä

Informaatioteoreettisen entropian käsitteen yhteensopivuudesta kvanttimekaniikan kanssa keskustellaan edelleen [3]. Termodynaamisen entropian käsitteen laajensi kvanttimekaniikkaan John von Neumann yleistämällä todennäköisyysjakauman kompleksiarvoiseksi kvanttitilojen tiheysmatriisiksi. Kvanttimekaanista entropiaa kutsutaankin von Neumann -entropiaksi.

3 Koodaus

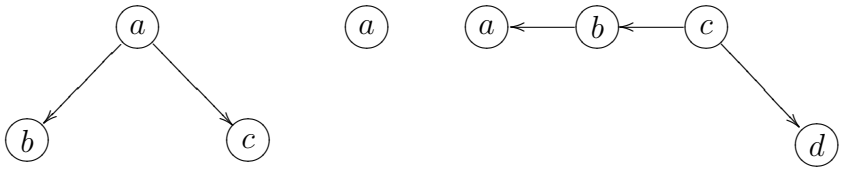
Sovelletaan seuraavaksi luvussa 2.1 esiteltyä entropian käsitettä häiriöttömän kanavan kautta välitettävien viestien mahdollisimman tehokkaaseen koodaukseen. Virhealttiutta käsitellään vasta luvuissa 4 ja 5, jolloin myös kanavan käsite määritellään tarkemmin. Nyt pyritään vain välittämään mahdollisimman suuri määrä viestejä annetun symbolimäärän puitteissa (eli olennaisesti annetussa ajassa). Tarkastelua varten on ensin tarpeen tutustua lyhyesti merkkijonojen ja puiden käsitteisiin.

3.1 Merkkijonot ja puut

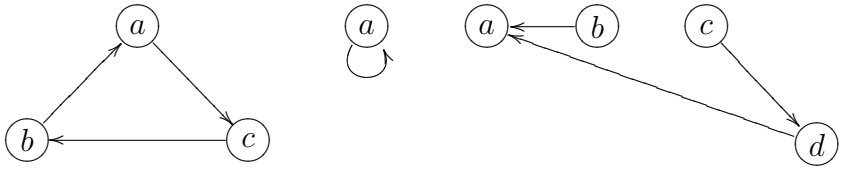
Merkkijono on jono $\alpha_1\alpha_2\ldots$ aakkostoksi kutsutun joukon A alkioita, eli merkkejä, $\alpha_i \in A$. Jos halutaan täsmentää merkkijonon olevan muodostettu tietyn aakkoston A merkeistä, voidaan merkkijonoa kutsua A-jonoksi.

Yhdistetty jono (engl. *juxtaposition*) muodostetaan merkkijonoista $\alpha_1\ldots\alpha_m$ ja $\beta_1\ldots\beta_n$ asettamalla ne peräkkäin $\alpha_1\ldots\alpha_m\beta_1\ldots\beta_n$. Merkkijonon $(\alpha_i)^k$ pituus $\mathrm{len}((\alpha_i)^k) = k$ on jonon merkkien määrä. Kaikkien n:n mittaisten merkkijonojen joukkoa merkitään A^n:llä ja kaikkien äärellisen mittaisten merkkijonojen joukkoa A^*:llä.

Puuksi kutsutaan yhtenäistä syklitöntä graafia [16]. Jatkossa tarvitaan ainoastaan juurellista puuta, joten graafiteoriaan ei syvennytä tarkemmin. Juurellinen puu (jatkossa puu viittaa juurelliseen puuhun) muodostuu epätyhjästä joukosta solmuja ja joukosta solmusta toiseen johtavia kaaria. Puulla on täsmälleen yksi juurisolmu ja vähintään yksi lehtisolmu. Kaikkiin solmuihin tulee täsmälleen yksi kaari, paitsi juurisolmuun, johon ei tule yhtään kaarta. Kaikista solmuista lähtee vähintään yksi kaari, paitsi lehtisolmuista, joista ei lähde yhtään kaarta.



Kuva 3.1: Esimerkkejä puista



Kuva 3.2: Esimerkkejä graafeista, jotka eivät ole puita

Kutsutaan puuta D-asteiseksi jos lehtisolmuja lukuun ottamatta jokaisesta solmusta lähtee D kaarta. Puun syvyys on suurin kaarien määrä juuren ja lehtisolmun välillä. Kuvassa 3.1 vasemmalta alkaen ensimmäinen puu on 2-asteinen ja syvyydeltään 1, toinen on 0-asteinen ja syvyydeltään 0. Kolmannen astelukua ei voi määrittää (juuresta c lähtee kaksi kaarta, ja solmusta b vain yksi). Sen syvyys on 2. Kuvassa 3.2 taas on esimerkkejä graafeista, jotka eivät ole puita. Vasemmalta alkaen ensimmäisessä ja toisessa graafissa ei ole juurisolmua eikä lehtisolmuja, toisessa on lisäksi kaari, jonka molemmat päät ovat samassa solmussa. Kolmannessa on kaksi juurisolmua (b ja c) eikä yksi niin kuin pitäisi ja solmuun a johtaa yhden sijasta kaksi kaarta.

3.2 Häiriötön koodaus

Häiriötön kanava voidaan ajatella laitteeksi, joka hyväksyy tietyn joukon mahdollisia syötteitä ja toistaa annetun syötteen toisessa paikassa virheettömästi jollain kiinteällä taajuudella (engl. *rate*). Tämän tarkempaa määrittelyä kanavalle ei tässä yhteydessä tarvitse.

Ei voida olettaa, että välitettäviä viestit ja kanavan hyväksymät syötteet vastaisivat suoraan toisiaan. Jotta viestit voitaisiin välittää kanavaa käyttäen, jokainen viesti on voitava esittää käyttäen ”merkkejä” kanavan hyväksymien syötteiden joukosta, eli ”aakkostosta”. Rajoitetaan mahdollisten viestien määrän äärelliseksi ja valitaan jokaista viestiä kohden yksikäsitteinen jono aakkoston merkkejä, eli ”koodisana”. Näin muodostuu ”koodi”, jolla muunnetut viestit voidaan välittää kanavaa käyttäen.

Esimerkki 3.1 Käsitteistö on arkinen ja myös toimii hyvin intuitiivisesti. Jos valitaan aakkostoksi suomen kielen kirjaimet ja välilyönti {’a’, . . . , ’ö’, ’ ’}, voidaan asettaa suusanallista viestiä ”hei maailma” vastaavaksi koodisanaksi merkkijono ’hei maailma’.

Koska kanava on häiriötön, ei tarvitse pohtia virheenkorjausta, vaan voidaan keskittyä välittämään annettu viesti mahdollisimman lyhyessä ajassa. Kanavan välitystaajuuden ollessa vakio, on koodisanojen lyhentäminen ainut mahdollisuus lyhentää välittämiseen käytettävää

aikaa. Suurilla viestimäärillä merkitsevää on koodisanojen keskimääräinen pituus, joka siten valitaan optimoitavaksi suureeksi.

Formalisoidaan vielä tarvittavat käsitteet. Olkoon $X\colon \Omega \mapsto \{x_1, \ldots, x_M\}$ diskreetti satunnaismuuttuja jakaumalla $p_1, \ldots, p_M$. Muuttujaa havainnoidaan toistuvasti. Nämä havainnot muodostavat jonon viestejä, jotka kanavaa käyttäen välitetään. Äärellinen epätyhjä joukko $A = \{a_1, \ldots, a_D\}$ on aakkosto (engl. *code characters, code alphabet*). A-sana on äärellinen A-jono. Koodisanat (engl. *code word*) muodostetaan valitsemalla jokaiselle viestille x_i yksikäsitteinen sana w_i. Koodisanojen kokoelma $C = \{w_1, \ldots, w_M\}$ on koodi. Koodin sanotaan olevan D-kantainen (engl. *D-ary*) koodi X:lle. Koodin kanta on $D = \#A$. 2-kantaista koodia kutsutaan binääriseksi (engl. *2-ary, binary*). Merkinnällä (u, n) viitataan mielivaltaiseen koodiin, jonka koodisanojen määrä on u ja sanapituus n. Häiriöttömän koodauksen tavoite on minimoida keskimääräinen sanapituus $l^* = \sum_{i=1}^{M} p_i \mathrm{len}(w_i)$.

3.3 Yksikäsitteinen purkautuvuus

On helppo havaita, että koodisanojen valinnalle täytyy asettaa joitain rajoituksia. Mielivaltaiset koodit eivät välttämättä ole lainkaan käyttökelpoisia.

Esimerkki 3.2 Olkoon $A = \{0, 1\}$ ja $X\colon \Omega \mapsto \{x_1, \ldots, x_4\}$. Muodostetaan koodi seuraavasti

$$\begin{array}{ll} x_1 & 0 \\ x_2 & 010 \\ x_3 & 01 \\ x_4 & 10 \end{array}$$

Nyt merkkijono 010 voi muodostua viestijonoista x_2, x_3x_1, x_1x_4. Kyseistä merkkijonoa ei siten voida yksikäsitteisesti purkaa takaisin viesteiksi.

Monitulkintaiset koodit halutaan rajata ulos tarkastelusta, koska ne eivät ole asetettujen tavoitteiden mukaisia. Rajauksessa auttaa seuraava määritelmä.

Määritelmä 3.3 (Yksikäsitteisesti purkautuva koodi) *Koodi on* yksikäsitteisesti purkautuva *(engl.* uniquely decipherable*), jatkossa UD-koodi, jos jokainen äärellinen koodisanojen jono voidaan muodostaa koodisanoista vain yhdellä tavalla, tai yhtäpitävästi jokainen äärellinen koodisanojen jono vastaa vain yhtä viestijonoa.*

Annetun koodin yksikäsitteinen purkautuvuus on mahdollista tarkistaa suhteellisen yksinkertaisella algoritmilla, jonka esittelivät Sardinas ja Patterson [15].

Eräs tapa varmistaa yksikäsitteinen purkautuvuus koodia konstruoitaessa, on valita koodisanat siten, että yksikään niistä ei ole toisen etuliite. Toisin sanoen ei ole olemassa koodisanaa w_i, joka voitaisiin muodostaa yhdistettynä jonona w_jC, jossa C on jokin epätyhjä merkkijono.

Määritelmä 3.4 (Viiveetön koodi) *Olkoon aakkosto A ja koodi $\{w_1, \ldots, w_M\}$. Koodi on* viiveetön *(engl.* instantaneous*), jos ei ole olemassa koodisanoja w_i ja w_j ($i \neq j$) ja epätyhjää merkkijonoa $C \in A^*$ siten, että $w_i = w_jC$.*

Termi viiveetön (*instantaneous*) on peräisin Abramsonilta [1] ja viittaa siihen, että viestit voidaan purkaa välittömästi niiden saavuttua. Jokaisen vastaanotetun merkin jälkeen tutkitaan vain, muodostaako jo vastaanotettu merkkijono koodisanan. Ellei muodosta, odotetaan seuraavaa merkkiä. Jos muodostaa, tiedetään jonon vastaavan kyseistä koodisanaa, koska kyseinen koodisana ei viiveettömyysominaisuuden nojalla ole minkään toisen koodisanan etuliite.

On osoitettu [5], että edellä mainittu UD-testi antaa lisäksi koodin viiveen. Samaa testiä voidaan siten käyttää myös viiveettömien koodien tunnistamiseen. Ne tosin on helppo tunnistaa myös suoraan määritelmän perusteella.

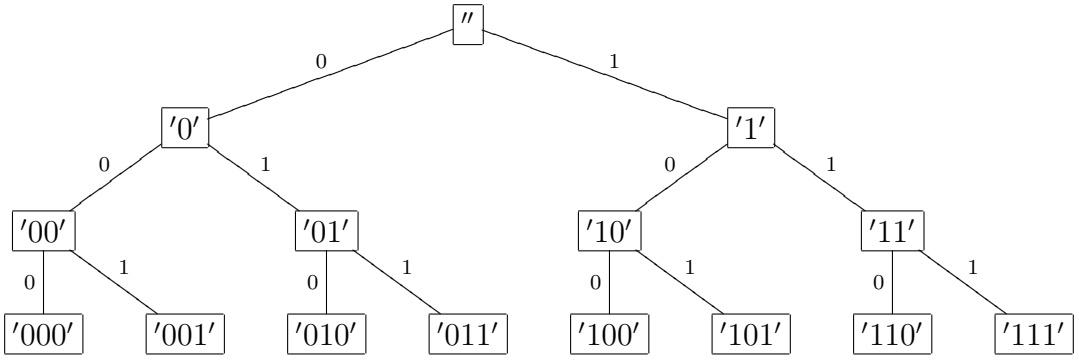
Esimerkki 3.5 Sähkötyksessä käytettävä Morse-koodi on 3-kantainen (engl. *3-ary, ternary*), $A = \{$'·', '–', ' '$\}$, yksikäsitteisesti purkautuva ja viiveetön koodi. Viiveettömyyden havaitsee helposti siitä, että tauko voi esiintyä ainoastaan koodisanan viimeisenä merkkinä, josta seuraa suoraan, etteivät koodisanat voi olla toistensa etuliitteitä. UD seuraa viiveettömyydestä, kuten seuraavaksi huomataan.

Antiteesillä nähdään helposti, että kaikki viiveettömät koodit ovat UD-koodeja.

Lause 3.6 *Olkoon C mielivaltainen koodi. Jos C on viiveetön, niin C on UD-koodi.*

Todistus: Tehdään antiteesi: C ei ole UD-koodi. Nyt C ei ole purettavissa ja erityisesti se ei ole purettavissa välittömästi viestin saavuttua. Se ei siten voi olla viiveetön. Näin ollen viiveetön koodi on aina myös UD-koodi. □

UD-koodi sen sijaan ei välttämättä ole viiveetön (esim. koodi $\{w_1, w_2\} = \{0, 01\}$, jossa välittömästi w_1:n vastaanottamisen jälkeen ei vielä voi varmasti tietää kumpi koodisanoista on kyseessä).



Kuva 3.3: 3 tasoa syvä puu aakkostolle $A = \{0, 1\}$ (2-asteinen)

Viiveetön koodi on yleistä UD-koodia suoraviivaisempi purettava. Lisäksi keskimääräisen sanapituuden minimoinnissa riittää tarkastella viiveettömiä koodeja (ks. [2, s. 40]).

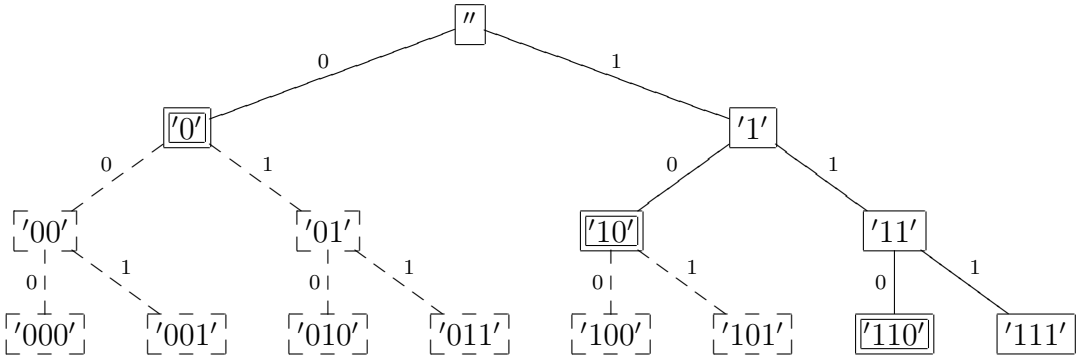
Lause 3.7 (Kraftin epäyhtälö) *Viiveetön koodi aakkostolle A on olemassa joss koodisanojen pituudet $l_i = \text{len}(w_i)$ toteuttavat ehdon $\sum_{i=1}^{M} D^{-l_i} \leq 1$, missä $D = \#A$.*

Todistus: Käytämme todistuksessa D-asteista k tasoa syvää puuta. Puun kaarten ajatellaan vastaavan aakkoston merkkejä. Jokaisesta solmusta lukuun ottamatta lehtisolmuja lähtee yksi kutakin aakkoston merkkiä vastaava kaari. Solmujen ajatellaan vastaavan merkkijonoja, jotka muodostuvat merkeistä, joita vastaavien kaarten kautta juuresta alkava reitti kyseiseen solmuun kulkee. Näin konstruoitu puu on havainnollistettu kuvassa 3.3.

Todistetaan ensin ”vain jos”-suunta. Olkoon C viiveetön D-kantainen koodi sanapituuksilla l_i. Aakkostoksi voidaan olettaa $A = \{0, 1, \ldots, D-1\}$. Voidaan myös olettaa sanapituuksien olevan järjestyksessä, $i < j \Rightarrow l_i \leq l_j$. Koodisanojen voidaan ajatella vastaavan solmuja D-asteisessa ja l_M tasoa syvässä puussa. Viiveettömyydestä seuraa, ettei yksikään koodisana voi sijaita alipuussa, jonka juurena on koodisana, joten l_k mittainen koodisana rajaa mahdottomina pois käytöstä maksimissaan $D^{l_M - l_k}$ potentiaalista koodisanaa (maksimi saavutetaan kun kaikki rajatut ovat lehtisolmuja). Kuvan 3.4 puuhun on kaksoiskehyksin merkitty koodin $C = \{0, 10, 110\}$ koodisanoja vastaavat solmut. Koodisanojen pois rajaamat solmut on merkitty katkoviivoin.

Annetun koodin pois rajaamien sanojen määrä on $\sum_{i=1}^{M} D^{l_M - l_i}$. Se ei voi ylittää lehtisolmujen kokonaismäärää, D^{l_M}, eli $\sum_{i=1}^{M} D^{l_M - l_i} \leq D^{l_M} \Rightarrow \sum_{i=1}^{M} D^{-l_i} \leq 1$.

Kääntäen (”jos”-suunta) oletetaan $l_1, l_2, \ldots, l_M \in \mathbb{Z}_+$, jotka on järjestetty, $i < j \Rightarrow l_i \leq l_j$ ja joille pätee $\sum_{i=1}^{M} D^{-l_i} \leq 1$. Haluttu viiveetön koodi voidaan nyt konstruoida valitsemalla aluksi D-asteisesta ja l_M tasoa syvästä puusta mielivaltainen tason l_1 solmu. Se rajaa pois $D^{l_M - l_1}$ lehtisolmua. Koska $\sum_{i=1}^{M} D^{-l_i} \leq 1 \Rightarrow \sum_{i=1}^{M} D^{l_M - l_i} \leq D^{l_M}$, pätee $D^{l_M - l_1} < D^{l_M}$,



Kuva 3.4: Koodia $C = \{0, 10, 110\}$ vastaava puu

kun $M > 1$, eli ainakin yksi lehtisolmu jää jäljelle, joten voimme jatkaa valitsemalla mielivaltaisen tason l_2 solmun. Koska l_N-mittaisen sanan valinnan jälkeen pois rajattujen lehtisolmujen määrä $\sum_{i=1}^{N} D^{l_M - l_i} < D^{l_M}$, aina kun $N < M$, voidaan prosessia jatkaa kunnes kaikki M koodisanaa on valittu. □

Merkitään jatkossa $\omega_C(l)$:llä l-mittaisten koodisanojen määrää koodissa C. Jos on kontekstista selvää minkä koodin suhteen sanamääräfunktion arvo määräytyy, voidaan kirjoittaa yksinkertaisesti $\omega(l)$. Merkitään lisäksi $l_{\max} = \max_{i \in [1,M]} l_i$. Nyt sanamääräfunktiota ω apuna käyttäen voidaan indeksoida uudelleen summaus

$$\begin{aligned} \sum_{i=1}^{M} D^{-l_i} &= \underbrace{D^{-1} + \cdots + D^{-1}}_{\omega(1) \text{ kpl}} + \underbrace{D^{-2} + \cdots + D^{-2}}_{\omega(2) \text{ kpl}} + \cdots + \underbrace{D^{-l_{\max}} + \cdots + D^{-l_{\max}}}_{\omega(l_{\max}) \text{ kpl}} \\ &= \sum_{j=1}^{l_{\max}} \omega(j) D^{-j}. \end{aligned} \tag{3.1}$$

Yleistetään seuraavaksi edellä todistettu tulos myös UD-koodeille.

Lause 3.8 *UD-koodi aakkostolle A on olemassa joss koodisanojen pituudet $l_i = \text{len}(w_i)$ toteuttavat ehdon $\sum_{i=1}^{M} D^{-l_i} \leq 1$, missä $D = \#A$.*

Todistus: Summaehdon $\sum_{i=1}^{M} D^{-l_i}$ riittävyys UD-koodin olemassaololle (eli "jos"-suunta) seuraa suoraan lauseista 3.7 ja 3.6.

Korotetaan yhtälön 3.1 oikea puoli potenssiin n. Saatavan summan termit ovat aukilaskettuna muotoa $\omega_{i_1} D^{-i_1} \omega_{i_2} D^{-i_2} \cdots \omega_{i_n} D^{-i_n}$, jossa $1 \leq i_k \leq l_{\max}$ kaikille k, eli

$$n \leq \sum_{j=1}^{n} i_j \leq n l_{\max},$$

joten

$$\left(\sum_{j=1}^{l_{\max}} \omega(j) D^{-j}\right)^n = \sum_{k=n}^{nl_{\max}} N_k D^{-k}, \tag{3.2}$$

jossa N_k on jokin kokonaisluku.

On mahdollista muodostaa yhteensä $\sum_{i_1+\cdots+i_n=k} \omega_{i_1}\omega_{i_2}\cdots\omega_{i_n}$ merkkijonoa, jotka alkavat i_1-mittaisella koodisanalla, jatkuvat i_2-mittaisella jne., päättyen i_n-mittaiseen ja joiden koodattu pituus on k. Huomataan, että

$$N_k = \sum_{i_1+\cdots+i_n=k} \omega_{i_1}\omega_{i_2}\cdots\omega_{i_n} \tag{3.3}$$

UD-ominaisuuden nojalla jokainen näistä merkkijonoista vastaa täsmälleen yhtä viestijonoa. Näin N_k ei voi ylittää kaikkien k:n mittaisten jonojen määrää, eli $N_k \leq D^k \Rightarrow N_k D^{-k} \leq D^k D^{-k} = 1$. Tämän ja yhtälöiden 3.1 ja 3.2 nojalla saadaan

$$\begin{aligned}
\left(\sum_{j=1}^{l_{\max}} \omega(j) D^{-j}\right)^n &\leq \sum_{k=n}^{nl_{\max}} 1 = nl_{\max} - n + 1 \leq nl_{\max} \\
\Leftrightarrow \sum_{j=1}^{l_{\max}} \omega(j) D^{-j} &\leq (nl_{\max})^{n^{-1}} \\
\xrightarrow[n\to\infty]{} \sum_{j=1}^{l_{\max}} \omega(j) D^{-j} &\leq 1 \\
\Leftrightarrow \sum_{i=1}^{M} D^{-l_i} &\leq 1
\end{aligned}$$

□

3.4 Koodauslause häiriöttömälle kanavalle

Palautetaan mieleen oletukset ja tavoite. Oletetaan häiriötön kanava. $X : \Omega \mapsto \{x_1, \ldots, x_M\}$ on diskreetti satunnaismuuttuja jakaumalla $p_1, \ldots, p_M$. Halutaan muodostaa kanavan syötteeksi kelpaavaa aakkostoa $A = \{a_1, \ldots, a_D\}$ käyttävä UD-koodi, joka minimoi keskimääräisen sanapituuden

$$l^* = \sum_{i=1}^{M} p_i l_i$$

.

Tavoitteeseen edetään kahdessa vaiheessa. Ensin osoitetaan, että keskimääräisellä sanapituudella on alaraja ja määritetään tuo raja. Toiseksi osoitetaan, että on mahdollista päästä mielivaltaisen lähelle alarajaa. Alarajan osoittaa seuraava lause. Ennen sitä huomioidaan vielä, että missä entropian kannalla ei ole ollut olennaista merkitystä, on merkitty lyhyesti $H(X) = H_b(X)$. Muistetaan tämä ja

$$\frac{H_b(X)}{\log_b D} = \sum_{i=1}^{M} p_i \frac{\log_b p_i^{-1}}{\log_b D} = \sum_{i=1}^{M} p_i \log_D p_i^{-1} = H_D(X).$$

Lause 3.9 (Koodauslause häiriöttömälle kanavalle) *$l^* \geq H_D(X)$ kaikille satunnaismuuttujan X ja häiriöttömän kanavan D-kantaisille UD-koodeille yhtäsuuruudella joss $p_i = D^{-l_i}$.*

Todistus: Muokataan lauseen ehto uuteen muotoon

$$\begin{aligned} l^* &\geq H_D(X) \\ \Leftrightarrow \sum_{i=1}^{M} p_i l_i &\geq \frac{H(X)}{\log D} \\ \Leftrightarrow \sum_{i=1}^{M} p_i \log(D^{-l_i})^{-1} &\geq \sum_{i=1}^{M} p_i \log p_i^{-1}. \end{aligned}$$

Saatu epäyhtälö eroaa lauseen 2.8 epäyhtälöstä vain siten, että termit D^{-l_i} eivät välttämättä summaudu ykköseen. Tästä päästään ohi määrittelemällä lauseen 2.8 $q_i = D^{-l_i} / \sum_{j=1}^{M} D^{-l_j}$. Näin $\sum_{i=1}^{M} q_i = 1$ ja lauseen 2.8 nojalla pätee

$$\sum_{i=1}^{M} p_i \log p_i^{-1} \leq \sum_{i=1}^{M} p_i \log \left(\frac{D^{-l_i}}{\sum_{j=1}^{M} D^{-l_j}} \right)^{-1} \tag{3.4}$$

yhtäsuuruudella joss $p_i = D^{-l_i} / \sum_{j=1}^{M} D^{-l_j}$ kaikille i. Epäyhtälön 3.4 nojalla

$$\begin{aligned} H(X) &\leq \sum_{i=1}^{M} p_i \log(D^{-l_i})^{-1} + \left(\sum_{i=1}^{M} p_i \right) \log \left(\sum_{j=1}^{M} D^{-l_j} \right) \\ H(X) &\leq l^* \log D + \log \left(\sum_{j=1}^{M} D^{-l_j} \right) \end{aligned} \tag{3.5}$$

yhtäsuuruudella joss $p_i = D^{-l_i} / \sum_{j=1}^{M} D^{-l_j}$. Lauseen 3.8 nojalla epäyhtälön 3.5 oikean puolen viimeinen termi on enintään 0, joten $H(X) \leq l^* \log D$.

Jos nyt $p_i = D^{-l_i}$, saadaan

$$H(X) = \sum_{i=1}^{M} p_i \log p_i^{-1} = \sum_{i=1}^{M} p_i l_i \log D = l^* \log D.$$

Pitää vielä todistaa, että jos $H(X) = l^* \log D$, niin $p_i = D^{-l_i}$ kaikille i. Epäyhtälön 3.5 nojalla oletuksesta $H(X) = l^* \log D$ seuraa

$$\log \left(\sum_{j=1}^{M} D^{-l_j} \right) \geq 0.$$

Lauseen 3.8 nojalla taas

$$\log \left(\sum_{j=1}^{M} D^{-l_j} \right) \leq 0,$$

joten

$$\log \left(\sum_{j=1}^{M} D^{-l_j} \right) = 0 \Leftrightarrow \sum_{j=1}^{M} D^{-l_j} = 1$$

Nyt epäyhtälön 3.5 nojalla $p_i = D^{-l_i}$ kaikille i. □

Koodauslause häiriöttömälle kanavalle antaa välittömänä sovelluksena teoreettisen rajan sille, kuinka hyviä häviöttömät datan pakkausmenetelmät voivat olla. Kyseessä on siis eräänlainen informaatioteorian perustavanlaatuinen säilymislaki. Analogia toimii siinäkin mielessä, että myös tätä lakia on aina ympäröinyt energian säilymislakien yhteydestä tuttu lain pätevyyden kiistävien huijarien ja pseudotieteilijöiden parvi [12].

3.5 Optimaaliset koodit

Seuraavaksi pureudumme keskimääräisen sanapituuden minimoinnin toiseen osaan. Osoitetaan, että on mahdollista päästä mielivaltaisen lähelle alarajaa. Tätä tulosta kutsutaan myös käänteiseksi koodauslauseeksi häiriöttömille kanaville.

Keskimääräisen sanapituuden minimoivaa koodia kutsutaan optimaaliseksi (engl. *optimal*). Koodia, joka saavuttaa lauseen 3.9 antaman keskimääräisen sanapituuden alarajan, kutsutaan absoluuttisesti optimaaliseksi (engl. *absolutely optimal*). Mielenkiintoinen havainto on, että tietyin oletuksin geneettinen koodi on viiveetön absoluuttisesti optimaalinen koodi [4].

Yleisessä tapauksessa, kun annettuna on mielivaltainen lähteen jakauma, ei absoluuttisesti optimaalista koodia yleensä löydy, koska lauseen 3.9 nojalla absoluuttisesti optimaaliselle koodille määräytyvät sanapituudet $l_i = \log_D p_i^{-1}$ eivät yleensä ole kokonaislukuja. Seuraavaksi paras vaihtoehto on valita koodisanat sanapituuksin $l_i = \lceil \log_D p_i^{-1} \rceil$. Näin valittua koodia kutsutaan Shannon-Fano-koodiksi. Sanapituuksien valinnasta seuraa, että sen keskimääräinen sanapituus on enintään yhden merkin päässä minimistä. Osoitetaan, että tällainen koodi on aina olemassa.

Lause 3.10 *Olkoon X diskreetti satunnaismuuttuja. On olemassa D-kantainen viiveetön koodi X:lle, jolle $H_D(X) \leq l^* < H_D(X) + 1$.*

Todistus: Valitaan

$$\begin{aligned} l_i &= \lceil \log_D p_i^{-1} \rceil \\ \Leftrightarrow \log_D p_i^{-1} \leq l_i &< \log_D p_i^{-1} + 1 \qquad \text{ja } l_i \in \mathbb{N} \end{aligned} \tag{3.6}$$

ja väitetään, että näillä sanapituuksilla voidaan muodostaa haluttu viiveetön koodi. Riittävä ehto halutun koodin olemassaololle, $\sum_{i=1}^{M} D^{-l_i} \leq 1$, saadaan lauseesta 3.7.

Epäyhtälön 3.6 vasemmasta puolesta seuraa $\log p_i \geq -n_i \log D \Leftrightarrow p_i \geq D^{-l_i}$. Siten $\sum_{i=1}^{M} D^{-l_i} \leq \sum_{i=1}^{M} p_i = 1$. Haluttu koodi on siis olemassa.

Keskimääräisen sanapituuden arvioimiseksi kerrotaan 3.6 puolittain p_i:llä ja summataan kun i käy 1:stä M:ään. Tuloksena on

$$\begin{array}{ccccc} \sum_{i=1}^{M} p_i \log_D p_i^{-1} \leq & \sum_{i=1}^{M} p_i l_i & < \sum_{i=1}^{M} p_i \log_D p_i^{-1} + \sum_{i=1}^{M} p_i \\ H_D(X) \leq & l^* & < H_D(X) + 1 \end{array}$$

□

Keskimääräinen sanapituus saadaan siis aina lähemmäs kuin yhden merkin päähän minimistä. Tulos on parannettavissa, mutta sitä varten tarvitaan kehittyneempi koodaustapa. Otetaan aina s kpl peräkkäisiä riippumattomia X:n havaintoja ja muodostetaan niistä satunnaisvektori $\overline{Y} = (X_1, \ldots, X_s)$. Koodisanat valitaan nyt $\overline{Y}$:n arvoille, joita on M^s kpl. Tätä tapaa kutsutaan lohkokoodaukseksi (engl. *block coding*), jossa lohkopituus on s. Näin muodostuu $\overline{Y}$:lle koodi, joka on olennaisesti myös koodi X:lle. Merkitään jatkossa l_X^* viitattaessa keskimääräiseen sanapituuteen yhtä X:n arvoa kohden.

Lause 3.11 (Käänteinen koodauslause häiriöttömälle kanavalle) *Olkoon X diskreetti satunnaismuuttuja. On olemassa D-kantainen viiveetön koodi, jolle $H_D(X) \leq l_X^* < H_D(X) + \epsilon$ kaikille $\epsilon > 0$.*

Todistus: Käytetään lohkokoodausta lohkopituudella s. Muistetaan lisäksi, että $H(\overline{Y}) = H(X_1, \ldots, X_s) = H(X_1) + \cdots H(X_s) = sH(X)$, koska X_i:t ovat riippumattomia ja jakaumaltaan samoja. Lauseen 3.10 nojalla on olemassa viiveetön koodi $\overline{Y}$:lle siten, että

$$\begin{aligned} H_D(\overline{Y}) \leq \; & l_{\overline{Y}}^* &< H_D(\overline{Y}) + 1 \\ \Leftrightarrow sH_D(X) \leq \; & l_{\overline{Y}}^* &< sH_D(X) + 1 \\ \Leftrightarrow H_D(X) \leq \; & \frac{l_{\overline{Y}}^*}{s} &< H_D(X) + \frac{1}{s} \end{aligned}$$

Huomataan, että $l^*_{\overline{Y}}/s = l^*_X$. Lohkopituutta s kasvattamalla saadaan l^*_X mielivaltaisen lähelle alarajaa $H_D(X)$. $\square$

4 Diskreetti kanava

Shannonin informaatioteorian perusasetelma kulminoituu tiedonsiirtokanava käsitteeseen. Asetelmassa on pyrkimyksenä toistaa yhdestä pisteestä lähetetty viesti toisaalla nopeasti ja virheettömästi.

4.1 Kanavan määritelmä

Viestintä voidaan (semantiikka sivuuttaen) yksinkertaistaa tilanteeseen, jossa valitaan olio (esim. symboli, sähkömagneettinen pulssi tai äänisignaali) mahdollisten lähetettävien viestien joukosta. Kanava on laite, joka reagoi lähetettävään viestiin toistamalla sen eri pisteessä aika-avaruutta. Kanavan virhealttius kuvataan antamalla jokaista lähetettyä viestiä vastaava todennäköisyysjakauma toistetulle viestille. Jakauma voi riippua lähetetyn viestin ohella myös kanavan lähetyksen aikaisesta sisäisestä tilasta. Termejä ”lähetetty” ja ”syöte” käytetään keskenään vaihdannaisina ja vastaavasti myös termejä ”vastaanotettu”, ”toistettu”, ”tuloste” ja ”ulostulo”.

Konstruoidaan seuraavaksi edellistä hahmottelua vastaava kanavan määritelmä. Rajaudutaan tässä luvussa ns. ”diskreettiin” tapaukseen, eli lähetettävät ja vastaanotettavat symbolit valitaan äärellisestä joukosta. Diskreetin kanavan käsittely toimii välivaiheena varsinaiseen päämäärään, eli jatkuvan kanavan ja erityisesti sen kapasiteetin tutkimiseen luvussa 5.

Määritelmä 4.1 (Diskreetti kanava) *Kanava muodostuu kahdesta äärellisestä symbolijoukoista R (vastaanotettavien symbolien joukko, engl.* received*) ja S (engl.* sent*), joukosta tiloja Γ ja joukosta todennäköisyysjakaumia $p_n(\beta_1, \ldots, \beta_n | \alpha_1, \ldots, \alpha_n; \gamma)$, jossa $\alpha_1, \ldots, \alpha_n \in S$, $\beta_1, \ldots, \beta_n \in R$, $\gamma \in \Gamma$ ja $n \in \mathbb{N}$.*

Määritelmässä esiintyvät todennäköisyydet ovat todennäköisyyksiä, joilla kanava toistaa symbolijonon $\beta_1, \ldots, \beta_n$, kun on lähetetty jono $\alpha_1, \ldots, \alpha_n$ ja kanavan tila ennen symbolin α_1 lähettämistä on ollut γ. Kanavan tila voi symboleita lähetettäessä muuttua. Tämä malli olettaa, että toistettavien symbolien jakauma määräytyy kanavan alkutilan ja lähetettävien symbolien jonon perusteella. Kanavan ei kuitenkaan oleteta olevan ”ennustava”, eli toistettavan symbolin β_n jakauma ei riipu myöhemmin lähetettävistä ja toistettavista symboleista α_m, jossa β_m, $m > n$.

Määritelmä 4.2 (Muistiton kanava) *Kanavan sanotaan olevan muistiton, jos*

1. *funktiot p_n eivät riipu tilasta γ, eli voidaan kirjoittaa $p_n(\beta_1, \ldots, \beta_n|\alpha_1, \ldots, \alpha_n)$ ja*
2. *$p_n(\beta_1, \ldots, \beta_n|\alpha_1, \ldots, \alpha_n) = \prod_{i=1}^{n} p_1(\beta_i|\alpha_i)$ kaikille $\alpha_1, \ldots, \alpha_n \in S$, $\beta_1, \ldots, \beta_n \in R$ ja $n \in \mathbb{N}$.*

Muistittomuuden ehdoista jälkimmäinen voidaan korvata kahdella muulla ehdolla. Olkoon $1 \leq k \leq n-1$. Merkitään

$$p_n(\beta_1, \ldots, \beta_{n-k}|\alpha_1, \ldots, \alpha_n) = \sum_{\beta_{n-k+1}, \ldots, \beta_n} p_n(\beta_1, \ldots, \beta_n|\alpha_1, \ldots, \alpha_n).$$

Kyseessä on siis todennäköisyys sille, että kun on lähetetty symbolit $\alpha_1, \ldots, \alpha_n$, ensimmäiset $n-k$ vastaanotettua symbolia ovat $\beta_1, \ldots, \beta_{n-k}$. Merkitään lisäksi

$$p_n(\beta_n|\alpha_1, \ldots, \alpha_n;\ \beta_1, \ldots, \beta_{n-1}) = \frac{p_n(\beta_1, \ldots, \beta_{n-1}, \beta_n|\alpha_1, \ldots, \alpha_n)}{p_n(\beta_1, \ldots, \beta_{n-1}|\alpha_1, \ldots, \alpha_n)}.$$

Tämä on ehdollinen todennäköisyys sille, että n:s vastaanotettu symboli on β_n, kun on lähetetty symbolit $\alpha_1, \ldots, \alpha_n$ ja $n-1$ ensimmäistä vastaanotettua symbolia ovat $\beta_1, \ldots, \beta_{n-1}$.

Lause 4.3 *Funktiot, jotka täyttävät muistittomuuden määritelmän 1. ehdon, täyttävät myös 2. joss seuraavat ehdot täyttyvät*

a. $p_n(\beta_n|\alpha_1, \ldots, \alpha_n;\ \beta_1, \ldots, \beta_{n-1}) = p_1(\beta_n|\alpha_n)$
kaikille $\alpha_1, \ldots, \alpha_n \in S$ ja $\beta_1, \ldots, \beta_n \in R$

b. $p_n(\beta_1, \ldots, \beta_{n-k}|\alpha_1, \ldots, \alpha_n) = p_{n-k}(\beta_1, \ldots, \beta_{n-k}|\alpha_1, \ldots, \alpha_{n-k})$
kaikille $\alpha_1, \ldots, \alpha_n \in S$, $\beta_1, \ldots, \beta_{n-k} \in R$ ja $1 \leq k \leq n-1$.

Todistus: Oletetaan, että määritelmän 4.2 2. ehto pätee. Nyt

$$\begin{aligned}
p_n(\beta_n|\alpha_1, \ldots, \alpha_n;\ \beta_1, \ldots, \beta_{n-1}) &= \frac{p_n(\beta_1, \ldots, \beta_{n-1}, \beta_n|\alpha_1, \ldots, \alpha_n)}{p_n(\beta_1, \ldots, \beta_{n-1}|\alpha_1, \ldots, \alpha_n)} \\
&= \frac{\prod_{k=1}^{n} p_1(\beta_k|\alpha_k)}{\sum_{\beta_n} p_n(\beta_1, \ldots, \beta_n|\alpha_1, \ldots, \alpha_n)} \\
&= \frac{\prod_{k=1}^{n} p_1(\beta_k|\alpha_k)}{\sum_{\beta_n} \prod_{k=1}^{n} p_1(\beta_k|\alpha_k)} \\
&= \frac{\prod_{k=1}^{n} p_1(\beta_k|\alpha_k)}{\prod_{k=1}^{n-1} p_1(\beta_k|\alpha_k) \sum_{\beta_n} p_1(\beta_n|\alpha_n)} \\
&= p_1(\beta_n|\alpha_n),
\end{aligned}$$

joten ehto *a* pätee. Huomataan edellisestä yhtälöstä erityisesti, että

$$\begin{aligned} p_n(\beta_1,\ldots,\beta_{n-1}|\alpha_1,\ldots,\alpha_n) &= \sum_{\beta_n} p_n(\beta_1,\ldots,\beta_n|\alpha_1,\ldots,\alpha_n) \\ &= \sum_{\beta_n}\prod_{k=1}^{n} p_1(\beta_k|\alpha_k) \\ &= \prod_{k=1}^{n-1} p_1(\beta_k|\alpha_k)\sum_{\beta_n} p_1(\beta_n|\alpha_n) \\ &= \prod_{k=1}^{n-1} p_1(\beta_k|\alpha_k) \\ &= p_{n-1}(\beta_1,\ldots,\beta_{n-1}|\alpha_1,\ldots,\alpha_{n-1}) \end{aligned}$$

Ehto *b* saadaan edellisestä induktiolla.

Kääntäen oletetaan nyt, että ehdot *a* ja *b* pätevät. Nyt yhtälön

$$p_n(\beta_1,\ldots,\beta_n|\alpha_1,\ldots,\alpha_n) = p_n(\beta_1,\ldots,\beta_{n-1}|\alpha_1,\ldots,\alpha_n)p_n(\beta_n|\alpha_1,\ldots,\alpha_n;\ \beta_1,\ldots,\beta_{n-1})$$

oikean puolen ensimmäinen termi on ehdon *b* nojalla $p_{n-1}(\beta_1,\ldots,\beta_{n-1}|\alpha_1,\ldots,\alpha_{n-1})$ ja ehdon *a* nojalla toinen termi on $p_1(\beta_n|\alpha_n)$. Induktiolla saadaan määritelmän 4.2 ehto 2. □

Edellisen lauseen ehto *a* takaa kanavan muistittomuuden ja ehto *b* sen, ettei kanavan tarvitse olla ennustaja (eli että ulostulo ei riipu myöhemmistä syötteistä). Muistittomuusominaisuus yksinkertaistaa huomattavasti kanavan määrittelyä. Lauseen 4.3 nojalla muistittoman kanavan määrittelyyn riittää $p_1(\beta|\alpha)$, joten voidaan kirjoittaa $p(\beta|\alpha)$.

Määritelmä 4.4 (Diskreetti muistiton kanava) *Kanava muodostuu kahdesta äärellisestä symbolijoukoista R (engl.* received*) ja S (engl.* sent*), sekä siirtymätodennäköisyysmatriisista (engl.* transition probability matrix*) $T = (p(\beta|\alpha))_{\alpha\beta}$, jossa $\forall\ \alpha \in S$, $\beta \in R$, $p(\beta|\alpha) \geq 0$ ja $\forall\alpha\ \sum_{\beta\in R} p(\beta|\alpha) = 1$.*

Siirtymätodennäköisyysmatriisia kutsutaan myös kanavamatriisiksi (engl. *channel matrix*). Luvussa 3 merkittiin viestejä $x_1,\ldots,x_M$ ja kanavan mahdollisia lähetettäviä symboleita (joukkoa S) eli aakkostoa $a_1,\ldots,a_D$. Jatkossa tätä käytäntöä muutetaan siten, että merkitään aakkostoa $x_1,\ldots,x_M$ ja tarpeen tullen koodaamattomia viestejä $m_1, m_2\ldots$, jolloin merkintää $H(X)$ voidaan luontevasti käyttää viittaamaan kanavan syötteen entropiaan.

Jatkossa riittää tarkastella muistittomia kanavia, koska osoittautuu, ettei kanavan tilaa tarvita käsiteltävänä olevien ominaisuuksien osoittamiseksi.

4.2 Kanavan kapasiteetti ja muita ominaisuuksia

Olkoon (S, R, T) diskreetti muistiton kanava, jossa $S = x_1, \ldots, x_M$, $R = y_1, \ldots, y_N$ ja $T = (p(y_j|x_i))_{ij}$. Olkoon lisäksi $X\colon \Omega \mapsto S$ satunnaismuuttuja jakaumalla $p(x_i)$. Nyt myös kanavan ulostulo on satunnaismuuttuja, merkitään sitä $Y\colon \Omega \mapsto R$. Syötteen jakauman ja kanavan siirtymätodennäköisyysmatriisin avulla voidaan laskea sekä muuttujien yhteisjakauma $p(x_i, y_j) = p(x_i)p(y_j|x_i)$, että ulostulon jakauma $p(y_j) = \sum_{i=1}^{M} p(x_i, y_j) = \sum_{i=1}^{M} p(x_i)p(y_j|x_i)$. Näistä jakaumista voidaan laskea syötteen entropia $H(X)$, tulosteen entropia $H(Y)$, yhteisentropia $H(X, Y)$, sekä ehdolliset entropiat $H(X|Y)$ ja $H(Y|X)$. Nyt on, luvun 2.3 johdattelun pohjalta, luontevaa määritellä *kanavan välittämän informaation määrä* kanavan syötteen ja tulosteen keskinäisinformaatioksi $I(X|Y) = H(X) - H(X|Y)$.

On hyvä huomata, että välitetyn informaation määrä riippuu syötteen entropiasta, joka taas riippuu syötteen jakaumasta. Välitetty informaatio voi olla pieni joko syötteestä tai kanavasta johtuvista syistä. Joko lähetetty symbolivirta ei sisällä informaatiota (syötteen entropia $H(X)$ on pieni) tai kanava hukkaa informaation (muuttujan Y havainnointi ei pienennä X:n entropiaa, eli $H(X|Y)$ ei ole olennaisesti pienempi kuin $H(X)$). Riippuvuus syötteen entropiasta voidaan purkaa etsimällä välitetyn informaation määrälle pienin yläraja syötteen jakauman suhteen. Tätä ylärajaa kutsutaan kanavan kapasiteetiksi.

Määritelmä 4.5 (Kapasiteetti) *Diskreetin muistittoman kanavan* (S, R, T) *kapasiteetti*

$$C = \sup_P I(X; Y),$$

missä $P = \{p\colon S \mapsto [0, 1] | p(x) \geq 0, \sum_{x \in S} p(x) = 1\}$.

Diskreetin muistittoman kanavan tapauksessa pienin yläraja on itseasiassa aito maksimi. Keskinäisinformaatio voidaan esittää syöteaakkoston todennäköisyyksien jatkuvana funktiona ja näiden todennäköisyyksien joukko on suljettu ja rajoitettu $\mathbb{R}^M$:n osajoukko, joten I saavuttaa tuossa joukossa maksiminsa. Jos $C = 0$, on kaikille syötejakaumille $I(X; Y) = 0 \Leftrightarrow H(X|Y) = H(X)$, eli X ja Y ovat riippumattomia.

Kapasiteetin määritelmän merkitys ei toistaiseksi esitetyn perusteella ole vielä nähtävissä. Luvussa 5 todistetaan kaksi kapasiteettia koskevaa merkittävää tulosta, jotka myös oikeuttavat suureen kutsumisen kapasiteetiksi. Karkeasti ottaen nämä tulokset kertovat, että informaatiota on mahdollista välittää kanavaa käyttäen mielivaltaisen pienellä virhealttiudella kunhan välitetyn informaation määrä (aikayksikköä kohden) on alle kapasiteetin. Ja käänteen, luotettava informaationvälitys ei ole mahdollista välitetyn informaation määrän ylittäessä kapasiteetin. Kanavan kapasiteetin määrittäminen yleisessä tapauksessa on vaikeaa.

Kanavan, jolle $H(X|Y) = 0$, sanotaan olevan *häviötön* (engl. *lossless*). Häviöttömän kanavan syöte tiedetään varmasti jos vastaava tuloste on tiedossa, eli virheen mahdollisuutta ei ole. Yhtäpitävästi R voidaan jakaa pistevieraisiin joukkoihin B_i siten, että $P(\{Y \in B_i | X = x_i\}) = 1$. Kanavan sanotaan olevan *deterministinen* jos siirtymätodennäköisyysmatriisissa on vain nollia ja ykkösiä, eli että syöte määrää vastaavan tulosteen yksikäsitteisesti. Kanava on *häiriötön* jos se on häviötön ja deterministinen.

4.3 Viestien vastaanotto

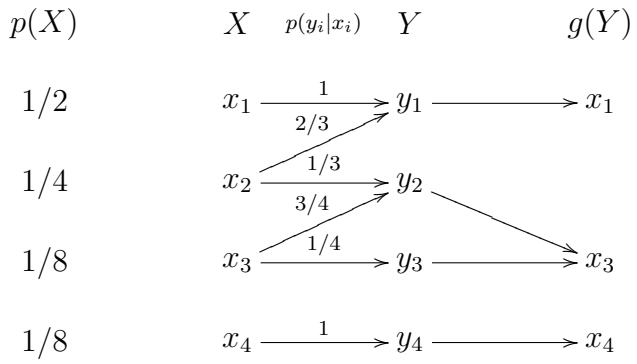
Tavoiteltaessa luotettavaa tiedonsiirtoa käyttäen virhealtista kanavaa, on vastaanotettujen viestien tulkinnalla keskeinen merkitys. Vastaanotetusta viestistä on voitava riittävän luotettavasti päätellä lähetetty. On pystyttävä valitsemaan "paras" tapa tehdä tämä päättely.

Olkoon (S, R, T) diskreetti muistiton kanava, jossa $S = x_1, \ldots, x_M$, $R = y_1, \ldots, y_N$ ja $T = (p(y_j|x_i))_{ij}$. Olkoon lisäksi $X \colon \Omega \mapsto S$ satunnaismuuttuja jakaumalla $p(x_i)$. Tarkastellaan aluksi yhden X:n arvon lähettämistä. Funktiota $g \colon R \mapsto S$ kutsutaan *purkajaksi* (engl. *decoder*) tai purkufunktioksi. Purkaja voidaan määritellä myös R:n pistevieraalla osituksella $B_1, \ldots, B_N$, jolloin $g(y) = x_i \Leftrightarrow y \in B_i$. Koska purkajan "hyvyyden" kriteeriksi valittiin luotettavuus, on paras purkaja sellainen, joka minimoi keskimääräisen virheen. Tällaista purkajaa kutsutaan *ideaalihavaitsijaksi* (engl. *ideal observer*). Merkitään keskimääräistä virhetodennäköisyyttä $p(e)$:llä ja vastaavasti virheettömän tiedonsiirron todennäköisyyttä $p(e') = 1 - p(e)$. Kun vastaanotettu viesti y on tunnettu, on virheettömän tiedonsiirron todennäköisyys sama kuin todennäköisyys sille, että lähetetty viesti $x = g(y)$. Näin voidaan kirjoittaa

$$p(e') = \sum_{j=1}^{N} p(y_j) p(e'|y_j) = \sum_{j=1}^{N} p(y_j) P(\{X = g(y_j)|y_j\}) \tag{4.1}$$

Yhtälössä 4.1 $p(y_j)$ ei riipu purkajasta. Jokaiselle y_j voidaan $g(y_j)$ valita vapaasti. Valitsemalla $g(y_j)$ siten, että se maksimoi $p(g(y_j)|y_j)$:n, se maksimoi samalla $p(e')$:n. Näin määritelty g on etsimämme ideaalihavaitsija.

Ideaalihavaitsijalla on joitakin ikäviä ominaisuuksia. Jos $p(g(y_j)|y_j)$:n maksimeja g:tä varioitaessa on useita, voidaan valita mikä hyvänsä sen vaikuttamatta virhetodennäköisyyteen. Tästä syystä ideaalihavaitsija ei ole yksikäsitteinen. Lisäksi ideaalihavaitsija riippuu syötteen jakaumasta. Jos syöte muuttuu, muuttuu yleensä myös ideaalihavaitsija. On myös tilanteita, joissa ideaalihavaitsija tekee aina virheen. Esimerkki tällaisesta on kuvassa 4.1. Kun lähetetään x_2, voi vastaanotettu olla y_1 tai y_2. Ideaalihavaitsijalle on kuitenkin $g(y_1) = x_1$ (koska $p(x_1|y_1) = 3/4$ ja $p(x_2|y_1) = 1/4$) ja $g(y_2) = x_3$ (koska $p(x_2|y_2) = 8/17$ ja



Kuva 4.1: Esimerkki kanavasta ja purkajasta

$p(x_3|y_2) = 9/17$).

Osoittautuu, että kun kanavien ominaisuuksia tarkastellaan riittävän pitkillä syötejonoilla, riittää virhetodennäköisyyksiä arvioitaessa tarkastella tasaisen jakauman ideaalihavaitsijaa (ks. [2, s. 63]). Tämä olennaisesti poistaa havaitsijan valinnan riippuvuuden syötteen jakaumasta, mutta toinen ongelmista säilyy. Virhetodennäköisyydelle ei voida antaa ylärajaa (vrt. kuvan 4.1 esimerkki tilanteesta, jossa tämä ”yläraja” on 1), jota pienempiä virhetodennäköisyydet kaikkien syötteiden kohdalla olisivat. Pelastajaksi osoittautuu lauseen 3.11 todistuksen yhteydessä esitelty lohkokoodaus. Jos yksittäisten symbolien sijasta koodataan ja lähetetään symbolijonoja $\overline{X} = (X_1, \ldots, X_n)$, joita jatkossa kutsutaan myös satunnaisvektoreiksi, on mahdollista konstruoida koodeja, joiden virhetodennäköisyydellä on yläraja. Konstruointi voidaan tehdä vieläpä niin, että virhe saadaan mielivaltaisen pieneksi.

Nyt voidaan käydä varsinaisen ongelman kimppuun.

5 Jatkuva kanava

Tässä luvussa käsittelen kanavia, jotka ovat jatkuvia kahdessa eri merkityksessä. Amplitudijatkuviksi kutsutaan kanavia, joiden sisäänmenon ja ulostulon symbolijoukot ovat ylinumeroituvia. Aikajatkuviksi taas kutsutaan kanavia, joiden tiedonsiirto on ”jatkuvaa” aika-akselilla. Formaalisti edellinen merkitsee sitä, että koodisanojen joukko on ylinumeroituva ja jälkimmäinen sitä, että signaaliavaruus on funktioavaruus, aika- ja amplitudijatkuvan kanavan tapauksessa reaalinen Hilbert-avaruus $\mathcal{L}_2[a, b]$.

Jatkuvassa tapauksessa käytämme käytännön syistä diskreetistä poiketen logaritmien kantalukuna luonnollista lukua e. Tällä ei ole vaikutusta itse tarkasteluihin, ainoastaan konkreettisista laskuista saatavien tulosten yksikköön.

5.1 Entropian yleistys

Jatkuvien kanavien tarkastelemiseksi on entropian käsite ensin yleistettävä jatkuva-arvoisille satunnaismuuttujille. Pintapuolisesti tarkasteltuna tämä näyttäisi onnistuvan triviaalisti korvaamalla epävarmuuden $H(X)$ määritelmässä 2.4 summa integraalilla. Käy kuitenkin ilmi, että näin saatu $H(X)$ poikkeaa olennaisesti diskreetistä versiosta. Tästä huolimatta triviaali yleistys toimii ja tässä luvussa todistamme sille diskreetin version (ks. luku 2.2) kanssa analogiset tulokset.

Määritelmä 5.1 (Jatkuvan satunnaismuuttujan entropia) *Olkoon $X\colon \Omega \mapsto \mathbb{R}$ jatkuva satunnaismuuttuja todennäköisyysjakaumalla $p(x)$. Määritellään jatkuvan satunnaismuuttujan X b-kantainen* entropia $H_b(X)$ *seuraavasti:*

$$H_b(X) = \int_{-\infty}^{\infty} p(x) \log_b \frac{1}{p(x)} \, \mathrm{d}x$$

Sillä edellytyksellä, että integraali on olemassa.

Jatkossa merkitään $H(X) = H_e(X)$. Diskreetistä tapauksesta poiketen $H(X)$ voi olla mielivaltaisen suuri tai pieni. Tämä nähdään valitsemalla X välille $[0, a]$ tasaisesti jakautuneeksi.

Nyt $H(X) = \log a$, jonka arvojoukko on koko $\mathbb{R}$ kun $0 < a < \infty$. Yhteisentropia, ehdollinen entropia ja keskinäisinformaatio määritellään jatkuvalle satunnaismuuttujalle edellisen kanssa analogisesti.

Määritelmä 5.2 (Jatkuvien satunnaismuuttujien yhteisentropia)

$$H(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{1}{p(x,y)}\,\mathrm{d}x\,\mathrm{d}y$$

Määritelmä 5.3 (Jatkuvien satunnaismuuttujien ehdollinen entropia)

$$H(X|Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{1}{p(y|x)}\,\mathrm{d}x\,\mathrm{d}y$$

Määritelmä 5.4 (Jatkuvien satunnaismuuttujien keskinäisinformaatio)

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

Seuraava lause on diskreetin tapauksen kanssa analogisten funktion H ominaisuuksien todistamisessa hyödyllinen aputulos.

Lause 5.5 *Olkoon $p(x)$ ja $q(x)$ mielivaltaisia tiheysfunktioita. Nyt*

1. *Jos $\int_{-\infty}^{\infty} p(x)\log\frac{1}{q(x)}\,\mathrm{d}x$ on äärellinen, on $\int_{-\infty}^{\infty} p(x)\log\frac{1}{p(x)}\,\mathrm{d}x$ olemassa ja lisäksi*

$$\int_{-\infty}^{\infty} p(x)\log\frac{1}{p(x)}\,\mathrm{d}x \le \int_{-\infty}^{\infty} p(x)\log\frac{1}{q(x)}\,\mathrm{d}x, \tag{5.1}$$

   jossa yhtäsuuruus pätee joss $p(x) = q(x)$ melkein kaikille x (Lebesque-mitan mielessä).

2. *Jos $\int_{-\infty}^{\infty} p(x)\log\frac{1}{p(x)}\,\mathrm{d}x$ on äärellinen, on $\int_{-\infty}^{\infty} p(x)\log\frac{1}{q(x)}\,\mathrm{d}x$ olemassa ja lisäksi epäyhtälö 5.1 pätee.*

Todistus: Molempien osien todistus on samankaltainen. Todistetaan ensimmäinen. Koska $\log b \le b - 1$ ja yhtäsuuruus pätee joss $b = 1$, pätee $p(x)\log(q(x)/p(x)) \le q(x) - p(x)$ yhtäsuuruudella joss $p(x) = q(x)$. Määritellään $q(x)/p(x) = 0$, kun $p(x) = q(x) = 0$, ja $0 \cdot \infty = 0$. Siten

$$\int_{-\infty}^{\infty} p(x)\log\frac{q(x)}{p(x)}\,\mathrm{d}x \le \int_{-\infty}^{\infty} q(x)\,\mathrm{d}x - \int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1 - 1 = 0$$

ja yhtäsuuruus pätee joss $p(x) = q(x)$ melkein kaikille x. Nyt

$$\begin{aligned} p(x) \log \frac{1}{p(x)} &= p(x) \left[\log \frac{1}{p(x)} + \log q(x) - \log q(x) \right] \\ &= p(x) \left[\log \frac{q(x)}{p(x)} - \log q(x) \right] \\ &= p(x) \log \frac{q(x)}{p(x)} - p(x) \log q(x) \end{aligned} \tag{5.2}$$

Huomaa, että ei voi olla olemassa positiivimittaista joukkoa, jossa $q(x) = 0$ ja $p(x) > 0$, koska muutoin $\int_{-\infty}^{\infty} p(x) \log \frac{1}{q(x)} \, \mathrm{d}x = +\infty$. Näin ollen yhtälön 5.2 oikea puoli voi olla muotoa $-\infty + \infty$ vain nollamittaisessa joukossa.

Edellisestä seuraa

$$\int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} \, \mathrm{d}x \leq \int_{-\infty}^{\infty} p(x) \log \frac{1}{q(x)} \, \mathrm{d}x$$

Lopuksi, jos yhtäsuuruus pätee yhtälössä 5.1, yllä olevan yhtälön oikean puolen äärellisyydestä seuraa yhtälön 5.2 nojalla

$$\int_{-\infty}^{\infty} p(x) \log \frac{q(x)}{p(x)} \, \mathrm{d}x = 0,$$

eli että $p(x) = q(x)$ melkein kaikille x. □

Tällä tuloksella on seuraavat välittömät seuraukset

Lause 5.6 *Olkoot X ja Y satunnaismuuttujia, joiden yhteisjakauma on $p(x, y)$. Oletetaan, että $H(X)$ ja $H(Y)$ ovat äärellisiä. Silloin*

1. *$H(X, Y)$ on olemassa ja $H(X, Y) \leq H(X) + H(Y)$ yhtäsuuruudella joss X ja Y ovat riippumattomat.*
2. *$H(X|Y)$ ja $H(Y|X)$ ovat olemassa ja $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$.*
3. *$H(Y|X) \leq H(Y)$ ja $H(X|Y) \leq H(X)$ yhtäsuuruudella joss X ja Y ovat riippumattomat.*
4. *$I(X; Y) = I(Y; X)$ ja $I(X; Y) \geq 0$ ja $I(X; Y) = 0$ joss X ja Y ovat riippumattomat.*

Todistus: Jottei lyhennysmerkintä $p(x)$ johtaisi harhaan, muistetaan, että satunnaismuuttuja, jonka jakauman mukaisesta todennäköisyydestä on kyse, ilmoitetaan implisiittisesti lyhennysmerkinnän muuttujan valinnalla. Siis jäljempänä $p(x) = P\{X = x\}$, eikä esim. $p(x) = P\{Y = x\}$.

Huomataan, että lauseen 5.5 ensimmäisen kohdan nojalla

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{1}{p(x,y)}\,\mathrm{d}x\,\mathrm{d}y \leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{1}{p(x)p(y)}\,\mathrm{d}x\,\mathrm{d}y$$

ja yhtäsuuruus pätee joss $p(x,y) = p(x)p(y)$ melkein kaikille (x,y). Oletuksesta, että $H(X)$ ja $H(Y)$ ovat äärellisiä, seuraa edellisen epäyhtälön oikean puolen äärellisyys, joten lausetta 5.5 voidaan soveltaa. Tämä todistaa ensimmäisen kohdan.

Toisen kohdan osoittamiseksi huomataan ensin, että

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{1}{p(y|x)}\,\mathrm{d}x\,\mathrm{d}y &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{1}{p(x,y)}\,\mathrm{d}x\,\mathrm{d}y \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{1}{p(x)}\,\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

Yhtälön oikean puolen ensimmäinen termi on ensimmäisen kohdan nojalla olemassa ja äärellinen. Toinen termi taas on äärellinen oletuksen nojalla. Siten yhtälön vasen puoli on on olemassa ja $H(Y|X) = H(X,Y) - H(X) \Leftrightarrow H(X,Y) = H(X) + H(Y|X)$. Lisäksi huomataan, että määritelmän nojalla $H(X,Y) = H(Y,X)$.

Kolmas kohta seuraa suoraan edellisistä, koska $H(X,Y) = H(X) + H(Y|X) \leq H(X) + H(Y)$ yhtäsuuruudella joss X ja Y ovat riippumattomat.

Toisesta kohdasta seuraa $H(X) - H(X|Y) = H(Y) - H(Y|X) \Leftrightarrow I(X;Y) = I(Y;X)$. Kolmannesta kohdasta seuraa suoraan $0 \leq H(X) - H(X|Y) = I(X;Y)$ joss X ja Y ovat riippumattomat. Näin myös neljäs kohta on todistettu. □

Seuraavaksi todistamme käänteisen koodauslauseen (lause 5.18) todistuksessa tarpeellisia aputuloksia.

Lause 5.7 *Olkoon X jatkuva satunnaismuuttuja jakaumalla, jonka tiheysfunktio on $p(x)$. Jos X:n varianssi σ^2 on äärellinen, niin $H(X)$ on olemassa ja $H(X) \leq \frac{1}{2}\log(2\pi e\sigma^2)$ yhtäsuuruudella joss $X \sim \mathrm{N}(a,\sigma^2)$, $a \in \mathbb{R}$.*

Todistus: Olkoon $E(X) = \mu$, $\mathrm{var}(X) = \sigma^2$ ja $q(x) = (2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/2\sigma^2}$. Nyt

$$\begin{aligned}\int_{-\infty}^{\infty} p(x)\log\frac{1}{q(x)}\,\mathrm{d}x &= \int_{-\infty}^{\infty} p(x)\left[\frac{1}{2}\log(2\pi\sigma^2) + \frac{(x-\mu)^2}{2\sigma^2}\,\mathrm{d}x\right] \\ &= \frac{1}{2}\log(2\pi\sigma^2) + \frac{\sigma^2}{2\sigma^2} = \frac{1}{2}\log(2\pi e\sigma^2).\end{aligned}$$

Tulos seuraa lauseen 5.5 ensimmäisestä osasta.

Tulemme seuraavassa luvussa tarkastelemaan aikadiskreettiä amplitudijatkuvaa gaussista kanavaa. Koska koodin koodisanojen määrä on äärellinen on näin ollen myös lähetettävä viesti. Edelliset entropiaa koskevat tulokset on tästä syystä yleistettävä tapauksiin, joissa lähetetty viesti on diskreetti ja vastaanotettu viesti (lähetetty viesti + kanavan jatkuva-arvoinen virhe) jatkuva.

Voidaan helposti osoittaa (mm. [2, s. 241]), että seuraavin määritelmin edellä todistetut tulokset yleistyvät äsken kuvailtuihin tilanteisiin, joissa diskreettejä ja jatkuvia satunnaismuuttujia tarkastellaan yhdessä.

Määritelmä 5.8 (Diskreettien ja jatkuvien muuttujien yhdistely) *Oletetaan diskreetti satunnaismuuttuja $X\colon \Omega \mapsto \{x_1, \ldots, x_M\}$ ja satunnaismuuttuja Y, jonka ehdollinen todennäköisyys kun $X = x_i$ on annettu, on $p(y|x_i)$. (Jatkossa kiinnostava tilanne on $Y = X + Z$, jossa Z on jatkuva gaussinen satunnaismuuttuja.) Nyt*

$$p(y) = \sum_{i=1}^{M} p(x_i)p(y|x_i).$$

Ja voidaan määritellä

$$\begin{aligned} H(Y|X = x_i) &= \int_{-\infty}^{\infty} p(y|x_i) \log p(y|x_i)^{-1} \,\mathrm{d}y \\ H(Y|X) &= \sum_{i=1}^{M} p(x_i) H(Y|X = x_i) \\ H(X|Y = y) &= \sum_{i=1}^{M} p(x_i|y) \log p(x_i|y)^{-1} \\ H(X|Y) &= \int_{-\infty}^{\infty} p(y) H(X|Y = y) \,\mathrm{d}y \end{aligned}$$

Lause 5.9 *Olkoon $X\colon \Omega \mapsto \{x_1, \ldots, x_M\}$ diskreetti satunnaismuuttuja ja Y satunnaismuuttuja, jonka ehdollinen todennäköisyys kun $X = x_i$ on annettu, on $p(y|x_i)$. Oletetaan, että $H(Y|X = x_i)$ on äärellinen kaikille $i \in \{1, \ldots, M\}$ ja $H_k(Y)$ on äärellinen kaikille $k \in \{1, \ldots, n\}$. Olkoon kaikille $k \in \{1, \ldots, n\}$ $p_k(\alpha)$ jokin X:n jakauma ($\alpha \in \{x_1, \ldots, x_M\}$). Olkoon $p_0(\alpha)$ näiden jakaumien konveksi lineaarikombinaatio $p_0(\alpha) = \sum_{k=1}^{n} a_k p_k(\alpha)$, missä $a_k \geq 0$ ja $\sum_{k=1}^{n} a_k = 1$. Nyt*

$$I_0(X;Y) \geq \sum_{k=1}^{n} a_k I_k(X;Y)$$

Edellä H_k ja I_k ovat vastaavat H ja I jakaumalla $p_k(\alpha)$.

Todistus: Lasketaan ensin $H_0(X|Y)$

$$H_0(X|Y) = \sum_{i=1}^{M} p_0(x_i) \int_{-\infty}^{\infty} p_0(y|x_i) \log p_0(y|x_i)^{-1} \, \mathrm{d}y.$$

Koska $p_0(y|x_i) = p(y|x_i)$,

$$\begin{aligned} H_0(X|Y) &= \sum_{k=1}^{n} \sum_{i=1}^{M} a_k p_k(x_i) \int_{-\infty}^{\infty} p(y|x_i) \log p(y|x_i)^{-1} \, \mathrm{d}y. \\ &= \sum_{k=1}^{n} a_k H_k(Y|X). \end{aligned} \tag{5.3}$$

Nyt on näytettävä, että

$$\sum_{k=1}^{n} a_k I_k(X;Y) - I_0(X;Y) \leq 0.$$

Yhtälön 5.3 nojalla

$$\sum_{k=1}^{n} a_k I_k(X;Y) - I_0(X;Y) = \sum_{k=1}^{n} a_k H_k(Y) - H_0(Y).$$

Oletuksen nojalla $\int_{-\infty}^{\infty} p_k(y) \log p_k(y)^{-1} \, \mathrm{d}y = H_k(Y)$ on äärellinen, joten lauseen 5.5 toisen kohdan nojalla $\int_{-\infty}^{\infty} p_k(y) \log p_0(y)^{-1} \, \mathrm{d}y$ on olemassa ja

$$\int_{-\infty}^{\infty} p_k(y) \log p_0(y)^{-1} \, \mathrm{d}y \geq \int_{-\infty}^{\infty} p_k(y) \log p_k(y)^{-1} \, \mathrm{d}y.$$

Kertomalla yllä olevan yhtälön a_k:lla ja summaamalla yli k:n ($k \in \{1, \ldots, n\}$), nähdään että $H_0(Y)$ on olemassa ja että $H_0(Y) \geq \sum_{k=1}^{n} a_k H_k(Y)$. □

Tämän luvun tulokset yleistyvät myös satunnaisvektoreille.

5.2 Aikadiskreetti amplitudijatkuva kanava

Amplitudijatkuviksi kutsutaan kanavia, joiden sisäänmenon ja ulostulon symbolijoukot S ja R voivat olla ylinumeroituvia.

Aikadiskreetti amplitudijatkuva kanava formalisoidaan vastaavalla tavalla kuin diskreetti kanava (vrt. määritelmä 4.4). Jatkuvalle kanavalle sisäänmenon ja ulostulon symbolijoukkoina

R ja S on reaaliluvut. Siirtymätodennäköisyysmatriisin tilalla on (Borel-mitallisten) tiheysfunktioiden $\overline{y} \mapsto p(\overline{y}, \overline{x})$, jossa $\overline{y}, \overline{x} \in \mathbb{R}^n$ joukko, jonka alkiot määräävät ulostulojen $\overline{y}$ todennäköisyydet jokaiselle syötteelle $\overline{x}$. Tällaisen kanavan (u,n)-koodi on n-vektorien (koodisanojen) joukko $\{\overline{w}_1, \ldots, \overline{w}_u\}$ yhdessä purkufunktion määrittelevän $\mathbb{R}^n$:n pistevieraan osituksen $\{B_1, \ldots, B_u\}$ (B_i Borel-joukkoja) kanssa. Kanavan (u,n,λ)-koodi on (u,n)-koodi, jonka virheherkkyys on enintään λ, eli

$$P\{\overline{y} \in B_i \mid \overline{x} = \overline{w}_i\} \geq 1 - \lambda$$

Useimmissa viestinnän sovelluksissa vastaanotetussa signaalissa on häiriöitä. Tämän ”häiriösignaalin” olettaminen normaalijakautuneeksi on fysikaalisista syistä perusteltua. Oletus pätee lähes kaikissa käytännön sovelluksissa ja on lisäksi (kuten jatkossa tulemme huomaamaan) suotavaa formalismin yksinkertaistamiseksi.

Määritelmä 5.10 (Aikadiskreetti gaussinen kanava) *Aikadiskreetti gaussinen kanava on aikadiskreetti ja amplitudijatkuva kanava, jonka tiheysfunktiot ovat muotoa*

$$p(\overline{y}|\overline{x}) = (2\pi N)^{-\frac{1}{2}n} \exp(-\sum_{j=1}^{n} \frac{(y_j - x_j)^2}{2N}) \tag{5.4}$$

jossa N on normaalijakaumaa noudattavan häiriösignaalin varianssi.

Näemme, että ellei lisärajoitteita aseteta, saadaan kanavan kapasiteetti mielivaltaisen suureksi samalla kun virheherkkyys voidaan pitää mielivaltaisen pienenä. Purkufunktion määräävän osituksen joukkojen B_i määrää voidaan lisätä rajatta, jolloin kapasiteetti kasvaa. Joukkojen kokoa voidaan samalla kasvattaa rajatta, jolloin virheherkkyys saadaan mielivaltaisen pieneksi.

Rajoitteita on kuitenkin mielekästä asettaa paitsi matemaattisen mielenkiinnon ylläpitämiseksi, myös jatkuvan gaussisen kanavan formalismin saamiseksi vastaamaan sitä fysikaalista tilannetta, johon formalismia aiotaan soveltaa.

Sovelluksissa kanavien sisäänmenolle ja ulostulolle asetetaan yleensä fysikaalisesti motivoituja rajoitteita. Rajoitteet mallinnetaan rajoittamalla kanavan määrittelevien tiheysfunktioiden määrittelyalue $\mathbb{R}^n$:n (Borel) osajoukkoon F_n. Kanavia, joille F_n on aito osajoukko $\mathbb{R}^n$:stä, kutsutaan jatkossa rajoitetuiksi kanaviksi tai F_n-rajoitetuiksi kanaviksi, mikäli rajoitteet on tarpeen spesifioida.

Haluamme rajoittaa ”signaalin voimakkuutta”, tarkalleen ottaen sen keskimääräistä voimakkuutta. Keskimääräisen tehon rajoitusta vastaava rajoitettu kanava saadaan asettamalla $F_n = \{\overline{x} : n^{-1} \sum_{i=1}^{n} x_i^2 \leq M\}$. Kutsumme tätä kanavaa jatkossa aikadiskreetiksi gaussiseksi M-tehorajoitetuksi kanavaksi.

5.3 Kapasiteetti

Shannonin kuuluisin tulos koskee amplitudijatkuvan kanavan kapasiteettia. Kapasiteetti on aiemmin määritelty keskinäisentropian kautta. Kapasiteetti on mahdollista määritellä siten, ettei määritelmään tarvitse sisällyttää tietoa kanavan purkufunktiosta ja siirtymätodennäköisyyksien tiheysfunktioista (ks. määritelmä 5.10). Merkinnällä (u, n, λ) viitataan mielivaltaiseen koodiin, jonka koodisanojen määrä on u, sanapituus n ja virhealttius λ.

Määritelmä 5.11 (Mahdollinen siirtonopeus) *Sanomme, että annettu R on mahdollinen siirtonopeus (engl.* premissible rate of transmission*) annetulle aikadiskreetille ja amplitudijatkuvalle kanavalle jos on olemassa koodi $(\lceil e^{nR} \rceil, n, \lambda_n)$, jolle $\lambda_n \to 0$ kun $n \to \infty$.*

Nyt voimme antaa kapasiteetille uuden, aiemman määritelmän kanssa yhtenevän tarkasteltavan kanavan purkufunktiosta ja siirtymätodennäköisyyksien tiheysfunktioista riippumattoman määritelmän.

Määritelmä 5.12 (Kanavan kapasiteetti (engl. *channel capacity*))

$$C = \sup \left\{ R \mid \exists \text{ koodi } (\lceil e^{nR} \rceil, n, \lambda_n), \text{ jolle } \lambda_n \to 0, \text{ kun } n \to \infty \right\}$$

Aikadiskreetin gaussisen kanavan kapasiteetin todistamiseksi tarvitaan joitakin apulauseita.

Lause 5.13 *Otetaan mielivaltainen aikadiskreetti, amplitudijatkuva, F_n-rajoitettu kanava. Kiinnitetään $n \in \mathbb{N}$. Olkoon $p(\overline{x}) : \mathbb{R}^n \to \mathbb{R}$ mielivaltainen todennäköisyystiheysfunktio. Nyt $\forall a \in \mathbb{R}$, merkitään*

$$A = \left\{ (\overline{x}, \overline{y}) \mid \log \frac{p(\overline{y}|\overline{x})}{p(\overline{y})} > a \right\}$$

Nyt $\forall u \in \mathbb{Z}_+$ $\exists$ (u, n, λ)-koodi, jonka koodisanat $\overline{w}_i \in F_n$, siten että

$$\begin{aligned} \lambda &\leq ue^{-a} + P\{(\overline{x}, \overline{y}) \notin A\} + P\{\overline{x} \notin F_n\} && (5.5) \\ &= ue^{-a} + \iint_{A^{\complement}} p(\overline{x}, \overline{y}) \, \mathrm{d}\overline{x} \, \mathrm{d}\overline{y} + \int_{F^{\complement}} p(\overline{x}) \, \mathrm{d}\overline{x} && (5.6) \end{aligned}$$

Todistus: Olkoon ϵ epäyhtälön 5.5 oikea puoli. Jos $\epsilon \geq 1$ pätee 5.5 triviaalisti. Voidaan olettaa, että $0 < \epsilon < 1$.

Merkitään joukon A $\overline{x}$-projektiota $A_{\overline{x}} = \{\overline{y} : (\overline{x}, \overline{y}) \in A\}$. Aloitetaan valitsemalla (jos mahdollista) $\overline{x}^{(1)} \in F_n$ koodisanaksi $\overline{w}_1$ siten, että

$$P\left\{ \overline{y} \in A_{\overline{x}^{(1)}} \mid \overline{x} = \overline{x}^{(1)} \right\} \geq 1 - \epsilon$$

ja valitaan $A_{\overline{x}^{(1)}}$ purkujoukoksi B_1. Jatketaan valitsemalla (jos mahdollista) $\overline{x}^{(k)} \in F_n$ koodisanaksi $\overline{w}_k$ siten, että

$$P\left\{\overline{y} \in A_{\overline{x}^{(k)}} \setminus \bigcup_{i=1}^{k-1} B_i \mid \overline{x} = \overline{x}^{(k)}\right\} \geq 1 - \epsilon$$

ja

$$B_k = A_{\overline{x}^{(k)}} \setminus \bigcup_{i=1}^{k-1} B_i.$$

Prosessi päättyy kun $\overline{x}^{(k)}$:n valinta ei enää onnistu, eli kun $\cup_{i=1}^{k-1} B_i$ on niin iso, ettei uutta $\overline{x}^{(k)}$:lle löydy riittävän pienen virheherkkyyden antavaa purkujoukkoa. Mikäli prosessi ei pääty äärellisen askelmäärän jälkeen, muodostavat koodisanat $\overline{w}_i$ ja purkujoukot B_i, $i = 1, \ldots, u$ halutun (u, n, λ)-koodin. Voidaan siis olettaa, että prosessi päättyy t:n valinnan jälkeen. Osoitamme, että $t \geq u$ osoittamalla, että $\epsilon \geq te^{-a} P\{(\overline{x}, \overline{y}) \notin A\} + P\{\overline{x} \notin F_n\}$.

Olkoon $B = \cup_{j=1}^{t} B_j$. Jos $t = 0$ asetetaan $B = \emptyset$. Nyt

$$\begin{aligned}
P\{(\overline{x}, \overline{y}) \in A\} &= \iint_{(\overline{x},\overline{y})\in A} p(\overline{x}, \overline{y})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} = \int_{\overline{x}} p(\overline{x}) \int_{\overline{y}\in A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} \\
&= \int_{\overline{x}} p(\overline{x}) \int_{\overline{y}\in B\cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} + \int_{\overline{x}} p(\overline{x}) \int_{\overline{y}\in B^{\complement}\cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} \\
&= \int_{\overline{x}} p(\overline{x}) \int_{\overline{y}\in B\cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} \\
&\quad + \int_{\overline{x}\in F} p(\overline{x}) \int_{\overline{y}\in B^{\complement}\cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} \\
&\quad + \int_{\overline{x}\notin F} p(\overline{x}) \int_{\overline{y}\in B^{\complement}\cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} \qquad (5.7)
\end{aligned}$$

Koska $B \cap A_{\overline{x}} \subset B$ ja $B_i \in A_{\overline{x}^{(i)}}$ lausekkeen 5.7 ensimmäiselle termille pätee

$$\begin{aligned}
\int_{\overline{x}} p(\overline{x}) \int_{\overline{y}\in B\cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} &\leq \int_{\overline{x}} p(\overline{x}) \int_{\overline{y}\in B} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} \\
&= P\{\overline{Y} \in B\} = \sum_{i=1}^{t} P\{\overline{Y} \in B_i\} \\
&\leq \sum_{i=1}^{t} P\{\overline{Y} \in A_{\overline{x}^{(i)}}\} \\
&\leq \sum_{i=1}^{t} e^{-a} = te^{-a}
\end{aligned}$$

Epäyhtälöistä viimeinen saadaan seuraavasti. Jos $\overline{y} \in A_{\overline{x}^{(i)}}$, niin $(\overline{x}^{(i)}, \overline{y}) \in A$. Silloin $\log(p(\overline{y}|\overline{x}^{(i)})/p(\overline{y})) > a$, eli $p(\overline{y}) < p(\overline{y}|\overline{x}^{(i)})e^{-a}$, joten

$$P\{\overline{Y} \in A_{\overline{x}^{(i)}}\} = \int_{\overline{y} \in A_{\overline{x}^{(i)}}} p(\overline{y})\,\mathrm{d}\overline{y} \leq e^{-a} \int_{\overline{y} \in A_{\overline{x}^{(i)}}} p(\overline{y}|\overline{x}^{(i)})\,\mathrm{d}\overline{y} \leq e^{-a}$$

Lausekkeen 5.7 toisen termin arvioimiseksi väitämme, että

$$\int_{\overline{y} \in B^{\complement} \cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{y} < 1 - \epsilon \quad \forall \overline{x} \in F_n \tag{5.8}$$

Oletetaan aluksi, että $\overline{x}$ on yksi koodisanoista $\overline{w}_k = \overline{x}^{(k)}$. Nyt $B^{\complement} \cap A_{\overline{x}} = \emptyset$. Näin on, koska jos $\overline{y} \in A_{\overline{w}_k}$, niin väistämättä $\overline{y} \in B$. Jos $\overline{y} \in A_{\overline{w}_k}$ ja $\overline{y} \notin \cup_{i=1}^{k-1} B_i$, niin $\overline{y} \in B_k \subset B$ (vrt. B_k:n valinta). Jos toisaalta $\overline{y} \in A_{\overline{w}_k}$ ja $\overline{y} \in \cup_{i=q}^{k-1} B_i$, niin $\overline{y} \in B$, joka on kaikkien purkujoukkojen B_i yhdiste. Nyt 5.8 on osoitettu kun $\overline{x}$ on yksi koodisanoista.

Oletetaan, että $\overline{x}$ ei ole koodisana. Jos käänteinen epäyhtälö 5.8 ($\geq$) pätee, niin (koska $\overline{x} \in F_n$ ja $B^{\complement} \cap A_{\overline{x}} = A_{\overline{x}} \setminus \cup_{i=1}^{t} B_i$) voidaan koodia laajentaa valitsemalla $\overline{x}^{(t+1)} = \overline{x}$ ja $B_{t+1} = B^{\complement} \cap A_{\overline{x}}$. Tämä taas on ristiriidassa sen oletuksen kanssa, että prosessi päättyisi t:n koodisanan valinnan jälkeen. Näin ollen lausekkeen 5.7 toiselle termille pätee

$$\int_{\overline{x} \in F} p(\overline{x}) \int_{\overline{y} \in B^{\complement} \cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} \leq 1 - \epsilon$$

Arvioidaan lausekkeen 5.7 kolmatta termiä seuraavasti muistaen, että $B^{\complement} \cap A_{\overline{x}} \subset \mathbb{R}^n$.

$$\int_{\overline{x} \notin F} p(\overline{x}) \int_{\overline{y} \in B^{\complement} \cap A_{\overline{x}}} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} \leq \int_{\overline{x} \notin F_n} p(\overline{x}) \int_{\overline{y} \in \mathbb{R}^n} p(\overline{y}|\overline{x})\,\mathrm{d}\overline{x}\,\mathrm{d}\overline{y} = P\{\overline{X} \notin F_n\}$$

Nyt saadaan

$$\begin{aligned} P\{(\overline{x}, \overline{y}) \in A\} &\leq te^{-a} + 1 - \epsilon + P\{\overline{X} \notin F\} \\ \Leftrightarrow \epsilon &\leq te^{-a} + (1 - P\{(\overline{x}, \overline{y}) \in A\}) + P\{\overline{X} \notin F\} \\ \Leftrightarrow \epsilon &\leq te^{-a} + P\{(\overline{x}, \overline{y}) \notin A\} + P\{\overline{X} \notin F\} \end{aligned}$$

Näin ollen $t \geq u$, eli haluttu koodi on olemassa myös kun koodisanojen valintaprosessi päättyy t:n valinnan jälkeen. □

5.3.1 Koodauslause

Seuraavaksi osoitamme, että aikadiskreetin M-tehorajoitetun gaussisen kanavan, jonka häiriön varianssi on N, kapasiteetti on

$$C = \frac{1}{2}\log(1 + \frac{M}{N}).$$

Todistus on kaksiosainen. Ensiksi todistamme koodauslauseen (engl. *coding theorem*) $C \geq C_0 = \frac{1}{2}\log(1 + \frac{M}{N})$, eli että kaikki siirtonopeudet $R < C_0$ ovat saavutettavissa. Toisessa vaiheessa todistamme käänteisen (engl. *converse*) koodauslauseen $C \leq \frac{1}{2}\log(1 + \frac{M}{N})$, eli että aina kun siirtonopeus R on saavutettavissa, pätee $R < C_0$.

Lause 5.14 (Koodauslause) $C \geq \frac{1}{2}\log(1 + \frac{M}{N})$

Todistus: Olkoon $R < \frac{1}{2}\log(1+\frac{M}{N})$. Jos nyt löytyy jono $(\lfloor e^{nR} \rfloor, n, \lambda_n)$-koodeja, joille $\lambda_n \to 0$ kun $n \to \infty$, on lause todistettu. Tällöin nimittäin määritelmän nojalla R on mahdollinen siirtonopeus kanavalle, joten $C = \sup\{R\}$ on vähintään R:n edellä annettu yläraja.

Valitaan $M_0 < M$ ja lauseen 5.13 $p(\overline{x})$ seuraavasti

$$p(\overline{x}) = (2\pi M_0)^{-\frac{1}{2}n} \exp(-\sum_{j=1}^{n} \frac{x_i^2}{2M_0}) \tag{5.9}$$

Toisin sanoen syötteen komponentit ovat riippumattomia ja normaalijakautuneita keskiarvolla 0 ja varianssilla $M_0 < M$ (jakauma $\mathrm{N}(0, M_0)$).

Nyt lausekkeen 5.5 kolmas termi

$$P\{\overline{X} \notin F_n\} = P\{n^{-1}\sum_{j=1}^{n} X_j^2 > M\} \xrightarrow[n\to\infty]{} P\{E(X_j^2) > M\} = P\{M_0 > M\} = 0$$

Muistetaan, että vastaanotettu viesti $\overline{Y} = \overline{X} + \overline{Z}$, jossa lähetetyn viestin $\overline{X}$ komponentit ovat riippumattomia ja jakaumaltaan $\mathrm{N}(0, M_0)$ (ks. lauseke 5.9) ja kohinan $\overline{Z} = \overline{Y} - \overline{X}$) komponentit ovat riippumattomia ja jakaumaltaan $\mathrm{N}(0, N)$ (ks. lauseke 5.4).

$\overline{Z}$:n ehdollinen jakauma kun $\overline{X}$ on annettuna on sama kuin $\overline{Z}$:n jakauma ilman ehtoa, joten $\overline{Z}$ ja $\overline{X}$ ovat riippumattomat. Koska riippumattomien normaalijakautuneiden muuttujien summa on edelleen riippumaton ja normaalijakautunut, ovat $\overline{Y}$:n komponentit riippumattomia ja jakaumaltaan $\mathrm{N}(0, M_0)$. Pidetään tämä mielessä ja tarkastellaan lausekkeen 5.5 toista

termiä.

$$\begin{aligned} a &\geq \log \frac{p(\overline{y}|\overline{x})}{p(\overline{y})} \qquad (5.10) \\ &= \log \frac{(2\pi N)^{-\frac{1}{2}n} \exp(-\sum_{j=1}^{n} \frac{(y_j - x_j)^2}{2N})}{(2\pi(N+M_0))^{-\frac{1}{2}n} \exp(-\sum_{j=1}^{n} \frac{(y_j - x_j)^2}{2(N+M_0)})} \\ &= \frac{1}{2} n \log\left(1 + \frac{M_0}{N}\right) + \frac{1}{2} \sum_{j=1}^{n} \left(\frac{y_j^2}{N+M_0} - \frac{(y_j - x_j)^2}{N} \right) \end{aligned}$$

Olkoon $W_j = Y_j^2/2(N+M_0) - Z_j^2/2N$. Koska $E(Y_j^2) = N + M_0$ ja $E(Z_j^2) = N$, on $E(W_j) = 0$. Nyt W_j:n varianssi voidaan laskea seuraavasti

$$\mathrm{var}(W_j) = E(W_j^2) = \frac{1}{4} \left\{ E\left(\left(\frac{Y_j}{\sqrt{N+M_0}} \right)^4 \right) + E\left(\left(\frac{Z_j}{\sqrt{N}} \right)^4 \right) - \frac{2E(Y_j^2 Z_j^2)}{N(N+M_0)} \right\}$$

Koska $Y_j/\sqrt{N+M_0} \sim \mathrm{N}(0,1)$ ja $Z_j/\sqrt{N} \sim \mathrm{N}(0,1)$,

$$E\left(\left(\frac{Y_j}{\sqrt{N+M_0}} \right)^4 \right) = E\left(\left(\frac{Z_j}{\sqrt{N}} \right)^4 \right) = 3.$$

Koska X_j ja Z_j ovat riippumattomat, pätee $E(Y_j^2 X_j^2) = E((Z_j^2 + 2X_j Z_j + Z_j^2) Z_j^2) = E(X_j^2)E(Z_j^2) + E(X_j)E(Z_j^3) + E(Z_j^4) = M_0 N + 0 + 3N^2$. Siten

$$\mathrm{var}(W_j) = \frac{1}{4} \left\{ 6 - \frac{2(M_0 N + 3N^2)}{N(N+M_0)} \right\} = \frac{M_0}{N+M_0}.$$

Olkoon $V_n = \sum_{j=1}^{n} W_j$. Nyt $E(V_n) = 0$ ja muuttujien W_j riippumattomuuden nojalla $\mathrm{var}(V_n) = \sum_{j=1}^{n} \mathrm{var}(W_j) = nM_0/(N+M_0)$. Nyt voimme kirjoittaa epäyhtälön 5.10 muotoon

$$a \geq \log \frac{p(\overline{y}|\overline{x})}{p(\overline{y})} = \frac{1}{2} n \log\left(1 + \frac{M_0}{N}\right) + V_n$$

Valitaan $a = \frac{1}{2} n \log(1 + M_0/N) - n\delta$, missä $\delta > 0$. Chebysevin epäyhtälön avulla saadaan nyt lausekkeen 5.5 toinen termi

$$\begin{aligned} P\{(\overline{x}, \overline{y}) \notin A\} &= P\left\{ (\overline{x}, \overline{y}) : \log \frac{p(\overline{y}|\overline{x})}{p(\overline{y})} \leq a \right\} \\ &= P\{V_n \leq -n\delta\} \\ &\leq \frac{E(V_n^2)}{n^2\delta^2} = \frac{M_0}{(N+M_0)\delta^2} \frac{1}{n} \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

Valitaan lauseen 5.13 muuttujaksi $u = \lceil e^{nR} \rceil$, jossa $R < \frac{1}{2}\log(1 + M/N)$. Kun nyt valitaan M_0 riittävän läheltä M:ää ja riittävän pieni δ, lausekkeen 5.5 ensimmäiselle termille pätee

$$ue^{-a} \leq e^{nR} e^{\frac{1}{2}n\log(1+M_0/N)} e^{n\delta} \xrightarrow[n\to\infty]{} 0$$

Nyt lauseen 5.13 nojalla löytyy jono $(\lceil e^{nR} \rceil, n, \lambda_n)$-koodeja, joille $\lambda_n \to 0$ kun $n \to \infty$. □

Lause 5.15 (Fanon epäyhtälö) *Oletetaan aikadiskreetti gaussinen kanava ja sille mielivaltainen koodaus (s, n), koodisanoilla $\overline{w}_1, \ldots, \overline{w}_s$. Olkoot välitetty viesti $\overline{X} = (X_1, \ldots, X_n)$ satunnaisvektori, joka saa arvon $\overline{w}_i$ todennäköisyydellä $p(\overline{w}_i)$ ja $\sum_{i=1}^{s} p(\overline{w}_i) = 1$, ja $\overline{Y}$ vastaava vastaanotettu viesti.*

Jos $p(e)$ on koodin kokonaisvirhetodennäköisyys annetulla välitettyjen viestien jakaumalla, pätee

$$H(\overline{X}|\overline{Y}) \leq H(p(e), 1 - p(e)) + p(e)\log(s-1)$$

Todistus: Olkoon g annetun koodin purkufunktio ja $\overline{y}$ vastaanotettu viesti. Ryhmittelyaksiooman (ks. luku 2.1) nojalla pätee

$$H(\overline{X}|\overline{Y} = \overline{y}) = H(q, 1-q) + qH(1) + (1-q)H(q_1, \ldots, q_{s-1})$$

missä $q = P\{\overline{X} = g(\overline{y})|\overline{Y} = \overline{y}\} = 1 - p(e|\overline{y})$ ja q_i:t ovat muotoa

$$\frac{p(\overline{x}|\overline{y})}{\sum_{\overline{x} \neq g(\overline{y})} p(\overline{x}|\overline{y})}$$

Koska $H(q_1, \ldots, q_{s-1}) \leq \log(s-1)$, pätee

$$H(\overline{X}|\overline{Y} = \overline{y}) \leq H(p(e|\overline{y}), 1 - p(e|\overline{y})) + p(e|\overline{y})\log(s-1) \tag{5.11}$$

Seuraavaksi on osoitettava, että todennäköisyyksien konveksin lineaarikombinaation (jatkuvassa tapauksessa tietysti summan sijaan integraali) epävarmuus on suurempi tai yhtäsuuri kuin epävarmuuksien konveksi lineaarikombinaatio.

Olkoon V satunnaismuuttuja, joka saa arvon 1, mikäli vastaanotettu ja purettu viesti poikkeaa alkuperäisestä (eli tapahtuu virhe) ja muutoin 0. Silloin pätee

$$\int_{-\infty}^{\infty} p(\overline{y}) H(p(e|\overline{y}), 1 - p(e|\overline{y}))\, \mathrm{d}\overline{y} = \int_{-\infty}^{\infty} p(\overline{y}) H(V|\overline{Y} = \overline{y})\, \mathrm{d}\overline{y} = H(V|Y).$$

Ja lauseen 5.6 ($H(X|Y) \leq H(X)$) nojalla

$$H(V|\overline{Y}) \leq H(V) = H(p(e), 1 - p(e)). \tag{5.12}$$

Yllä käytettyä tulosta ja luvun 5.1 tuloksia voidaan soveltaa suoraan, koska gaussisille kanaville kaikki H:t ovat aina äärellisiä.

Jos kerrotaan epäyhtälö 5.11 $p(\overline{y})$:llä ja integroidaan yli $\overline{y}$:n, huomataan epäyhtälön 5.12 avulla, että $H(\overline{X}|\overline{Y}) \leq H(p(e), 1 - p(e)) + p(e)\log(s-1)$, juuri lauseen edellyttämällä tavalla. □

Tarvitaan vielä kaksi apulausetta ennen kuin pääsemme käänteisen koodauslauseen kimppuun.

Lause 5.16 *Oletetaan aikadiskreetti amplitudijatkuva gaussinen kanava. Olkoot äärellisen arvojoukon omaava satunnaisvektori $\overline{X} = (X_1, \ldots, X_n)$ lähetetty viesti ja $\overline{Y}$ vastaava vastaanotettu viesti. Keskinäisinformaatiolle pätee*

$$I(\overline{X}; \overline{Y}) \leq \sum_{i=1}^{n} I(X_i|Y_i)$$

Todistus: Gaussisuuden nojalla H:t ovat äärellisiä. Koska $p(\overline{x}|\overline{y}) = p(x_1|y_1)\cdots p(x_n|y_n)$, pätee $H(\overline{Y}|\overline{X}) = \sum_{i=1}^{n} H(Y_i|X_i)$. Lisäksi lauseen 5.6 kohdan a perusteella pätee $H(\overline{Y}) = H(Y_1, \ldots, Y_n) \leq \sum_{i=1}^{n} H(Y_i)$. Nyt

$$\begin{aligned} I(\overline{X}; \overline{Y}) &= H(\overline{Y}) - H(\overline{Y}|\overline{X}) \\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) = \sum_{i=1}^{n}(H(Y_i) - H(Y_i|X_i)) = \sum_{i=1}^{n} I(X_i|Y_i) \end{aligned}$$

□

Lause 5.17 *Aikadiskreetin M-tehorajoitetun gaussisen kanavan, jonka häiriön varianssi on N, kaikille koodeille (s, n) pätee*

$$\log s < \frac{nC_0 + \log 2}{1 - E(p(e))}$$

missä $C_0 = \frac{1}{2}\log(1 + M/N)$ ja $E(p(e))$ on koodauksen keskimääräinen virhetodennäköisyys.

Todistus: Olkoot koodisanat $\overline{w}_i = (\overline{w}_{j1}, \ldots, \overline{w}_{jn})$, $\overline{X} = (X_1, \ldots, X_n)$ tasaisella jakaumalla (kaikkien sanojen todennäköisyys sama) satunnaisesti valittu koodisana ja $\overline{Y}$ vastaanotettu sana kun $\overline{X}$ on lähetetty. Lauseen 5.15 nojalla

$$\begin{aligned} H(\overline{X}|\overline{Y}) &\leq H(E(p(e)), 1 - E(p(e))) + E(p(e)) \log(s-1) \\ &\leq \log 2 + E(p(e)) \log s. \end{aligned} \tag{5.13}$$

Koska koodisanojen todennäköisyydet ovat samat saadaan

$$I(\overline{X}; \overline{Y}) = H(\overline{X}) - H(\overline{X}|\overline{Y}) = \log s - H(\overline{X}|\overline{Y}) \tag{5.14}$$

Nyt riittää näyttää, että

$$\sum_{i=1}^{n} I(X_i|Y_i) \leq nC_0. \tag{5.15}$$

Tästä, lauseesta 5.16 ja kaavoista 5.13 ja 5.14 seuraa $nC_0 \geq \log s - \log 2 - E(p(e))$, eli väite.

Osoitetaan, että 5.15 pätee. Valitaan X satunnaisesti (tasainen jakauma) koodisanojen sisältämien symbolien $\overline{w}_{ij}$ joukosta ($1 \leq i \leq s$ ja $1 \leq j \leq n$). X voidaan valita valitsemalla ensin satunnaisesti indeksi j ja sitten koodisana. Tarkastelemalla sanan j:ttä merkkiä huomataan $P\{X = \alpha\} = n^{-1} \sum_{j=1}^{n} P\{X_j = \alpha\}$. Tästä ja lauseesta 5.9 seuraa, että

$$I(X|Y) \geq n^{-1} \sum_{j=1}^{n} I(X_j|Y_j) \tag{5.16}$$

Tehorajoituksesta seuraa $n^{-1} \sum_{j=1}^{n} x_{ij}^2 \leq M$. Nyt voimme arvioida X:n varianssia seuraavasti.

$$\mathrm{var}X = E(X^2) - E(X)^2 \leq E(X^2) = (sn)^{-1} \sum_{i=1}^{s} \sum_{j=1}^{n} x_{ij}^2 \leq s^{-1} \sum_{i=1}^{s} M = M$$

Muistetaan, että $Y = X + Z$, missä $Z \sim \mathrm{N}(0, N)$ ja X ja Z ovat riippumattomia (ks. lauseen 5.14 todistus). Täten $\mathrm{var}Y = \mathrm{var}X + \mathrm{var}Z \leq M + N$ ja lauseen 5.7 nojalla $H(Y) \leq \frac{1}{2} \log(2\pi e(M+N))$. $H(Y|X) = H(Z)$ ja edelleen lauseen 5.7 nojalla $H(Z) = \frac{1}{2} \log(2\pi eN)$, joten $I(X|Y) = H(Y) - H(Y|X) \leq \frac{1}{2} \log(1 + M/N) = C_0$. Nyt epäyhtälön 5.16 nojalla pätee $\sum_{i=1}^{n} I(X_i|Y_i) \leq nC_0$. □

5.3.2 Käänteinen koodauslause

Kauniiksi lopuksi luvulle täydennämme jatkuvan kanavan kapasiteetin todistuksen todistamalla käänteisen koodauslauseen, eli että kun siirtonopeus pidetään yli C_0:n kanavan virhetodennäköisyys konvergoi ykköseen koodisanojen pituuden kasvaessa. Seuraava tulos on olennaisesti edellinen sillä lisämausteella, että $E(p(e))$ saadaan painettua mielivaltaisen pieneksi, jolloin oletuksiin voidaan ottaa mukaan myös mielivaltaisen pieni virheraja δ.

Lause 5.18 (Käänteinen koodauslause) *Jos $\epsilon > 0$ ja $0 \leq \lambda < 1$, niin riittävän suurella n kaikille koodeille (s, n, λ) pätee $log(s) < n(C_0 + \epsilon)$, missä $C_0 = \frac{1}{2}\log(1 + \frac{M}{N})$.*

Todistus: Olkoot $\overline{w}_1, \ldots, \overline{w}_s$ koodisanat ja $B_1, \ldots, B_s$ vastaavat purkujoukot. Arvioidaan B_i:tä rajoitetuilla joukoilla. Kiinnitetään positiivinen luku δ ja määritellään $B_i^* = B_i \cap D_n = \mathcal{B}_n(0, \sqrt{n(M + N + \delta)})$, eli n-ulotteisen euklidisen avaruuden pallo, jonka keskipiste on origossa ja säde on $\sqrt{n(M + N + \delta)}$. Tästä seuraa, että jos lähetetään $\overline{X} = \overline{w}_i$, todennäköisyys että vastaanotettu $\overline{Y}$ on B_i^*:n ulkopuolella on

$$\begin{aligned} P\{\overline{Y} \notin B_i^* | \overline{X} = \overline{w}_i\} &\leq P\{\overline{Y} \notin B_i | \overline{X} = \overline{w}_i\} + P\{\overline{Y} \notin D_n | \overline{X} = \overline{w}_i\} \\ &\leq \lambda + P\{\overline{Y} \notin D_n | \overline{X} = \overline{w}_i\} \end{aligned} \tag{5.17}$$

Osoitetaan, että

$$P\{\overline{Y} \notin D_n | \overline{X} = \overline{w}_i\} \xrightarrow[n\to\infty]{} 0 \tag{5.18}$$

Olkoon $\overline{w}_i = (\overline{w}_{i1}, \ldots, \overline{w}_{in})$ ja $\overline{Y} = (Y_1, \ldots, Y_n)$. Nyt $\overline{Y} = \overline{w}_i + \overline{Z}$, missä $Z_j \sim \mathrm{N}(0, N)$. Oletuksista seuraa, että

$$n^{-1} \sum_{k=1}^{n} Y_k^2 = n^{-1} \sum_{k=1}^{n} x_{ik}^2 + n^{-1} \sum_{k=1}^{n} Z_k^2 + 2n^{-1} \sum_{k=1}^{n} x_{ik} Z_k \tag{5.19}$$

Ensimmäinen termi on tehorajoitusoletuksen nojalla $\leq M$. Kun n kasvaa rajatta, lähestyy toinen termi N:ää ja kolmannen termin varianssi

$$\frac{4}{n^2} \sum_{k=1}^{n} x_{ik}^2 N \leq \frac{4NM}{n} \to 0.$$

Chebysevin epäyhtälön nojalla kolmas termi lähestyy nollaa. Siten $P\{n^{-1} \sum_{k=1}^{n} Y_k^2 > M + N + \delta\} \to 0$, kun $n \to \infty$. Mutta koska $P\{n^{-1} \sum_{k=1}^{n} Y_k^2 > M + N + \delta\} = P\{\overline{Y} \notin D_n | \overline{X} = \overline{w}_i\}$, pätee yhtälö 5.18.

Näin saamme uuden kokoelman purkujoukkoja B_i^*. Joukot eivät tosin tarkalleen ottaen muodosta purkujoukkojen kokoelmaa, koska ne eivät muodosta ositusta koko viestiavaruudelle

F_n. Äsken osoittamamme (yhtälö 5.18) nojalla kokoelma voidaan jatkaa F_n:n situkseksi mielivaltaisella tavalla tämän todistuksen siitä miksikään muuttumatta.

Riittävän suurilla n uuden koodin (s, n, λ^*) virhetodennäköisyys on kaavojen 5.17 ja 5.18 nojalla mielivaltaisen lähellä λ:a. Erityisesti voidaan olettaa, että $\lambda^* \leq (1 + \lambda)/2 < 1$.

Seuraavaksi haetaan alaraja purkujoukkojen B_i^* tilavuudelle $\mathrm{V}(B_i^*)$. Alaraja saadaan siitä, että joukkojen on oltava riittävän suuria takaamaan jokaiselle koodisanalle virheetön siirto vähintään todennäköisyydellä $1 - \lambda^*$.

Olkoon ϵ_n pienin positiivinen luku, jolle

$$P\left\{ n^{-1} \sum_{k=1}^{n} Z_k^2 > N(1 - \epsilon_n) \right\} \geq \frac{1 + \lambda}{2}.$$

Huomaa, että $\epsilon_n \to 0$, kun $n \to \infty$, koska $n^{-1} \sum_{k=1}^{n} Z_k^2 \to N$.

Olkoon G_i $\sqrt{nN(1 - \epsilon_n)}$-säteinen $\overline{w}_i$-keskinen avoin pallo. Väitetään, että

$$\mathrm{V}(B_i^*) \geq \mathrm{V}(G_i), \qquad i = 1, \ldots, s. \tag{5.20}$$

Osoitetaan tämä antiteesillä. Oletetaan siis, että $\exists i \in \{1, \ldots, s\}$ siten, että $\mathrm{V}(B_i^*) < \mathrm{V}(G_i)$. Tulemme osoittamaan, että

$$P\{\overline{Y} \in B_i^* | \overline{X} = \overline{w}_i\} < P\{\overline{Y} \in G_i | \overline{X} = \overline{w}_i\}. \tag{5.21}$$

Aiempien oletusten nojalla kuitenkin

$$\begin{aligned} P\{\overline{Y} \in B_i^* | \overline{X} = \overline{w}_i\} &\geq 1 - \frac{(1 - \lambda)}{2} \geq P\left\{ n^{-1} \sum_{k=1}^{n} Z_k^2 > N(1 - \epsilon_n) \right\} \\ &\geq P\{\overline{Y} \in G_i | \overline{X} = \overline{w}_i\} \end{aligned}$$

Tämä on ristiriidassa epäyhtälön 5.21 kanssa, joten epäyhtälö 5.20 pätee.

Osoittaaksemme, että epäyhtälö 5.21 seuraa antiteesistä, teemme seuraavan tarkastelun. Olkoon $\overline{y}_0 \in \partial G_i$. Koska normaalijakauman tiheysfunktio $p_n(\overline{y}|\overline{x})$ (ks. lauseke 5.4) on pisteiden $\overline{x}$ ja $\overline{y}$ etäisyyden suhteen aidosti vähenevä funktio, pätee

$$\begin{aligned} P\{\overline{Y} \in B_i^* \setminus G_i | \overline{X} = \overline{w}_i\} &= \int_{B_i^* \setminus G_i} p_n(\overline{y}|\overline{w}_i)\, \mathrm{d}\overline{y} \\ &\leq \int_{B_i^* \setminus G_i} p_n(\overline{y}_0|\overline{w}_i)\, \mathrm{d}\overline{y} = p_n(\overline{y}_0|\overline{w}_i)\mathrm{V}(B_i^* \setminus G_i) \end{aligned} \tag{5.22}$$

Samaan tapaan saadaan myös

$$\begin{aligned} P\{\overline{Y} \in G_i \setminus B_i^* | \overline{X} = \overline{w}_i\} &= \int_{G_i \setminus B_i^*} p_n(\overline{y}|\overline{w}_i)\,\mathrm{d}\overline{y} \\ &\geq \int_{G_i \setminus B_i^*} p_n(\overline{y}_0|\overline{w}_i)\,\mathrm{d}\overline{y} = p_n(\overline{y}_0|\overline{w}_i)\mathrm{V}(G_i \setminus B_i^*) \quad (5.23) \end{aligned}$$

Summaamalla epäyhtälöt 5.22 ja 5.23 saadaan

$$\begin{aligned} & P\{\overline{Y} \in G_i \setminus B_i^* | \overline{X} = \overline{w}_i\} - P\{\overline{Y} \in B_i^* \setminus G_i | \overline{X} = \overline{w}_i\} \\ \geq\ & p_n(\overline{y}_0|\overline{w}_i)(\mathrm{V}(G_i \setminus B_i^*) - \mathrm{V}(B_i^* \setminus G_i)) \\ \Leftrightarrow\ & P\{\overline{Y} \in G_i | \overline{X} = \overline{w}_i\} - P\{\overline{Y} \in G_i \cap B_i^* | \overline{X} = \overline{w}_i\} \\ & -(P\{\overline{Y} \in B_i^* | \overline{X} = \overline{w}_i\} - P\{\overline{Y} \in G_i \cap B_i^* | \overline{X} = \overline{w}_i\}) \\ \geq\ & p_n(\overline{y}_0|\overline{w}_i)(\mathrm{V}(G_i) - \mathrm{V}(G_i \cap B_i^*) - (\mathrm{V}(B_i^*) - \mathrm{V}(G_i \cap B_i^*))) \\ \Leftrightarrow\ & P\{\overline{Y} \in G_i | \overline{X} = \overline{w}_i\} - P\{\overline{Y} \in B_i^* | \overline{X} = \overline{w}_i\} \\ \geq\ & p_n(\overline{y}_0|\overline{w}_i)(\mathrm{V}(G_i) - \mathrm{V}(B_i^*)) \quad (5.24) \end{aligned}$$

Näin ollen antiteesin nojalla epäyhtälö 5.21 seuraa äsken johdetusta epäyhtälöstä 5.24.

Äsken todistamastamme väitteestä (ks. epäyhtälö 5.20) ja siitä, että $B_i^* \subset D_n$ seuraa

$$\sum_{i=1}^{s} \mathrm{V}(G_i) \leq \mathrm{V}(\bigcup_{j=1}^{s} B_j^*) \leq \mathrm{V}(D_n) \quad (5.25)$$

Jos nyt $V_n = \mathrm{V}(\mathcal{B}_n(0,1))$, seuraa epäyhtälöstä 5.23

$$\begin{aligned} sV_n(nN(1-\epsilon_n))^{n/2} &\leq V_n(n(M+N+\delta))^{n/2} \\ \Leftrightarrow \log s &\leq \frac{n}{2}\log\frac{M+N+\delta}{N(1-\epsilon_n)} \end{aligned}$$

Kun nyt valitaan riittävän pieni δ ja n on riittävän suuri, seuraa

$$log(s) < n\left(\frac{1}{2}\log\left(1+\frac{M}{N}\right)+\epsilon\right)$$

□

Lause 5.19 (Jatkuvan kanavan kapasiteetti) *Aikadiskreetin M-tehorajoitetun gaussisen kanavan kapasiteetti*

$$C = \frac{1}{2}\log\left(1+\frac{M}{N}\right).$$

Todistus: Lauseen 5.18 ja kapasiteetin määritelmän nojalla

$$\begin{aligned} R = \frac{log(s)}{n} &< \frac{1}{2}\log\left(1+\frac{M}{N}\right)+\epsilon \\ \sup\left\{R \mid \exists\,(\lceil e^{nR}\rceil, n, \lambda_n),\ \text{jolle } \lambda_n \xrightarrow[n\to\infty]{} 0\right\} &\leq \frac{1}{2}\log\left(1+\frac{M}{N}\right) \\ C &\leq \frac{1}{2}\log\left(1+\frac{M}{N}\right) \end{aligned}$$

Lisäksi koodauslauseen 5.14 nojalla

$$C \geq \frac{1}{2}\log\left(1+\frac{M}{N}\right)$$

□

5.4 Aika- ja amplitudijatkuva kanava

Amplitudijatkuviksi kutsutaan kanavia, joiden sisäänmenon ja ulostulon symbolijoukot S ja R ovat ylinumeroituvia. Aikajatkuviksi taas kutsutaan kanavia, joiden tiedonsiirto on ”jatkuvaa” aika-akselilla, eli signaaliavaruus on tietty funktioavaruus. Aika- ja amplitudijatkuvaksi kutsutun kanavan signaaliavaruus on reaalinen Hilbert-avaruus $\mathcal{L}_2[-T/2, T/2]$. Funktiot on rajoitettu äärelliselle aikavälille T, koska muuten välitetty viesti ei koskaan saapuisi perille.

Shannon oli erityisesti kiinnostunut jatkuvista gaussisista kaistarajoitetuista kanavista, koska niiden avulla oli mahdollista mallintaa tietoliikenteessä varsin tavallisia tilanteita. Jatkuvaa kanavaa sanotaan gaussiseksi, jos sen signaaliin indusoima virhe on mallinnettavissa stationaarisella gaussisella satunnaisprosessilla. Kaistarajoitetuksi kanavaa sanotaan kun sen signaaliavaruuden kaikille funktioille taajuusjakauma (engl. *spectral density*) $N(\omega)$ on identtisesti nolla äärellisen välin $[-2\pi W, 2\pi W]$ ulkopuolella, jossa W on kaistaleveys.

Ikävä kyllä $\mathcal{L}_2[-\infty, \infty]$-funktio (muu kuin nollafunktio) ei voi olla sekä aika- että kaistarajoitettu. Shannon kiersi ongelman käyttämällä kaistarajoitettua mutta aikarajoittamatonta signaalia ja kohinaa ja huomioimalla välitettynä signaalina signaalin ja kohinan summan rajoittuman välille $[-T/2, T/2]$. Tässä tarkastelussa jo välitetyn signaalin ”häntä” ei katoa minnekään (koska funktio ei ole aikarajoitettu), vaan jää periaatteessa sotkemaan tulevia signaaleja. Shannon pystyi tätä malliaan käyttäen kapasiteettilauseen toisen puolen

$$C \geq W\log\left(1+\frac{M}{NW}\right).$$

Toinen puoli on peräisin Wyneriltä [22], joka erilaista kaistarajoitetun kanavan mallia käyttäen pystyi osoittamaan yllä olevan lausekkeen olevan täsmälleen kapasiteetti.

6 Päätäntö

Informaatioteorian ja koodausteorian käsitteitä ja tuloksia on nyt käyty läpi siinä laajuudessa kuin on ollut tarpeellista aikadiskreetin amplitudijatkuvan gaussisen kanavan kapasiteetin osoittamiseksi. Lukija on toivottavasti hahmottanut käsitteiden ja tulosten lisäksi myös joitakin informaatioteorian kytkentöjä muille aloille. Lopuksi hahmotetaan vielä Claude Shannonin osuutta informaatioteorian kehitykseen.

Informaatioteorian voidaan tutkimusalana katsoa syntyneen Shannonin klassikkotyön [17] myötä. Jättiläiset, joiden harteilta Shannon näki edeltäjiään kauemmas, olivat kehitelleet tai intuitiivisesti käyttäneet monia informaatioteorian kannalta keskeisiä abstraktioita. Koodauksen idea on lähes yhtä vanha kuin kirjoitettu kieli ja redundanssin poisto koodauksella oli tuttua jo Samuel Morselle. Entropian käsite oli termodynamiikassa peräisin Rudolf Clausiukselta niinkin kaukaa kuin 1850-luvulta. Sitä kehittivät Ludvig Boltzmann ja tilastollisen mekaniikan tarpeisiin Willard Gibbs. Shannon ei informaation mittaansa muotoillessaan tiettävästi ollut tietoinen termodynamiikan ja tilastollisen mekaniikan entropiasta, johon hän käytännössä oli päätynyt. John von Neumann kuitenkin oli. Hänen kerrotaan [20] sanoneen informaation mittansa nimeämistä pohtineelle Shannonille

> *You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.*

Shannonin todistukset eivät olleet kovin täsmällisiä. Useita niistä voi paremminkin kutsua todistushahmotelmiksi. Tästä syystä matemaatikot vierastivat hänen töitään. Toisaalta töiden matemaattisuus (klassikkoartikkelissa on 23 lausetta ja niiden ”todistukset”) sai aikalaisinsinöörit lähes sivuuttamaan ne. Kuitenkin kaikilla Shannonin esittämillä tuloksilla on ennen pitkää ollut sovelluksia ja kaikki väittämät ovat ennen pitkää osoittautuneet oikeiksi ja vieläpä niin, että hänen esittämänsä todistusten hahmotelmat on voitu täydentää aukottomiksi todistuksiksi. [9]

Pian Shannonin tulosten julkaisun jälkeen käynnistyi tiivis koodien metsästys. Kesti kuitenkin puoli vuosisataa ennen kuin Shannonin tulosten antama lupaus häiriöisen kanavan kapasiteetista pystyttiin lunastamaan. 1960-luvulla kyettiin noin puoleen kapasiteetista. Ja 80-luvulla oltiin jo hyvin lähellä ns. ”Shannonin rajaa”, mutta tieto siitä, että vielä parempaan

pystytään, ei jättänyt tutkijoita rauhaan. Viimein vuosituhannen vaihteessa lupausten lunastajaksi kruunattiin LDPC-lohkokoodien (jotka tunnetaan myös Gallager-koodeina) moderni versio [14].

Informaatioteoria on poikkitieteellinen tutkimusala, joka ammentaa piirteitä matematiikasta, tilastotieteestä, tietotekniikasta, fysiikasta ja elektroniikasta. Moderni häiriöt hallitseva tietoliikenne, satelliitit ja kännykät, ovat informaatioteorian voimannäytteitä. Koodausteorian sovellukset ulottuvat kaikkialle CD-levyjen virheenkorjauksesta (CIRC, cross-interleaved Reed-Solomon coding) tiedon pakkaukseen (esim. LZW, Lempel-Ziv-Welch -koodaus). Kaiken tämän jälkeenkin informaatioteoria on kaikkea muuta kuin loppuun kaluttu tutkimusala.

7 Lähteet

[1] N. M. Abramson: *Information Theory and Coding*, New York: McGraw-Hill Book Co., 1963.

[2] Robert Ash: *Information Theory*, Interscience Publishers, 1967.

[3] Caslav Brukner ja Anton Zeilinger: "Conceptual inadequacy of the Shannon information in quantum measurements", *Physical Review A (Atomic, Molecular, and Optical Physics)*, 63(2), s. 022 113, 2001.
URL http://link.aps.org/abstract/PRA/v63/e022113

[4] G. Cullmann ja J. M. Labouygues: "Le code génétique, code instantané absolument optimal", *Comptes rendus de l'Académie des sciences. Série III, Sciences de la vie*, 301(5), ss. 157–60, 1985, ISSN 0764-4469.

[5] S. Even: "Tests for unique decipherability", *IEEE Transactions on Information Theory*, 9(2), ss. 109–112, 1963.

[6] Michael P. Frank: "The Physical Limits of Computing", *IEEE Computing in Science & Engineering*, (ss. 16–26), 2002.
URL http://www.cise.ufl.edu/research/revcomp/physlim/plpaper.html

[7] Michael P. Frank: "Physical Limits of Computing", luentokalvot, WWW, 2006, viitattu 20.6.2007.
URL http://www.eng.fsu.edu/~mpf/PhysLim/PL-sp06-m03-InfoComm.ppt

[8] Stanford Goldman: *Information Theory*, Dover Publications, 1968.

[9] Solomon W. Golomb *ym.*: "Claude Elwood Shannon (1916–2001)", *Notices of the AMS*, 49(1), 2002.

[10] P. M. Lee: "On the Axioms of Information Theory", *The Annals of Mathematical Statistics*, 35(1), ss. 415–418, mar 1964, ISSN 0003-4851.
URL http://links.jstor.org/sici?sici=0003-4851%28196403%2935%3A1%3C415%3AOTAOIT%3E2.0.CO%3B2-B

[11] Ari Lehtonen: “Informaatioteoria”, 2002, muistiinpanot Ari Lehtosen Jyväskylän yliopiston Matematiikan ja tilastotieteen laitoksella pitämistä luennoista.

[12] Jean loup Gailly: “comp.compression FAQ – item 9, Compression of random data”, WWW-sivusto, viitattu 14.6.2007.
URL `http://www.faqs.org/faqs/compression-faq/part1/section-8.html`

[13] David J. C. MacKay: *Information Theory, Inference, and Learning Algorithms*, 4. painos, Cambridge University Press, 2003.

[14] T. Richardson, A. Shokrollahi ja R. Urbanke: “Design of provably good low-density parity check codes”, teoksessa “Proceedings of the IEEE International Symposium on Information Theory”, (s. 199), 2000.

[15] A. A. Sardinas ja G. W. Patterson: “A Necessary and Sufficient Condition for the Unique Decomposition of Coded Messages”, teoksessa “IRE Convention Record”, (ss. 104–108), 1953, osa 8.

[16] Vesa Savolainen: *Verkkoteoria*, Docendo, 2001.

[17] Claude Elwood Shannon: “A Mathematical Theory of Communication”, *The Bell System Technical Journal*, 27, ss. 379–423, 623–656, 1948.

[18] Claude Elwood Shannon ja Warren Weaver: *Mathematical Theory of Communication*, University of Illinois press, 1949.

[19] Kyle Siegrist: “Virtual Laboratories in Probability and Statistics”, WWW, viitattu 21.6.2007.
URL `http://www.math.uah.edu/stat/`

[20] M. Tribus ja E. C. McIrvine: “Energy and information”, *Scientific American*, (224), syyskuu 1971.

[21] Norbert Wiener: *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, New York: John Wiley and Sons, 1949.

[22] Aaron D. Wyner: “The Capacity of the Band-Limited Gaussian Channel”, *The Bell System Technical Journal*, 45, ss. 359–395, 1966.